Trinity College Dublin, the University of Dublin

Essays in the Political Economy and Industrial Organisation of Media: Infrastructure, Identity, and Influence

Michael McRae

January 20, 2026

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

_____

Michael McRae

Acknowledgements

First and foremost, I am grateful to my family for their unwavering encouragement, particularly my wife, Michelle, for her patience, love, and support throughout this journey. I am also deeply thankful to my daughters, Isabelle and Ayla, whose joy and curiosity have been a constant grounding and source of inspiration.

I am forever grateful to my supervisors, Carol Newman and Tara Mitchell, for their guidance, encouragement, and intellectual support throughout this project. I also thank the members of my committee and colleagues at Trinity College Dublin for their invaluable feedback and support.

This research benefited from helpful comments and conversations with Benjamin Arold, Melena Djourelova, Nicola Mastrorocco, Elliot Motte, James Snyder Junior, and Tianyi Wang, and from the institutional support of the Grattan and Scholarship and Suzannah McAleese, TRISS Research Fellowships, and TRISS Travel Bursary.

I am also grateful to participants at seminars and conferences for their thoughtful feedback and encouragement. In particular, I thank audiences at the 3rd IFN Workshop in Economic History, the Text-as-Data Workshop at Lancaster, the Monash-Paris-Warwick-Zurich Text-as-Data Workshop, the Irish Economics Association Annual Conference (2022, 2023, and 2025), and the Irish Postgraduate and Early Career Economics Workshop.

I acknowledge the Irish Centre for High-End Computing (ICHEC) and for the provision of computational facilities and support. This work was also supported by RIT (Research IT, Trinity College Dublin).

Summary of Methods and Findings

This thesis investigates how economic and institutional forces shape political discourse by analysing the behaviour of media intermediaries: social media platforms, legacy newspapers, and religious radio networks, under varying structural constraints. Across three empirical studies, I apply causal inference methods to evaluate how media respond to incentives and how such responses affect symbolic representation, speech norms, and political behaviour.

The first paper examines infrastructure-led content moderation by exploiting a previously undocumented policy change by Google in July 2022, which required apps to restrict threatening and misinformation-related content to remain on the Play Store. The policy asymmetrically constrained Gettr and, to a lesser extent, Truth Social, while Gab remained unaffected. Using a novel panel of over 28 million posts from 62,000 users and a triple-differences (DDD) design, I find sharp, persistent declines in threatening content among constrained platforms, especially among high-risk users. No such decline occurred on Gab. I also find no evidence of harmful content being displaced to Gab or other forms of reactive user migration, suggesting that the policy reduced threatening discourse without triggering the spillovers commonly associated with public deplatforming. The findings suggest that upstream platform governance can significantly reshape online discourse, even in the absence of public enforcement or deplatforming.

The second paper studies how economic incentives shaped Southern newspapers' adoption of the term "Black" instead of "Negro" during the civil rights era. I leverage a coordinated campaign launched by civil rights leaders in June 1966 and track label diffusion using a corpus of 58 million digitised articles from 450 Southern newspapers. Adoption coincided with a positive shift in portrayals of Black actors and was predicted by market competition and white racial conservatism. To isolate the economic channel, I compare adoption rates across co-owned newspapers with shared editorial teams but different local readerships. Finally, I show that early adoption is associated with subsequent increases in Black political representation, suggesting that symbolic norm change facilitated substantive incorporation.

The third paper evaluates the political impact of Christian-conservative broadcasting by studying the expansion of Salem Communications following deregulatory shifts in the 1990s. Salem grew rapidly by acquiring stations and monetising through reli-

gious programming insulated from commercial pressures. Using plausibly exogenous variation in signal exposure from topography, I estimate the causal impact of Salem's entry on Republican vote share in presidential elections. I find persistent increases in Republican support, especially in evangelical counties, and evidence of long-run partisan realignment.

Together, these papers demonstrate how structural features of media markets, i.e. ownership constraints, platform governance, audience composition, and business models, mediate the relationship between identity, discourse, and democratic participation. The thesis contributes to political economy, media economics, economic history and the study of symbolic politics in institutional contexts.

Contents

Declaration **3**

Acknowledgements **5**

Summary of Methods and Findings **7**

Introduction **11**

1 Paper 1 (Social Media): Vertical Governance of Online Speech: Evidence from Google's Moderation Mandate **17**

2 Paper 2 (Newspapers): Black by Popular Demand: Media, Competition and the Evolution of a Social Norm **99**

3 Paper 3 (Radio): From the Pulpit to the Polls: The Electoral Consequences of Christian Talk Radio **181**

Conclusion **227**

Introduction

Over the past century, the institutions responsible for shaping political discourse have undergone profound transformation. While traditional gatekeepers such as newspapers and broadcast networks once served as the dominant intermediaries between political actors and the public, the modern information ecosystem is now characterised by a complex interplay of distributed platforms, commercial infrastructure providers, and identity-based media systems. This thesis examines how economic incentives and institutional structures shape the production, diffusion, and political consequences of discourse in such an environment. Across three empirical settings: digital platform governance, racial label adoption in legacy media, and religious broadcasting, I investigate how shifts in the structure of media markets affect the content of political communication and, in turn, political behaviour and representation.

Each of the three papers that comprise this dissertation engages a common set of theoretical concerns. First, they explore how economic intermediaries, platforms, newspapers, radio networks, respond to incentives generated by their position within vertically or horizontally structured media markets. Second, they examine how these responses shape downstream patterns of expression, information, or symbolic representation. And third, they test whether these shifts in content meaningfully influence political attitudes or participation, especially among historically underrepresented groups. The overarching goal is to provide causal evidence on how profit-motivated actors in information markets mediate the relationship between social identity, political voice, and institutional power.

The first paper studies the role of infrastructure-level governance in shaping content moderation practices on alternative social media platforms. I exploit a silent but binding change in Google's Play Store developer policy, announced in July 2022, which required platforms to remove threatening and misinformation-related content to remain eligible for distribution. The policy applied asymmetrically across the three largest alternative to mainstream social media networks: Gettr was already on the Play Store and directly constrained, Truth Social was applying and faced indirect pressure, while Gab remained excluded. Using a novel panel of over 28 million posts from 62,000 users and a triple-differences design, I find that threatening content declined sharply and persistently on the treated platforms, particularly among users with high baseline risk. These changes are not explained by self-censorship, contemporaneous shocks, or

user sorting. I find no evidence of content displacement to Gab or reactive migration by users, indicating that the policy reduced harmful speech without triggering the spillovers often observed in more visible deplatforming episodes. I further document substantial declines in lawful but politically sensitive narratives, such as election denial and commentary on the January 6 Capitol insurrection, suggesting that upstream enforcement can reshape not only the boundaries of permissible speech, but also the broader ideological terrain of discourse. This paper contributes to literatures on platform governance, vertical restraints in digital markets, and the political economy of private regulation.

The second paper examines the role of economic incentives in shaping the diffusion of a new racial identity label, "Black" instead of "Negro", in Southern U.S. newspapers during the civil rights era. Leveraging a coordinated in-group campaign launched by Black civil rights leaders in June 1966, I trace how the label diffused across a corpus of 58 million digitised newspaper articles from 450 newspapers. I show that adoption of the "Black" label coincided with more agentic and positive portrayals of Black people and groups in the press, and that variation in adoption was predicted by newspapers' exposure to white racial conservatism and local market competition. To isolate economic mechanisms, I exploit within-firm comparisons among newspapers produced by the same editorial teams but circulated in demographically distinct counties. Finally, I link label adoption to county-level increases in Black political representation, finding that earlier adoption of "Black" in Voting Rights Act-covered counties predicted faster growth in Black elected officials. The paper highlights how economic pressures facing white-oriented media outlets shaped the diffusion of a symbolic norm, and how these symbolic shifts influenced real political incorporation.

The third paper investigates how religious broadcasting mobilises partisan support by examining the political consequences of Salem Communications' rapid expansion in the post-deregulation era. Between 1994 and 2005, Salem, a Christian-conservative radio network, grew from 18 to over 100 stations, enabled by two major deregulatory shifts: the repeal of the Fairness Doctrine in 1987 and the Telecommunications Act of 1996. These changes allowed Salem to scale nationally without adhering to balanced reporting norms or ownership caps. Unlike traditional commercial broadcasters, Salem derived a substantial share of its revenue from religious programming purchased by ministries, insulating it from advertiser pressure and enabling sustained ideological messaging. Exploiting quasi-random variation in signal strength driven by local topography, I estimate the causal effect of Salem exposure on Republican vote share in U.S. presidential elections. Results show persistent increases in Republican support, particularly in counties with high evangelical adherence, and long-run partisan realignment in areas first exposed. This paper contributes to literatures on media effects, religious mobilisation, and the political consequences of structural change in media regulation.

Taken together, these three papers demonstrate how the structural and economic

conditions under which media operate shape both the content and consequences of political communication. By focusing on settings where expressive norms, identity-based messaging, or moderation policies are in flux, the dissertation offers new evidence on how media intermediaries both reflect and reinforce prevailing power structures. The findings speak to ongoing debates about the role of digital platforms in democratic life, the political function of symbolic representation, and the consequences of ideological segmentation in modern information environments.

The remainder of this thesis is structured as follows. Chapter 1 presents the first paper on infrastructure-led content moderation. Chapter 2 presents the second paper on racial label adoption and its political effects. Chapter 3 presents the third paper on religious radio and conservative voting.

Chapter I

Vertical Governance of Online Speech: Evidence from Google's Moderation Mandate

Abstract

This paper provides the first causal evidence that upstream infrastructure providers can reshape social media discourse by enforcing content moderation through access-based leverage. I exploit a 2022 update to Google's Play Store policy requiring stricter removal of violent threats and misinformation, along with variation in platform exposure across three similar 'Alt-Tech' social media platforms, within a triple-differences design. Using a novel panel of over 28 million posts from 62,000 users, I find that threatening content declined sharply and persistently on the exposed platforms, particularly among high-risk users. These effects are not explained by user self-censorship, public awareness, contemporaneous events, or selective data loss. I also document significant reductions in lawful but politically sensitive narratives, including election denial and January 6 insurrection commentary. The findings show how infrastructure-level enforcement can durably alter the boundaries of permissible speech across platforms, contributing to literatures on platform governance, vertical restraints in digital markets, and the institutional foundations of online discourse.

1 Introduction

In digital markets, a small number of private infrastructure providers control the primary channels through which users access applications, information, and communication tools. While governments continue to regulate online speech to varying degrees, the actions of private infrastructure firms such as mobile app stores and cloud service networks have the potential to reshape the boundaries of discourse in ways that are neither fully transparent nor democratically accountable.

While prior work has shown that stronger moderation policies on social media platforms can reduce harmful content (Allcott and Gentzkow, 2017; Klonick, 2018; Roberts, 2019), little is known about how such moderation is shaped by upstream infrastructure providers acting through market access conditions. This paper studies one such enforcement mechanism: a dominant mobile app distributor's imposition of stricter moderation requirements on platforms. In July 2022, Google announced a new policy requiring that all apps on the Play Store implement robust systems to remove threats of violence and misinformation related to sensitive events, such as public health emergencies or allegations of electoral fraud. Apps perceived as endorsing or profiting from denial of such events could be suspended. Importantly, the policy applied both to existing apps and to those seeking admission but not to apps with no ties to the Google marketplace, generating a quasi-experimental setting in which platforms faced differential exposure to the new enforcement regime.

This dynamic reflects a form of vertical control in which upstream intermediaries impose behavioural constraints on downstream platforms through access-based leverage rather than ownership or pricing. In contrast to classical models of vertical integration or exclusive dealing (Rey and Tirole, 2007; Lafontaine and Slade, 2007; Rey and Vergé, 2004; Hart and Tirole, 1990; Klein and Murphy, 1988), the enforcement studied here operates through infrastructure bottlenecks in digital distribution, allowing dominant firms to condition downstream conduct by controlling access to users. Much like franchisors enforcing uniform service standards or retailers requiring quality certification for shelf space, these non-price vertical restraints shape platform behaviour without formal ownership or pricing mechanisms. This logic is especially salient in two-sided markets (Rochet and Tirole, 2003; Armstrong, 2006), where platforms must balance user engagement and advertiser monetisation while remaining compliant with upstream distribution constraints. This paper provides the first causal evidence that vertical governance, exercised through mobile distribution control, can reshape platform moderation.

Identifying the causal effect of moderation policies on user content is challenging, as platform decisions to moderate are typically endogenous to factors such as ideology, user base composition, and growth strategy. Assessing the role of upstream enforcement adds further difficulty due to the opacity of policy implementation and

the prevalence of 'deplatforming'[1], which often creates discontinuities in data and induces user migration (Agarwal et al., 2022). I leverage a sharp and largely unpublicised shift in Google's developer policy that, crucially, did not result in any 'deplatforming' and, to the best of my knowledge, received no media coverage. This setting enables continuous observation of content across platforms without confounding spillovers from public awareness or shocks to the overall availability of the platforms. My identification strategy exploits variation along three dimensions: platform exposure to Google's policy, user-level pre-treatment threat intensity, and time; within a triple-differences framework. Specifically, I compare changes in threatening content between high- and low-risk users before and after the policy across treated platforms (Gettr and Truth Social) and a structurally untreated counterfactual (Gab), which remained excluded from the Play Store throughout.

The analysis draws on a novel panel dataset of over 28 million user posts from 62,000 users across the three platforms, spanning February 2021 to December 2024. I measure the 'threateningness' of posts using scores from the Perspective API (Wulczyn et al., 2017; Dixon et al., 2018) and track the prevalence of content categories explicitly targeted by the policy, such as election denial, anti-vaccine misinformation, and discourse related to the January 6 Capitol insurrection.

To validate this design, I begin by showing that Gab serves as a credible counterfactual, assessing pre-treatment linguistic similarity and parallel trends in outcome evolution both within and across platforms. I then systematically address several additional identification threats.

First, I rule out mechanical explanations. Although the policy applied only to Android apps, users post across platforms, and my crawler captures only web-visible content. The sharp drop in threatening posts at the treatment date is inconsistent with Android-only or retroactive enforcement, implying a conservative estimate of the true effect. Second, I address concerns about sample selection and attrition. High-threat users are more likely to remain active, especially on Gab, which could attenuate moderation effects. However, treatment effects remain stable across both balanced and unbalanced samples, suggesting that panel composition does not drive the results. Third, I examine selection into platforms. If users with higher threat levels sorted disproportionately into treated or untreated platforms before the policy, platform differences could reflect composition rather than treatment. This risk is mitigated by the DDD design, which includes user fixed effects and relies on within-user changes over time. I further confirm overlapping pre-treatment threat distributions across platforms and replicate the main results using matched subsamples. Fourth, I rule out alternative explanations tied to contemporaneous U.S. events (e.g., the Mar-a-Lago raid, midterm primaries, or Apple policy updates) through dynamic placebo tests and a cross-country falsification analysis using a large sample of Brazilian users. Fifth, I test

[1]The practice of removing apps from the digital infrastructure entirely

for user self-censorship and find no evidence of anticipatory behaviour, no detectable changes in platform moderation policy language, and treatment effects that increase monotonically with user-level risk, especially in the immediate aftermath of the policy. These patterns are inconsistent with purely user-driven adjustment. Finally, I test for spillovers. Among users active on both Gettr and Gab, I find no evidence that exposure to treated platforms altered behaviour on Gab or disrupted aggregate trends.

The results show that a one standard deviation increase in a user's pre-policy level of threatening content reduced such content by 40–60 percent of the pre-treatment mean on Gettr and Truth Social relative to Gab, with effects emerging immediately after the policy change and persisting over the entire study period, the end of 2024. Results are monotonically increasing in user 'threateningness' and are robust to user fixed effects, platform-specific time trends, and a wide range of alternative outcome and user exposure definitions. Taken together, the evidence supports a causal interpretation: upstream enforcement by Google induced measurable and durable changes in downstream moderation on exposed platforms.

Beyond the moderation of threatening content, I also document a distinct and independent shift in the prevalence of politically sensitive narratives, including election fraud claims, anti-vaccine discourse, and January 6 commentary. Users who were previously active in discussing these topics significantly reduced such engagement following the policy change. Given the timing and consistency of these declines across affected platforms, the evidence, again, points to platform-enforced moderation as the most plausible mechanism. This broader contraction of salient but largely lawful discourse highlights how infrastructure-based enforcement can compel platforms to reshape the contours of acceptable speech, not only by suppressing unlawful or violent content but by dampening controversial narratives prevalent in the ecosystem. These effects highlight how vertical governance by upstream infrastructure providers can significantly alter the landscape of discourse across the entire digital ecosystem, superseding fragmented or inconsistent platform-level rules.

A central concern in platform governance is whether content moderation suppresses harmful behaviour or simply pushes it elsewhere. Prior studies have shown that public, high-profile moderation efforts, such as coordinated bans or app store removals, often trigger backlash, user migration, and content reconstitution on less regulated platforms (Rizzi, 2024; Agarwal et al., 2022; Horta Ribeiro et al., 2023). To assess whether Google's more silent, infrastructure-level policy shift produced similar spillover dynamics, I examine three channels of behavioural adjustment. First, I test whether threatening users disengaged or reduced visibility on treated platforms. Second, I evaluate whether users reacted to perceived enforcement by discussing moderation before exiting. Third, I assess whether users displaced threatening content to Gab after leaving Gettr. Across all three domains, the evidence points to internal adjustment within treated platforms, not reactive redistribution or cross-platform reallocation. This sug-

gests that infrastructure enforcement can reduce harmful content without triggering the backlash dynamics observed in more visible moderation episodes.

The findings point to several important consequences for platform governance. First, control over distribution can translate into control over what information circulates online. When a single infrastructure provider conditions access to the market, it effectively sets rules that platforms must follow, even in the absence of formal regulation. Second, enforcement actions by one distributor often lead platforms to implement policy changes globally, affecting all users regardless of how they access the platform. Third, this asymmetry creates incentives for free-riding by other distributors who benefit from enforcement without bearing its costs. It also complicates decision-making for platforms that must navigate conflicting expectations across different gatekeepers. Finally, the combination of economic intermediation and content-based enforcement blurs the traditional line between market infrastructure and regulatory authority. This calls for a policy framework that takes seriously the role of infrastructure providers in shaping online speech.

These dynamics raise broader concerns about legitimacy, accountability, and welfare in digital speech markets. While private enforcement can reduce harmful content, it also shifts the power to define acceptable expression from public institutions to private firms. In this case, Google was able to suppress a wide range of political narratives that were neither illegal nor factually false. These were precisely the types of speech that alternative platforms were designed to support. Yet the threat of losing market access was enough to remove them from the ecosystem. If a small number of infrastructure providers can reshape platform behaviour through commercial pressure alone, the marketplace of ideas may reflect the risk preferences of a few firms rather than democratic or legal standards. The result may be excessive caution, leading platforms to over-comply and suppress controversial but lawful discourse. Understanding how these governance dynamics operate through market structure is essential for assessing the broader effects of infrastructure control on digital expression.

This paper contributes to three strands of literature. First, it adds to the growing body of work on content moderation and governance in digital media platforms. Existing research has examined how platforms regulate speech internally, balancing user engagement with reputational, legal, and normative pressures (Liu et al., 2021; Madio and Quinn, 2024; Kominers and Shapiro, 2024; Beknazar-Yuzbashev et al., 2024). Scholars have shown that platform architecture and recommendation systems shape users' exposure to misinformation and political polarisation (Allcott et al., 2020; Levy, 2021), and influence levels of hate speech, incivility, and partisan sorting (Mosquera et al., 2020; Müller and Schwarz, 2021; Cao et al., 2023). Recent work by Kalra (2025) finds that depersonalised content feeds reduce toxicity but at the cost of user engagement. Within this literature, one prominent line of empirical research focuses on platform self-governance. For example, Müller and Schwarz (2023) document re-

ductions in toxic language after Twitter removed Donald Trump's account, while Rizzi (2024) shows that Twitter's hate speech policy reduced hate on-platform but induced cross-platform spillovers to Parler. A related stream examines the role of public regulation, such as Germany's NetzDG law, which has been shown to reduce hate speech and associated offline violence (Duran et al., 2022; Andres and Slivko, 2021).

A closely related but still nascent literature examines infrastructure-level content governance, where upstream intermediaries, such as app stores or cloud hosts, shape downstream platform behaviour by conditioning market access. Existing work focuses primarily on deplatforming events, showing that app store removals can trigger user migration to fringe alternatives with higher levels of misinformation and toxicity (Agarwal et al., 2022; Horta Ribeiro et al., 2023). However, these studies leave open whether infrastructure pressure alters behaviour within platforms, rather than simply shifting users across them. This paper addresses that gap by providing the first causal evidence that vertical enforcement by infrastructure providers can induce moderation changes within affected platforms.

Moreover, unlike prior deplatforming episodes that triggered measurable spillovers to alternative platforms (Rizzi, 2024; Agarwal et al., 2022; Horta Ribeiro et al., 2023), I find no evidence of large-scale redistribution of harmful content to less regulated environments. Cross-platform users did not significantly increase threatening discourse on Gab following exit from moderated platforms, and moderation-related discourse rose steadily across all platforms, suggesting broad and diffuse engagement with content rules rather than concentrated backlash. This distinction highlights the importance of enforcement visibility: by operating silently and within the infrastructure, Google's policy appears to have reduced harmful content without provoking mass migration or content displacement, a dynamic that challenges conventional narratives of reactive user spillover. In doing so, the paper reframes infrastructure intermediaries not just as gatekeepers of market access, but as de facto regulators capable of shaping online discourse at scale. This advances the literature by uncovering a previously underexplored channel of governance, non-price vertical control, with implications for the political economy of digital speech, platform competition, and regulatory design.

This paper also contributes to the industrial organisation literature on vertical restraints in digital markets. Classic models emphasise how upstream firms can influence downstream behaviour through non-price mechanisms such as shelf-space access, franchising, or certification requirements (Klein and Murphy, 1988; Rey and Tirole, 2001; Lafontaine and Slade, 2007). I extend this logic to the governance of political speech in digital platforms, showing how mobile app stores act as infrastructure bottlenecks that condition downstream behaviour without direct ownership or pricing. In doing so, the paper builds on recent work by Madio et al. (2025), who document how financial intermediaries enforce content standards in adult entertainment. My findings generalise this mechanism to a broader regulatory context, demonstrating how vertical control

operates in mainstream political discourse and scales across platform architectures.

The third and final strand relates to the institutional foundations of political communication and digital governance. Foundational work in political economy shows how institutional actors influence the flow of information and political accountability, often via control over media ownership or content distribution (Besley and Prat, 2006; Prat, 2018). I extend this logic to infrastructure providers who, while not themselves content producers or governments, exercise substantial de facto authority over discourse by determining which platforms can access distribution channels and under what conditions. The enforcement regime I study thus resembles the kind of informal institutional power described by Acemoglu and Robinson (2008). Digital infrastructure, like legacy media systems, constitutes a locus of political power, capable of shaping expressive incentives and altering the effective rules that govern online speech.

This paper is structured as follows. Section 2 provides background and institutional details of mobile app distribution, content-moderation policies, and the growth of alternative platforms. Section 3 describes the data. Sections 4 discusses identification and explores the causal effect of the Google announcement on online behaviour. Section 5 discusses policy implications and Section 6 concludes.

2 Background

2.1 Private Governance, Infrastructure, and the Rise of Alt-Tech Platforms

In recent years, social media platforms have come under increasing pressure to moderate harmful or extreme content, including hate speech, misinformation, and incitements to violence. These pressures originate from governments, advertisers, civil society groups, and users, and reflect mounting concern about the downstream consequences of digital speech (Muller and Schwarz, 2019; Duran et al., 2022). In some settings, governments have implemented legal frameworks requiring platforms to remove specific types of content, Germany's 2017 Network Enforcement Act (*NetzDG*) being the most prominent example (Duran et al., 2022). But outside a few jurisdictions, content moderation remains primarily a matter of private governance, shaped by reputational concerns, platform policies, and advertiser preferences.

One consequence of this evolving regulatory environment has been the emergence of a parallel ecosystem of alternative social media platforms, often referred to as "Alt-Tech."[2] These platforms, including Truth Social, Gab, Gettr, and Parler, among others, have positioned themselves as ideological counterweights to mainstream platforms,

[2]The term "alt tech" is a portmanteau of "alt right" and "Big Tech," a phrase adopted by some prominent conservatives and their supporters beginning in 2015, following bans from mainstream platforms. Roose, Kevin. "The Alt-Right Created a Parallel Internet. It's an Unholy Mess." *The New York Times*, December 11, 2017 (`https://www.nytimes.com/2017/12/11/technology/alt-right-internet.html`).

promising to uphold "free speech" and resist what they characterise as politically biased or excessive moderation. Their appeal lies in their willingness to host content disallowed elsewhere, and their user bases often include groups disillusioned with the enforcement policies of larger platforms. Empirical research has documented the prevalence of polarising, conspiratorial, or hateful content on these sites (Zannettou et al., 2018; Aliapoulios et al., 2021), emphasising the challenges of moderating harmful speech in a fragmented online environment.

Although these platforms market themselves as unregulated spaces, their viability depends on access to key infrastructure. Many operate under business models reliant on advertising or subscriptions, which require sustained engagement and a sizeable user base, conditions that, in practice, often depend on discoverability through app stores and accessibility across devices. To reach users at scale, Alt-Tech platforms rely on third-party services for web hosting, payment processing, and mobile app distribution, dependencies that expose them to external forms of governance.

A growing number of infrastructure providers have adopted terms of service prohibiting the facilitation of hate speech or violent content, and have exercised this authority to deny access to noncompliant platforms. Parler, for example, was removed from both the Apple and Google app stores in January 2021 and later reinstated on Apple following adherence to moderation policies.[3] Gab was delisted from both major app stores in 2017. While these are headline Alt-Tech examples, smaller-scale enforcement is commonplace for all apps. Developer forums such as the iOS Developer and Android Developer subreddits contain numerous reports of apps being rejected or removed for failing to adhere to moderation-related policies, particularly those governing user-generated content.[4]

In addition to app store removals, infrastructure enforcement has taken several forms. For example, Amazon Web Services suspended Parler's hosting in January 2021, rendering the site inaccessible until it migrated to alternative providers. Cloudflare has withdrawn Distributed Denial of Service (DDoS) protection from multiple fringe platforms, including The Daily Stormer (2017), 8chan (2019), and Kiwi Farms (2022), each time citing violations of anti-abuse or hate speech policies. Payment processors such as PayPal and Stripe also cut off services to Gab in 2018 after it was linked to extremist violence. A full timeline of these and a non-exhaustive list of other examples is provided in Appendix Table C.1.

These cases illustrate the potential for infrastructure providers to exert meaningful control over platform access, not by removing content directly, but by conditioning the availability of critical technical services on compliance with moderation standards. Among these intermediaries, mobile app stores occupy a particularly influential position. As smartphone usage has become ubiquitous, most users now access social media

[3]Clayton, James. 2021. "Parler set to return to Apple's App Store." *BBC News*, April 19, 2021

[4]`https://www.reddit.com/r/iOSProgramming/`; `https://www.reddit.com/r/androiddev/`

platforms primarily through mobile apps rather than web browsers.[5] Two firms, Apple and Google, control the only official app stores available on iOS and Android devices, respectively, and require that platforms comply with a range of technical and content-related standards in order to remain listed. These policies prohibit the promotion of violence, hate speech, and other forms of harmful content, and, as mentioned, have been cited in multiple app removals or rejections.

2.2 Google's 2022 Enforcement Shock

On July 27, 2022, Google updated its Developer Program Policy to clarify and expand its expectations under the "Inappropriate Content" clause. It reiterated Google's stance against apps containing threatening or violent content and introduced a revised provision on "Sensitive Events" (see Figure C.2). The update prohibits apps from capitalising on or being insensitive toward events with "significant social, cultural, or political impact," including civil unrest, public health emergencies, natural disasters, and deaths. While content related to such events may be allowed if it holds educational, documentary, scientific, or artistic (EDSA) value, apps must not, for example, deny the occurrence of "well-documented, major tragic events" or appear to profit from such events without discernible benefit to victims. Compared to the previous version, the revised policy broadened the scope of sensitive events and made enforcement language more precise. Although framed as a clarification, the update was immediately enforceable and applied to both new and existing apps without a grace period.[6] This shift has important implications for user-generated content platforms, which may now face heightened liability not only for individual posts but for broader patterns of content or monetisation. Platforms hosting controversial or conspiratorial speech must implement more robust moderation systems to assess factual accuracy, contextual sensitivity, and commercial exploitation, raising compliance costs and the likelihood of content takedowns. The July update also included stricter provisions on misleading health claims, impersonation, ad placement, subscription cancellation, and VPN usage.[7]

This update was part of a broader sequence of policy revisions across 2022 that progressively expanded the scope and specificity of Google's content moderation requirements. On April 6, Google announced revised rules for user-generated content (UGC), scheduled to take effect on October 11, which placed system-level responsibility on platforms to implement proactive and ongoing moderation (see Figure C.3). Google also issued a clarification of its hate speech policy, reaffirming its prohibition

[5] According to Statista, over 95% of U.S. social media users access platforms primarily via mobile apps. See: `https://www.statista.com/topics/4689/mobile-social-media-usage-in-the-united-states`

[6] Archived version retrieved from the Internet Archive: `https://web.archive.org/web/20220727163419/https://support.google.com/googleplay/android-developer/answer/12253906`

[7] Some categories had specific deadlines, e.g., 30 days or a date within three months; the provisions relevant to this study were immediate.

on incitement and outlining expectations for compliance with local legal standards in different jurisdictions.[8]

By late July, enforcement activity had intensified, and the immediate applicability of the sensitive events clause introduced a new source of compliance exposure for platforms operating in sensitive political or social domains. A rejection of Truth Social from the Play Store on August 19, citing inadequate moderation systems and threatening content, which is discussed in the proceeding section, provided clear evidence that monitoring and enforcement was already being carried out in practice. Table A.1 summarises the key policy milestones over this period.

Unlike other provisions introduced earlier in the year, which included future compliance deadlines, the sensitive events clause took effect without delay. This makes July 27 a natural threshold for treatment exposure. From that point forward, platforms distributing on Android were subject to heightened enforcement risk and regulatory scrutiny. The empirical analysis that follows adopts this timing convention, comparing treated and untreated platforms to assess how this shift in enforcement salience influenced strategic moderation responses.

2.3 Gettr - Existing Google Play App

Gettr is a social media microblogging platform operating in the Alt-Tech space, presenting itself as a free speech alternative to mainstream platforms. Launched on July 4, 2021, by former Trump administration spokesperson Jason Miller, it quickly attracted a user base composed largely of conservatives and others who viewed traditional platforms as overly restrictive or politically biased in their moderation practices. While Gettr promised fewer restrictions on speech, it also emphasised a baseline commitment to user safety. As of early 2022, the app had been downloaded 6.5 million times globally across the Apple App Store and Google Play.[9] When Google's 2022 policy update introduced stricter content moderation requirements for apps distributing user-generated content, Gettr, as an existing Play Store application operating in a politically sensitive niche, faced heightened compliance expectations despite the absence of any publicly disclosed sanctions.

The potential removal of Gettr from the Google Play Store would entail substantial costs across multiple dimensions. First, from a user acquisition standpoint, the Play Store serves as a primary gateway for Android users to discover and install applications.[10] De-platforming would necessitate reliance on alternative distribution methods,

[8] See `https://web.archive.org/web/20220406190711/https://support.google.com/googleplay/android-developer/answer/11899428`

[9] David Klepper and Barbara Ortutay, "A Year After Trump Purge, 'Alt-Tech' Platform Offers Far-Right Refuge," *Associated Press*, February 5, 2022. https://apnews.com/article/donald-trump-technology-business-social-media-joe-rogan-83bdedc0163473e393918f95f889d594

[10] The Play Store provides a central platform for users to browse, search, and find apps. It's the official and pre-installed app store for Android devices, and includes features which ensures apps are kept up-to-date with the latest versions and security patches, making it the most reliable and secure

such as direct APK downloads or third-party app stores, which historically have shown significantly lower adoption rates due to user friction and security concerns (Goodwin and Woolley, 2022). Second, user engagement and retention could suffer; without access to the Play Store, users would not receive automatic updates, leading to a fragmented user experience and potential security vulnerabilities. Third, reputational risks are considerable. The removal of an app from major platforms often garners media attention and can be perceived as a signal of non-compliance or association with harmful content, potentially deterring both users and advertisers. The case of Parler illustrates these challenges: following its removal from the Google Play Store in January 2021 due to inadequate moderation policies, Parler experienced a significant decline in daily active users and faced difficulties in maintaining its infrastructure and user base.[11] [12]

2.4 Truth Social - Prospective Google Play App

Truth Social is a similar Alt-Tech micro-blogging platform, launched on February 21, 2022, by Trump Media & Technology Group (TMTG), a company founded by U.S. President Donald Trump. Marketed as an alternative to mainstream social media platforms, the platform positioned itself as a "haven of free speech," explicitly catering to users who believed major tech companies unfairly censored conservative viewpoints. Truth Social's launch followed Trump's permanent suspension[13] from Twitter in January 2021, which spurred efforts to create an independent platform free from the content moderation policies of Big Tech.

Although Truth Social was available to iOS users from its launch and accessible via web browser, it was initially unavailable on Google Play. As noted in Section 2, Google's new policy explicitly required all user-generated content platforms, whether existing or seeking listing, to demonstrate "effective systems for moderating" sensitive or violent content. On August 19, 2022, Google notified TMTG that Truth Social's Android app submission violated these requirements, reportedly sharing screenshots of posts that incited physical violence. In a statement reported at the time, a Google spokesperson reiterated that an effective moderation system was a condition for any app to be listed on Google Play. Truth Social responded that it was working to address these issues and was committed to compliance "without compromising our promise to be a haven for free speech."[14]

The incident came at a sensitive period for TMTG, whose planned merger with

way to get apps.

[11] Chen, S. (2021, January 9). *Google suspends Parler from app store after deadly Capitol violence.* Axios. Retrieved from https://www.axios.com/2021/01/09/capitol-mob-parler-google-ban

[12] Hosseinmardi, H., Ghasemian, A., Clauset, A., Mobius, M., Rothschild, D., & Watts, D. J. (2023). *Deplatforming did not decrease Parler users' activity on fringe social media.* PNAS Nexus, 2(3), pgad035. https://doi.org/10.1093/pnasnexus/pgad035

[13] This suspension was lifted in November 2022, following Elon Musk's acquisition of Twitter. See https://www.bbc.com/news/world-us-canada-63692369.

[14] Nico Grant, "Google Says Trump's Truth Social Must Scrub Violent Content to Join Play Store," *The New York Times*, August 30, 2022.

Digital World Acquisition Corp. (a SPAC that had raised nearly $300 million) was under federal investigation. Delays in app store approval threatened the platform's growth and investor confidence.[15] On October 12, 2022, Google reversed its initial decision and approved Truth Social for distribution on Google Play, contingent on its adherence to the same requirement to remove incitements to violence. This listing significantly expanded Truth Social's reach among Android users, who represent roughly 42% of the U.S. smartphone market.[16]

By contrast with Gettr, already on Google Play and never facing a public enforcement announcement, Truth Social underwent a period of outright exclusion. Other Alt-Tech platforms, such as Rumble, had already secured placement on Google Play, while Parler was in the process of regaining access,[17] creating further pressure for Truth Social to comply.

While Truth Social publicly maintained its ideological commitments, its eventual accommodation of Google's moderation requirements raises a broader question: did the threat of exclusion from a key distribution channel induce substantive changes in platform behaviour? Some reporting suggests that Truth Social may already have had internal moderation infrastructure in place.[18] Whether these actions reflected an internal policy shift, an effort to preempt enforcement, or a response to other pressures remains unclear.

2.5 Gab - Google Play Excluded

Gab serves as a critical test case for distinguishing between Google-led enforcement effects and broader ecosystem-wide changes in user behaviour or platform moderation. Founded in 2016, Gab was created as a free speech-oriented alternative to mainstream platforms, explicitly targeting users who felt marginalised by existing content moderation practices. The platform features a micro-blogging format similar to Truth Social and Gettr, with short posts ("gabs"), reposts, and a follower network.

Gab officially launched on Google Play in May 2017 but was removed by Google on August 17, 2017, for failing to demonstrate a sufficient level of moderation, including for content that encourages violence and advocates hate against groups of people."[19] In response, Gab filed an antitrust lawsuit against Google, which it dropped in October 2017 in favour of lobbying efforts targeting alleged monopolistic practices by Big Tech.[20]

[15] See Matthew Goldstein, "Google Approves Trump's Truth Social for Its Play Store," *The New York Times*, October 13, 2022. `https://www.nytimes.com/2022/10/13/business/truth-social-google-digital-world.html`

[16] `https://gs.statcounter.com/os-market-share/mobile/united-states-of-america/`.

[17] https://www.pewresearch.org/journalism/2022/10/06/alternative-social-media-appendix-a-detailed-tables-for-audit-of-seven-alternative-social-media-sites/

[18] Anecdotal evidence from mid-2022 indicates the removal of content critical of Donald Trump. See https://www.vanityfair.com/news/2022/06/truth-social-moderation-trump

[19] See Google faces lawsuit over removing Gab from Play Store," *BBC News*, September 18, 2017. https://www.bbc.com/news/technology-41306437

[20] See ibid.

Unlike other platforms in the Alt-Tech ecosystem, Gab has maintained an explicit ideological commitment to resisting external pressures to moderate user content. Its leadership has consistently framed content moderation requirements by major platforms and app stores as violations of free speech principles.[21] As a result, Gab has relied primarily on web-based access and third-party installations, significantly altering its infrastructure compared to peers like Gettr and Truth Social that remained dependent on app store distribution. Gab earns revenue through premium subscriptions, donations, and affiliate advertising.[22]

Despite these differences in infrastructure and access, Gab's user base and ideological positioning are highly similar to those of Gettr and Truth Social. This similarity makes Gab an important counterfactual: because it was already excluded from Google Play prior to the 2022 policy update and had no ambition of rejoining the distribution network, any subsequent changes in user behaviour or moderation practices on Gab cannot plausibly be attributed to new distribution-related pressures. Comparing Gab to other Alt-Tech platforms thus helps isolate the role of app store enforcement from broader endogenous shifts in the Alt-Tech ecosystem.

3 Data

My analysis draws on newly constructed datasets from three social media platforms: Truth Social, Gettr, and Gab. I assemble detailed post-level, comment-level, and user-level datasets from each platform to study the evolution of threatening and 'sensitive' content and platform level responses to Google's 2022 content moderation policy update.

3.1 Truth Social, Gettr, and Gab

I collect comprehensive user activity data from the three studied platforms, capturing the full period from each platform's public launch through December 2024. Specifically, the data begin in July 2021 for Gettr (its launch month) and in February 2022 for Truth Social (following its release), ensuring complete coverage from inception. I assume that content moderation is implemented platform-wide and collect data from the publicly accessible web interfaces of each platform. Of course, if changes in moderation were implemented only for Android-based app users (e.g., in response to Google Play policy), such changes would not appear in my dataset, an important consideration when interpreting platform-wide effects.

To ensure comparability across platforms, I apply a consistent sampling frame. I first compile extensive lists of known user accounts for each platform, then randomly

[21] Coldewey, D. (2017, August 17). *Alt-social network Gab booted from Google Play Store for hate speech.* TechCrunch.

[22] See Gab 2020 SEC Annual Report Filing

sample users who were active in both the pre- and post-policy periods. This ensures that the final samples focus on users with sustained engagement, making behaviour comparable across platforms and over time. Table A.3 provides detailed summary statistics for the samples used, including user counts, post counts, timelines of data collection, and descriptive statistics for engagement measures. The estimates show broadly similar post-level characteristics across platforms in the pre-treatment period. Average threat scores and incidence of key topic flags (e.g., election fraud, anti-vaccine, QAnon) are of comparable magnitude, suggesting that the baseline distribution of harmful or conspiratorial content is relatively consistent. One notable exception is the substantially higher engagement on Truth Social which likely reflects dynamics during the platform's onboarding phase. This may be partly organic, given concentrated early activity among prominent users, but could also reflect developer-side amplification or algorithmic design aimed at boosting visible engagement, a pattern observed historically during the growth phases of some social platforms.[23]

For each platform, I begin with a pool of approximately 2 million known user accounts from which I randomly sample users. These user lists are assembled by seeding with a randomly selected user and recursively collecting all followers and followings, continuing the process until the network expansion yields no new users for an extended period. This approach produces a near-complete snapshot of the active user network at the time of collection, excluding only those accounts that were closed, suspended, or otherwise inaccessible. I then randomly select 20,000 users, from each platform and maintain the users with English language content from before and after July 27, 2022. The final sample is 16,449 users on Gettr, 16,891 on Truth Social, and 9,155 from Gab.

In a robustness exercise, for the Gettr platform, I identify a subsample of US based users who list their location at the county level in their profiles resulting in 14,430 users. I also create a secondary dataset of 14,639 known Brazilian users. These two samples allow me to explore country specific contemporaneous shocks which may confound estimates and assess for platform led moderation of various sensitive events which are geo-specific.

In addition to the raw counts of posts, comments, and replies, I collect metadata on each user's likes per post, repost frequencies, and other engagement behaviours where available. These data enrich the analysis by enabling investigation not only of content production but also of patterns of user engagement and amplification across platforms. I omit reporting or including user's follower or following counts in any analysis, as I do not have data on pre-treatment counts of followers and any post treatment following count is surely endogenous to content moderation.

[23]See Thomas Barrabi, "Meta exec's frantic warning about Instagram's alarming 'fake' activity numbers in spotlight at FTC trial," *New York Post*, May 14, 2025. Available at: `https://nypost.com/2025/05/14/business/meta-executive-warned-fake-engagement-on-instagram-app-could-be-in-range-of-40-court-docs/`.

3.2 Measurement of Threatening Content using the Perspective API

To evaluate the extent to which user-generated content contains threatening language, I use Google's Perspective API (Wulczyn et al., 2017; Dixon et al., 2018). The API produces a machine-learning-based prediction for several dimensions of perceived harmfulness. My primary measure is the API's "threat" score, which ranges from 0 (not threatening) to 1 (most threatening). According to the API documentation, threatening language is defined as "an intention to inflict pain, injury, or violence against an individual or group." I provide examples along the continuum of threat scores in Table A.4 for reference.

Each post and comment from Truth Social, Gettr, and Gab is scored consistently using the same API version, classification settings, and post-processing pipeline to ensure comparability across platforms and over time.

In addition to the threat score, the Perspective API returns related scores including toxicity, severe toxicity, identity attack, insult, and profanity.[24] Although the analysis focuses primarily on the threat score, I report robustness checks using these alternative dimensions.

I repeat the exercise on the alternate Brazilian sample using the Portuguese version of perspectives API.

3.3 Measurement of Discourse around Sensitive Events

To examine the prevalence of "sensitive-events" in user content, I focus on five politically salient topics frequently implicated in content moderation debates: 1) January 6 commentary, 2) election fraud claims, 3) anti-vaccine discourse, 4) conspiratorial narratives, and 5) climate change skepticism.[25]

The five topics were selected because they align with the policy's definition of sensitive events and reflect issue areas that have triggered scrutiny by both economic researchers and platform governance bodies (Guess et al., 2020; Berlinski et al., 2023; Aridor et al., 2024).

Content was classified using a rule-based dictionary method. For each topic, I constructed a hand-curated list of regular expressions designed to match distinctive phrases and terms representative of the discourse. The dictionaries draw from domain expertise, prior literature, and manual inspection of platform-specific vernacular. Each post was then flagged with a binary indicator for each topic based on whether any of its associated patterns were detected. This approach enables high-recall classification of topic-specific discourse at scale while maintaining transparency and interpretability in the construction of labels.

[24] See https://developers.perspectiveapi.com/s/about-the-api-attributes-and-languages?language=en_US for documentation on all available attributes.

[25] Definitions of each category are provided in Table C.2.

I aggregate three of these categories: election fraud, anti-vaccine content, and QAnon-related discourse, into a single composite indicator for conspiratorial narratives:

$$\texttt{conspiracies}_i = \mathbb{1}\{\texttt{election_fraud}_i = 1 \vee \texttt{anti_vax}_i = 1 \vee \texttt{qanon}_i = 1\}$$

Sample details and descriptive statistics on the distribution of threat and topics scores from before the policy change are provided for each platform in Table A.3.

4 The Effect of Google's Policy on Platform Content Moderation

This section begins by addressing the identification challenges in estimating the causal effect of Google's policy change on platform content moderation. I then outline the empirical strategy and present within-platform results, robustness checks, and causal estimates.

4.1 Identification

Google's July 27, 2022 Play Store bulletin provides a unique quasi-experimental lever to study whether upstream distributors can compel downstream platforms to moderate speech. At the moment of the announcement, the three largest "Alt-Tech" social platforms occupied markedly different positions in the Android distribution chain: Gettr was already listed; Truth Social had a pending application whose approval now hinged on compliance; and Gab, having been permanently delisted in 2017, remained entirely outside the Play Store. As a result, the announcement triggered a sharp, platform-specific change in regulatory exposure that was orthogonal to the behaviour of any individual user. While Gettr retained its listing and Truth Social was eventually admitted, whether either platform actually altered its moderation practices remains an empirical question.

To identify the causal effect of the Google policy on moderation, measured via changes in users threatening content, I implement a triple-differences (DDD) design. This approach leverages variation along three dimensions: between users (those with high vs. low pre-policy threat tendencies), over time (before vs. after the policy), and across platforms (those affected by the policy vs. Gab, which was unaffected). This design isolates the policy's effect on users exposed to it through their activity on covered platforms, relative to comparable users on an untreated platform. The validity of this strategy rests on standard DDD assumptions, including parallel trends across treatment groups in the absence of intervention, and no behavioural response prior to the announcement (no anticipation), which I address by way of event study analyses both within and between platforms. In addition, several identification threats

complicate interpretation. These include: mechanical threats comprising incomplete exposure to treatment and data loss due to retroactive content takedown; selection and attrition; contemporaneous shocks; user selection into platforms; voluntary user self-censorship; and spillovers across platforms.

Mechanical threats. Although the Google Play Store policy applied specifically to Android apps, user content is accessible across multiple interfaces, including web and iOS, where platform incentives to moderate may differ. If developers implemented content moderation only for Android clients while allowing more permissive content on web or iOS, then threat levels observed through web scraping could understate the true effect. However, such selective moderation is technically complex, resource-intensive, and uncommon in practice, as it requires maintaining divergent infrastructure across platforms. Moreover, platform moderators can retroactively remove policy-violating posts from the global content pool. Since the web crawler captures only surviving posts from the web based version of the platform, the estimated decline in threatening content is a conservative, lower-bound measure of the platform's full moderation response.

Selection and attrition. A key identification concern is non-random user selection around the time of treatment. The main specification includes only users active both before and after the Google policy change, which risks selection into the analysis panel. If users with higher pre-treatment threat scores are more likely to remain, the post-policy sample will overrepresent high-threat users, potentially attenuating moderation effects by limiting scope for observed declines. Conversely, if higher-threat users disproportionately exit before the policy, the remaining sample will underrepresent those most likely to be moderated, biasing estimates upward.

Second, even conditional on inclusion, there may be selective attrition if users with different threat levels exit at different rates after the policy. If higher-threat users disproportionately leave after treatment, the observed decline in threatening content may partly reflect compositional shifts rather than platform enforcement or remaining behavioural change.

To directly assess both threats, Table A.5 presents two sets of results. For each platform, I make use of the full random sample, which includes users who posted before and after the policy shift, as well as users who presented only before. Columns (1)–(6) show that users with higher pre-treatment threat scores are significantly more likely to remain in the post-policy panel on all three platforms. The same is replicated for Gab in Table A.6 This selection is strongest on Gab, moderate on Truth Social, and smallest on Gettr, but consistently positive and significant. Because these users are more likely to post threatening content, their overrepresentation in the post period may exaggerate the estimated moderation effect, biasing the coefficient downward (i.e., making it more negative than the true effect).

Columns (7)–(10) then assess how long users remain active after the policy. On Gettr, higher-threat users tend to exit sooner, suggesting either enforcement targeting

or voluntary disengagement. On Truth Social and Gab, by contrast, higher-threat users remain active longer, which might be interpreted as limited banning or greater persistence among the most extreme users. Sample and post-treatment heterogeneity in attrition across platforms implies the need to test whether sample selection and attrition distort the main estimates.

I implement two such robustness checks in Section 4.3.2. First, I re-estimate the main specification using a balanced panel of users observed throughout the entire period (i.e., those present at endline) and then test whether estimates are stable across rolling monthly user exit bins. Second, I compare estimates from the full sample including users not present after treatment to those in the full panel, excluding user fixed effects to retain entrants. In both cases, I find no meaningful differences in estimated treatment effects.

Contemporaneous Shocks. Another central challenge, given the design is the possibility that concurrent events, such as political developments, parallel policy changes (e.g., from Apple), or broader competitive pressures within the platform ecosystem, may have occurred around the same time as Google's announcement, confounding the interpretation of any observed content moderation. In particular, changes in threatening content might reflect a general shift in online discourse rather than a response specific to Google's distribution policy. My empirical strategy addresses this concern in three ways.

First, I use Gab as an untreated counterfactual. Gab shares the same language environment and U.S.-oriented user base as the treated platforms but was entirely excluded from the Google Play Store. Any broad shocks to online speech, such as national political events or shifts in discourse norms, should plausibly affect Gab as well. To validate this assumption, I assess both behavioural trends and linguistic similarity across platforms using pre-treatment event studies and platform-level comparisons. To evaluate content similarity, I extract weekly n-gram frequencies by platform and compute pairwise similarity metrics (Jaccard and cosine) to track convergence or divergence in language use over time. As shown in Figure B.2, Gab and Gettr exhibit consistently high pre-treatment similarity (cosine $\approx$ 0.5–0.6), while Truth Social remains linguistically distinct from both, likely reflecting differences in platform culture and agenda-setting. Around the policy announcement, both the Gab–Gettr and Gettr–Truth Social pairs show a sharp, short-lived drop in similarity, consistent with a sudden divergence in content likely driven by platform-specific moderation responses. Post-treatment, similarity gradually increases, with Truth Social in particular converging toward Gab and Gettr in the run-up to the 2024 U.S. presidential election, suggesting a re-alignment or normalisation of discourse over time. Taken together, these patterns reinforce the use of Gab as a valid counterfactual and reduce concern that observed changes on the treated platforms are driven by idiosyncratic or coincidental shifts in language or topic salience.

Second, the event-study design allows for a precise assessment of the timing of content changes relative to the Google policy announcement. By estimating dynamic effects across the full study period, the design enables the detection of both anticipatory patterns and delayed responses, while also providing a built-in placebo for events that occur at other points in time. I use this framework to assess three high-salience events that might plausibly influence threatening content: (1) the FBI's August 8 raid on Donald Trump's residence and the August 19 congressional letters sent to platform CEOs; (2) Elon Musk's acquisition of Twitter in late October; and (3) the U.S. midterm elections, which unfolded over several months. In addition, Table C.3 documents Apple App Store policy changes throughout the year. The only potentially relevant update, a revision on October 25 to the "objectionable content" clause, occurred well after the period of interest and closely mirrors Google's earlier policy. This timing is consistent with a leader–follower dynamic in distributor enforcement and is further ruled out by placebo tests showing no discontinuity around the Apple update.

Third, I conduct a cross-country falsification test using a sample of 14,639 Portuguese-speaking users in Brazil on Gettr. These users operate within a distinct political and linguistic environment, largely insulated from U.S.-specific news cycles and cultural cues. If the observed changes were driven by U.S. political shocks or broader shifts in English-language discourse, we would not expect to see similar patterns among Brazilian users. By testing whether any discontinuity occurs around the Google policy announcement within this independent context, the analysis provides a strong check on the language- and country-specific validity of the identification strategy.

Selection into platforms. A further concern is that user selection into platforms may interact with the DDD structure in a way that affects interpretation. In particular, if users with higher pre-treatment threat levels disproportionately sort into treated or untreated platforms, the second and third differences (user-level threat and platform exposure) may not be orthogonal. This would threaten the assumption that differences in threat evolution across platforms can be attributed to treatment rather than underlying user composition. This concern is mitigated by the fact that the DDD design incorporates user fixed effects, ensuring that within-user changes are isolated. Since the identifying variation comes from how users with different pre-treatment threat intensities adjust over time within each platform, and how these changes differ across platforms, baseline differences in user composition do not mechanically bias the results. Nonetheless, to assess the plausibility of this assumption, I examine the distribution of pre-policy threat scores across platforms. All platforms contain substantial variation in user-level threat scores, with broadly overlapping distributions in both raw and standardised scores (Figure B.3). Supporting this interpretation is the high degree of pre-treatment cosine similarity in linguistic content across platforms (see previous subsection). As an additional check, I replicate the DDD results using matched subsamples with equivalent threat-score distributions across platforms (see robustness checks

in Section 4.4).

User self-censorship. Further, attributing any observed change in content to platform moderation requires ruling out the possibility that users adjusted their behaviour in anticipation of increased enforcement, independently of any formal platform response. Distinguishing between platform-initiated and user-initiated changes is difficult without access to internal moderation logs, but behavioural shifts by users are nonetheless a function of platform incentive structures. With this in mind, I expect that user behaviour may gradually adapt in response to enforcement, especially in the longer term, as users internalise new platform constraints. Such adaptation is likely to shape longer-run dynamics and is examined separately in the user behaviour analysis. However, the key threat to identification is anticipatory self-censorship that precedes or coincides with the treatment, which could confound estimates of platform-led moderation. While it is theoretically possible that some users engaged in self-moderation, either voluntarily or in response to implicit signals, this explanation is unlikely for several reasons.

First, users would need to have known that stricter enforcement was forthcoming. This would require either active monitoring of Google's Developer Program policy, exposure to media coverage, or direct platform communication. Yet, a search of mainstream and technology-focused media sources, including blogs, using Google News and newspapers.com yields no evidence of coverage of the July 27 policy update. While a few articles mention contemporaneous changes to app policies, none refer to the "sensitive events" clause or required platform moderation at all. Google Trends data likewise show no spike in public interest in content moderation around the update date, but instead reveal increased search activity only in late October, when Elon Musk's takeover of Twitter sparked broader public debate over content moderation (see Figure B.1).

Second, the platforms in question explicitly marketed themselves as "free speech" alternatives to mainstream platforms, making it unlikely that they would signal to users either explicitly or implicitly that stricter content controls were imminent. While covert enforcement mechanisms like 'shadow-banning' are possible, there is no evidence of public-facing campaigns urging users to moderate their own content. I compare the public moderation policies of Truth Social and Gettr before and after July 27 using the Internet Archive's Wayback Machine,[26] and find no documented changes in either case. Third, I analyse user-level posts for any rise in discussions around moderation during this period and find no immediate increase in references to moderation or enforcement in user discourse.

I empirically assess the plausibility of self-censorship in several ways. First, the event study design allows me to examine whether content began declining prior to the policy date, which would indicate anticipatory self-moderation. This approach is particularly relevant to the Truth Social case. Since there were no public announce-

[26] https://archive.org/

ments regarding Truth Social's pre-approval violation of Google policies until August 19, 2022, nor to the best of my knowledge any media coverage prior to this date, I can check for content adjustments which preceded public communication at a granular (i.e. daily) level. If users were responding to platform signals rather than direct enforcement, any shift in content prior to August 19 would be difficult to attribute to user self-censorship, especially considering non of these users are Android users before at least October 22, 2022, when Truth Social was admitted to the Google Play store.

Second, I assess the distribution of changes in content intensity across users, examining whether reductions are concentrated among the most threatening users or distributed over all users. This approach is particularly useful to assess adjustments in the immediate period after the policy announcement. Third, I leverage cross-country variation: if users were moderating in response to Google's policy, effects should be larger in markets where Android dominates. I compare the magnitude and timing of responses between the U.S. (where Android usage is around 45%) and Brazil (where it exceeds 90%).

Spillovers. Lastly, there must be no spillovers across platforms. Moderation activity on Google-exposed platforms should not influence user behaviour on Gab. Spillovers could occur if users migrate to Gab following moderation, or if users active on multiple platforms adjust their content on Gab in response to moderation received elsewhere. I test for spillovers utilising a sub sample of users known to be present on both Gettr and Gab.

4.2 Empirical Strategy

My empirical approach begins by focusing within platforms to examine whether there is evidence of moderation following the Google policy signal. Specifically, I assess changes in the level of threatening content posted by users who, in the pre-announcement period, exhibited particularly high levels of such content. I compare content adjustments as a function of users pre-treatment 'threateningness' on each platform before and after Google's announcement. If moderation occurred, either by the platform or through user self-censorship, we would expect to see a decline in the average threat level of posts from higher threat users.

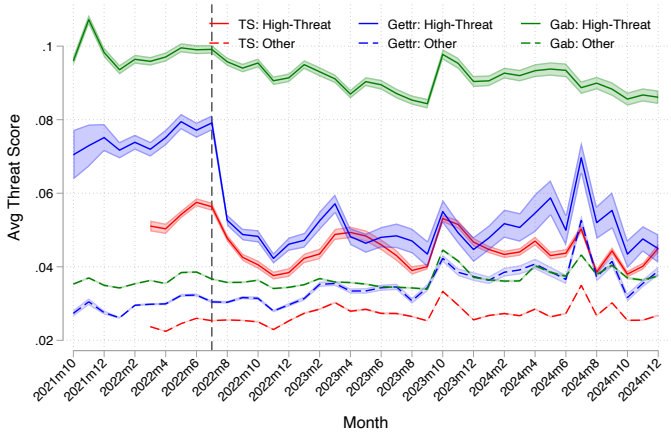
I begin by plotting the evolution of average post threat scores, comparing users whose pre-announcement posts were above the 90th percentile of the threat score distribution ("High-Threat") to all other users. Figure 4.1 shows that, prior to the announcement, threatening content on each platform evolved in parallel across user groups. More importantly, similar groups evolved in parallel between platforms.

Following the announcement, however, average threat scores among high-threat users decline sharply on both Gettr and Truth Social after the July 27 Google announcement, continuing to fall for another month before returning to parallel trends with other users on each platform around September 30. Truth Social starts from a

lower pre-treatment average and exhibits a more modest decline. It also shows periods of rising threat scores that move in parallel with Gettr, though they never return to pre-treatment levels. Unlike the divergence seen between July and September 2022, these post-treatment movements consistently track those of low-threat users.

In absolute terms, the post-treatment mean for high-threat users converges to a statistically similar level across both Google-exposed platforms, suggesting a common moderation threshold required for compliance. Among non-high-threat users, there is an overall upward trend in threatening content during the same period. Finally, no comparable reduction is observed among high-threat users on Gab. While there is a slight, gradual convergence toward the threat levels of other users, there is no sharp decline at any point in the series. This is the first indication that the changes seen on Gettr and Truth Social are attributable to Google's policy shift.

Figure 4.1: Threat score of users posts across platforms before and after Google announcement



Notes: This figure plots the average threat score of posts across Gettr, Truth Social and Gab. I split users within platform based on the threat score of all of that platforms users posts in the pre-exposure period. "High-Threat" users are those whose posts before the Google announcement were on average above the 90th percentile of the threat score distribution by platform. "Other" are all remaining users. The shaded areas indicate 95% confidence intervals for the means.

To formally test for a shift in threat content within the covered platforms, I implement a difference-in-differences design that exploits variation across users in their pre-announcement threat levels and across time relative to the policy signal. The empirical specification is:

$$Y_{pit} = \beta(Post_t \times \overline{Threat}_i) + \alpha_i + \delta_t + \epsilon_{it} \tag{1}$$

Here, Y_{pit} denotes the threat score of post p by user i on date t; $Post_t$ is an indicator equal to 1 for dates after July 27, 2022 (the public announcement), and 0 otherwise; $\overline{Threat}_i$ is the user's average threat score in the pre-announcement period, standardised within platform to allow effect sizes to be interpreted in standard deviation units and compared across platforms; and α_i and δ_t are user and year-day fixed effects,

respectively.

To capture the average treatment effect on the treated (ATT), I extend the design to a triple-differences framework, comparing shifts in threatening content by pre-treatment user threat scores between covered platforms (Gettr and Truth Social) and the uncovered platform (Gab).

$$\begin{aligned} Y_{pit} = \beta &\left(Post_t \times \overline{Threat}_i \times Exposed_i\right) \\ &+ \gamma_1 \left(Post_t \times \overline{Threat}_i\right) \\ &+ \gamma_2 \left(Post_t \times Exposed_i\right) \\ &+ \gamma_3 \left(\overline{Threat}_i \times Exposed_i\right) \\ &+ \alpha_i + \delta_t + \epsilon_{it} \end{aligned} \tag{2}$$

Where, $Exposed_i$ is a binary variable equal to 1 for Gettr and Truth Social, and 0 for Gab. In the most restrictive version of this specification I include user-platform fixed effects, and platform-week fixed effects.

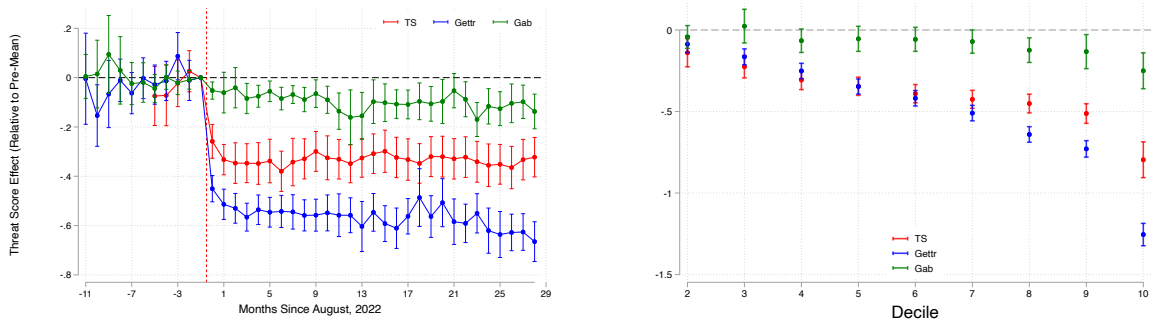
4.3 Results

I begin with a dynamic version of Equation 1, in Figure 4.2a, which replaces the Post indicator variable with dummies for the months around the Google announcement. This specification enables assessment of evidence of within platform parallel trends between users, the timing of any observed changes in content and the direct comparison between platforms and provides compelling evidence that the observed decline in threatening content on Gettr and Truth Social was a direct response to the Google policy update. The figure indicates that before Google's announcement, users with high and low pre-treatment threat scores followed similar trends in how threatening their posts were across all three platforms. Immediately following the announcement, 'threateningness' dropped sharply and remained consistently below its pre-announcement level for the entire study period on both Truth Social and Gettr as a function of user pre-treatment 'threateningness', but not on Gab. For each one standard deviation in increase in pre-treatment 'threateningness', the likelihood of a post being classified as threatening after July 27 declines by between 50% - 60% on Gettr and 30% - 40% on Truth Social, relative to the pre-treatment mean on each platform respectively. There is a relatively small and marginally significant decline in threatening content on Gab over the longer term but this does not begin until four months after the Google announcement and suggests trend rather than moderation. Importantly, over the entire period, there are no other sharp adjustments following the policy shift, indicating that this was a unique and temporally isolated shock to platform moderation, with no comparable shifts in threatening content occurring later in the study period. Weekly versions of these event

studies in Figure B.6 highlight the temporal proximity of the reduction in threatening content to the policy shift on Google exposed platforms. Within one week of the announcement, both platforms see a sharp, statistically significant and persistent decline compared to the week before the announcement. This highlights how quickly these platforms are able to moderate when incentivised to do so and rules out potential confounders, which occur later than July 27, driving the reduction in covered content.

In Figure B.5a, I estimate a flexible specification that interacts each decile of the pre-treatment threat score distribution with the post-Google-announcement indicator, using users in the lowest decile (bottom 10th percentile) as the reference group. The results show that the effect of the announcement increases monotonically with user 'threateningness' for both Gettr and Truth Social. Relative to users in the lowest decile, users in the 10th percentile (i.e., average pre-treatment content threat scores between [5.5–66] on Gettr and [4–55] on Truth Social) experience a reduction in threat score of 0.043 and 0.022, respectively, corresponding to decreases of approximately 125% and 80% relative to each platform's pre-announcement mean (see Figure B.4 for level estimates). On Gab there is no monotonicity in the effect: there is an economically small and roughly constant, marginally significant decline across deciles until the highest decile, which exhibits a decline of just under 25% of the pre-treatment mean.

Figure 4.2: Effect of Google Announcement on User Threatening Content Across Platforms



(a) Event Studies

(b) Decile Regressions

Notes: Panel a) shows platform-specific regressions of user post threat scores on the interaction of pre-treatment user average standardised threat scores with a set of month indicators relative to Google's announced policy change. The omitted category is July 2022. The dependent variable is scaled relative to each platform's pre-treatment mean and can be interpreted as proportional changes compared to the mean. Panel b) shows platform-specific regressions of the change in users' daily threat scores compared to the platform specific pre-treatment mean on a set of indicators for users' cross-sectional pre-treatment decile of average threat score. The omitted category is the first decile. In both panels, regressions include user control variables interacted with year-day fixed effects, user fixed effects, and year-day fixed effects. Standard errors are clustered at the user level. *** p<0.01, ** p<0.05, * p<0.10.

In Figure B.4b, I estimate the same specification but restrict the post-treatment period to the first two months (i.e., August–September), capturing only the immediate platform response to the Google policy. The results again show a clear monotonic

pattern on the Google-exposed platforms, with the magnitude of the effect slightly smaller than in the full-period estimates. In contrast, Gab shows no such monotonicity, and the overall magnitude remains small and flat across deciles.

This short-term pattern is difficult to reconcile with anticipatory self-censorship, which would likely produce a more diffuse reduction across user types, if anything, affecting lower- and mid-threat users more uniformly. Instead, the disproportionate decline concentrated among high-threat users strongly suggests targeted platform enforcement, consistent with either manual or automated moderation directed at policy-violating content. The absence of a comparable response on Gab further supports the interpretation that this initial moderation response is driven by platform behaviour, not user-led adjustment.

I now turn to the benchmark difference-in-differences estimates from Equation 1. Each column of Table 4.1 presents the estimated effect of a one-standard-deviation increase in a user's pre-announcement threat score on the average 'threateningness' of their visible (i.e., unremoved) content after Google's announcement. The left panel are estimates for Gettr, while the right are for Truth Social. In each panel, column (1) includes user fixed effects, column (2) adds year-day fixed effects, column (3) adds user specific pre-treatment level controls $\times$ year-day fixed effects, and column (4) adds a user specific linear trend, which relaxes the assumption that users with different pre-treatment threat average scores follow parallel trends. All specifications indicate a significant reduction in the 'threateningness' of posts on Google exposed platforms after the July 27 announcement. The results are very stable over all specifications even with the inclusion of user-specific linear trends. The estimates imply that the announcement was associated with a reduction in average threat scores of approximately 52% on Gettr and 32% on Truth Social for users who were one standard deviation above the mean in pre-announcement threat levels, consistent with increased moderation targeting high-risk users in response to the policy change.

Estimates for Gab in Table 4.2 highlight a small (<4% compared to pre-treatment mean) and significant overall average decline in threatening content on Gab as a function of user 'threateningness' after the Google announcement. Results are noticeably smaller than other estimates and less robust to the inclusion of user trends than Google distributed app estimates.

Table 4.1: Google Announcement and Online User Threatening Content

	Gettr				Truth Social			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\overline{Threat}$	-0.0184***	-0.0181***	-0.0179***	-0.0174***	-0.0084***	-0.0082***	-0.0083***	-0.0083***
× Post	(0.0005)	(0.0005)	(0.0004)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
N	2,968,879	2,968,879	2,968,879	2,968,879	11466215	11466215	11466215	11466215
User FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Day FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
User Control × Time FE	No	No	Yes	Yes	No	No	Yes	Yes
User Time Trend	No	No	No	Yes	No	No	No	Yes
# of Users	25952	25952	25952	25952	13515	13515	13515	13515
DV Pre-Mean	0.0336	0.0336	0.0336	0.0336	0.0265	0.0265	0.0265	0.0265
Treatment SD	0.0819	0.0819	0.0819	0.0819	0.0657	0.0657	0.0657	0.0657

Note: The table presents the results of estimating Equation 1. The dependent variable is the threat score of the users posts on platforms Gettr and Truth Social. $\overline{Threat}_i$ is the pre-treatment average threat score of user i's posts. User post threat score is standardised within platform. $Post_t$ is an indicator variable = 1 following Google's announcement on July 27, 2022, and 0 otherwise. The level of observation is user × year-day. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

Table 4.2: Google Announcement and Online User Threatening Content (Gab)

	Gab			
	(1)	(2)	(3)	(4)
$\overline{Threat}$	-0.0036***	-0.0035***	-0.0035***	-0.0018**
× Post	(0.0011)	(0.0011)	(0.0011)	(0.0009)
N	13204938	13204938	13204938	13204938
User FE	Yes	Yes	Yes	Yes
Year-Day FE	No	Yes	Yes	Yes
User Controls × Time FE	No	No	Yes	Yes
User Time Trend	No	No	No	Yes
# of Users	7374	7374	7374	7374
DV Pre-Mean	0.0448	0.0448	0.0448	0.0448
Treatment SD	0.0979	0.0979	0.0979	0.0979

Note: The table presents the results of estimating Equation 1. The dependent variable is the threat score of the users posts on Gab. $\overline{Threat}_i$ is the pre-treatment average threat score of user i's posts. User post threat score is standardised within platform. $Post_t$ is an indicator variable = 1 following Google's announcement on July 27, 2022, and 0 otherwise. The level of observation is user × year-day. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

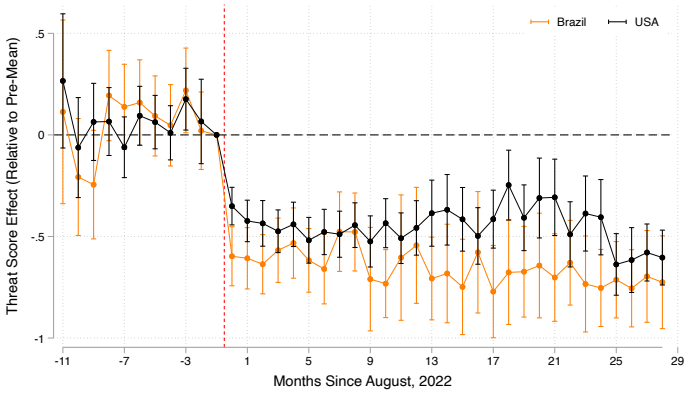
4.3.1 Cross-Country Falsification: Brazil versus the United States

A central identification threat is that some contemporaneous, U.S.-specific shock, for example, the FBI search of Donald Trump's Mar-a-Lago home on August 8, 2022, rather than the Google bulletin, drove the post–July 27 decline in threatening content.

A related concern is that the observed effect could reflect Android-only self-censorship: if U.S. Android users independently toned down their rhetoric while iOS and desktop users did not, the average drop might be a compositional artefact rather than true moderation. The fact that we observe any sharp discontinuity suggests that moderation occurs server-side rather than within a specific channel build (e.g., Android only), but it is still possible that some Android users responded independently to the policy shift.

Event study estimates in Figure 4.3 compare threatening content between the U.S. and Brazilian samples. The results highlight three key points: (1) the estimates for the U.S. sample are similar in both magnitude and timing to those in the full English-speaking sample, suggesting that the main results are unlikely to be driven by U.S.-specific dynamics; (2) this interpretation is reinforced by the Brazilian estimates, which track closely with the U.S. series and thus rule out contemporaneous U.S. political or social shocks as primary drivers; and (3) Brazilian users, who disproportionately access these platforms via Android, exhibit no statistically significant divergence from U.S. users, which suggests an Android-specific self-censorship mechanism is unlikely.

Figure 4.3: Effect of Google Announcement on User Threatening Content on Gettr Across Countries



Notes: This figure plots regressions of country specific user post threat scores on Gettr on the interaction of pre-treatment user average threat scores with a set of month indicators relative to Google's announced policy change. The omitted category is July 2022. The dependent variable is scaled relative to each countries pre-treatment mean and can be interpreted as proportional changes compared to the mean. The bars indicate 95% confidence intervals.

4.3.2 Robustness

I conduct a range of robustness checks to assess the credibility of the benchmark results. Each set of tests examines a distinct modelling or measurement choice to ensure the estimated effect is not an artefact of a specific specification.

In Appendix Table A.7, I re-estimate the main specification from Equation 1 with alternative codings of the dependent and treatment variables. Column (1) uses the log of the threat score rather than the level to account for skewness. Columns (2)–(4) replace the continuous measure of pre-treatment 'threateningness' with categorical bins: quartiles in column (2), a top quartile dummy in column (3), and a top decile dummy

in column (4). These tests assess whether the effect is concentrated among the most extreme users or more broadly distributed. Across all variants, I continue to find a significant post-policy decline in threatening content on platforms exposed to Google's enforcement. Appendix Table A.8 probes the sensitivity of results to different thresholds for defining high-threat users. Instead of quartiles or deciles, I use a range of cutoff values based on pre-policy average threat scores, including more stringent definitions of low-threat control groups. The negative post-policy shift remains robust across all cutoffs. Appendix Table A.9 tests whether results hold when exposure is defined using a user's maximum pre-policy threat score, rather than their average. I implement this both as a continuous variable and through dummies at various thresholds (e.g., top 10%, top 20%). This alternative characterisation of user "extremity" yields patterns consistent with the baseline, including strong effects even when the control group is restricted to users with very low threat scores.

In Appendix Table A.10, I assess whether effects vary by content type. I separately estimate treatment effects for original posts (column 1), replies (column 2), and reposts (column 3). The largest declines are observed for original posts, followed by replies, with smaller but still significant effects for reposts.

Next, Appendix Table A.11 examines whether the observed effects extend to other dimensions of harmful speech using alternative scores from the Perspective API, specifically, toxicity, severe toxicity, identity attack, insult, and profanity. On Gettr, I observe consistent and significant declines across all measures. On Truth Social, results are similarly negative for most categories, though effects for severe toxicity and profanity are smaller and statistically insignificant, likely reflecting the platform's pre-existing restrictions on such content. Finally, I examine an independent hate speech measure derived from a fine-tuned BERT model (Aluru et al., 2020) in Appendix Figure B.5. I again find a sharp drop in hate speech beginning after July 27 on the two platforms exposed to Google's policy, with no comparable change on Gab.

Lastly, to assess the consequences of selective attrition and panel sample selection for the estimated treatment effects, I conduct three sets of robustness checks. First, reported in Table A.12, columns (1)–(6) test whether differential attrition after treatment biases the results. These specifications splits the samples into i) users who remain active through the end of the study period, eliminating variation in exposure duration, and ii) users who exit before the last month of the sample period in December 2024, capturing those with shorter observed engagement. I compare estimates from the full sample (columns 1 and 4) to those from the balanced "endline" panel (columns 2 and 5) and those from the attrited sample (columns 3 and 6). I find no meaningful differences in the post-treatment interaction effect, suggesting that differential dropout after treatment does not drive the main results.

In addition, in Figure B.7, I plot the estimated post-treatment interaction effect separately for users exiting before each quarterly cutoff, as well as within non-overlapping

quarterly exit cohorts. The top row shows estimates for cumulative exit samples, while the bottom row reports results for disjoint quarterly exit buckets. For both Truth Social and Gettr, the treatment effects remain consistently negative and relatively stable across cohorts. In contrast, the estimates for Gab are smaller in magnitude and either marginally significant or statistically indistinguishable from zero. The absence of clear temporal divergence in any of the platforms, and the stability of the effect for those subject to Google's policy, provides further support that selective user attrition does not account for the main findings.

Columns (7)–(10) of Table A.12 then test for selection into the panel, relaxing the requirement that users be active both before and after the policy change. These specifications omit user fixed effects to accommodate those exiting before treatment and compare average treatment effects in the full sample (columns 7 and 9) to those in the subsample of "stayers", who are active users on both sides of the treatment (columns 8 and 10). Again, the estimated effects are similar in size and precision, indicating that the results are not being driven by selection into the panel based on pre-treatment threat. Taken together, these findings suggest that the observed decline in threatening content is not primarily due to compositional changes in the user base.

To benchmark these results, I replicate all robustness tests on Gab (Appendix Table A.13 through Table A.17), where the Google policy does not apply. Across specifications, the estimated effects are consistently small, statistically insignificant, or weakly signed, reinforcing the interpretation that the observed shifts on Gettr and Truth Social were driven by policy exposure rather than broader trends. Assessment of the consequences of selective attrition and panel sample selection in Table A.18 suggest no selection or attrition bias on Gab.

4.3.3 Other External Events

The event studies provide strong evidence that Google's July 27, 2022 policy announcement triggered a platform-specific reduction in threatening content on Truth Social and Gettr. The patterns are sharp, temporally aligned with the announcement, and absent on Gab, supporting both the parallel trends assumption and the exclusion of major confounders. Nevertheless, before continuing to the causal estimation, it is worth considering whether any external events could have selectively influenced threatening content on the Google-exposed platforms but not Gab. For an alternative explanation to threaten identification, it would need to (1) disproportionately affect Gettr and Truth Social, (2) begin precisely around July 27, (3) induce a discontinuity in threat scores among high-risk users, independent of the Google policy, and (4) affect users in English speaking countries and Brazil equally. I discuss some possibilities and explain why none plausibly account for the results.

Twitter Takeover. Elon Musk's acquisition of Twitter in late October 2022 has been cited as a potential driver of migration from Alt-Tech platforms, possibly lowering

threat levels via selective user exit.[27][28] However, the timing is inconsistent with the results. The largest and only discontinuity in threatening content appears in late July, not October. Daily and weekly threat scores show no structural breaks around the Twitter acquisition, and Gab again shows no change, ruling out a sector-wide shift. If extreme users had exited en masse following the acquisition, a second drop would be visible in October, yet none appears. This is further supported by the robustness of main results to specifications which test the effects of selective attrition in the previous section.

U.S. Political Events: Mar-a-Lago Raid. One potential confound is the FBI search of Donald Trump's residence on August 8, 2022, which triggered a surge in threatening rhetoric and prompted a congressional inquiry into platform moderation on August 19.[29] However, several facts make this explanation implausible. Most importantly, results from the Portuguese-language Brazil sample show a nearly identical drop in threatening content beginning on July 27, despite the event being neither politically or linguistically relevant in that context. Topic modelling confirms that the event was salient in user discourse on all three platforms in English but not the Brazilian context on Gettr. Moreover, the timing is inconsistent: the observed decline begins precisely on July 27 and is largely complete by mid-August, ruling out the possibility of delayed effects from the Mar-a-Lago raid.

4.4 Causal Estimates

Having established economically large, consistent and robust estimates for within firm reduction in 'threateningness' of visible content, and having ruled out mechanical and contemporaneous threats to identification, I now turn to the causal estimates for Google's policy on platform side moderation of user content. I begin by assessing evidence of third difference parallel trends and ruling out spillovers onto the counterfactual platform Gab. I then turn to estimation and robustness.

To assess the assumption of parallel trends and any anticipation effects, I provide dynamic DDD estimates in Figure B.8. The first thing to note is that there is no statistically significant difference in high and low-threat users between platforms before the policy shift. While there is a slight upward pre-trend, its direction suggests that the estimated ATT may represent a conservative (lower-bound) estimate of the policy's effect on threatening content. Secondly, there is a statistically significant difference in

[27]Kate Conger and Ryan Mac, "How Elon Musk Changed the Meaning of Twitter for Users," New York Times, October 27, 2023, `https://www.nytimes.com/2023/10/27/technology/elon-musk-twitter-year.html`.

[28]Sheila Dang, "Twitter is losing its most active users, internal documents show," Reuters, October 26, 2022, `https://www.reuters.com/technology/exclusive-where-did-tweeters-go-twitter-is-losing-its-most-active-users-internal-2022-10-25/`.

[29]Drew Harwell and Cat Zakrzewski, "Lawmakers Demand Data about Online Threats against Law Enforcement," *Washington Post*, August 19, 2022. See Figure C.4 for a copy of the letter sent to CEOs.

the adjustment in platform level exposure (user pre-treatment threat score) between the Google based apps and Gab beginning at the time of the Google policy shift and persistent over the entire period.

I directly test the plausibility of the no-spillovers assumption by examining two groups: a sub-sample of 2,555 users active on both Gab and Gettr before the policy shift[30] and users who join Gab shortly after treatment. Column (1) of Table A.23 estimates the effect of the policy on Gab users who are not known to be present on Gettr, and finds a small but statistically significant post-treatment decline in threat scores. Column (2) restricts the sample to users known to be active on both platforms and yields a slightly larger significant reduction. In Column (3), I formally test for heterogeneity in this effect by including an interaction for cross-platform users; the coefficient is statistically indistinguishable from zero, indicating no significant difference in post-treatment reductions between the two groups. In columns (4) and (5) I restrict the sample to users present on both Gettr and Gab and tests whether changes in Gab content correlate with these users' pre-treatment threat levels on Gettr. If the moderation on Gettr affects their behaviour on Gab, then this will influence the interpretation of Gab as the counterfactual. In column (4), I interact the users pre-treatment Gettr threat average with the post treatment dummy and find no significant interaction effect. In column (5) I include a third interaction, whether the Gettr user exited Gettr over the prior or not and I, again, find no statistically significant effects. These results suggest that users exposed to Google's policy and active on Gab, did not behave any differently from other existing Gab users on the Gab platform after the policy.

Second, I assess differences in posting behaviour between users who joined Gab after the policy shift and were active on Gettr versus those who joined Gab after the policy but were not known to be active on Gettr. While the primary analysis focuses on users active both before and after the policy change, new joiners may still affect the behaviour of incumbent Gab users through interaction, potentially biasing the estimates. Column (6) of Table A.23 presents average differences between these two groups, showing no statistically significant difference in threat content. In Column (7), I interact group membership with the number of months since the policy change, including time-to-policy fixed effects. I find that Gettr users who joined Gab immediately after the policy shift exhibited significantly lower threat scores than other joiners of the same month, suggesting a potential short-term spillover of lower threatening content across platforms. This effect diminishes over time. This could be interpreted as an influx of users who naturally sorted towards Gettr and have a preference for less threatening content more consistent with the Gettr user base than Gab, but then shift following distaste for moderation. Another interpretation may be that users exposed to moderation of their threatening content on Gettr are less likely to post threatening content on Gab in the immediate aftermath of the experienced moderation. I explore

[30]Users are identified as being on both platforms by the presence of identical names across platforms, a method practised in Mitts (2021), Jones et al. (2017), and Rizzi (2024).

these dyanmics in more detail in Section 4.6. While new Gab joiners with prior Gettr activity initially posted less threatening content, this does not reflect a baseline difference between Gab and Gettr users. As shown in Column (3), cross-platform users active pre-treatment on Gab do not exhibit significantly different pre-to-post trends in threat content. Thus, the temporary difference observed among new joiners in Column (6) likely reflects short-term behavioural spillovers rather than permanent selection. Because the main estimates rely on users active before the policy shift, the validity of Gab as a counterfactual is not compromised.

Table 4.3 provides the triple difference specification, capturing variation across time, users, and platform-level exposure to Google's policy. The results show that, relative to Gab, threat scores declined significantly more on platforms exposed to Google's enforcement incentives. For each one standard deviation increase in a user's pre-policy 'threateningness', average post-policy threat scores fell by approximately 44 percent more on Gettr and 26 percent more on Truth Social compared to Gab, corresponding to an average reduction of 32 percent across the treated platforms. These differences represent substantial relative changes in content conditional on prior risk level, and point to a clear discontinuity in moderation response linked to infrastructure pressure. The preferred specification includes user-platform fixed effects, which control for time-invariant differences in baseline threat levels and platform usage patterns across users, and platform-week fixed effects, which absorb common shocks and policy changes affecting all users on a given platform in a given week. Results are remarkably stable across all specifications.

Robustness. I subject the triple difference specification to the same set of robustness checks applied in the platform-specific models, adapting them to the three-way interaction. As detailed in Table A.19 through Table A.22, the results remain consistent across alternative functional forms of the pre-treatment threat measure, different definitions of high-risk users, restrictions by post type, and across alternative toxicity measures from the Perspective API. Across all specifications, the estimated differences between treated platforms and Gab remain large, statistically significant, and directionally stable. In another specification, I repeat the exercise in a difference in differences specification excluding the sample to just the top 90th percentile users from each platform, and test the difference between platforms, after the announcement. Results in Figure B.9 show the same sharp reductions in Google based apps at the announcement date. Finally, I replicate the triple difference analysis on a matched sample constructed by aligning the distribution of pre-treatment threat scores across platforms using coarsened bins; results in Figure B.10 confirm that the estimated treatment effects remain consistent.

Table 4.3: DDD - Google Platforms vs Gab Users Threat Scores After Google Announcement

	Platforms				**Google**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\overline{Threat}$	-0.0529***	-0.0529***	-0.0522***	-0.0523***	-0.0529***	-0.0529***	-0.0522***	-0.0523***
× Post	(0.0157)	(0.0157)	(0.0155)	(0.0156)	(0.0157)	(0.0157)	(0.0155)	(0.0156)
Gettr ×	-0.4423***	-0.4423***	-0.4424***	-0.4411***				
$\overline{Threat}$ × Post	(0.0227)	(0.0227)	(0.0226)	(0.0225)				
TS ×	-0.2564***	-0.2564***	-0.2557***	-0.2547***				
$\overline{Threat}$ × Post	(0.0299)	(0.0299)	(0.0298)	(0.0298)				
Google ×					-0.3164***	-0.3164***	-0.3159***	-0.3145***
$\overline{Threat}$ × Post					(0.0236)	(0.0236)	(0.0235)	(0.0235)
N	25217118	25217118	25217118	25217118	25217118	25217118	25217118	25217118
User FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Platform FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
User × Platform FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
User Controls × Time FE	No	No	Yes	Yes	No	No	Yes	Yes
Platform × Year-Week FE	No	No	No	Yes	No	No	No	Yes
# of Users	46106	46106	46106	46106	46106	46106	46106	46106
DV Pre-Mean	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Treatment SD	2.3541	2.3541	2.3541	2.3541	2.3541	2.3541	2.3541	2.3541

Note: The table presents the results of estimating Equation 1. The dependent variable is the threat score of the users posts on Truth Social. $\overline{Threat}_i$ is the pre-treatment average threat score of user i's posts. User post threat score is standardised. $Post_t$ is an indicator variable = 1 following Google's announcement on August 19, 2022, and 0 otherwise. The level of observation is user × year-day. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

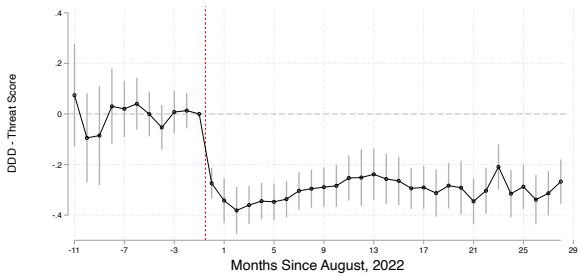
4.5 Topic Specific Responses to Infrastructure Moderation

While Google's 2022 Play Store policy formally targeted violent or inciting content, its revised language extended enforcement to a broader class of "sensitive events," defined as those with "significant social, cultural, or political impact", including civil emergencies, public health crises, and political conflicts. This expansion created the possibility that platforms would not only suppress explicitly threatening speech but also reduce the visibility of topics perceived to carry reputational or regulatory risk, even when such topics were not inherently unlawful or false. To investigate this possibility, I examine how enforcement pressure shaped the prevalence of user posts on five politically salient topics frequently implicated in content moderation debates.

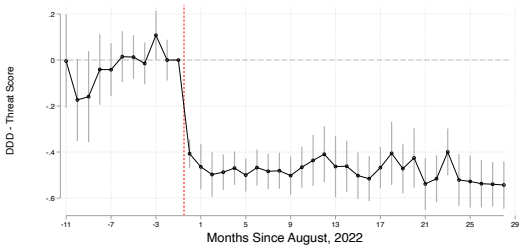
Empirical Approach. I replicate the design introduced in earlier sections and construct a user level measure of pre treatment topic intensity, defined as the proportion of posts made by a user on a given topic during the six months preceding the policy announcement. This proportion is standardised to facilitate interpretation. I then estimate dynamic difference in differences specifications, where the outcome is a binary indicator for whether a user posts on the topic in a given month, and the treatment

Figure 4.4: DDD Threat Google apps vs Gab

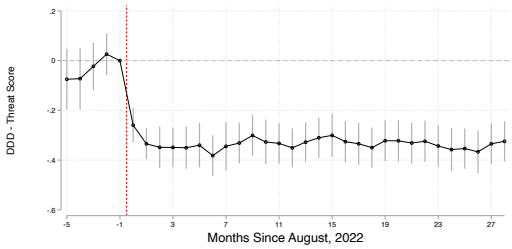
(a) Google vs Gab



(b) Gettr vs Gab



(c) TS vs Gab



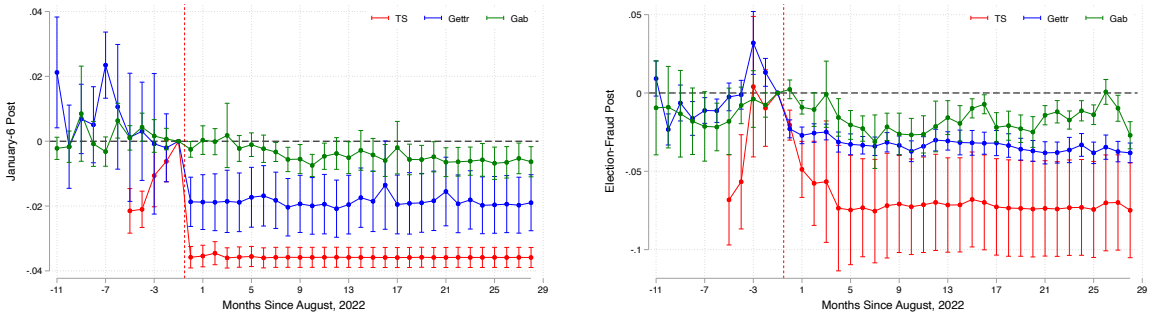
Note: Each panel reports dynamic triple-differences (DDD) estimates from regressions of post-level threat scores on the interaction between a user's pre-treatment average standardised threat score, month indicators relative to Google's policy announcement in late July 2022, and platform indicators. The omitted category is July 2022. This captures differential changes in threatening content on Google-exposed platforms (Gettr and Truth Social) relative to the untreated platform (Gab). Panel (a) aggregates both Gettr and Truth Social into a single treatment group, while panels (b) and (c) report disaggregated estimates for Gettr and Truth Social, respectively. All models include fixed effects for users, platforms, platform-by-month, and user-by-platform combinations. The sample includes posts from August 2021 to January 2025. Estimates are scaled relative to platform-specific pre-treatment means. Standard errors are clustered at the user level.

variable is the interaction between the standardised pre treatment share and time relative to the policy change.

This approach allows me to examine whether users with greater prior engagement on a topic reduced such engagement differentially following the introduction of Google's policy.

Results. The results in Figures 4.5, 4.6 and B.11 reveal a consistent pattern. Among platforms exposed to Google's enforcement regime, users with higher baseline engagement in a topic were significantly less likely to post about that topic after the policy announcement. The effects emerge immediately in August 2022 and persist throughout the post period.

Figure 4.5: Topic-specific event study estimates by pre-treatment user topic average posts proportion.



(a) January 6 commentary

(b) Election fraud claims

Note: Each panel shows the interaction effect between user-level pre-policy topic engagement (standardised) and time relative to Google's July 2022 policy update, using the dynamic difference-in-differences specification from Equation 1. Bars are 95% confidence intervals. Models include user fixed effects and platform-by-month trends. Standard errors clustered at the user level.

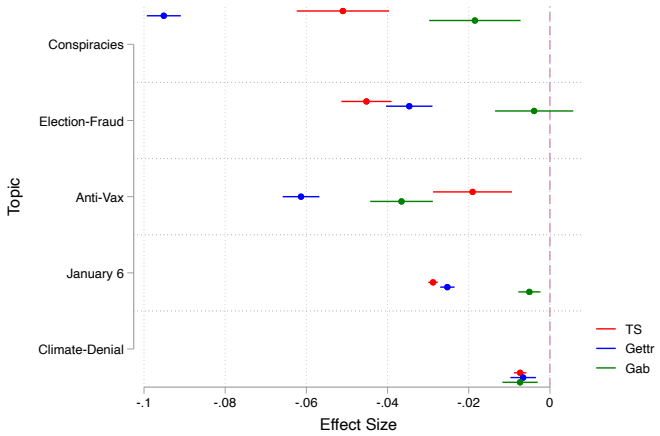
The largest and most sustained effects appear for January 6 commentary, election fraud claims, and conspiracy theories. On both Truth Social and Gettr, users with high prior engagement reduced topic related posting sharply in the first month following the policy change. The magnitude of the decline is greater on Truth Social for January 6 and election fraud[31], and greater on Gettr for anti-vaccine discourse and conspiracy theories.

By contrast, Gab displays a slightly downward trajectory in topic content, similar to the pattern noticed in threatening content. There is no statistically significant sharp reduction in the first few months after the policy change on any topic. With the possible exception of climate change skepticism, where there is a delayed decline around eight months after the announcement, topic related engagement on Gab follows a relatively stable or gradually declining trend. This divergence strengthens the interpretation that observed shifts on Truth Social and Gettr were induced by infrastructure exposure,

[31]This is not surprising given the close ties of Truth Social to Donald Trump, and his connections to both topics.

rather than endogenous user behaviour or platform wide drift.

Figure 4.6: Comparison of salient 'sensitive Events' discourse across platforms after Google policy announcement



Notes: This figure plots estimated coefficients from user-by-date fixed effects regressions examining the interaction between a post-policy indicator for August 2022 onwards and the user's pre-policy average topic engagement, standardised within each platform. Each panel corresponds to a distinct topic: Election-Fraud, Anti-Vax, Climate-Denial, January-6, and a composite "Conspiracies" variable (election fraud, anti-vax, and QAnon). The dependent variable is a binary indicator for topic presence in a post, and the key independent variable is the interaction term capturing whether more engaged users shifted content more strongly post-policy. Estimates are shown separately for Truth Social (TS, red), Gettr (blue), and Gab (green), with 95% confidence intervals clustered at the author level.

Falsification. As a falsification test, I assess whether users' pre-treatment levels of threatening speech can account for the decline in topic-related posting observed in the main analysis. This exercise evaluates whether platform responses reflect a broader suppression of high-risk users, irrespective of content, rather than targeted enforcement aligned with the specific issue areas referenced in Google's policy. To implement this test, I repeat the dynamic difference in differences specification from Section 4.5, replacing the topic-specific exposure measure with each user's standardised pre-policy share of threatening posts. The outcomes remain unchanged: monthly binary indicators for whether a user posted on each of the five topical categories. The estimates in Figure B.12 reveal no significant association between baseline 'threateningness' and changes in topic-specific posting behaviour. Across all topics and platforms, interaction effects are consistently small and statistically indistinguishable from zero, suggesting that the observed declines in topic engagement are not driven by general enforcement against high-risk users but by issue-specific responses tied to content exposure.

4.6 Spillovers and Behavioural Responses

The previous sections establish that threatening and sensitive content declined substantially on platforms exposed to Google's enforcement policy. While Section 4.4 presents evidence that spillovers are unlikely to bias the estimated treatment effects, this section

investigates whether the policy generated behavioural responses or unintended side effects. Specifically, I assess three channels: (1) declines in posting activity and user visibility; (2) changes in user discourse surrounding moderation prior to exit; and (3) behavioural shifts on Gab among users who exited Gettr, as a test of cross-platform spillovers. The findings suggest that the policy reduced harmful content through a combination of platform enforcement and user disengagement, without triggering backlash, coordinated migration, or reallocation of threatening discourse to less regulated spaces.

To assess user engagement on platforms, I assess exit rates across platforms by pre-treatment threateningness quartiles. I then construct a panel version of the dataset at the user-day level for each platform and repeat the dynamic version of Equation 1, replacing the outcome variable with: i) an indicator for whether a user posted original content on a given day, and ii) the total number of posts a user made on a given day. To measure 'virality', I replace the outcome variable with i) each posts total like count and ii) the total number of likes a user received on a given day.

As exhibited in Figure B.13, the largest proportion of exiters on Gettr occurs in the first few months following the policy shift, with the effect slightly more concentrated above the first quartile of pre-treatement 'threateningness'. There is a modest increase in user exit on Truth Social in the months following the policy, although this effect is overshadowed by a much larger spike in exits during the final quarter of 2023. I interpret this latter decline in users as a response to disclosures of substantial losses in November 2023, an interpretation supported by a second wave of user disengagement following first-quarter reports of over $327 million in losses and just $770,000 in revenue.[32] There is a gradual incline in the share of disengagement on Gab over the entire period with no sharp rise around July 2022. Overall, the evidence points to a short-term rise in exits on treated platforms following the policy, with longer-run patterns shaped by platform-specific events.

Results in Figure B.14 show similarly distinct patterns of behavioural change across platforms following the policy shift. On Gettr, users become significantly less likely to post on a given day and reduce overall post volume by roughly 50% relative to the pre-treatment platform average in the long run, though there is some indication of pre-existing trends. Engagement, as measured by likes, also declines steadily among higher-threat users throughout the post-treatment period. This suggests that the policy may have disproportionately reduced the visibility and engagement of threatening users, either through demotion or self-censorship. On Truth Social, there is an initial 10% decline in posting following the policy announcement, which gradually reverts to pre-treatment levels over the subsequent 12 months. Likes remain stable in the short run

[32]Company performance driven user exit may be explained by declining user confidence in the platform's viability, reduced incentives for marketers or influencers to remain, negative press coverage affecting reputation-sensitive users, and/or anticipatory disengagement by low-activity accounts; Weprin, Alex (November 13, 2023). "Trump's Truth Social Lost Tens of Millions Since Launch, New Filing Shows." *The Hollywood Reporter*; Goldstein, Matthew (May 20, 2024). "Trump Media Reports Lackluster Revenues and Large Losses." *The New York Times*.

but increase over time, particularly among high-threat users, in line with a resurgence in posting. In contrast, Gab exhibits a steady decline in posting among higher-threat users, even before the policy shift, while likes remain relatively flat for several months before entering a gradual downward trend. This could be interpreted as endogenous disengagement or shifting user dynamics, but the post-policy dip in likes may also indicate latent spillover effects or audience fragmentation as the ecosystem adjusts.

While these trends show aggregate behavioural changes, they do not reveal whether users perceived or reacted to moderation pressure in ways that prompted exit or disengagement. To explore this possibility, I next examine user discourse in the days surrounding exit. Figure B.15 examines whether high-threat users became more likely to discuss moderation in the period immediately preceding exit and whether this pattern changed after the policy. I identify all users who exited in the three months before and after the July 2022 policy announcement on Gettr and flag whether they mentioned moderation in the 30 days prior to their last post. I then regress this outcome on pre-treatment threat deciles, separately by platform and period. Prior to the policy, all three platforms exhibit a similar pattern: moderation discourse rises with threat decile but levels off in the upper range. This suggests a baseline tendency for more threatening users to express concerns about moderation before exit. In the three months after the policy, however, platform trajectories diverge. On Gettr, there is a sharp and monotonic increase: users in the highest deciles are 13–15 percentage points more likely than low-threat users to mention moderation before leaving. On TS, this pattern is weaker and only marginally significant in the top decile. Gab shows no such relationship. When the post-treatment window is extended to include all users exiting through October 2024, the pattern on Gettr and TS remains strong and significant, while the effects on Gab flatten further. These findings suggest that only platforms exposed to Google's enforcement pressure exhibited a post-policy rise in threat-linked moderation discourse preceding user exit.

Finally, to assess whether exiting users transferred harmful content to less regulated platforms, I examine changes in their Gab behaviour around the time of their Gettr exit. While Table A.23 already shows no significant shift in Gab content among cross-platform users after the policy, this analysis focuses on individual-level dynamics in the days surrounding exit. I estimate separate event study models for users who exited Gettr before and after the policy, interacting event time with policy-period status and controlling for user and date fixed effects. As shown in Figure B.16, users who exited after the policy display a short-lived, statistically significant increase in threatening content on Gab just before exit, followed by an immediate return to baseline. This pattern is not observed among pre-policy exiters. I also find no post-exit increase in moderation-related discourse, nor any divergence between groups. These results provide further evidence that the policy did not trigger the kind of user migration or content reallocation observed in earlier studies of public deplatforming.

Taken together, the results suggest that Google's policy reduced harmful content through a combination of platform enforcement and selective user disengagement, particularly among more threatening users on Gettr. I observe a clear short-term rise in user exit following the policy, but this exit was not accompanied by a measurable increase in threatening content or moderation-related discourse on Gab. Nor do I find evidence of backlash or coordination typical of previous deplatforming episodes. Instead, the response appears fragmented and individualised, marked by reduced posting, lower engagement, and exit from treated platforms without detectable reallocation of harmful discourse to alternative sites. While I cannot observe all potential destinations, the evidence within the observable ecosystem points to reduced activity rather than strategic migration.

One possible explanation for this muted response lies in the nature of enforcement. Unlike prior episodes involving coordinated bans or high-profile moderation crackdowns (Agarwal et al., 2022; Horta Ribeiro et al., 2023; Rizzi, 2024), the Google policy relied on infrastructure-level levers, quietly conditioning app store access rather than directly removing users. This relatively silent enforcement may have limited public backlash, reducing the visibility and ideological salience that often catalyses mass user exits. Moreover, by restricting availability rather than enacting outright bans, the policy may have encouraged gradual disengagement rather than coordinated migration. In this sense, infrastructure-led governance can constrain harmful content without provoking the redistributive dynamics seen in prior moderation campaigns.

5 Policy Implications

My findings reveal how dominant app distributors, those that control a key access point between users and platforms, can act as de facto regulators of content moderation. Through their control of mobile distribution infrastructure, these firms can compel platforms to adjust their moderation practices, not by directly managing platform content, but by conditioning continued access to users on compliance with content standards. This mode of private governance has far-reaching implications for the regulation of digital markets and online speech.

First, market power in distribution can translate into governance power over information flow, even in the absence of formal regulatory authority. When access to users is bottlenecked through a single intermediary, that intermediary gains leverage to impose content standards with potentially global effects. This raises fundamental concerns about the legitimacy of private governance, particularly when such enforcement lacks transparency or democratic oversight. At the same time, there is growing evidence that moderating online content can yield real-world benefits: hate speech online has been shown to increase offline hate crimes (Müller and Schwarz, 2021), while content moderation can reduce such violence (Duran et al., 2022). The exercise of

private governance thus poses a tradeoff, between concentrated, opaque enforcement and potentially substantial public benefits.

Second, enforcement by a single distributor can produce system-wide effects. When a platform modifies its moderation policies in response to distributor pressure, these changes often apply across the platform, affecting all users; not just those accessing the service through the distributor's channel. As a result, enforcement by one actor can reshape behaviour across the broader digital ecosystem.

Third, this dynamic creates the potential for free-riding among other distributors. Platforms may incur the costs of compliance to retain access to a dominant distribution channel, while users accessing the same platform through other means, such as web or alternative app stores, benefit from improved moderation without any corresponding enforcement mechanism. This asymmetry may distort platform incentives and lead to inefficiencies in content governance.

Finally, the role such infrastructure firms play blurs the boundary between market intermediaries and regulators. Distributors with the capacity to enforce content standards function simultaneously as infrastructure and as normative enforcers. This hybrid role challenges existing legal and policy frameworks, which often presume a clear division between infrastructure provision and content regulation.

In sum, the findings suggest that dominant distribution intermediaries can exert substantial influence over platform behaviour, not through direct involvement in content production or moderation, but through their gatekeeping position in the digital economy. This influence extends beyond structural incentives to shape the boundaries of political speech itself: platforms may suppress controversial but lawful content, such as election fraud claims or January 6 commentary, not because of regulatory obligation, but to avoid commercial risk. Regulatory frameworks that address not only platform-level decisions, but also the market structure and power of the intermediaries that govern access to users are therefore necessary to preserve a pluralistic marketplace of ideas.

6 Conclusion

This paper has shown that infrastructure providers, through their control of key distribution channels, can exert substantial influence over the governance of online speech. By conditioning platform access on compliance with content moderation standards, dominant app distributors like Google shape not only which platforms can reach users, but also the behaviour of users and the content that circulates within these ecosystems.

Using the July 2022 update to Google's Play Store policies as a quasi-experimental setting, I find that threatening content and politically sensitive misinformation declined sharply on platforms exposed to the new enforcement regime, while no comparable change occurred on an unexposed platform. These effects are concentrated among

users who previously exhibited high levels of noncompliant content, and they persist over time. The results are robust to a range of alternative specifications and falsification exercises, strengthening the causal interpretation.

The findings have important implications for the political economy of digital markets. First, they demonstrate that private enforcement mechanisms can meaningfully shape the boundaries of online discourse even absent state intervention. Infrastructure providers, by leveraging control over market access, effectively set behavioural expectations for platforms and users, operating as de facto regulators without the procedural safeguards that typically constrain public authority.

Second, the governance-by-infrastructure model raises concerns about transparency, accountability, and legitimacy. Enforcement decisions are often opaque, driven by internal policies that are neither democratically ratified nor consistently applied. While infrastructure pressure can reduce the prevalence of harmful or dangerous speech, it also creates risks of overcorrection and inequitable treatment across platforms and communities.

Third, this form of regulation is mediated by market structure: platforms more dependent on dominant distributors have stronger incentives to comply, while those less reliant on mainstream infrastructure may resist. This uneven exposure complicates both enforcement outcomes and policy responses, suggesting that interventions focused solely on platform behaviour may miss a crucial upstream source of governance.

Finally, the findings invite a broader reconsideration of how we conceptualise governance in digital markets. Rather than treating platforms as isolated sites of decision-making, we must recognise them as embedded in complex supply chains where infrastructure providers, often operating behind the scenes, play a central role in shaping information flows.

Future research should further investigate the conditions under which private enforcement complements or substitutes for public regulation, the distributional consequences of infrastructure-led governance, and the trade-offs between effectiveness, legitimacy, and innovation in moderating online speech. As control over digital pathways continues to consolidate, understanding how market power and normative power interact will be essential for designing policy frameworks that preserve the openness, accountability, and pluralism of the online public sphere.

References

Acemoglu, D. and Robinson, J. A. (2008). The role of institutions in growth and development. Technical Report Working Paper No. 10, Commission on Growth and Development.

Agarwal, S., Ananthakrishnan, U. M., and Tucker, C. E. (2022). Content moderation at the infrastructure layer: Evidence from parler. *Available at SSRN 4232871*.

Aliapoulios, M., Bevensee, E., Blackburn, J., De Cristofaro, E., and Zannettou, S. (2021). An early look at the parler online social network. arXiv preprint arXiv:2101.03820. https://arxiv.org/abs/2101.03820.

Allcott, H., Braghieri, L., Eichmeyer, S., and Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, 110(3):629–676.

Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–236.

Aluru, S. S., Mathew, B., Saha, P., and Mukherjee, A. (2020). Deep learning models for multilingual hate speech detection. *arXiv preprint arXiv:2004.06465*.

Andres, R. and Slivko, O. (2021). Combating online hate speech: The impact of legislation on twitter. Technical report, ZEW Discussion Papers.

Aridor, G., Jiménez-Durán, R., Levy, R., and Song, L. (2024). The economics of social media. *Journal of Economic Literature*, 62(4):1422–1474.

Armstrong, M. (2006). Competition in two-sided markets. *RAND Journal of Economics*, 37(3):668–691.

Beknazar-Yuzbashev, G., Jiménez Durán, R., and Stalinski, M. (2024). A model of harmful yet engaging content on social media. *Available at SSRN*.

Berlinski, N., Doyle, M., Guess, A. M., Levy, G., Lyons, B., Montgomery, J. M., Nyhan, B., and Reifler, J. (2023). The effects of unsubstantiated claims of voter fraud on confidence in elections. *Journal of Experimental Political Science*, 10(1):34–49.

Besley, T. and Prat, A. (2006). Handcuffs for the grabbing hand? media capture and government accountability. *American Economic Review*, 96(3):720–736.

Cao, B., Lindo, J. M., and Zhong, J. (2023). Can social media rhetoric incite hate incidents? evidence from trump's "chinese virus" tweets. *Journal of Urban Economics*, 137:103580.

Dixon, L., Li, J., Sorensen, J., Thain, N., and Vasserman, L. (2018). Measuring and mitigating unintended bias in text classification. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pages 67–73.

Duran, V. J., Muller, K., and Schwarz, C. (2022). The effect of content moderation on online and offline hate: Evidence from germany's netzdg. Working Paper. Forthcoming .

Goodwin, C. N. and Woolley, S. (2022). Sideloading: An exploration of drivers and motivations. In *35th International BCS Human-Computer Interaction Conference*, pages 1–6. BCS Learning & Development.

Guess, A., Nyhan, B., and Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 us election. *Nature Human Behaviour*, 4(5):472–480.

Hart, O. and Tirole, J. (1990). Vertical integration and market foreclosure. *Brookings Papers on Economic Activity: Microeconomics*, 1990:205–285.

Horta Ribeiro, M., Hosseinmardi, H., West, R., and Watts, D. J. (2023). Deplatforming did not decrease parler users' activity on fringe social media. *PNAS nexus*, 2(3):pgad035.

Jones, J. J., Bond, R. M., Bakshy, E., Eckles, D., and Fowler, J. H. (2017). Social influence and political mobilization: Further evidence from a randomized experiment in the 2012 us presidential election. *PloS one*, 12(4):e0173851.

Kalra, A. (2025). Hate in the time of algorithms: Evidence on online behavior from a large-scale experiment. *arXiv preprint arXiv:2503.06244*.

Klein, B. and Murphy, K. M. (1988). Vertical restraints as contract enforcement mechanisms. *Journal of Law and Economics*, 31(2):265–297.

Klonick, K. (2018). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review*, 131(6):1598–1670.

Kominers, S. D. and Shapiro, J. M. (2024). Content moderation with opaque policies. Technical report, National Bureau of Economic Research.

Lafontaine, F. and Slade, M. (2007). Vertical integration and firm boundaries: The evidence. *Journal of Economic Literature*, 45(3):629–685.

Levy, R. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review*, 111(3):831–870.

Liu, Y., Yildirim, P., and Zhang, Z. J. (2021). Social media, content moderation, and technology. *arXiv preprint arXiv:2101.04618*.

Madio, L., Mitchell, M., Quinn, M., and Reggiani, C. (2025). Asymmetric content moderation in search markets: The case of adult websites. Technical report, CESifo Working Paper.

Madio, L. and Quinn, M. (2024). Content moderation and advertising in social media platforms. *Journal of Economics & Management Strategy*.

Mitts, T. (2021). Banned: How deplatforming extremists mobilizes hate in the dark corners of the internet.

Mosquera, R., Galletta, S., and Tavits, M. (2020). Does social media reinforce mobilization? facebook, public events, and the 2016 us presidential election. *American Political Science Review*, 114(4):1325–1341.

Muller, K. and Schwarz, C. (2019). Fanning the flames of hate: Twitter and anti-minority violence. *American Economic Review*, 109(10):3594–3642.

Müller, K. and Schwarz, C. (2023). The effects of online content moderation: Evidence from president trump's account deletion. *Available at SSRN 4296306*.

Müller, K. and Schwarz, C. (2021). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association*, 19(4):2131–2167.

Prat, A. (2018). Media power. *Quarterly Journal of Economics*, 133(3):1331–1383.

Rey, P. and Tirole, J. (2001). Integration, delegation and the nature of contracts. *Quarterly Journal of Economics*, 116(4):1063–1100.

Rey, P. and Tirole, J. (2007). A primer on foreclosure. In Armstrong, M. and Porter, R., editors, *Handbook of Industrial Organization*, volume 3, pages 2145–2220. Elsevier.

Rey, P. and Vergé, T. (2004). Bilateral control with vertical contracts. *RAND Journal of Economics*, pages 728–746.

Rizzi, M. (2024). Self-regulation of social media and the evolution of content: a cross-platform analysis. *Available at SSRN 5018309*.

Roberts, S. T. (2019). *Behind the Screen: Content Moderation in the Shadows of Social Media*. Yale University Press, New Haven, CT.

Rochet, J.-C. and Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4):990–1029.

Wulczyn, E., Thain, N., and Dixon, L. (2017). Ex machina: Personal attacks seen at scale. In *Proceedings of the 26th international conference on world wide web*, pages 1391–1399.

Zannettou, S., Bradlyn, B., De Cristofaro, E., Sirivianos, M., Stringhini, G., and Blackburn, J. (2018). What is Gab? a bastion of free speech or an alt-right echo chamber? In *Companion Proceedings of the The Web Conference 2018*, pages 1007–1014. `https://dl.acm.org/doi/10.1145/3184558.3191531`.

Appendix

A Tables

Table A.1: Timeline of Google Play Policy Changes Relevant to Moderation Enforcement

Date	Policy	Key Provisions
Pre-2022	UGC Policy Baseline	Required moderation, terms of service compliance, in-app reporting and blocking systems.
Pre-2022	Hate Speech Baseline	Prohibited incitement and hate-based content; provided clarified examples.
Apr 6, 2022	Hate Speech Reiteration	Reaffirmed scope of incitement bans and compliance with local legal regimes, effective May 11 [Archive]
Apr 6, 2022	Expanded UGC Rules	Introduced system-level UGC requirements, effective October 11. [Archive]
Jul 27, 2022	Sensitive Events Clause	Enforceable immediately; required platforms to avoid insensitivity toward tragic events. [Archive]
Aug 19, 2022	Truth Social Rejection	Google formally rejected the app for inadequate content moderation. [News]
Oct 11, 2022	UGC Rule Enforcement	Formal deadline for compliance with the April UGC requirements.
Oct 12, 2022	Truth Social Acceptance	Google publicly announces acceptance of the app citing adequate content moderation. [News]

Table A.2: Comparison of Google's Sensitive Events Policy Before and After July 27, 2022

Policy Component	Before (Jan 17, 2022)	After (July 27, 2022)
Opening Clause	"lack reasonable sensitivity towards or capitalize on a natural disaster, atrocity, health crisis, conflict, death, or other tragic event"	"capitalize on or are insensitive toward a sensitive event with significant social, cultural, or political impact"
Definition of Sensitive Events	Examples included: natural disaster, atrocity, health crisis, conflict, death, or tragic event	Expanded and clarified: civil emergencies, natural disasters, public health emergencies, conflicts, deaths, or other tragic events
EDSA Clause	Allowed if content had Educational, Documentary, Scientific, or Artistic (EDSA) value or aimed to raise awareness	Same: EDSA clause retained verbatim
Event Denial Clause	"Denying a major tragic event"	"Denying the occurrence of a well-documented, major tragic event"
Profiteering Clause	Profiting from a tragic event without benefit to victims	Profiting from a sensitive event without benefit to victims

Table A.3: Descriptive statistics

Panel A: User Post Data	Gettr	TruthSocial	Gab
Users in Sample	25952	13586	7194
Likes (per user post)	10.900	3095.498	11.136
	[131.070]	[1.7e+04]	[97.994]
Reposts (per user post)	3.767	1123.812	5.769
	[53.188]	[4988.941]	[44.359]
Replies (per user post)	1.013	219.964	1.474
	[13.435]	[2061.993]	[10.747]
N_{pre}	877794	2687183	4160391
N_{tot}	1757312	11785676	13156911
Panel B: User Panel Data			
% Users post on a given day	0.127	0.296	0.343
	[0.332]	[0.456]	[0.475]
% Users comment on a given day	0.061	0.027	0.343
	[0.239]	[0.161]	[0.475]
Daily Posts	0.281	1.716	2.306
	[1.236]	[6.760]	[9.176]
Daily Likes of a Users posts	1.895	5310.825	25.652
	[58.978]	[3.0e+04]	[284.404]
Daily Replies to a Users posts	0.176	377.383	3.395
	[5.894]	[2934.263]	[28.819]
Daily Reposts of a Users posts	0.655	1928.080	13.288
	[23.341]	[9708.583]	[126.677]
N_{pre}	5050095	1566267	1806241
N_{tot}	28297340	13698512	8238494
Panel C: Post Measures			
Threat score	0.034	0.026	0.045
	[0.082]	[0.065]	[0.098]
Max Threat score	0.444	0.535	0.637
	[0.169]	[0.137]	[0.128]
Toxicity	0.175	0.144	0.193
	[0.210]	[0.175]	[0.215]
Severe toxicity	0.020	0.013	0.028

Continued on next page

	Gettr	TruthSocial	Gab
	[0.066]	[0.046]	[0.084]
Identity attack	0.043	0.032	0.069
	[0.091]	[0.070]	[0.132]
Insult	0.122	0.096	0.125
	[0.194]	[0.163]	[0.193]
Profanity	0.085	0.066	0.100
	[0.160]	[0.122]	[0.172]
Election fraud	0.020	0.039	0.019
	[0.139]	[0.194]	[0.137]
Anti-vaccine	0.044	0.024	0.067
	[0.204]	[0.153]	[0.251]
QAnon	0.018	0.023	0.044
	[0.132]	[0.151]	[0.205]
January 6	0.004	0.003	0.004
	[0.065]	[0.059]	[0.065]
Climate denial	0.001	0.001	0.002
	[0.032]	[0.033]	[0.040]

Note: This table summarises user-level content characteristics and content across platforms from before July 27, 2022. Panel A captures user level posts conditional on posting. Panel B utilises the panel dataset, which is at the user-day level. Panel C are average pre-treatment measures from the continuous Perspectives API scores and binary measures of sensitive event flags.

Table A.4: Examples of User Posts Threat Scores

Date	**Score**	**Post Content**
30 Jun 2022	0.88	I would've killed them all
01 Jul 2022	0.85	I say kill this communist!
26 Jul 2022	0.80	I will go personally kill the invaders.
16 Jul 2022	0.75	THIS MOTHER FUCKER NEEDS TO DIE!
08 Jun 2022	0.73	Why hasn't someone killed this piece of sht yet? Nazi trash deserves to be sent straight to hell.
07 Apr 2022	0.73	He needs a bullet in his head
07 Jun 2022	0.70	Soros is PURE EVIL. He needs to DIE!
19 Feb 2022	0.70	It's about this time I would have yanked him off the horse and kicked him in the face repeatedly
02 Mar 2022	0.60	"Wish I could of jumped through the screen and beat those wankers too a pulp, vile bastards"
19 Feb 2022	0.50	They don't want the pesky media there filming when they incite violence and start shooting people.
06 Jul 2022	0.40	Give up your guns at your peril people.
21 Mar 2022	0.30	Bill is one of the dumbest fcking people on television. He should go back to sticking his thumb up his ass
20 Nov 2021	0.20	Sudden Deaths of healthy young athletes. Vaccines are killing off Americans

Note: This table provides examples of users text contents from posts across Gab, Truth Social and Gettr and their coinciding 'Threat Score' as designated by the Perspectives API.

Table A.5: Selection Testing

	Gettr			**Truth Social**			**Gettr**		**Truth Social**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Threat	0.5454***	0.5142***	0.7285***	0.6654***	0.6095***	1.3059***	-7.9373***	-4.6341**	8.9951***	8.5292***
	(0.0668)	(0.0664)	(0.1219)	(0.1868)	(0.1779)	(0.2384)	(2.3016)	(2.2478)	(2.7252)	(2.6825)
N	47,249	47,249	47,249	18,562	18,562	18,562	26,033	26,033	13,586	13,586
# Leavers	16501	16501	16501	4886	4886	4886	-	-	-	-
# Stayers	30748	30748	30748	13676	13676	13676	30748	30748	13676	13676
User Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes
Weights	No	No	Yes	No	No	Yes	No	No	No	No

Note: The table presents two sets of regressions examining user selection and attrition relative to the policy change. Columns (1)–(6) estimate the likelihood that a user posts after July 27, 2022, as a function of pre-treatment threat scores. Columns (7)–(10) estimate the number of days users remained active after the policy. Specifications progressively add controls for user posting behaviour and exposure. Columns (3) and (6) apply analytic weights. All models use one observation per user. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table A.6: Selection Testing (Gab)

	Gab				
	(1)	(2)	(3)	(4)	(5)
Threat	2.4201***	1.7977***	2.9157***	25.1216***	19.6488***
	(0.2049)	(0.1837)	(0.1976)	(4.7898)	(4.7119)
N	14,860	14,860	14,860	7,533	7,533
# Leavers	7327	7327	7327	-	-
# Stayers	7533	7533	7533	7533	7533
User Controls	No	Yes	Yes	No	Yes
Weights	No	No	Yes	No	No

Note: The table presents two sets of regressions examining user selection and attrition relative to the policy change. Columns (1)–(3) estimate the likelihood that a user posts after July 27, 2022, as a function of pre-treatment threat scores. Columns (4)–(5) estimate the number of days users remained active after the policy. Specifications progressively add controls for user posting behaviour and exposure. Column (3) applies analytic weights. All models use one observation per user. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table A.7: Robustness I - Google Apps

	Gettr				**Truth Social**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\overline{Threat}$	-0.2196***				-0.1064***			
× Post	(0.0045)				(0.0105)			
$\overline{Threat}$		-0.0175***				-0.0052***		
(Top Quartile) × Post		(0.0004)				(0.0005)		
$\overline{Threat}$			-0.0272***				-0.0105***	
(Top Decile) × Post			(0.0009)				(0.0014)	
$\overline{Threat}$				-0.0065***				-0.0044***
(Q2) × Post				(0.0004)				(0.0006)
$\overline{Threat}$				-0.0137***				-0.0067***
(Q3) × Post				(0.0004)				(0.0006)
$\overline{Threat}$				-0.0268***				-0.0110***
(Q4) × Post				(0.0005)				(0.0008)
N	2,968,879	2,968,879	2,968,879	2,968,879	11466215	11466215	11466215	11466215
User FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of Users	25952	25952	25952	25952	13515	13515	13515	13515
DV Pre-Mean	0.0336	0.0336	0.0336	0.0336	0.0265	0.0265	0.0265	0.0265
Treatment SD	0.0819	0.0819	0.0819	0.0819	0.0657	0.0657	0.0657	0.0657

Note: The table presents the results of estimating alternative specifications supporting robustness of Equation 1 results. Column (1) dependent variable is the natural logarithm of the user content threat score while columns (2) through (4) are level. In column (2), I replace the main variable of interest with pre-treatment average threat score quartiles compared to the lowest quartiles, and then a dummy for the top quartile and decile respectively in columns (3) and (4). The level of observation is user × year-day. All regressions control for user and year-day fixed effects. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

Table A.8: Robustness II - Google Apps

	Truth Social				Gettr			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
$\overline{Threat}$× Post (01apr2022)	-0.0080***				-0.0262***			
	(0.0012)				(0.0007)			
$\overline{Threat}$× Post (01may2022)		-0.0078***				-0.0219***		
		(0.0013)				(0.0007)		
$\overline{Threat}$× Post (01jun2022)			-0.0062***				-0.0171***	
			(0.0013)				(0.0010)	
$\overline{Threat}$× Post (01jul2022)				-0.0042***				-0.0131***
				(0.0012)				(0.0010)
N	4,747,170	4,733,070	4,689,190	4,588,967	716,627	687,676	646,510	594,740
User FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of Users	10,533	10,019	8,813	7,532	13,445	11,766	10,203	8,591
DV Pre-Mean	0.0278	0.0278	0.0278	0.0278	0.0334	0.0334	0.0334	0.0334
Treatment SD	0.0633	0.0633	0.0633	0.0633	0.0805	0.0805	0.0805	0.0805

Note: The table presents the results of estimating Equation 1 with different cut off dates for calculating the pre-treatment average user threat score. The level of observation is user × year-day. All regressions control for user and year-day fixed effects. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

Table A.9: Robustness III - Google Apps

	Gettr					Truth Social				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Max Threat	-0.0082***					-0.0026***				
× Post	(0.0002)					(0.0003)				
Max Threat		-0.0150***					-0.0077***			
(≥ 0.2)× Post		(0.0003)					(0.0003)			
Max Threat			-0.0118***					-0.0043***		
(≥ 0.4)× Post			(0.0003)					(0.0003)		
Max Threat				-0.0103***					-0.0015***	
(≥ 0.6)× Post				(0.0008)					(0.0005)	
Max Threat					0.0078					-0.0024
(≥ 0.8)× Post					(0.0126)					(0.0016)
N	2,968,879	2,968,879	2,968,879	2,968,879	2,968,879	11466215	11466215	11466215	11466215	11466215
User FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of Users	25952	25952	25952	25952	25952	13515	13515	13515	13515	13515
DV Pre-Mean	0.0336	0.0336	0.0336	0.0336	0.0336	0.0265	0.0265	0.0265	0.0265	0.0265
Treatment SD	0.1698	0.1698	0.1698	0.1698	0.1698	0.1366	0.1366	0.1366	0.1366	0.1366

Note: The table presents the results of estimating Equation 1 with alternative definitions for how threatening users were in the pre-treatment period based on their maximum threat score, rather than on their average pre-treatment threat score. Column (1) uses a continuous variable while columns (2) though (4) are dummy variables equal to 1 at and above the stated threshold and 0 otherwise. The level of observation is user × year-day. All regressions control for user and year-day fixed effects. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

Table A.10: Robustness IV - Google Apps

	Gettr			**Truth Social**		
	(1)	(2)	(3)	(4)	(5)	(6)
$\overline{Threat}$	-0.0144***	-0.0137***	-0.0167***	-0.0072***	-0.0069***	-0.0029***
× Post	(0.0008)	(0.0006)	(0.0008)	(0.0007)	(0.0013)	(0.0010)
Source	Original	Reply	Repost	Original	Reply	Repost
N	710,427	1,217,429	1,035,971	7,964,599	256,713	3,249,490
User FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
# of Users	20572	20971	19163	13347	1590	6965
DV Pre-Mean	0.0322	0.0346	0.0330	0.0263	0.0288	0.0262
Treatment SD	0.0778	0.0880	0.0751	0.0651	0.0753	0.0622

Note: The table presents the results of estimating Equation 1 split by the types of content. Columns (1) are original posts content (not comments or replies); columns (2) are replies/comments; and columns (3) are re-posts of other users material. The level of observation is user × year-day. All regressions control for user and year-day fixed effects. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

Table A.11: Robustness V - Google Apps

	(1)	(2)	(3)	(4)	(5)	(6)
$\overline{Threat}$	-0.0181***	-0.0936***	-0.0066***	-0.0121***	-0.0569***	-0.0355***
× Post	(0.0005)	(0.0034)	(0.0004)	(0.0010)	(0.0027)	(0.0020)
N	2,968,879	2,968,879	2,968,879	2,968,879	2,968,879	2,968,879
User FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Control × Time FE	Yes	Yes	Yes	Yes	Yes	Yes
# of Users	25952	25952	25952	25952	25952	25952
DV Pre-Mean	0.0336	0.1749	0.0200	0.0431	0.1216	0.0848
Treatment SD	0.0819	0.2103	0.0658	0.0911	0.1943	0.1598
	(1)	(2)	(3)	(4)	(5)	(6)
$\overline{Threat}$	-0.0082***	-0.0052***	-0.0004	-0.0038***	-0.0048***	-0.0006
× Post	(0.0006)	(0.0018)	(0.0013)	(0.0008)	(0.0016)	(0.0016)
N	11466215	11466215	11466215	11466215	11466215	11466215
User FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Control × Time FE	Yes	Yes	Yes	Yes	Yes	Yes
# of Users	13515	13515	13515	13515	13515	13515
DV Pre-Mean	0.0265	0.1465	0.0133	0.0324	0.0978	0.0670
Treatment SD	0.0657	0.1772	0.0471	0.0711	0.1648	0.1236

Note: The table presents the results of estimating Equation 1 split by the types of content. Columns (1) are original posts content (not comments or replies); columns (2) are replies/comments; and columns (3) are re-posts of other users material. The level of observation is user × year-day. All regressions control for user and year-day fixed effects. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

Table A.12: Robustness VI - Google Apps

	Gettr		**Truth Social**		**Gettr**		**Truth Social**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\overline{Threat}$	-0.0202***	-0.0226***	-0.0103***	-0.0101***	-0.0269***	-0.0269***	-0.0211***	-0.0211***
× Post	(-37.96)	(-17.02)	(-13.19)	(-9.06)	(-60.67)	(-60.61)	(-14.76)	(-14.76)
N	3,490,683	754,169	11,914,040	8,488,525	3,493,325	3,136,067	11,915,521	11,785,676
# of Users	36305	5091	16891	4000	38944	26033	18371	13586
User FE	Yes	Yes	Yes	Yes	No	No	No	No
Year-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	No	No	No	No
Endline Sample	No	Yes	No	Yes	-	-	-	-
Panel Sample	-	-	-	-	No	Yes	No	Yes

Note: This table reports robustness checks addressing potential bias from selective attrition and panel sample selection. Columns (1)–(4) restrict the sample to users observed through the end of the study period ("endline" sample), testing for bias due to differential post-treatment attrition. Columns (5)–(8) estimate a simplified specification without user fixed effects to include users not present throughout, comparing the full sample to the subset of "stayers" who are active both before and after the policy change. All regressions include year-day fixed effects. Standard errors clustered at the user level. *** p<0.01, ** p<0.05, * p<0.10.

Table A.13: Robustness I - Gab

	(1)	(2)	(3)	(4)
$\overline{Threat}$	-0.0399***			
× Post	(0.0120)			
$\overline{Threat}$		-0.0049***		
(Top Quartile) × Post		(0.0011)		
$\overline{Threat}$			-0.0066***	
(Top Decile) × Post			(0.0020)	
$\overline{Threat}$				-0.0014
(Q2) × Post				(0.0013)
$\overline{Threat}$				-0.0018
(Q3) × Post				(0.0012)
$\overline{Threat}$				-0.0064***
(Q4) × Post				(0.0015)
N	13501666	13501666	13501666	13501666
User FE	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
# of Users	7384	7384	7384	7384
DV Pre-Mean	0.0441	0.0441	0.0441	0.0441
Treatment SD	0.0980	0.0980	0.0980	0.0980

Note: The table presents the results of estimating alternative specifications supporting robustness of Equation 1 results. Column (1) dependent variable is the natural logarithm of the user content threat score while columns (2) through (4) are level. In column (2), I replace the main variable of interest with pre-treatment average threat score quartiles compared to the lowest quartiles, and then a dummy for the top quartile and decile respectively in columns (3) and (4). The level of observation is user × year-day. All regressions control for user and year-day fixed effects. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

Table A.14: Robustness II - Gab

	(1)	(2)	(3)	(4)
$\overline{Threat}$	-0.0022**			
(01apr2022) × Post	(0.0011)			
$\overline{Threat}$		-0.0019*		
(01may2022) × Post		(0.0011)		
$\overline{Threat}$			-0.0015	
(01jun2022) × Post			(0.0012)	
$\overline{Threat}$				-0.0016
(01jul2022) × Post				(0.0012)
N	13369475	13215525	13104253	12910241
User FE	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
# of Users	6410	6129	5693	4968
DV Pre-Mean	0.0441	0.0441	0.0441	0.0441
Treatment SD	0.0980	0.0980	0.0980	0.0980

Note: The table presents the results of estimating Equation 1 with different cut off dates for calculating the pre-treatment average user threat score. The level of observation is user × year-day. All regressions control for user and year-day fixed effects. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

Table A.15: Robustness III - Gab

	(1)	(2)	(3)	(4)	(5)
Max Threat	-0.0008				
× Post	(0.0007)				
Max Threat		-0.0046***			
(≥ 0.2)× Post		(0.0006)			
Max Threat			-0.0020**		
(≥ 0.4)× Post			(0.0010)		
Max Threat				-0.0001	
(≥ 0.6)× Post				(0.0009)	
Max Threat					-0.0010
(≥ 0.8)× Post					(0.0022)
N	13501666	13501666	13501666	13501666	13501666
User FE	Yes	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
# of Users	7384	7384	7384	7384	7384
DV Pre-Mean	0.0441	0.0441	0.0441	0.0441	0.0441
Treatment SD	0.1230	0.1230	0.1230	0.1230	0.1230

Note: The table presents the results of estimating Equation 1 with alternative definitions for how threatening users were in the pre-treatment period based on their maximum threat score, rather than on their average pre-treatment threat score. Column (1) uses a continuous variable while columns (2) though (4) are dummy variables equal to 1 at and above the stated threshold and 0 otherwise. The level of observation is user × year-day. All regressions control for user and year-day fixed effects. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

Table A.16: Robustness IV - Gab

	(1)	(2)
$\overline{Threat}$	0.0000	-0.0035***
× Post	(.)	(0.0011)
Source	Original	Reply
N	226	13501191
User FE	Yes	Yes
Year-Day FE	Yes	Yes
Controls	Yes	Yes
# of Users	11	7375
DV Pre-Mean	0.0405	0.0441
Treatment SD	0.0668	0.0980

Note: The table presents the results of estimating Equation 1 split by the types of content. Columns (1) are original posts content (not comments or replies); columns (2) are replies/comments; and columns (3) are re-posts of other users material. The level of observation is user × year-day. All regressions control for user and year-day fixed effects. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

Table A.17: Robustness V - Gab

	(1)	(2)	(3)	(4)	(5)	(6)
$\overline{Threat}$	-0.0035***	-0.0053***	-0.0023	0.0001	-0.0043***	-0.0035**
× Post	(0.0011)	(0.0019)	(0.0017)	(0.0017)	(0.0014)	(0.0017)
N	13204938	13204938	13204938	13204938	13204938	13204938
User FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Control × Time FE	Yes	Yes	Yes	Yes	Yes	Yes
# of Users	7374	7374	7374	7374	7374	7374
DV Pre-Mean	0.0448	0.1928	0.0285	0.0689	0.1255	0.1005
Treatment SD	0.0979	0.2152	0.0842	0.1326	0.1932	0.1724

Note: The table presents the results of estimating Equation 1 split by the types of content. Columns (1) are original posts content (not comments or replies); columns (2) are replies/comments; and columns (3) are re-posts of other users material. The level of observation is user × year-day. All regressions control for user and year-day fixed effects. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

Table A.18: Robustness VI - Gab

	Gab			
	(1)	(2)	(3)	(4)
$\overline{Threat}$	-0.00336**	-0.00318***	-0.00564***	-0.00569***
× Post	(-2.52)	(-3.77)	(-4.88)	(-4.92)
N	16128850	11899570	16130864	15659334
# of Users	12846	3187	14860	7533
User FE	Yes	Yes	No	No
Year-Day FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	No	No
Endline Sample	No	Yes	-	-
Panel Sample	-	-	No	Yes

Note: This table reports robustness checks addressing potential bias from selective attrition and panel sample selection. Columns (1)–(2) restrict the sample to users observed through the end of the study period ("endline" sample), testing for bias due to differential post-treatment attrition. Columns (3)–(4) estimate a simplified specification without user fixed effects to include users not present throughout, comparing the full sample to the subset of "stayers" who are active both before and after the policy change. All regressions include year-day fixed effects. Standard errors clustered at the user level. *** p<0.01, ** p<0.05, * p<0.10.

Table A.19: Robustness I - DDD

	Platforms				Google			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gettr $\overline{Threat}$ × Post	-0.1814***							
	(0.0089)							
TS $\overline{Threat}$ × Post	-0.0738***							
	(0.0120)							
Gettr $\overline{Threat}$ (Top Quartile) × Post		-0.3860***						
		(0.0246)						
TS $\overline{Threat}$ (Top Quartile) × Post		-0.0980***						
		(0.0287)						
Gettr $\overline{Threat}$ (Top Decile) × Post			-0.6045***					
			(0.0435)					
TS $\overline{Threat}$ (Top Decile) × Post			-0.2769***					
			(0.0637)					
Gettr $\overline{Threat}$ (Q2) × Post				-0.1690***				
				(0.0278)				
TS $\overline{Threat}$ (Q2) × Post				-0.1477***				
				(0.0346)				
Gettr $\overline{Threat}$ (Q3) × Post				-0.3383***				
				(0.0251)				
TS $\overline{Threat}$ (Q3) × Post				-0.2030***				
				(0.0321)				
Gettr $\overline{Threat}$ (Q4) × Post				-0.6193***				
				(0.0314)				
TS $\overline{Threat}$ (Q4) × Post				-0.2819***				
				(0.0398)				
Google $\overline{Threat}$ × Post					-0.1050***			
					(0.0097)			
Google $\overline{Threat}$ (Top Quartile) × Post						-0.1744***		
						(0.0229)		
Google $\overline{Threat}$ (Top Decile) × Post							-0.3689***	
							(0.0465)	
Google $\overline{Threat}$ (Q2) × Post								-0.1820***
								(0.0232)
Google $\overline{Threat}$ (Q3) × Post								-0.2523***
								(0.0213)
Google $\overline{Threat}$ (Q4) × Post								-0.3953***
								(0.0284)
N	25217223	25217118	25217118	25217118	15540919	15540824	15540824	15540824
User FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Platform FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
User × Platform FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Platform × Year-Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of Users	46109	46106	46106	46106	45142	45139	45139	45139
DV Pre-Mean	0.0349	0.0349	0.0349	0.0349	0.0349	0.0349	0.0349	0.0349
Treatment SD	0.0826	0.0826	0.0826	0.0826	0.0826	0.0826	0.0826	0.0826

Note: The table presents the results of estimating alternative specifications supporting robustness of Equation 1 results. Column (1) dependent variable is the natural logarithm of the user content threat score while columns (2) through (4) are level. In column (2), I replace the main variable of interest with pre-treatment average threat score quartiles compared to the lowest quartiles, and then a dummy for the top quartile and decile respectively in columns (3) and (4). The level of observation is user × year-day. All regressions control for user and year-day fixed effects. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

Table A.20: Robustness II - DDD

	Platforms				Google			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gettr	-0.4705***							
$\overline{Threat}$ (April 2022) × Post	(0.0276)							
TS	-0.2155***							
$\overline{Threat}$ (April 2022) × Post	(0.0342)							
Gettr		-0.3484***						
$\overline{Threat}$ (May 2022) × Post		(0.0305)						
TS		-0.2114***						
$\overline{Threat}$ (May 2022) × Post		(0.0362)						
Gettr			-0.2208***					
$\overline{Threat}$ (June 2022) × Post			(0.0254)					
TS			-0.1420***					
$\overline{Threat}$ (June 2022) × Post			(0.0354)					
Gettr				-0.1107***				
$\overline{Threat}$ (July 2022) × Post				(0.0225)				
TS				-0.0327				
$\overline{Threat}$ (July 2022) × Post				(0.0288)				
Google					-0.3348***			
$\overline{Threat}$ (April 2022) × Post					(0.0270)			
Google						-0.3017***		
$\overline{Threat}$ (May 2022) × Post						(0.0275)		
Google							-0.1896***	
$\overline{Threat}$ (June 2022) × Post							(0.0267)	
Google								-0.0641***
$\overline{Threat}$ (July 2022) × Post								(0.0231)
N	24893092	24639339	24252769	23366504	15394356	15282411	15078100	14582049
User FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Platform FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
User × Platform FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Platform × Year-Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of Users	41740	39269	35149	28503	41535	39126	35051	28461
DV Pre-Mean	0.0349	0.0349	0.0349	0.0349	0.0349	0.0349	0.0349	0.0349
Treatment SD	0.0826	0.0826	0.0826	0.0826	0.0826	0.0826	0.0826	0.0826

Note: The table presents the results of estimating alternative specifications supporting robustness of Equation 1 results. Column (1) dependent variable is the natural logarithm of the user content threat score while columns (2) through (4) are level. In column (2), I replace the main variable of interest with pre-treatment average threat score quartiles compared to the lowest quartiles, and then a dummy for the top quartile and decile respectively in columns (3) and (4). The level of observation is user × year-day. All regressions control for user and year-day fixed effects. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

Table A.21: Robustness III - DDD

	Platforms			Google						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Gettr Max Threat × Post	-0.2026***									
	(0.0153)									
TS Max Threat × Post	-0.0702***									
	(0.0177)									
Gettr Max Threat (≥ 0.2) × Post		-0.3218***								
		(0.0166)								
TS Max Threat (≥ 0.2) × Post		-0.1849***								
		(0.0179)								
Gettr Max Threat (≥ 0.4) × Post			-0.2912***							
			(0.0290)							
TS Max Threat (≥ 0.4) × Post			-0.1274***							
			(0.0290)							
Gettr Max Threat (≥ 0.6) × Post				-0.2793***						
				(0.0312)						
TS Max Threat (≥ 0.6) × Post				-0.0410						
				(0.0268)						
Gettr Max Threat (≥ 0.8) × Post					0.1092					
					(0.3582)					
TS Max Threat (≥ 0.8) × Post					-0.0506					
					(0.0778)					
Google Max Threat × Post						-0.1105***				
						(0.0141)				
Google Max Threat (≥ 0.2) × Post							-0.2344***			
							(0.0145)			
Google Max Threat (≥ 0.4) × Post								-0.1710***		
								(0.0285)		
Google Max Threat (≥ 0.6) × Post									-0.0758***	
									(0.0210)	
Google Max Threat (≥ 0.8) × Post										-0.0450
										(0.0679)
N	25217118	25217118	25217118	25217118	25217118	15540824	15540824	15540824	15540824	15540824
User FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Platform FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
User × Platform FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Platform × Year-Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of Users	46106	46106	46106	46106	46106	45139	45139	45139	45139	45139
DV Pre-Mean	0.0349	0.0349	0.0349	0.0349	0.0349	0.0349	0.0349	0.0349	0.0349	0.0349
Treatment SD	0.1537	0.1537	0.1537	0.1537	0.1537	0.1537	0.1537	0.1537	0.1537	0.1537

Note: The table presents the results of estimating alternative specifications supporting robustness of Equation 1 results. Column (1) dependent variable is the natural logarithm of the user content threat score while columns (2) through (4) are level. In column (2), I replace the main variable of interest with pre-treatment average threat score quartiles compared to the lowest quartiles, and then a dummy for the top quartile and decile respectively in columns (3) and (4). The level of observation is user × year-day. All regressions control for user and year-day fixed effects. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

Table A.22: Robustness V - DDD

	(1)	(2)	(3)	(4)	(5)	(6)
$\overline{Threat}$	-0.0473***	-0.0229***	-0.0350	-0.0128	-0.0325***	-0.0272***
× Post	(0.0150)	(0.0069)	(0.0256)	(0.0165)	(0.0085)	(0.0085)
Gettr	-0.4386***	-0.4533***	-0.2812***	-0.2511***	-0.3862***	-0.3577***
$\overline{Threat}$ × Post	(0.0227)	(0.0226)	(0.0372)	(0.0343)	(0.0261)	(0.0292)
TS	-0.2533***	0.0031	0.0117	-0.0752**	0.0062	0.0443*
$\overline{Threat}$ × Post	(0.0292)	(0.0132)	(0.0956)	(0.0301)	(0.0178)	(0.0242)
N	18267806	18267806	18267806	18267806	18267806	18267806
User FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Platform FE	Yes	Yes	Yes	Yes	Yes	Yes
User × Platform FE	Yes	Yes	Yes	Yes	Yes	Yes
Platform × Year-Week FE	Yes	Yes	Yes	Yes	Yes	Yes
# of Users	45701	45701	45701	45701	45701	45701
DV Pre-Mean	0.0349	0.1694	0.0202	0.0493	0.1126	0.0827
Treatment SD	0.0826	0.1989	0.0668	0.1052	0.1818	0.1496

	(1)	(2)	(3)	(4)	(5)	(6)
$\overline{Threat}$	-0.0473***	-0.0229***	-0.0350	-0.0128	-0.0325***	-0.0272***
× Post	(0.0150)	(0.0069)	(0.0256)	(0.0165)	(0.0085)	(0.0085)
Google	-0.3155***	-0.0095	-0.0695	-0.1082***	-0.0076	0.0207
$\overline{Threat}$ × Post	(0.0230)	(0.0132)	(0.0684)	(0.0270)	(0.0177)	(0.0232)
N	18267806	18267806	18267806	18267806	18267806	18267806
User FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Platform FE	Yes	Yes	Yes	Yes	Yes	Yes
User × Platform FE	Yes	Yes	Yes	Yes	Yes	Yes
Platform × Year-Week FE	Yes	Yes	Yes	Yes	Yes	Yes
# of Users	45701	45701	45701	45701	45701	45701
DV Pre-Mean	0.0349	0.1694	0.0202	0.0493	0.1126	0.0827
Treatment SD	0.0826	0.1989	0.0668	0.1052	0.1818	0.1496

Note: The level of observation is user × year-day. All regressions control for user and year-day fixed effects. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

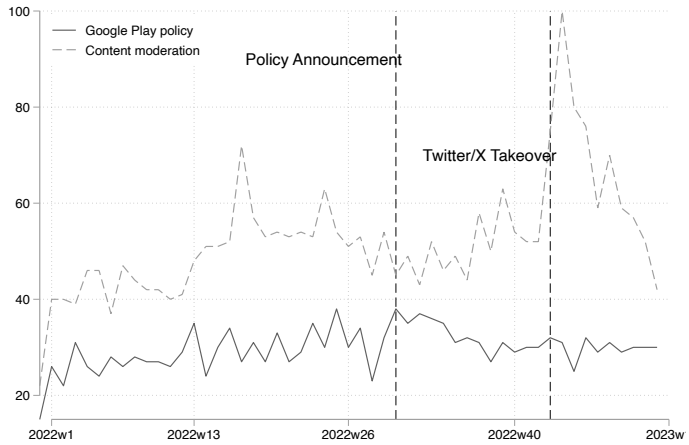
Table A.23: Spillovers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\overline{Threat}$	-0.0039***	-0.0054***	-0.0039***				
× Post	(0.0014)	(0.0016)	(0.0014)				
Cross-Platform			-0.0022***				
× Post			(0.0007)				
Cross-Platform			-0.0015				
× $\overline{Threat}$ × Post			(0.0021)				
Gettr				-0.0005	0.0014		
$\overline{Threat}$ × Post				(0.0006)	(0.0023)		
Gettr Leaver					0.0009		
× Post					(0.0014)		
Gettr					-0.0025		
$\overline{Threat}$ × Gettr Leaver × Post					(0.0024)		
Cross-Platform						0.0061	-0.0153***
						(0.0074)	(0.0058)
Cross-Platform							0.0001
× Months Since Post							(0.0004)
N	11253525	2,004,577	13258102	950,085	950,085	539,085	504,137
User FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls × Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of Users	4642	2555	7197	1516	1516	978	957
DV Pre-Mean	0.0456	0.0405	0.0448	0.0405	0.0405	.	.
Treatment SD	0.0990	0.0922	0.0979	0.0922	0.0922	.	.

Note: The table presents the results of estimating Equation 1. The dependent variable is the threat score of the users posts on Truth Social. $\overline{Threat}_i$ is the pre-treatment average threat score of user i's posts. User post threat score is standardised. $Post_t$ is an indicator variable = 1 following Google's announcement on August 19, 2022, and 0 otherwise. The level of observation is user × year-day. Standard errors clustered by user. *** p<0.01, ** p<0.05, * p<0.10.

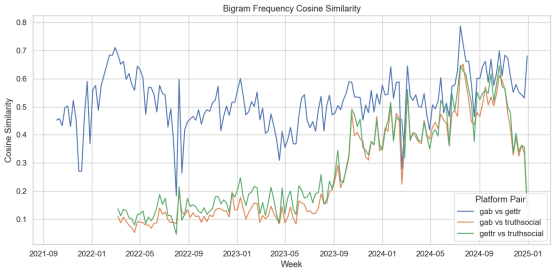
B Figures

Figure B.1: Google Trends - Google Policy and Content Moderation

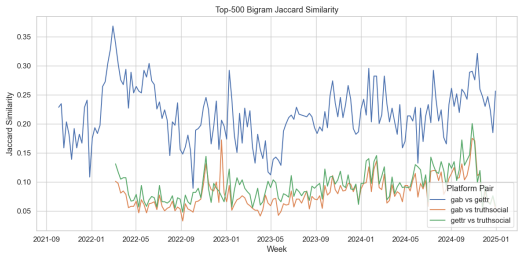


Note: Weekly Google Trends search interest for "Google Play policy" and "content moderation" worldwide. The Y-axis is a search pressure odds ration where 100 is the week with the ihghest search of the associated term, while a score of 50 means 50% of the search pressure compared to the highest search pressure.

Figure B.2: Similarity in content between platforms over time



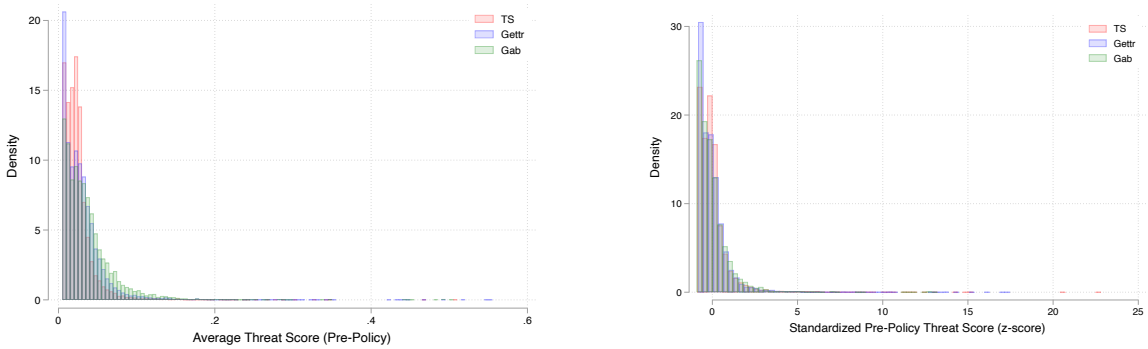
(a) Cosine Similarity



(b) Jaccard Similarity

Notes: This figure shows weekly pairwise similarity in language use between Gab, Gettr, and TruthSocial, based on n-gram frequency distributions. Panel (a) reports cosine similarity and panel (b) reports Jaccard similarity, computed from platform-specific n-gram distributions for each week. Higher values indicate greater overlap in the linguistic content of posts across platforms. The vertical line marks July 27, 2022, corresponding to Google's announced moderation policy change. Post-treatment convergence in similarity is visible, particularly between TruthSocial and the other platforms.

Figure B.3: Distribution of Pre-treatment Threat Scores Across Platforms

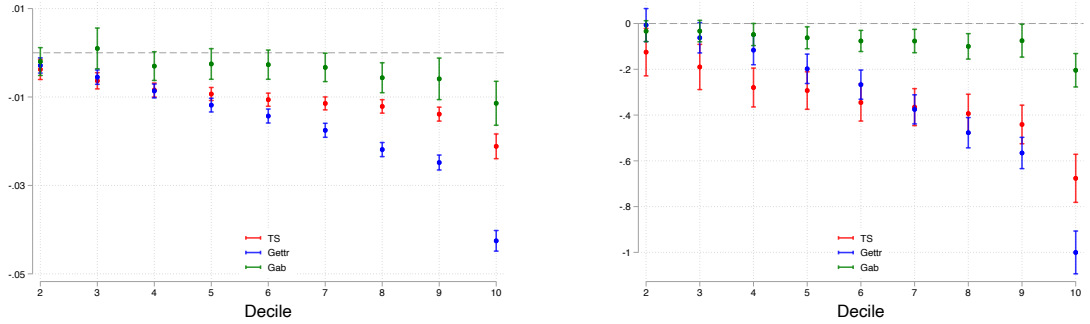


(a) Raw Distribution

(b) Z-Score Distribution

Notes: This figure compares the distribution of pre-treatment threat scores across platforms. Panel (a) shows the raw distribution of average threat scores by user, while panel (b) standardises scores within each platform to highlight relative dispersion and tail behaviour.

Figure B.4: Google Announcement on User Threatening Content Across Deciles Within Platforms

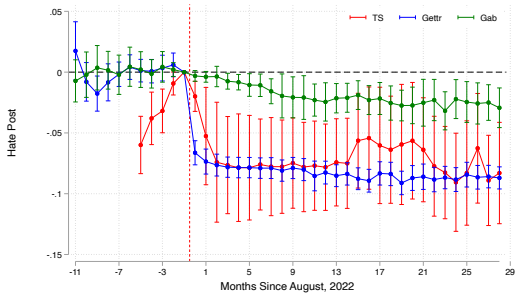


(a) Absolute Magnitude

(b) Mean Relative - Short Term Effect

Notes: Figure shows platform-specific regressions of the change in users' daily threat scores on a set of indicators for users' cross-sectional pre-treatment decile of average threat score. The omitted category is the first decile. In both panels, regressions include user control variables interacted with year-day fixed effects, user fixed effects, and year-day fixed effects. Standard errors are clustered at the user level. *** p<0.01, ** p<0.05, * p<0.10.

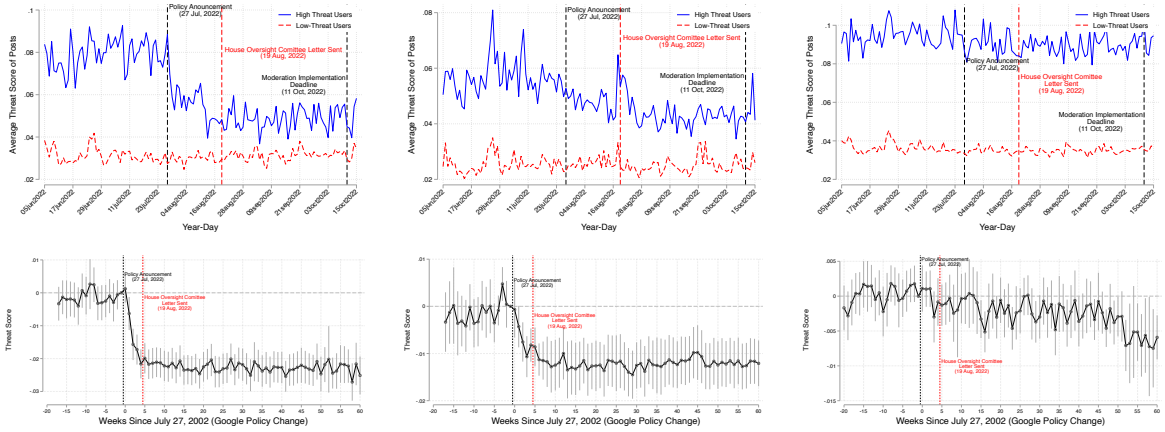
Figure B.5: Effect of Google Announcement on User Hate Speech Across Platforms



(a) Decile Regressions

Notes: Figure shows platform-specific regressions of users' posts classified as hate speech on the interaction of standardised pre-treatment user average hate speech prevalence with a set of month indicators relative to Google's announced policy change. The omitted category is July 2022. Standard errors are clustered at the user level. *** p<0.01, ** p<0.05, * p<0.10.

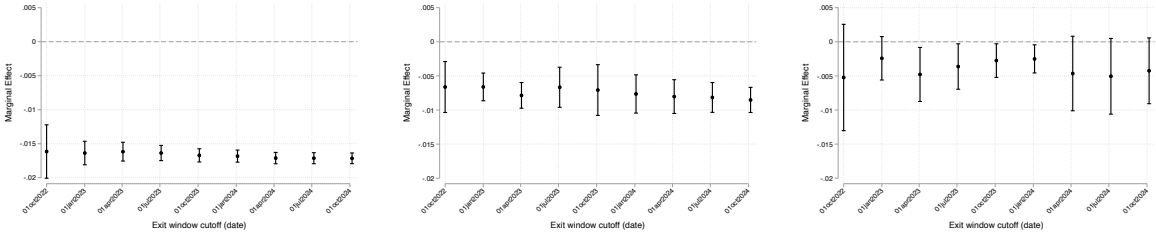
Figure B.6: User Threat Scores in days/weeks around July 27, 2022.



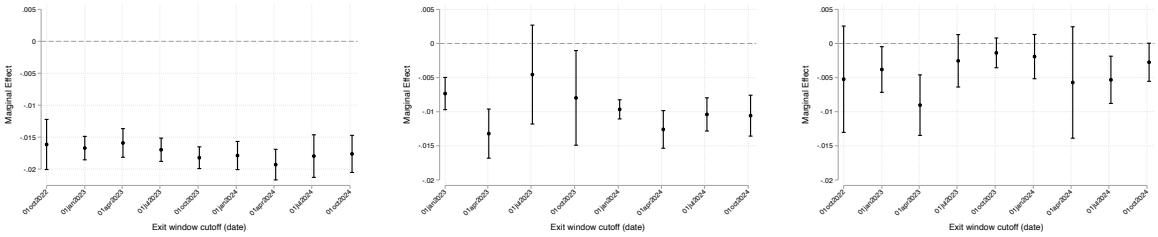
(a) Gettr **(b)** TS **(c)** Gab

Note: Across platform user threatening posts before and after Google announcement on July 27, 2022. The upper panels are the average threat score of posts by platform users, with users split based on the threat score of all of the users posts on that platform in the pre-exposure period. "High-Threat Users" are those whose posts before the Google announcement were on average above the 90th percentile of the threat score distribution. "Low-Threat Users" are all remaining users. The shaded areas indicate 95% confidence intervals for the means. Lower panels are platform specific regressions of user post threat score on the interaction of pre-treatment user average threat score cross-sections with a set of indicators for week pre-/ post- Google announcement. The omitted category is the week before the July 27, 2022. Standard errors are clustered at the user level. Regressions contain user control variables × year-day FE, user FE, year-day FE. Standard errors are clustered at the user level.

Figure B.7: Robustness - Selective Attrition



(a) Gettr - Accumulative **(b)** TS - Accumulative **(c)** Gab - Accumulative



(a) Gettr - Buckets **(b)** TS - Buckets **(c)** Gab - Buckets

Note: This figure evaluates robustness to selective attrition by estimating the marginal effect of the post-policy interaction with pre-treatment threat level across exit cohorts. Each point represents the coefficient on the interaction term from a separate regression of user 'threateningness' on post × threat, restricted to users who exited the platform before a given quarterly cutoff date (top row), or who exited within non-overlapping 3-month buckets preceding each cutoff (bottom row). Vertical lines show 95% confidence intervals. All regressions are restricted to users observed both before and after treatment (i.e., "stayers"), and include user fixed effects, day fixed effects, and time-varying controls. Standard errors are clustered at the user level.

Figure B.8: Cross platform spillover event studies

(a) Gettr vs Gab (b) Gettr users on Gab

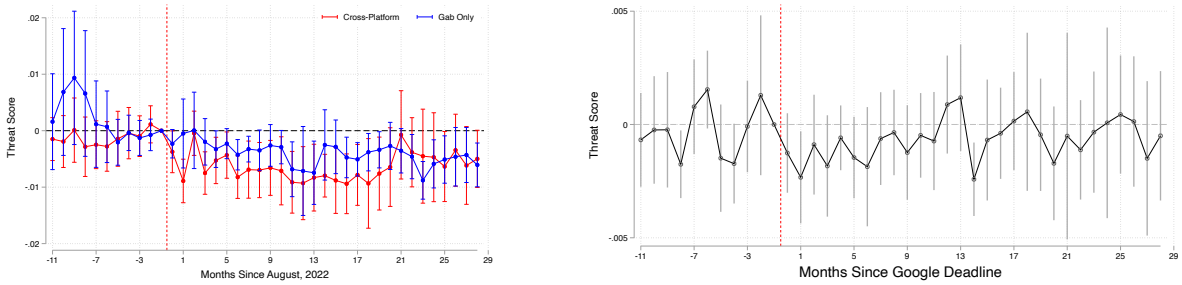
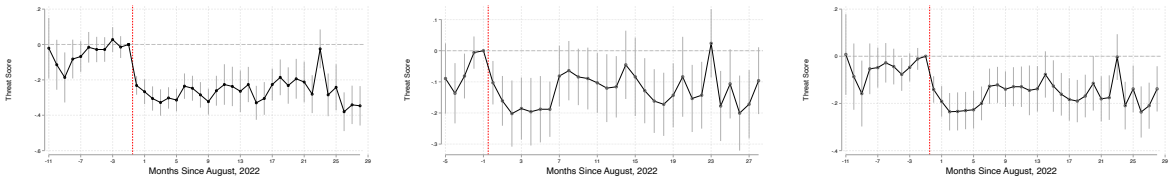


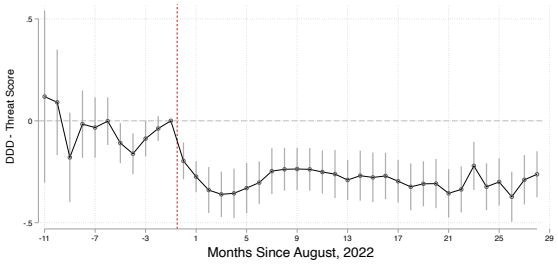
Figure B.9: DiD High Threat Users between platforms.



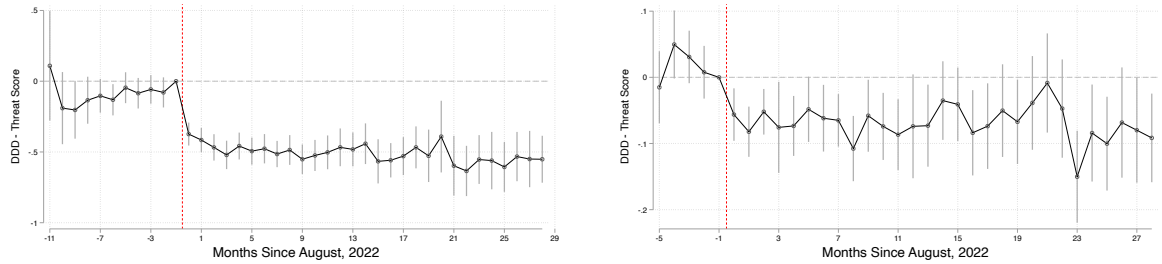
(a) Gettr **(b)** TS **(c)** Google

Note: Panels are regressions of user threatening score on the interaction of the Google exposed platform indicator with Gab as the omitted category, interacted with a set of indicators for month pre-/ post- Google announcement. The omitted category is July, 2022. Standard errors are clustered at the user level. Regressions contain user FE, platform-year-day FE and user-platform FE. Standard errors are clustered at the user level.

Figure B.10: DDD Estimates on Matched Sample by Pre-Treatment Threat Scores



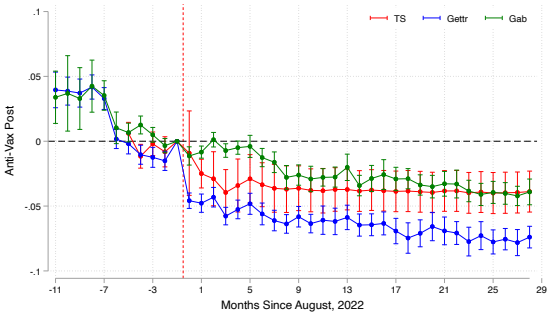
(a) Google



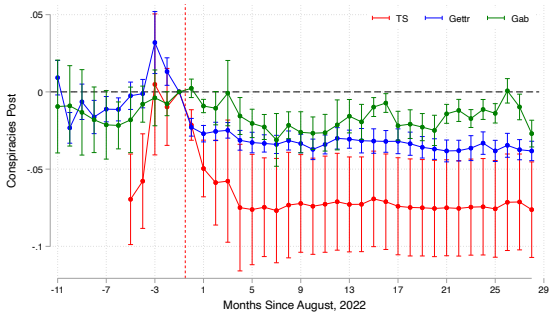
(b) Gettr **(c)** TS

Note: Each panel reports estimates from a difference-in-differences regression comparing changes in user threat scores between Gab and either the pooled Google-exposed platforms (Panel a), Gettr (Panel b), or (Panel c), using a sample matched on the distribution of pre-treatment threat scores. Users were matched by coarsened bins of standardised pre-treatment threat scores to ensure comparable threat distributions across platforms. Estimates show deviations from the pre-policy baseline (July 2022), with user and platform-day fixed effects included. Standard errors are clustered at the user level.

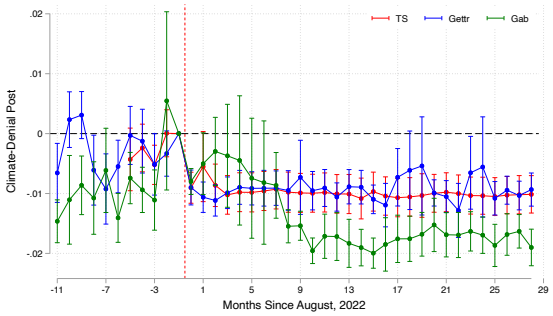
Figure B.11: Topic-specific event study estimates by pre-treatment topic intensity.



(a) Anti-vaccine content



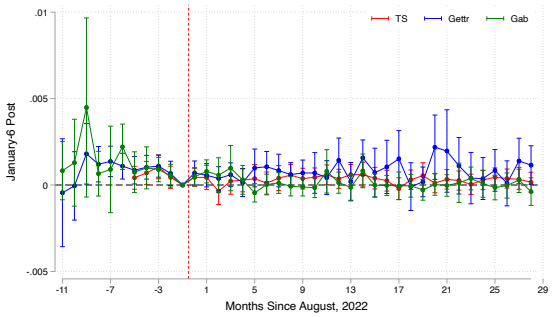
(b) Conspiracy theories



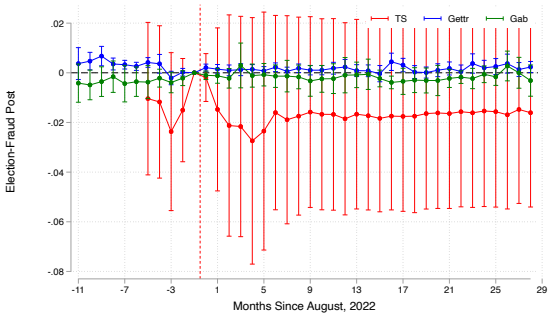
(c) Climate change denial

Note: Each panel shows the interaction effect between user-level pre-policy topic engagement (standardised) and time relative to Google's July 2022 policy update, using the dynamic difference-in-differences specification from Equation 1. Bars are 95% confidence intervals. Models include user fixed effects and platform-by-month trends. Standard errors clustered at the user level.

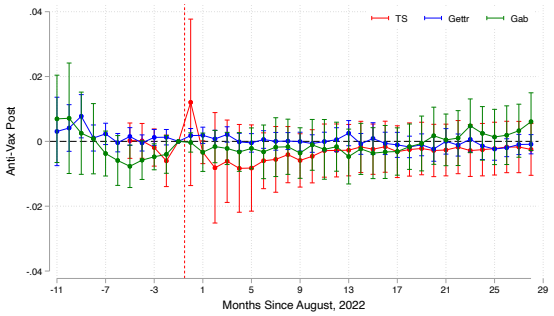
Figure B.12: Topic-specific event study falsification estimates by pre-treatment threatening average score.



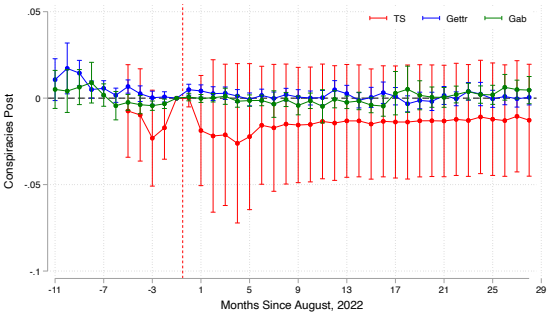
(a) January 6 commentary



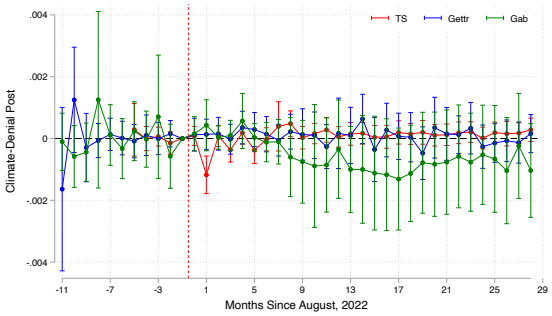
(b) Election fraud claims



(c) Anti-vaccine content



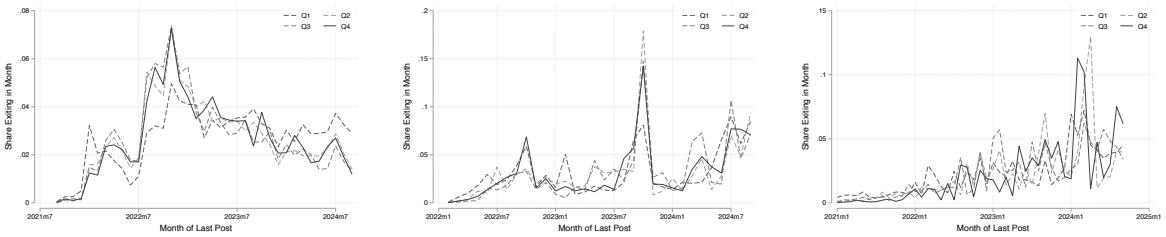
(d) Conspiracy theories



(e) Climate change denial

Note: Each panel shows the interaction effect between user-level pre-policy topic engagement (standardised) and time relative to Google's July 2022 policy update, using the dynamic difference-in-differences specification from Equation 1. Bars are 95% confidence intervals. Models include user fixed effects and platform-by-month trends. Standard errors clustered at the user level.

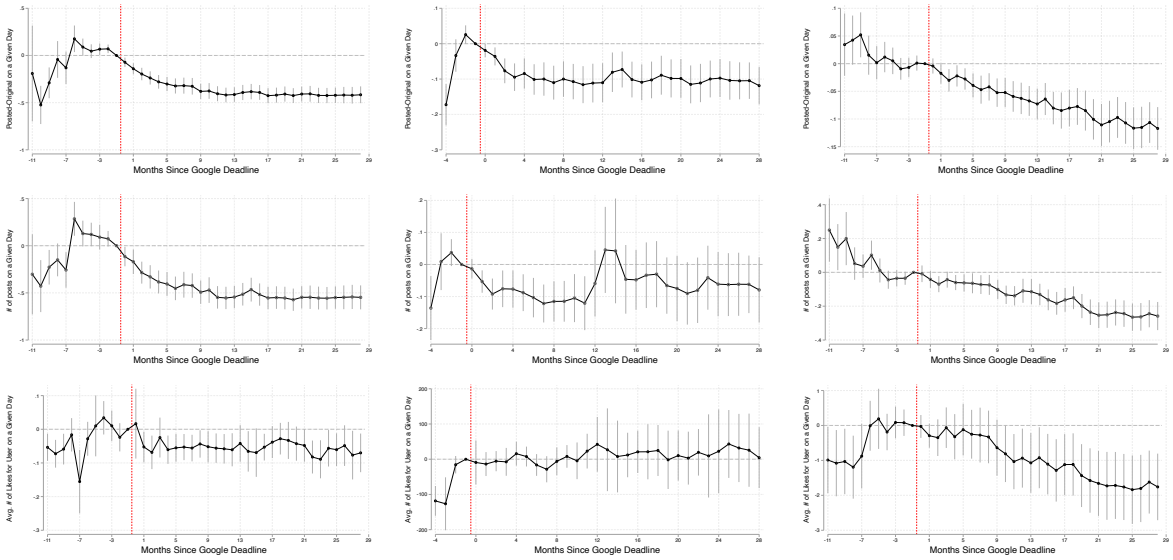
Figure B.13: Time Series of Platform Exiters by User 'Threateningness'



(a) Gettr **(b)** TS **(c)** Gab

Note: Each panel plots the monthly share of users who exited the platform (i.e., made their final post) by quartile of pre-treatment threateningness. Users are assigned to quartiles based on their average threatening score prior to the Google announcement. For each quartile, the share of users exiting in a given month is calculated as the number of users whose final post falls in that month divided by the total number of users in that quartile. The sample is restricted to users whose last post was before October 2024. Line types distinguish quartiles, with Q4 representing the highest baseline threat level.

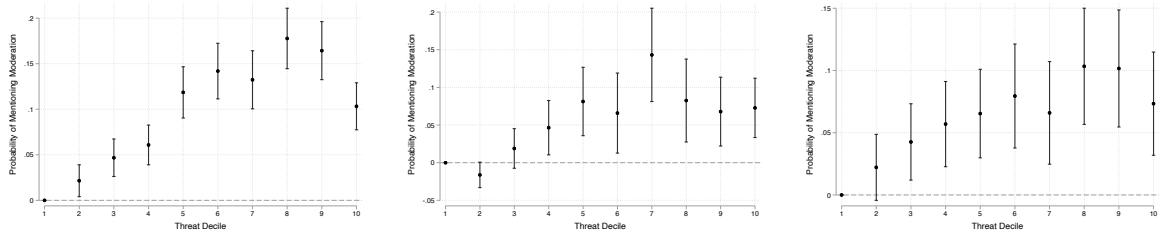
Figure B.14: User Engagement by Pre-Treatment Threat Intensity



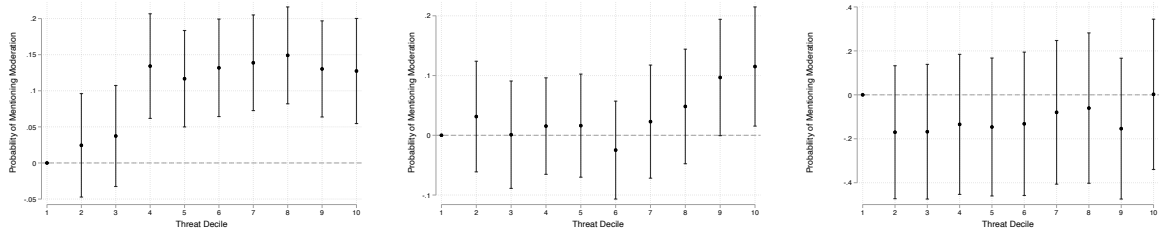
(a) Gettr **(b)** TS **(c)** Gab

Note: Panels are regressions of outcomes on user pre-treatment threatening score interacted with a set of indicators for month pre-/ post- Google announcement. The omitted category is July, 2022. The top panel outcome variable is an indicator for whether the user posted original content on a given day. The middle panel outcome is the number of posts a user makes on a given day (conditional on having joined the platform). The lower panel outcome variable is the number of like a post receives. Standard errors are clustered at the user level. Regressions contain user FE and year-day fixed effects. Standard errors are clustered at the user level.

Figure B.15: Moderation Discourse Before User Disengagement Before and After Policy Change



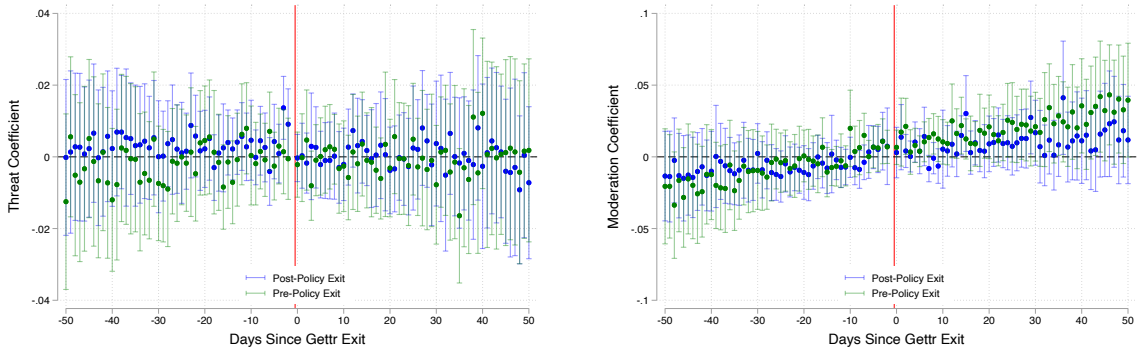
(a) Gettr - Before **(b)** TS - Before **(c)** Gab - Before



(a) Gettr - After **(b)** TS - After **(c)** Gab - After

Note: This figure shows the probability that users mentioned content moderation in the 30 days prior to their disengagement from the platform (i.e., last post date), plotted by deciles of pre-treatment threatening content. Panels (a)–(c) report estimates for the pre-policy period (users who exited before July 27, 2022), while Panels (d)–(f) show the immediate post-policy period (users active after July 2022 but exiting before November). For each platform, I regress a binary indicator for whether the user ever mentioned moderation in their final 30 days on pre-treatment threat deciles, controlling for users' average post frequency and timing relative to the policy announcement. 95% confidence intervals shown.

Figure B.16: Event study of Gab discourse before and after Gettr exit, by policy period

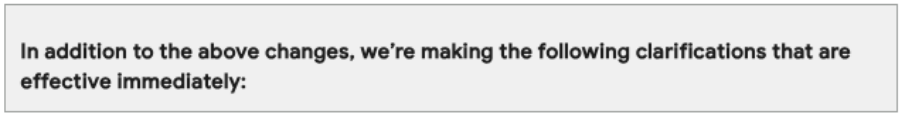


(a) Threatening content (b) Moderation-related discourse

Note: Each panel plots event study estimates of Gab user behaviour in the days surrounding their last post on Gettr. Treated users (blue) are those who exited Gettr after Google's July 2022 policy change; control users (green) exited Gettr prior to the policy. Outcomes are (a) the threat score of Gab posts, and (b) the likelihood of moderation-related language. Estimates are from separate regressions of the outcome variable on event-time dummies (days relative to Gettr exit), absorbed by user and date, with standard errors clustered at the user level. The baseline period is day –1.

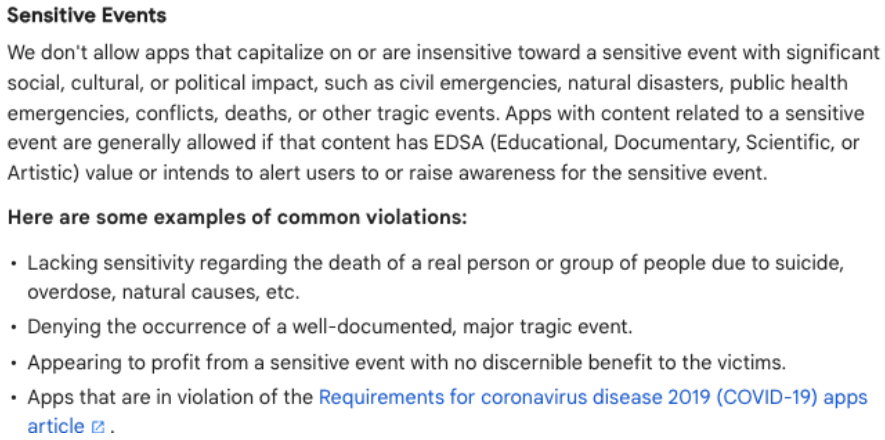
C Other Content

Figure C.1: Google announcement - Sensitive Events



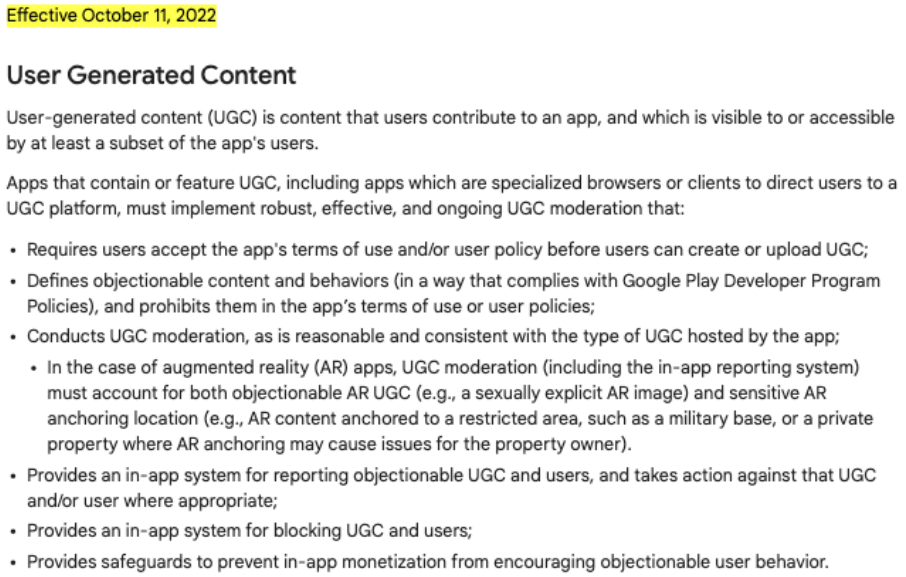
Note: Excerpt from Google developer policy update on July 27,2022. See [Archive]

Figure C.2: Google announcement - Sensitive Events



Note: Excerpt from Google developer policy update on July 27,2022. Appeared under the heading in Figure C.1. See [Archive]

Figure C.3: Google announcement - User Generates Content Moderation



Note: Excerpt from Google developer policy update on April 6,2022. See [Archive]

Table C.1: Major Infrastructure-Based Enforcement Actions Against Alt-Tech and Fringe Platforms

Platform	**Year**	**Infrastructure Provider**	**Type of Infrastructure**	**Reason for Enforcement**	**Outcome**
Gab	2017	Google Play Store	App Store	Violated hate speech policy; insufficient moderation	Removed from Play Store; later dropped lawsuit against Google
Gab	2018	Joyent, GoDaddy, PayPal/Stripe	Hosting, Domain, Payment Processor	Links to extremist violence (Pittsburgh synagogue shooting)	Hosting and payment services suspended; domain removed; platform migrated
Parler	2021	Apple and Google App Stores	App Store	Failure to moderate violent content post-Jan 6 Capitol riot	Removed from both app stores
Parler	2021	Amazon Web Services	Hosting Provider	Hosting of violent content; inadequate content moderation systems	Hosting suspended; site went offline temporarily
TruthSocial	2022	Google Play Store	App Store	Failure to moderate violent content; screenshots of inciting posts shared by Google	Initially rejected; approved after improving moderation systems
8chan (later 8kun)	2019	Cloudflare, Voxility	CDN / DDoS Protection	Hosting manifestos related to mass shootings	DDoS protection revoked; site went offline temporarily

Continued on next page

Table C.1 – continued from previous page

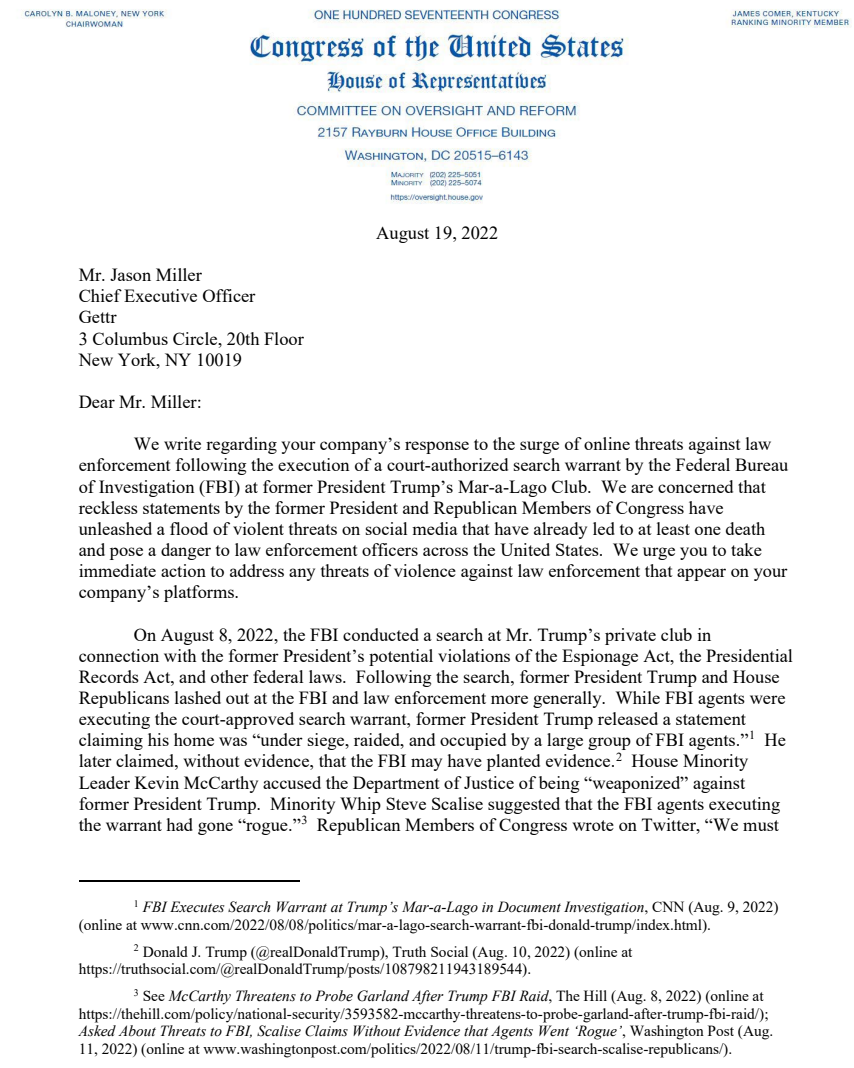
Platform	Year	Infrastructure Provider	Type of Infrastructure	Reason for Enforcement	Outcome
The Daily Stormer	2017	GoDaddy, Google Domains, Cloudflare	Domain, CDN / DDoS Protection	Hate speech and involvement in Charlottesville rally	Domain and CDN services revoked; site forced to migrate
Kiwi Farms	2022	Cloudflare	CDN / DDoS Protection	Targeted harassment and doxxing campaigns	DDoS protection revoked; site went offline temporarily
Wimkin	2021	Apple and Google App Stores	App Store	Violent content and failure to moderate post-Jan 6	Removed from app stores
Al-Manar	2012	Apple and Google App Stores	App Store	Content promoting hate and ties to designated terrorist organisations	Removed from add stores

Table C.2: Examples of Dictionary Terms by Topic

Topic	**Representative Patterns / Terms**
Election Fraud	`ballot`, `fraud`, `2000 mules`, `stop the steal`, any combination of `steal + election`
Anti-Vaccine	`hoax`, `plandemic`, `scamdemic`, `Bill Gates`, `Pfizer`, `vax`
QAnon	`WWG1WGA`, `The storm is coming`, `Q drop`, `Deep state`, `Adrenochrome`, `Save the children`
Climate Change Denial	`climate hoax`, `geoengineering`, `Agenda 21`, `green tyranny`, `chemtrails`
January 6 Commentary	`J6 patriot`, `Free the J6`, `Ray Epps`, `False flag`, `Political prisoners`, `Ashli Babbitt`

Note: Full regular expression lists are available upon request and included in the replication package.

Figure C.4: House Oversight Committee to Gettr - August 19, 2022



CAROLYN B. MALONEY, NEW YORK
CHAIRWOMAN

ONE HUNDRED SEVENTEENTH CONGRESS

JAMES COMER, KENTUCKY
RANKING MINORITY MEMBER

Congress of the United States

House of Representatives

COMMITTEE ON OVERSIGHT AND REFORM

2157 RAYBURN HOUSE OFFICE BUILDING

WASHINGTON, DC 20515–6143

MAJORITY (202) 225–5051
MINORITY (202) 225–5074

https://oversight.house.gov

August 19, 2022

Mr. Jason Miller
Chief Executive Officer
Gettr
3 Columbus Circle, 20th Floor
New York, NY 10019

Dear Mr. Miller:

We write regarding your company's response to the surge of online threats against law enforcement following the execution of a court-authorized search warrant by the Federal Bureau of Investigation (FBI) at former President Trump's Mar-a-Lago Club. We are concerned that reckless statements by the former President and Republican Members of Congress have unleashed a flood of violent threats on social media that have already led to at least one death and pose a danger to law enforcement officers across the United States. We urge you to take immediate action to address any threats of violence against law enforcement that appear on your company's platforms.

On August 8, 2022, the FBI conducted a search at Mr. Trump's private club in connection with the former President's potential violations of the Espionage Act, the Presidential Records Act, and other federal laws. Following the search, former President Trump and House Republicans lashed out at the FBI and law enforcement more generally. While FBI agents were executing the court-approved search warrant, former President Trump released a statement claiming his home was "under siege, raided, and occupied by a large group of FBI agents."[1] He later claimed, without evidence, that the FBI may have planted evidence.[2] House Minority Leader Kevin McCarthy accused the Department of Justice of being "weaponized" against former President Trump. Minority Whip Steve Scalise suggested that the FBI agents executing the warrant had gone "rogue."[3] Republican Members of Congress wrote on Twitter, "We must

[1] *FBI Executes Search Warrant at Trump's Mar-a-Lago in Document Investigation*, CNN (Aug. 9, 2022) (online at www.cnn.com/2022/08/08/politics/mar-a-lago-search-warrant-fbi-donald-trump/index.html).

[2] Donald J. Trump (@realDonaldTrump), Truth Social (Aug. 10, 2022) (online at https://truthsocial.com/@realDonaldTrump/posts/108798211943189544).

[3] See *McCarthy Threatens to Probe Garland After Trump FBI Raid*, The Hill (Aug. 8, 2022) (online at https://thehill.com/policy/national-security/3593582-mccarthy-threatens-to-probe-garland-after-trump-fbi-raid/); *Asked About Threats to FBI, Scalise Claims Without Evidence that Agents Went 'Rogue'*, Washington Post (Aug. 11, 2022) (online at www.washingtonpost.com/politics/2022/08/11/trump-fbi-search-scalise-republicans/).

Note: This is a letter sent to Gettr CEO Jason Miller by the House Oversight Committee on August 19, 2022. Identical versions were also sent to the CEO's of Facebook, Twitter, Truth Social, Gab, Rumble, Telegram, and TikTok.

Table C.3: Timeline of Apple App Store Review Policy Updates in 2022

Date	**Summary of Policy Changes**
2022-04-05	Apple permits "reader" apps (e.g., for books, audio, video) to include an in-app link to a developer-controlled website for account creation and management. This relaxation of 3.1.3(a) is formalized via the *External Link Account Entitlement*. However, such apps still cannot promote alternative payment methods within the app.

Date	Summary of Policy Changes
2022-06-06	Guidelines updated ahead of upcoming OS releases: • Reinforced rules on overtly sexual content (1.1.4) and background service use (2.5.4). • Clarified that law enforcement involvement is required for apps reporting criminal activity. • Strengthened rules against use of health-related APIs (e.g., HealthKit) for advertising or data mining. • Emphasized that in-app purchases cannot be used to buy credit for real-money gaming (5.3.3). • Required clear developer contact information in all apps, especially those used in classrooms. • Reiterated restrictions on using Apple interfaces, emoji, or music assets without authorization (5.2.5).

Date	Summary of Policy Changes
2022-10-25	Major clarification and expansion of content governance and monetization rules: • New prohibition on apps exploiting **sensitive events** (e.g., conflicts, epidemics) for profit (1.1.7). • Expanded 1.1.4 to explicitly mention human trafficking and exploitation as grounds for rejection. • NFTs: Apps may use in-app purchase to sell NFTs and related services (minting, listing, transferring), but cannot link to external marketplaces. • Cryptocurrencies: Apps may enable exchange-based transactions only in licensed jurisdictions. • Display advertising must remain within the main app binary and avoid sensitive targeting (2.5.18). • Matter-supported apps must use Apple's Matter SDK or certified alternatives. • Clarified demo access expectations: developers must offer a working demo account or a full-feature demo mode. • Specified that in-app purchases are required to unlock functionality (no external QR codes, crypto wallets, etc.). • New exemption for advertising campaign management apps from in-app purchase requirements (3.1.3(g)).

Chapter II

Black by Popular Demand: Media, Competition and the Evolution of a Social Norm

Abstract

What role do economic incentives play in social change? I study how media, responding to market pressures, shapes the evolution of an identity-based norm. Leveraging an exogenous shift in language preferences for the term 'Black' instead of 'Negro,' triggered by Black Civil Rights leaders in the mid-1960s, I examine what drives local newspapers' responses. First, I show that diffusion of the "Black" label in US Southern newspapers began immediately after this shock and that when the "Black" label is adopted, media representation of Black entities is immediately and persistently aligned with the identity intended by the "Black" campaign leaders. I then provide evidence that newspapers' demand-side exposure to White racial conservatism predicts both lower and slower adjustment to the new label, and that this effect is exacerbated by the intensity of local newspaper market competition. I show that the effect is not driven by local journalists, editors or companies' policies. Finally, I show that counties whose papers embraced "Black" earlier, in areas covered by Section 5 of the Voting Rights Act, experienced significantly faster growth in Black local officeholding between 1966 and 1974, suggesting that symbolic shifts in white-oriented media translated into tangible gains in Black political incorporation. This paper highlights how economic incentives within media markets can slow the diffusion of evolving norms, particularly when these norms challenge entrenched attitudes within the majority.

Keywords: Media economics, Social Norms, Economic History, Leadership, Coordination
JEL: L82, Z13, D72

1 Introduction

Labels play an important role in shaping identity (Tajfel and Turner, 1979; Hogg and Abrams, 1988; Giles and Johnson, 1987; Turner et al., 1987). In recent decades, many communities have increasingly redefined the labels which characterise them. Consider the LGBTQ+ community's evolving terminology, expanding "gay" to "LGBTQ" reflected broader inclusivity and recognition of diverse identities. Similarly, Indigenous peoples have increasingly advocated for the use of tribal names over colonial labels, reclaiming historical and cultural pride, while disability advocacy has embraced terminology that highlights nuanced perspectives on identity and ability. Given that identity influences intergroup perceptions and interactions (Becker, 1963; Phinney, 1990), as well as self-perception and economic outcomes (Akerlof and Kranton, 2000), particularly in the context of discrimination-driven misallocation of human capital (Hsieh et al., 2019), these label shifts are likely to play a significant role in both social and economic change. Yet, achieving societal acceptance of new labels and the underlying identities for which they represent, requires diffusion beyond the initiating group (Acemoglu and Jackson, 2015). This is especially true in the case of minority groups who by definition represent a small portion of society. Media outlets, as powerful intermediaries, likely play a central role in these norm evolutions by shaping public discourse and creating common knowledge (Morris and Shin, 2002; Chwe, 2013). At the same time, media face economic incentives on both the supply and demand side of their objective functions which shape their profit maximising choices (Gentzkow and Shapiro, 2006; Djourelova, 2023). Thus, studying media's label choices can shed light on how economic incentives shape processes of social change.

In this paper, I study how economic forces shaped the adoption of a new identity label for a minority group during a period of significant social change. I focus on the diffusion of the term "Black" as a replacement for "Negro" in U.S. newspaper coverage of Black communities during the civil rights era. The empirical setting is defined by a sharp, in-group-led campaign in June 1966, when Black civil rights leaders publicly called on the public to replace the term "Negro" with "Black." I leverage this intervention to trace the diffusion and political impact of the new label in three parts. First, I explore the diffusion of the label campaign by documenting the speed and scope of linguistic change, using keyword searches in two online newspaper archives and a corpus of 58 million digitised articles from 450 Southern newspapers between 1960 and 1973. I further show how adoption varies across time and space, and examine whether it was accompanied by systematic shifts in the narrative framing of Black individuals and communities in the press, as reflected in measures of agency and sentiment. Second, I test whether variation in adoption can be explained by economic incentives. I digitise historical newspaper circulation audits from 1964 to construct a novel dataset of newspaper-county circulation flows and local level newspaper market competition and match it to measures of White (i.e out-group) racial conservatism. This allows me to estimate how exposure to out-group preferences and local market structure shaped media behaviour. To isolate supply-side effects, I exploit a natural feature of the setting: a subsample of the observed newspapers were operated by the same editorial teams but circulated in distinct geographic markets, allowing for within-firm comparisons under differing demand

and competitive conditions. Third, I examine the downstream political effects of label adoption, asking whether regions exposed to newspapers that adopted the term "Black" earlier also saw greater subsequent increases in local Black political representation. Together, these three components provide a comprehensive account of how language change unfolds in the media, the economic forces that facilitate or constrain its diffusion, and its potential consequences for political incorporation.

Measuring the media's role in the evolution of a social norm presents substantial empirical challenges. Norm changes typically unfold gradually, without a clear onset or endpoint, complicating causal inference. Moreover, media outlets are not passive conduits of social change: they both shape and respond to evolving norms, creating a dynamic, symbiotic relationship between media behaviour and norm evolution. I address these challenges by exploiting a discrete, plausibly exogenous shock to the norm: a coordinated national campaign by Black Civil Rights leaders in June 1966 calling for the replacement of the term "Negro" with "Black." This intervention, launched publicly during the March Against Fear and led by Stokely Carmichael, created sudden and widespread pressure to adopt the new label, generating variation at the newspaper level plausibly independent of local trends in media behaviour (Garrow, 1986; Smith, 1992).

A second key challenge lies in isolating the incentives driving newspaper adoption decisions. Newspapers often serve heterogeneous audiences across multiple dimensions, while also balancing internal editorial preferences with external economic pressures. Disentangling these forces is difficult. I address this challenge in three ways. First, by focusing on White-owned Southern newspapers, where readership was overwhelmingly White (Staples, 2021) and Black citizens more often relied on radio than print media (Torres, 2003; Acham, 2005; Boroghkozy, 2012), I reduce the dimensionality of audience heterogeneity. Second, I combine detailed newspaper-county circulation data with county-level measures of White racial conservatism to quantify newspapers' exposure to out-group consumer preferences across different markets along a single dimension (i.e. racism). Third, I exploit cross-sectional variation in local newspaper market concentration and in the extent to which individual newspapers relied on specific markets for circulation revenue. This allows me to test whether adoption patterns varied with a newspaper's financial exposure to different audience segments, providing a way to distinguish between economic incentives tied to out-group preferences and other potential drivers of adoption. Finally, I leverage a natural feature of the setting: during the period studied, a first wave of newspaper consolidations resulted in several newspapers being operated by the same editorial and journalistic teams but serving distinct geographic markets. This allows me to compare adoption decisions across newspapers facing different consumer and competitive environments while holding supply-side decision-makers constant.

To measure diffusion of the new label, I create an an index which measures the proportional references to Black people using the label "Black" out of all references to Black people using either of the labels "Black" or "Negro". I count instances of the use of the two labels by newspaper×month using the two online newspaper archives search features. Since "Black" referencing Black people may be confused with many other uses of the label which are not referencing Black people, I develop a machine learning pipeline to identify the most common references to Black entities in the full text sub-sample to create a searchable "Black"

dictionary.

To assess whether shifts in racial labels alter how Black individuals are represented in the media, I examine changes in the language surrounding Black entities before and after the adoption of the term "Black." If label adoption is associated with systematic shifts in agency and sentiment, it would suggest that linguistic choices influence not just terminology, but also underlying portrayals of group status and behaviour. To test this, I use a large language model (ChatGPT) to classify the linguistic agency and sentiment of Black entities in 2.5 million newspaper sentences.

In the first and most descriptive part of the analysis, I show that the "Black Power" campaign launched in June 1966 marks a significant turning point in newspapers' labelling practices. The adjustment in the average use of dictionaries begins with an increase in the use of the term Black in the period 1966-1968 and continues with both an increase in "Black" and reduction in "Negro" in the period 1968-1973. On average, the relative use of the term increases from $\sim$12.5% to $\sim$60% from before to after the second semester of 1966. I find no evidence of a reduction in overall coverage of stories which include Black entities in the first 5 years after the campaign, with a general decline in references to Black entities thereafter.

In regards to narratives, I find that sentences about "Black" entities are statistically significantly more likely to represent the entities with high agency and positive sentiment than sentences about "Negro" entities, beginning in the second semester of 1966. This result is persistent over the entire post-treatment period. While I cannot fully rule out confounding from broader discursive change, I address key threats to identification through a series of complementary research designs. To account for the possibility that variation in sentiment reflects differences in authorship or article tone, I compare "Black" and "Negro" entities within the same article. To address concerns that shifts in tone reflect the emergence of new organisations (e.g., Black Power, Black Panthers), I disaggregate entities into individuals, groups, and organisations, showing that the results are primarily driven by references to group entities such as "Black people" or "Black protestors." Finally, to ensure results are not driven by the entry of new entities, I restrict attention to recurring named entities and show that changes in representation are concentrated among those that appear both before and after the campaign. Taken together, these approaches provide strong evidence that the observed shift in tone reflects a broader redefinition of Black representation associated with the adoption of the new label.

In the second part of the paper, I explore how newspaper exposure to the dominant out-group's preferences and competitive pressures determine newspaper diffusion. I conceptualise the dynamics in a framework where firms must balance an incentive to compete for demand by aligning content with consumers priors, as seen in models by Gentzkow and Shapiro (2006, 2007) and Mullainathan and Shleifer (2005), and an incentive to conform to the evolving norms within the media institution (Banerjee, 1992). I assume that increasing levels of competition strengthens the incentive to confirm consumer priors, which may explain short term variation in adoption, while herd behaviour predicts longer term convergence. Besides the social and economic pressures that drive the label evolution, I also allow within firm biases to affect output decisions.

To proxy newspapers' exposure to out-group preferences, I use the county vote share

for the racially conservative and segregationist States' Rights Democrats candidate, Strom Thurmond, in the 1948 presidential election. To estimate competitive pressure, I calculate county level daily newspaper market concentration from 1964 by means of a Herfindahl-Hirschman Index (HHI). This measure is relevant to a sub-sample of 165 daily newspapers. To precisely measure newspaper exposure to these measures, I calculate each newspaper's weighted exposure, using weights based on the proportional circulation in the newspapers' most circulated counties. I combine these data with newspaper outlet and county level pre-treatment controls from multiple sources.

I employ a generalised difference in differences approach, where I exploit the plausibly exogenous temporal variation in newspaper consumer preferences for the "Black" label, as well as spatial variation in newspapers' pre-treatment exposure to White racial conservatism and competition.

Estimation results indicate that newspaper exposure to higher levels of pre-treatment White racially conservative consumers predicts lower relative use of the racially liberal label, "Black", following the shock. For example, newspapers exposed to consumers with a majority vote share for representative Strom Thurmond in the 1948 presidential election had a 10% lower average use of the term "Black" when referring to Black people after the shock compared to newspapers exposed to a minority vote share.

Newspapers facing increasing competitive pressures are significantly more sensitive to out-group pressure both at the extensive and intensive margins, i.e. increasing White racial conservatism predicts lower adoption of the "Black" label, with the magnitude increasing in the presence and magnitude of competition. I find little evidence of competition driven product differentiation, challenging the conventional view that competition inherently fosters diversity in media content, suggesting instead that it may deepen entrenched biases when minority preferences diverge from majority tastes. These findings suggest a complex interplay between market forces and social dynamics, revealing how competition not only sharpens the sensitivity of newspapers to the dominant ideologies in their readership but also amplifies their role in either resisting or accelerating cultural change.

Results are robust to a large set of controls, including county-level covariate fixed effects, within state- and DMA$\times$month fixed effects. Results hold with multiple alternative measures of White racial conservatism and measures of market competition taken from different pre-treatment years. As a placebo test, I repeat the competitive analysis for non-daily newspapers who are not subject to the same daily newspaper market competition as the sub-sample of daily newspapers, but are operating in the same county of commercial residence. Estimates are highly statistically insignificant, suggesting that there are no spatially varying confounders which correlate with my main variables of interest and drive results. While I cannot rule out the potential that there is some other unmeasured confounder, in a series of statistical horse races, I find no evidence that other likely demand- or supply-side forces, including Black preferences, political alignment of consumers or suppliers, or subscription to newswires, are driving the results.

To explore supply side-effects, I utilise a subsample of newspaper pairs, owned by the same companies, operating in the same building, with the same team of journalists and editor, but facing varying levels of demand-side racial conservatism and competition due to

heterogeneity in spatial circulation. Estimates hold in this restrictive specification suggesting that the observed label diffusion appears to be driven primarily by external market pressures and local consumer ideologies, rather than journalists, editors or companies policies.

In the final part of the paper, I move beyond media behaviour to investigate whether these linguistic shifts had material political consequences. I aggregate the racial-terminology index to the county level and combine this with newly digitised panel data on Black elected officials from Bernini et al. (2023b). The analysis estimates whether counties exposed to newspapers that adopted the label "Black" earlier or more consistently saw faster growth in local Black officeholding following enfranchisement. I find that they did: such counties experienced significantly faster growth in Black officeholding after enfranchisement. A fully saturated triple-difference event-study design interacts the racial index with Voting Rights Act coverage and year, identifying from differential changes in representation across structurally similar counties exposed to different media cues. While not randomly assigned, the rich set of controls, along with robustness checks using residualised language variation, help suggest a causal interpretation. A subsequent mechanism analysis is consistent with the interpretation that these patterns are not driven by changes in turnout or demographics, but rather reflect norm change among white readership, potentially reinforcing the political legitimacy of Black officeholding following the removal of formal voting barriers.

This paper contributes four different strands of literature. First, it contributes to the literature on media bias, particularly the measurement and determinants of slant in media language and the relationship between media competition and media output. Firstly, previous research has found that competition increases ideological diversity offsetting the incentive to confirm the priors of dominant consumer groups (Gentzkow et al., 2014). My findings challenge the conventional view that competition inherently fosters diversity in media content, suggesting instead that it may deepen entrenched biases when minority preferences diverge from majority tastes. Secondly, in a large cross-section of U.S. newspapers, Gentzkow and Shapiro (2010) demonstrate that newspapers systematically align their ideological slant with the political preferences of their local readership, highlighting how demand-side partisanship shapes media language. On the supply side, Djourelova (2023) shows that a top-down editorial policy, namely, the Associated Press's decision to ban the term "illegal immigrant", led to a widespread shift in language use across outlets. My paper contributes to this literature by showing that grassroots political actors, specifically civil rights leaders, can also shape language choices in the media. I further demonstrate that competition and market power condition how outlets respond to such pressure, revealing how structural features of media markets mediate the influence of both demand- and supply-side forces on language adoption.

This paper also contributes to the literature on the evolution of norms by showing how media behaviour, driven by local demand and market power, influences the diffusion of a new social label within the media landscape. While existing research has theorised both the role of prominent agents in disrupting entrenched norms (Acemoglu and Jackson (2015)), and the role of media in enhancing coordination and the creation of common knowledge (Chwe (2013)), while others have empirically measured how media behaviour is influenced by audience preferences (Gentzkow and Shapiro (2006); George and Waldfogel (2003)), to the best of my knowledge, no studies to date, have directly traced this process as it unfolds.

I show how Black civil rights leaders acted as prominent and observable agents driving the shift from "Negro" to "Black." My findings reveal how newspapers, shaped by economic incentives, either facilitated or resisted this label change and that this adjustment coincides with greater political representation of the initiating group. This empirical analysis provides unique insight into how media institutions, acting as both intermediaries and gatekeepers, influence the pace and extent of norm evolution in real-time. Understanding this dynamic is important because it shows how media behaviour can shape the discourse around social identities and reflect broader societal shifts. These insights remain relevant today, as media continues to play a pivotal role in shaping social and political identities in response to modern social movements and evolving cultural norms.

Third, this paper speaks to the evolution of racial attitudes in the U.S. South by examining the media's role in racial attitude formation and reinforcement (Key, 1949; Carmines and Stimson, 1981; Kuziemko and Washington, 2018; Ottinger and Winkler, 2022). My paper contributes to this literature by highlighting how media market competition influences the persistence of racial labels and the strategic use of agenda setting and tone, which may perpetuate prejudice in the Southern Black community. Importantly, this analysis also highlights how Black people's own chosen representations of themselves shaped these narratives, challenging external portrayals and contributing to shifts in societal norms. The broader implication is that media markets can act as powerful agents of norm preservation or change, maintaining or adjusting narratives that hinder progress toward racial equality, particularly when economic incentives align with maintaining the status quo.

My final contribution is to the applied-micro literature on Black descriptive representation. Existing work shows that removing formal voting barriers reveals sizeable effects on public spending and welfare (e.g., Cascio and Washington, 2014; Logan, 2018; Bernini et al., 2023a; Bernini et al., 2024;). I complement this institutional perspective by documenting a symbolic channel: conditional on identical franchise rules and baseline demographics, counties whose white-oriented newspapers adopted the self-chosen label "Black" sooner experienced larger subsequent gains in Black elected officeholding. Because Southern dailies primarily served white readers, the patterns are most plausibly driven by shifts in majority beliefs about the legitimacy of Black candidacies, rather than by direct mobilisation of Black voters. In this way, the paper links economic incentives in information markets to political incorporation, showing that profit-motivated editorial choices about words may meaningfully influence who ultimately wields local power.

The rest of the paper is structured as follows. Section 2 and Section 3 provide background and data details respectively. Section 4 explores diffusion of the new label and Black representation. Section 5 provides estimates for the effect of economic incentives on diffusion. Section 6 explores downstream effects of label adjustment on local Black government appointments and Section 7 concludes.

2 Background

In this section, I introduce the setting. I discuss the cultural significance, the label campaign and heterogeneity in preferences of the shift from "Negro" to "Black". I then motivate my

analysis by providing background regarding conflicting support for racial liberalism among White consumers and the market forces faced by newspapers.

2.1 The Campaign for "Black"

Throughout US history, various terms have been used to refer to Black people, with the standard preferential term transitioning from "Colored" in the nineteenth and early twentieth century to "Negro" until the late 1960s, then to "Black," and now perhaps to "African American"[1]. This is not dissimilar to the evolving labelling of other groups who have been the subject of prejudice[2].

Until the late 1960s, the term "Negro" was widely accepted by Black and White communities and commonly used to describe Black people. Martin Luther King Jr., for example, self-identified as "Negro" in his iconic "I Have a Dream" speech delivered in 1963, in which he used the term "Negro" 15 times and "Black" as an adjective four times. However, by mid-1966, the Student Nonviolent Coordinating Committee (SNCC) had begun planning a concerted ideological campaign to promote the use of "Black" as a new self-referential label, one they viewed as rejecting the legacy of paternalism and inferiority associated with "Negro."

SNCC's internal shift toward a Black nationalist frame can be traced to early 1966, when members of the Atlanta Project and the Vine City Project circulated internal memoranda critiquing white liberal influence and calling for autonomous Black leadership. These internal discussions culminated in position papers such as "The Basis of Black Power," which laid the ideological groundwork for SNCC's subsequent public realignment and helped frame the label "Black" as a political identity rather than a racial descriptor.[3] These documents echoed earlier discussions by Vine City and Atlanta projects field staff, who argued that terms like "Negro" were artefacts of white paternalism and should be abandoned in favour of language that emphasised collective dignity and agency.[4]

The Meredith March Against Fear in June 1966 offered an opportunity to launch the label shift. Originally initiated by James Meredith to promote voter registration, the march was reappropriated by SNCC and other civil rights organisations after Meredith was shot on the second day. SNCC organisers, including field secretary Willie Ricks, used the march as a platform to test and propagate the "Black Power" slogan in local meetings and speeches across Mississippi. Ricks later recalled, "We were building up to Greenwood. I had been

[1]Each of these terms to describe Black people have existed in the English language since US settlement (Smith, 1992).

[2]Other examples include, the evolution of nomenclature for the disabled community, "crippled", "handicapped", "disabled", "differently-abled", "disabled persons", "person with a disability"; the Indigenous peoples of North America, "Indian", "Indian American", and "Native American"; and American Arctic peoples of southern and western Alaska, "Yupik"; and northern and eastern Alaska and all of Canada, "Inuit".

[3]See "The Basis of Black Power," SNCC Position Paper, 1966. Accessible at: `https://www.crmvet.org/docs/6604_sncc_atlanta_race.pdf`. This document was also published in abridged form in *The New York Times*, August 5, 1966.

[4]See interview with Gwendolyn Zoharah Simmons in the SNCC Digital Gateway: `https://snccdigital.org/people/gwen-zoharah-simmons-robinson/`. See also Clayborne Carson, *In Struggle: SNCC and the Black Awakening of the 1960s* (Harvard University Press, 1981), pp. 213–214.

hollering 'Black Power' in all those little towns before we got there."[5]

On June 16, 1966, after being arrested for trespassing and released the same day, SNCC Chairman Stokely Carmichael addressed a nighttime rally in Greenwood, Mississippi. In a speech that would mark a rhetorical break from earlier civil rights discourse, he declared: "We been saying freedom for six years and we ain't got nothin'. What we get to start saying now is Black Power!" This moment, widely recognised as the public debut of the slogan, marked the culmination of SNCC's linguistic shift.

Both United Press International (UPI) and the Associated Press (AP) covered the speech, and their wire reports were syndicated nationally, appearing in hundreds of newspapers across the country. Yet even within this ostensibly uniform coverage, local editors exercised discretion over racial identifiers and political framing. For instance, two Southern newspapers, the *Odessa American* (Texas) and the *Tampa Tribune* (Florida), ran versions of the same UPI article on June 18, 1966. One described Carmichael as a "Negro leader," while the other labeled him a "Black nationalist leader." Such divergence, despite a common source text, illustrates that newspapers were already making discretionary linguistic choices at a critical early moment in the campaign. **??** presents the opening paragraphs of each article.

Coverage of the Greenwood speech contributed to national exposure of the "Black Power" slogan and the movement it symbolised. While the speech itself does not appear to have been broadcast live, its message and imagery reached wide audiences through follow-up reporting and television analysis. CBS News aired a special report later that year titled *Black Power, White Backlash*, which featured Stokely Carmichael and examined the ideological shift within the civil rights movement.[6] This national coverage helped cement the phrase in public discourse. Mark Whitaker argues that the slogan's emergence catalysed a broader cultural shift: 1966 marked a tipping point in Black self-presentation, as younger African Americans embraced Afros, dashikis, and a rejection of the term "Negro" in favour of more assertive and politically conscious identifiers.[7]

The adoption of "Black" was not merely rhetorical but was actively organised and disseminated. In the weeks following Greenwood, SNCC and its allies circulated newsletters, press statements, and community pamphlets that framed "Black Power" as a comprehensive cultural and political program. These materials, such as the July 1966 brochure *Black Power and the American Dream*, explicitly rejected the term "Negro," providing linguistic guidelines and encouraging solidarity under a newly assertive racial identity.[8] The campaign was visual as well: promotional posters, buttons, and newspaper columns featured the raised fist icon and slogans encouraging Black ownership of institutions and language.

This rebranding effort was explicitly strategic. SNCC and affiliated organisations urged

[5]Quoted in *Reporting Civil Rights, Part Two: American Journalism 1963–1973* (Library of America, 2003), p. 735. See also Clayborne Carson (1981), *In Struggle: SNCC and the Black Awakening of the 1960s*, Harvard University Press.

[6]See *Black Power, White Backlash*, CBS News Special Report, 1966. Available at: `https://www.youtube.com/watch?v=otMQeJFOAzM`.

[7]See Mark Whitaker, interview with NPR, February 8, 2023. "Stokely Carmichael and the story behind the phrase 'Black Power.'" Available at: `https://www.npr.org/2023/02/08/1155093955/mark-whitaker-black-panthers-stokely-carmichael-civil-rights-saying-it-loud-1966`.

[8]See *Black Power and the American Dream*, SNCC pamphlet, July 1966. Available at: `https://www.crmvet.org/docs/660000_bpw_stokeley.pdf`.

Black newspapers, churches, and student associations to abandon "Negro" in favour of "Black" as an act of psychological liberation and political resistance. While some local Black newspapers and student groups embraced the term and adopted internal style changes, uptake was uneven. In this sense, the label was not organically adopted but was the subject of coordinated persuasion, embedded in broader debates over integration, nationalism, and institutional control.[9]

This attempted redefinition was more than a linguistic shift; it was part of a broader project of self-determination, aimed at instilling pride, militancy, and resistance to White-dominated cultural narratives. While such ideals had antecedents in the work of W.E.B. Du Bois, Marcus Garvey, and Malcolm X, the 1966 campaign likely gained broader influence due to both the contemporaneous political moment and the media saturation of the Meredith March.[10]

Like many in-group labelling efforts, the campaign faced internal divisions. Martin Luther King Jr. cautioned that the "Black Power" slogan evoked "connotations of violence and separatism," and distanced himself from its implications in an October 1966 speech.[11] Others argued that the term's radical associations slowed adoption among moderate Black and White communities (Carmines and Stimson, 1981). Still, evidence suggests growing identification with "Black" among younger and more politically engaged African Americans. A survey conducted after the 1967 Newark uprising found that 50 percent of Black male rioters aged 15–35 identified as "Black," compared to 33 percent of non-rioters (Caplan and Paige, 1968). In a 1969 *Newsweek* poll of 977 Black adults, "Black" was more popular among younger respondents (Bennett, 1970; Poll, 1969). On the 1982 General Social Survey (GSS), 61 percent of those born after 1942 favoured "Black," compared to 40 percent of those born before 1933. Meanwhile, the 1979–80 National Black Survey found that disapproval of "Negro" and "Colored" increased steadily across younger cohorts. According to a 1968 *Newsweek* poll, "Negro" remained the majority preference, but by 1974, "Black" had become dominant.

2.2 Southern White Consumers and Newspapers

Why Might Southern White Consumers Care About Racial Labels?

From the Jim Crow era to the 1960s, racial conservatism dominated Southern society, reflecting a desire to maintain control and resist progressive change. The identity of Black citizens in the South was associated with historical connotations of hierarchy and subordination, while the shift to "Black" symbolised agency and empowerment by Black leaders. The 1960s forced progressive racial reforms upon the South, including the Civil Rights Act of 1964 and the Voting Rights Act of 1965. These acts dismantled segregation and enfranchised Black citizens but also exacerbated anxieties among racially conservative White Southerners. The political ramifications of these anxieties were significant. By 1980, the Southern exodus of Democratic voters was almost entirely driven by White racially conservative individuals, who abandoned the Democratic Party as it became associated with civil rights and racial liberalism (Kuziemko and Washington, 2018).

[9] See Joseph (2007), esp. Ch. 5.

[10] On the broader diffusion of these ideas, see Smith (1992), pp. 409.

[11] Speech available at: `https://www.crmvet.org/docs/6610_mlk_power-poverty.pdf`.

Why Might Southern Newspapers Care About Southern White Consumers?

Southern newspapers appear to have historically prioritised the preferences of their White audiences. With Black citizens representing a smaller proportion of the population and median White incomes in 1960 being over double those of Black citizens (USD 17,088 vs. USD 8,686, U.S. Census), it is plausible that newspapers had strong commercial incentives to cater to White readers.

Apologies issued by Southern newspapers since the early 2000s suggest that these outlets played a role in upholding racial conservatism. These apologies often reference institutional White supremacy, neglect in covering the civil rights movement, and the denial of basic courtesies, such as addressing Black citizens with titles like "Mr." or "Mrs." (Staples, 2021). Such actions may have reflected an effort to align with the racial attitudes of the dominant readership.

Black citizens, by contrast, were more likely to turn to Black-oriented radio for news, rather than newspaper or television (Torres, 2003; Acham, 2005; Boroghkozy, 2012). Analysis of data from the 1964 and 1968 American National Election Surveys, between which the Voting Rights Act was passed and newly enfranchised Black citizens may have been seeking more political information, suggests that Black newspaper readership was low, relatively stable and unaffected by enfranchisement.

In this context, it is worth considering whether racial labels themselves became a commercial consideration for newspapers. The decision to adopt or resist the term "Black" may not have been purely linguistic but potentially reflected strategic choices shaped by the preferences and prejudices of the White readership. Whether newspapers balanced their role as profit-driven businesses with the broader political and cultural implications of their editorial choices is a central question of this study (George and Waldfogel, 2003; Strömberg, 2004).

3 Data

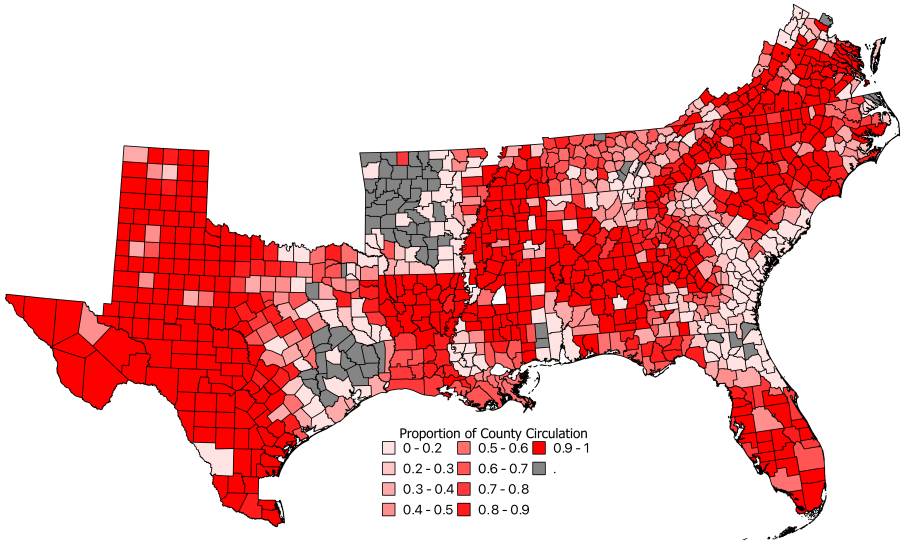
This section describes the construction of the newspaper-level panel dataset used to analyse media adoption of racial labels. The dataset combines article-level text from historical Southern newspapers with newspaper-county circulation data and measures of local White racial conservatism and market structure. I first detail the construction of the newspaper text data, including the dictionaries used to identify references to Black individuals and the creation of my main outcome variable. I then describe the sources and construction of the circulation-weighted measures of exposure to racial conservatism and competition, as well as additional county and newspaper-level covariates.

3.1 Newspaper text data

To measure newspaper adjustment in racial labels, I collect text data from two digital collections of historical US newspapers stored with: *newspapers.com* and *newspaperarchives.com*. These archives collectively provide the largest coverage of US historical newspapers currently available, featuring more than 1 billion pages from 30,000 different newspapers including urban and rural daily and less frequent newspapers. In the South, with coverage from before

and after June of 1966, the archives have data from 450 newspapers with headquarters in 265 counties, comprising 170 daily newspapers and 280 non-daily newspapers. This represents approximately 42% of all daily newspapers circulating at the time and more than 50% of total circulation for 70% of all Southern counties. Figure 3.1 shows the proportion of county circulation that these newspapers cover in the Southern 11 confederacy states[12]. These archives carry a key limitation from an analysis perspective in that the archiver scans newspapers generating text using optical character recognition (OCR). This allows the user to search for specific phrases but not access all of the text in a digital format for deeper quantitative textual analysis. As a result, the primary measure of racial ideology in this study is based on the relative frequency of racially descriptive phrases across newspapers. I supplement this with full-text data from a sub-sample of articles to enable deeper linguistic analysis.

Figure 3.1: Proportion of total county daily newspaper circulation provided by newspapers archived on *newspapers.com* and *newspaperarchive.com*.



Full-text corpus. I introduce a novel corpus of 58 million newspaper articles covering 267 newspaper to: 1) generate dictionaries for counting label frequencies and 2) conduct deeper textual analysis. This corpus was created using a deep learning pipeline that incorporates layout detection, legibility classification, OCR, and the association of article texts spanning multiple bounding boxes. It is part of a larger forthcoming novel corpus of more than 200 million articles covering local newspapers in the USA from 1960 - 1975, and the subject of a forthcoming paper, which provides significantly more detail. In short, the pipeline is as follows: (1) Identify headline, author, article, image, and other bounding boxes using a pre-trained deep learning model (Dell et al., 2024); (2) combine article bounding boxes which may span multiple columns with correct headlines which may span multiple columns; (3) extract article text (Tesseract) and check for legibility; (4) calculate total number and physical space allocated for each type of bounding box on the page (e.g. advertisements, articles, etc.); (5)

[12]The 11 Southern states of the confederacy include South Carolina, Mississippi, Florida, Alabama, Georgia, Louisiana, Texas, Virginia, Arkansas, Tennessee, and North Carolina

article topic classification using a combination of BERT Topic and LDA; (6) source (i.e. AP, UPI, local, etc.) classification using using a fine tuned BERT model, and (7) identification of Black entities using a fine tuned BERT model. An example of bounding box and article-headline detection can be see in **??** and resulting article level output can be seen in Section **??**.

Dictionaries. While "Negro" does not require complementary language to disseminate meaning, mentions of the word "Black" referencing Black people may be confused with many other meanings. To reduce noise and maximise the likelihood of obtaining all references in the sample, I utilise the large corpus to build a "Black" dictionary and conduct some more sophisticated textual analysis.

The pipeline for dictionary creation is as follows: I, firstly conduct a pre-processing exercise to identify articles which are about Black people and to reduce noise[13]. I then take a random sample of 3000 sentences which contain the word "black" from 2000 articles from the pre-processing stage and manually code each story as either relevant (1) or not relevant (0) to a Black person, community or group. I then fine tune a Bidirectional Encoder Representations from Transformers (BERT) model, which is pre-trained on general language understanding tasks for sequence classification, to specifically identify articles related to Black people and Black issues using the pre-coded training data. The model demonstrated a final validation accuracy of approximately 96.4%. I then release the model on the remaining articles. I check for consistency by randomly selecting 1000 articles and manually determine accuracy of approximately 98%. There are 1 million articles in the sub-corpus which make reference to a "Black" or "Negro" entity.

I identify the top 100 unique bi-gram and tri-gram noun phrases and adjective-noun combinations containing the word 'black' using a combination of TF-IDF vectorisation and the RAKE (Rapid Automatic Keyword Extraction) algorithm to create the "Black" dictionary. In a robustness exercise, I exclude phrases such as "Black Power" and "Black Panthers" to ensure that the dictionary captures references to Black people, communities, and institutions rather than ideological or political references to the Black Power movement. This helps assess the extent to which results are driven by broader racial labelling rather than discourse specific to the movement. The complete dictionaries used are presented in **??** and n-gram frequency in **??**.

I count the frequency of these 100 phrases as well as all mentions of the word 'negro' per newspaper×month from the universe of newspapers in the available archive in the 11 Southern states for which there are articles from before and after the second semester of 1966 over the period 1960 – 1973.

Index. My main outcome variable is an index measuring the relative frequency of the "Black" dictionary out of all references to Black people using either "Black" or "Negro". An increase in this index represents an increase in the relative use of the term "Black" compared to "Negro". The index is calculated according to the equation:

[13]I filter out all articles which do not contain the words black or negro. Using the Black subset, by article, I remove instances of black preceded by mr or mrs (caps and period ignored), instances of blacks followed by any other letter, to remove surnames and other uses, i.e. Blackwood, blacksmith, etc., and any instance preceded by a list of the 1000 most common first names in the USA, e.g. John Black. If the article then still contains the word black, black- or blacks, it is maintained.

$$\widetilde{black}_{nt} = \frac{\sum_{p_b=1}^{100} black_{p_b nt}}{\sum_{p_b=1}^{100} black_{p_b nt} + \sum_{p_g=1}^{2} negro_{p_g nt}} \tag{1}$$

Where $black_{p_b nt}$ is the frequency of Black dictionary phrase $p_b = 1, 2, \ldots, 100$ in newspaper n in period t; and $negro_{p_g nt}$ is the frequency of Negro dictionary phrase $p_g = 1, 2$ in newspaper n in period t. A relatively lower score of $\widetilde{black}_{nt}$ is interpreted as newspaper n using more racially conservative language in period t [14].

Black Narratives. The linguistic shift from "Negro" to "Black" was not merely symbolic; it was part of a larger act of cultural and political resistance, aimed at reshaping how Black identity was perceived both within the community and in broader society. Scholars have highlighted how shifts in terminology can signify broader changes in societal attitudes, with specific terms carrying distinct connotations of agency, identity, and respect (Fairclough, 1989; Omi and Winant, 2014; Hill, 2008). Firms may not only choose which label to use, they may choose how to represent Black entities in the narrative formation of articles. Further, firms may choose to employ strategic framing to balance adherence to evolving norms with consumer preferences. To explore this, I look for nuanced linguistic changes by exploring whether newspapers vary their use of "Negro" and "Black" in ways that reflect or diminish the agency of the Black community and whether the sentiment associated with these terms portrays the community in a positive or negative light.

Agency, in linguistic terms refers to how much control, power, or autonomy is attributed to a group or individual through language (Fairclough, 1992). The "Black Power" movement emphasised autonomy, self-determination, and pride in Black identity, actively rejecting narratives of passivity imposed by dominant societal structures. This ideological shift may have influenced the linguistic representation of Black individuals and communities, encouraging portrayals that reflected greater agency and empowerment. To systematically analyse variation in linguistic agency, I assess sentence structures in the large corpus of Black articles to see if Black individuals, communities or groups are placed as the subject (agents) in Subject-Verb-Object (SVO) constructions. For example, phrases like "Black activists organised a protest" (affirming agency) versus "A protest was organised by Black activists" or "A protest occurred among Black communities" (diminishes agency).

Representing an entity with more agency does not immediately mean that it is a positive representation. Agents can be represented as doing positive or negative actions as well as having positive or negative actions done to them. To ensure I am capturing this phenomenon, I also classify the sentiment around the entity.

I dissect the article level Black entity corpus to the sentence level, and keep only the sentences which reference Black entities. There are 2.5 million sentences in this sample. I then ask Chat-GPT-4o to: 1) identify the Black entity in a sentence; 2) determine whether

[14] Due to resource constraints, it was necessary in some instances to collect the number of pages which mention each phrase rather than the total number of mentions of each phrase. To maintain consistency, estimates capture the number of pages for all newspapers.). To explore the potential measurement differences between using the number of mentions and the number of pages, I collected both mentions and pages counts per semester on a sub-sample from the 11 Southern states from *newspapers.com*. This allowed me to directly measure systematic differences between the two measures. My sample consists of 227 newspapers covering 7792 newspaper semesters. The two measures have a total correlation of 0.986, with average correlation of 0.909. **??** presents the scatter plot of the two different measures.

the entity is a person, group, organisation or object; 3) to code the entities linguistic agency in the sentence as either high, neutral or low (i.e. exhibiting affirming, neutral or diminishing linguistic agency); and 4) to code the entities sentiment in the sentence as either positive, neutral or negative. I also ask ChatGPT do give a rating on how confident it is with the overall classification as "High", "Medium" or "Low" and then restrict the sample to only "High" confidence classifications (see Appendix **??** for Chat GPT prompt). Using ChatGPT-4o for this task provides unparalleled scalability and accuracy compared to traditional manual coding or pre-trained machine learning models, as it combines nuanced linguistic understanding with adaptive contextual reasoning, while drastically reducing the time and resources required for sentence-level annotation. Assessment of the accuracy of a random sample of Chat-GPT "High" confidence classifications revealed a $> 99\%$ match to human coding with the same instructions. Examples of specific classifications can be seen in **??**.

It is important to acknowledge the potential for biases in large language models (LLMs), as these models inherently reflect the political, cultural, and ideological perspectives embedded in their training data.[15] If Chat-GPT is aware of the historical relevance of the shift to "Black" then it may attach meaning beyond what is provided in the sentence. I mitigate this two ways: 1) I mask the date and any newspaper metadata for the input and only provide the sentence and instructions; and 2) I replace references to groups with an alternative reference name, e.g. "Black" with 'XXX', "Negro" with 'YYY' and White with 'ZZZ', so that the AI is unaware of which race or group of people are being referred to and with which label.

3.2 Newspaper Incentives Data

Demand - White racial conservatism. To measure newspaper exposure to White racial views, I exploit the county vote share for the States' Rights Democrats (or 'Dixiecrats') candidate Strom Thurmond in the 1948 Presidential election (see Figure 3.2 for spatial vote share distribution). There are some clear indications that a vote for the States' Rights Democrats represented a vote for White racial conservatism in 1948. Firstly, considering that in the Southern states Black citizens were effectively restricted from voting highlights that this vote represents the vote share of the White population. Secondly, the States' Rights Democrats was formed to directly combat federal government intervention in state affairs regarding Jim Crow laws and the segregation of Black and White citizens (Carmines and Stimson, 1981). Thirdly, Strom Thurmond regularly projected fiercely segregationist sentiment, at least during the period for which he was running (Bass and Thompson, 2006)[16]. While it is likely that the individual level sentiment towards segregation and civil rights may have evolved between 1948 and the period studied, given the long history of racial conservatism in the South, it is likely a good proxy for average between county variation in White racial conservatism. I am not the first to use this measure to proxy White Racial

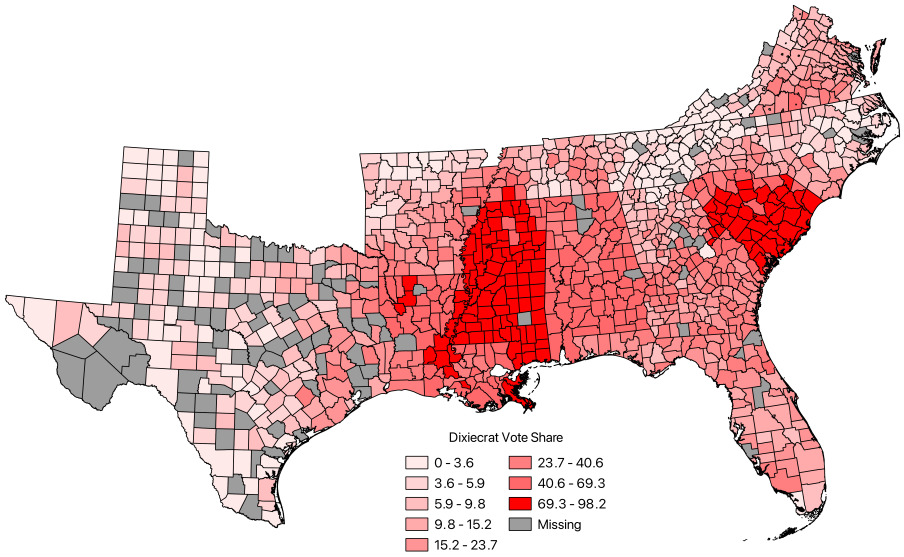
[15]The following studies investigate the political and ideological views embedded within LLMs: (Miotto et al., 2022; Fischer et al., 2023; Santurkar et al., 2023; Ren et al., 2024; Choudhary, 2024; Retzlaff, 2024; Rozado, 2024; Röttger et al., 2024; Moore et al., 2024).

[16]In his acceptance speech in Birmingham, Alabama, Strom Thurmond stated, "I want to tell you, ladies and gentlemen, that there's not enough troops in the army to force the Southern people to break down segregation and admit the Nigra' race into our theatres, into our swimming pools into our homes and into our churches." (Bass and Thompson, Strom, 117)

conservatism (see Kuziemko and Washington (2018)). The data are taken from the Inter-university Consortium for Political and Social Research (ICPSR) General Election Data for the United States, 1950-1990.

Given the potential for temporal measurement error, I perform robustness tests utilising several other variables to proxy for county level racial conservatism: 1) whether the county made use of racially discriminatory voter registration test or device in the 1964 presidential election[17], estimated county proportion of slaves from 1860 and, from Acharya et al. (2016), local measures of the agricultural suitability to grow cotton which are predictive of slave-based production of cotton as well as contemporary differences in political attitudes across counties in the American South.

Figure 3.2: Variation in county vote share for the 'Dixiecrat' party (1948)



Competition. To measure circulation, I digitise a novel dataset of daily newspaper circulation collected by the Audit Bureau of Circulation (ABC) from 1964. Since the ABC does not collect geographically disaggregated data for low-circulation newspapers, in similar fashion to Djourelova (2023), I impute these observations with data on total circulation from the USNP (Gentzkow et al., 2014), assuming that small newspapers circulate mainly in the county of their headquarters.[18] I aggregate daily newspaper circulation from 1964 to the county level and then calculate the market share for each newspaper outlet-county pair. I use these shares to calculate county newspaper market concentration using a Herfindahl-Hirschman Index (HHI).[19] Spatial variation in county level HHI are in Figure 3.3. As a

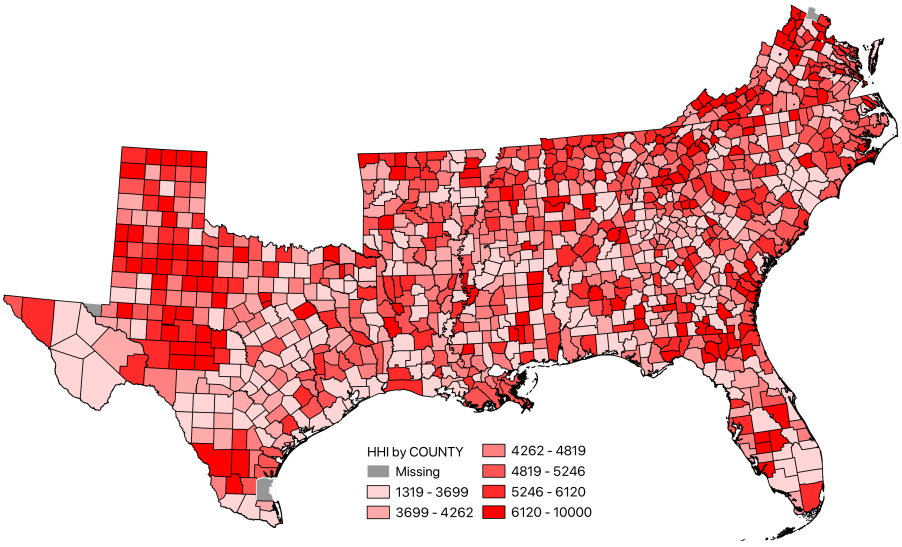
[17]Considering that most congressional districts mapped precisely to county borders during the period and that large proportions of districts which implemented tests or devices are clustered and map directly to states borders (see **??**), one can think of this variable as being largely representative of a state level treatment representing more widespread and systemic racial conservatism.

[18]I assess this assumption for all newspapers for which I have both the ABC and USNP data. The proportion of newspapers with the highest proportion of circulation in their county of commercial residence is 0.93. The correlation between total circulation and the number of counties circulated in is 0.84.

[19]$\text{HHI} = \sum_{i=1}^{N} s_i^2, \quad 0 \leq s_i \leq 100, \quad 0 \leq \text{HHI} \leq 10,000$, where N is the total number of firms in

second measure, I calculate the total number of daily newspapers with headquarters in a given county in 1964. This allows one to, at least crudely, measure the extensive margin of competition in the county of commercial residence.

Figure 3.3: HHI by county in 1964



Weighting Newspaper Exposure. To precisely measure newspapers' exposure to WRC and competition, I proxy newspaper exposure to county level variables by weighting them by the proportional circulation of that county-newspaper pair out of each newspaper's total circulation. For analysis, I focus on the top three counties with the highest circulation for each newspaper, considering that for more than 95% of the newspapers in my sample, more than 90% of circulation is in the top three circulated counties[20]. Circulation weighting is calculated as follows:

$$M_n = \sum_{i=1}^{3} \left(\frac{\text{Circulation}_{nc_i}}{\sum_{j=1}^{3} \text{Circulation}_{nc_j}} \right) M_{c_i}$$

where: M_n is the weighted measure for newspaper n; $\text{Circulation}_{nc_i}$ is the circulation of newspaper n in county c_i, weighted by its proportion within the top 3 counties; M_{c_i} is the county-level measure (e.g., 'Dixiecrat' vote share) in county c_i, with $i = 1, 2, 3$ representing the top 3 counties circulated in by newspaper n.

Additional county data. To account for time-varying county level differences which may correlate with the White racial conservatism and drive newspaper labelling decisions, I supplement my analyses with additional data from various sources including an indicator for the presence of an National Association for the Advancement of Colored People (NAACP) chapter using data from Matthews and Prothro (1966), interacted

the market and s_i is the market share of firm i.

[20] In a robustness exercise I show that estimates using both the vote share from the county with the highest share as well as the weighted average of all circulated counties, does not effect the interpretation of results.

with age/race demographics from the 1960 US census, White and Black education and literacy rates, county median income levels from from the Bureau of Labor Statistics, county presidential vote shares, the proportion of church goers, and the proportion of agricultural land. Appendix A provides an overview of all data sources.

Additional newspaper data. To account for differences between newspaper which may confound estimates, I collect data on whether a newspaper prints in the morning or evening, the proportion of a newspaper's articles which cover Black entities in 1965, and the proportion of a newspaper's articles written by newswires services vs. local journalists in 1965.

4 Label and Narrative Diffusion

Label Diffusion. I adopt two approaches to test for the presence of a shock to newspapers' label choice. The first makes use of the Black and Negro corpus already described but limited to newspapers in the South. The approach is to ask which words and phrases are predictive for whether an article was published before or after June 16, 1966. Let $f_{p,l,\text{before}}$ and $f_{p,l,\text{after}}$ denote the total number of times phrase p of length l (one to three words) is used before and after June 16, 1966, respectively. Let $f_{\sim p,l,\text{before}}$ and $f_{\sim p,l,\text{after}}$ denote the total occurrences of length-l phrases that are not phrase p, before and after June 16, 1966, respectively. The χ^2 statistic for phrase p of length l is then:

$$\chi^2_{p,l} = \frac{\left(f_{p,l,\text{before}} f_{\sim p,l,\text{after}} - f_{p,l,\text{after}} f_{\sim p,l,\text{before}}\right)^2}{\left(f_{p,l,\text{before}} + f_{p,l,\text{after}}\right)\left(f_{p,l,\text{before}} + f_{\sim p,l,\text{before}}\right)\left(f_{p,l,\text{after}} + f_{\sim p,l,\text{after}}\right)\left(f_{\sim p,l,\text{before}} + f_{\sim p,l,\text{after}}\right)}.$$

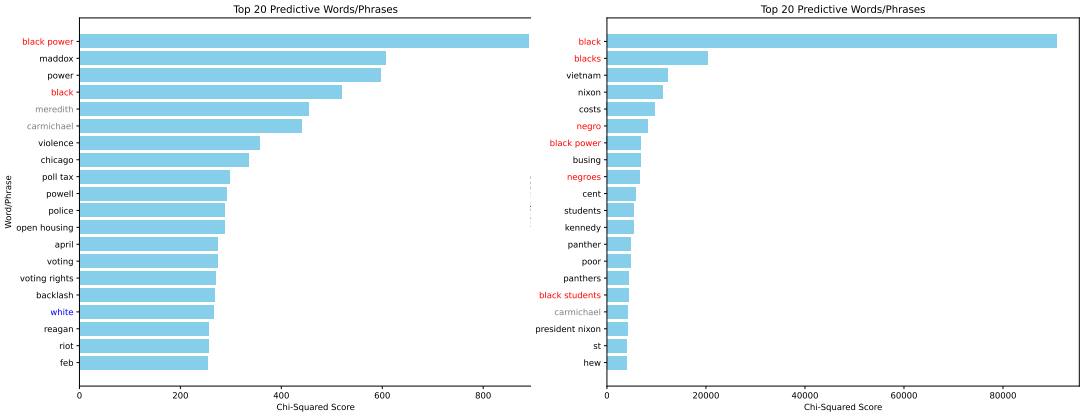
Figure 4.1a limits the analysis to two months prior and post Stokeley Carmichael's first Black Power speech. "Black Power", "Black" and "Carmichael" emerge as some of the most predictive of the articles publication date. In Figure 4.1b, I extend the sample from 1960 to 1973. Not surprisingly, "Black" and "Blacks" are by far the most predictive n-grams from before to after the campaign launch as are "Negro" and "Negroes" signalling the longer term substitution.

Secondly, I assess the Black dictionary and racial index for trend breaks in June of 1966. From May to July of 1966, the Black dictionary and racial index increase by 300% and 200%, respectively. A test for no structural breaks at that point in time using the methods introduced by Chow (1960) is rejected at the 1% significance level for both the Black dictionary and racial index measures. The evolution of the dictionaries and racial index are presented in Figure 4.2 and Figure 4.3.

Between January 1960 and May 1966, the average proportion of the "Black" dictionary's use is 0.1076. In the period between June 1966 and June 1973, the average monthly proportional use increases to 0.897 with an average over the entire post-period

Figure 4.1: Phrases predictive of pre- versus post-campaign launch

(a) May - August (1966) (b) Whole period



Note: Changes in the words and phrases used by the Southern newspapers coverage of Black people, communities, and groups before versus after the launch of the campaign for "Black". Phrases most predictive for whether an article covering Black people, communities, or groups is published before or after June 16, 1966. Top 20 n-grams ($n \in 1, 2, 3$) based on χ^2 test statistic.

of 0.568. The rise in "Black" dictionary usage drives the change in the racial index between the second half of 1966 and the second half of 1968, while it is both an increase in the "Black" dictionary and a decrease in the "Negro" dictionary, from after the first half of 1968 onwards. "Black" dictionary usage in the period before June of 1966, is rare, with an average of 9.4 mentions per newspaper per month increasing to 73.6 after. The use of the term "Negro" fluctuates in the years before 1966 as one might expect, cycling in the prevalence of national news covering Black issues. For example, the largest spikes to the use of the "Negro" dictionary occur in June of 1963, following President Kennedy's address to the nation publicly asserting his support for racial equality on moral grounds; and in July of 1964 and August of 1965 following the signing of the Civil Rights Act and Voting Rights Act, respectively. Importantly, the same fluctuations are not seen in the "Black" dictionary.

This evidence appears conclusive that the adoption of the term "Black" is triggered by the public national call from civil rights leaders beginning in June of 1966. The continued use of the term "Negro" until the latter half of 1968 can be attributed to significant cultural and institutional inertia, i.e. the persistence of norms (Acemoglu and Jackson, 2015). As Acharya et al. (2016) and Woodward (1955) illustrate, deeply entrenched racial norms and practices in the South have historically been slow to evolve, with societal change often lagging behind calls for reform. Moreover, the political acceptance of the "Black Power" movement as a non-violent and justifiable position for Black citizens was not widespread until after 1968 (Jeffries, 2006). This suggests that newspapers may have been responding cautiously to evolving societal attitudes, collectively dropping the term "Negro" only once broader political and cultural legitimacy for the term "Black" had been established. However, as shown in **??**, the pace of adop-

tion varied significantly across Southern counties, highlighting the need to investigate whether the incentives facing newspapers also varied spatially.

Figure 4.2: The evolution of racial dictionaries in Southern local newspapers (1960 - 1973)

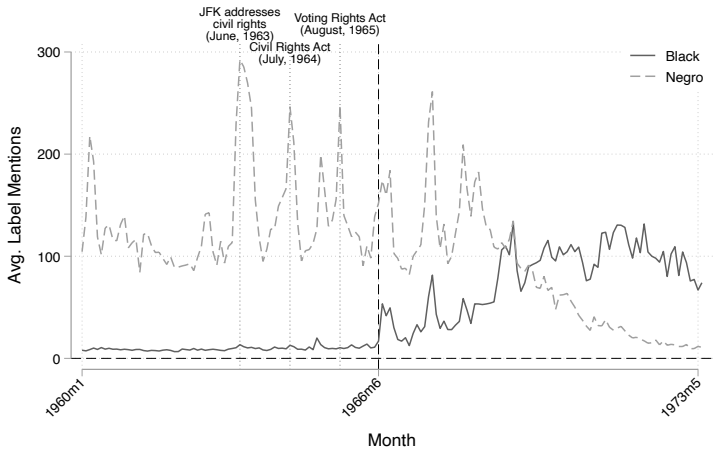
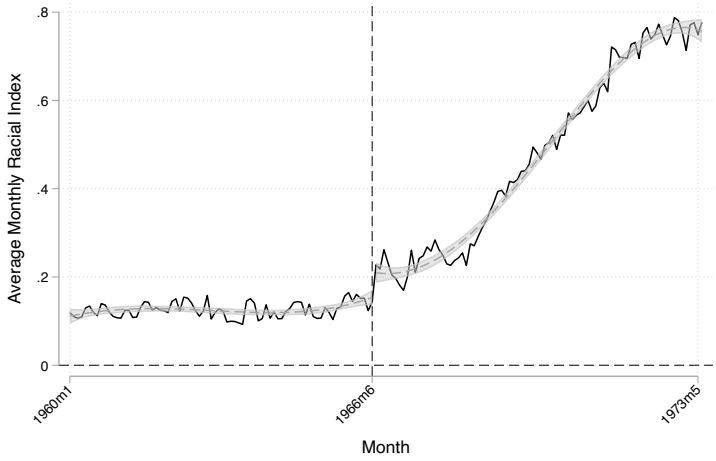


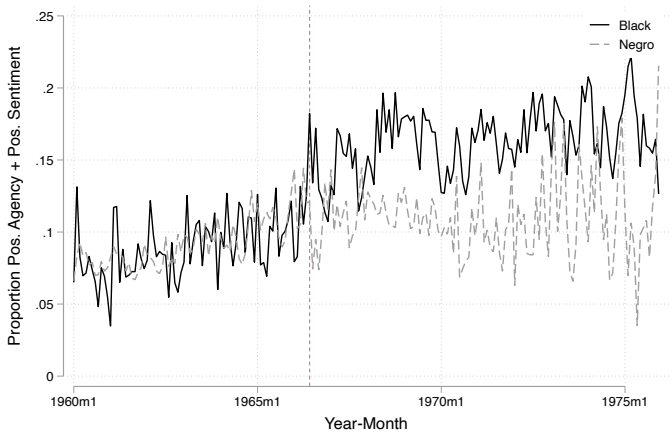
Figure 4.3: The evolution of racial index in Southern local newspapers (1960 - 1973)



Narrative Diffusion. I next examine whether the adoption of the term "Black" led to systematic changes in how Black individuals and communities were represented in newspapers. If the choice of label affects the agency and sentiment associated with described entities, this would suggest that language change is not merely symbolic, but plays a substantive role in shaping social norms. **??** plots the evolution of average agency and sentiment over time. Agency scores for Black and Negro entities rise steadily throughout the period, while sentiment remains flat until around 1966, after which it increases consistently through the early 1970s. Figure 4.4 highlights a sharp divergence in the likelihood that a Black entity is represented as both high-agency (actively "doing") and associated with positive sentiment, compared to a Negro entity. The timing of the divergence aligns closely with the launch of the Black Power

campaign.

Figure 4.4: The evolution of Black entities represented positively while 'doing' in Southern Newspaper sentences (1960 - 1975)



Empirical Strategy. To estimate the effect of the Black Power movement on media representation of Black entities through the channel of label adoption, I implement a generalised dynamic difference in differences strategy:

$$(Ag - Sent)_{ejnt} = \sum_{s=-k}^{k} \beta_s \, Black_{ejnt} \times Semester_s + \delta_{n,t} + \epsilon_{nct} \tag{2}$$

Where $(Ag - Sent)_{ejnct}$ is an indicator variable equal to 1 if there is both high agency and positive sentiment surrounding entity e in sentence j of newspaper n on day t; $Black_{ejnt}$ is an indicator variable equal to 1 if the entity is represented using "Black," and 0 otherwise; $Semester_s$ are indicators for each semester relative to 1965h2; and $\delta_{s,t}$ are Newspaper×Day fixed effects.

The identifying assumption is that, in the absence of the Black campaign, representation of Black entities would have evolved in the same manner with either the "Black" or "Negro" label. The time series analysis in Figure 4.4 is highly supportive of this assumption.

I further estimate two complementary specifications designed to identify whether the adoption of the "Black" label was associated with a broader shift in representational tone. The first restricts the sample to 850,576 sentences from 260,513 articles containing both "Black" and "Negro" entities, and compares the average agency and sentiment of those entities between sentences within the same article (see **??** for examples). This isolates variation in label choice within a fixed journalistic context, holding constant outlet, author, and topic, and mitigates concerns that more progressive journalists may self-select into adopting the "Black" label. The second strategy follows 3,017 named entities that appear both before and after the campaign, and estimates the differential change in their representation depending on whether they are labeled

"Black" or "Negro" over time within the same newspaper.[21] This design holds constant the identity of the entity and outlet, and avoids conflating representational change with the appearance of new movement-specific actors like "Black Power."

Both strategies help distinguish linguistic adoption from broader discursive shifts. By excluding movement-linked phrases (e.g., "Black Power", "Black Panthers") from the dictionary and isolating variation within articles and within recurring entities, the analysis shows that the observed shifts in narrative tone are not simply a function of movement discourse. Rather, they reflect more general changes in how Black people, communities, and groups were portrayed in local news coverage.

The identifying assumption is that, absent the campaign, entities labeled "Black" would have experienced similar trends in representational tone as those labeled "Negro." The time-series evidence in Figure 4.4 supports this, showing parallel trends prior to the campaign and a sharp divergence beginning in mid-1966.

Results. Results in Figure 4.5 indicate that, within a newspaper on a given day, i.e. a newspaper issue, beginning in the semester directly following June of 1966, there is a statistically significant increase in the probability of a "Black" entity being represented with high agency and positive sentiment above a "Negro" entity. The effect is persistent over the entire period. Estimates in **??** indicate that the effect is largely driven by references to group entities (e.g. "Negroes" or "Black protestors"), rather than individual people (e.g. "black man" or "Stokely Carmichael"), or organisations (e.g. "the Black caucus", "Black Panthers" or "The Black Power Movement").

To assess whether this shift in linguistic agency was primarily associated with local editorial decisions or reflected broader national norms in journalistic language, I compare sentences written by national newswire services to those written by local journalists. Results in **??** show no statistically significant difference between the two groups. This suggests that the observed shift in agency was not isolated to local editorial choices, but reflected a broader discursive change in how Black entities were represented across both national and local coverage.

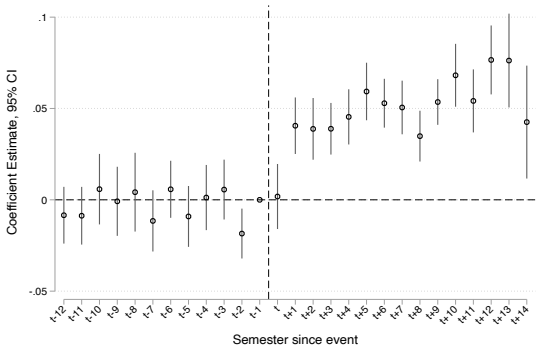
Utilising the restricted within-article sample of articles featuring both "Black" and "Negro" entities in the same journalistic context, as shown in **??**, and a panel of recurring newspaper-entity pairs (**??**), I estimate how representation differs across labels while holding constant outlet, topic, and entity identity. Results from both designs are statistically similar to the full sample and indicate that journalists were more likely to associate the term "Black" with higher agency and positive sentiment in the wake of the 1966 campaign. This pattern is consistent with the interpretation that the campaign served as a discursive turning point, coinciding with a shift in how Black individuals and communities were portrayed.

While the analysis isolates representational change associated with the adoption of the "Black" label, a key identifying assumption is that, in the absence of the 1966

[21] Examples include "black man", "blacks", "the black player", "negroes".

campaign, entities labeled "Black" and "Negro" would have followed similar trends in agency and sentiment. The parallel trends observed in the time series prior to mid-1966 lend support to this assumption. Nevertheless, it is important to acknowledge that the 1966 shift could coincide with other structural or discursive changes, such as evolving editorial norms, broader civil rights developments, or national political events, that may have differentially influenced the representation of entities labeled "Black." Although the within-article and within-entity designs help rule out compositional bias and self-selection into label use, they cannot fully eliminate the possibility of coincident shocks. As such, while the timing and pattern of the divergence strongly implicate the campaign as a focal point for discursive change, I interpret the evidence as identifying the effects of a broader narrative shift catalysed by, but not necessarily uniquely caused by, the campaign.

Figure 4.5: Dynamic DiD - High agency and positive sentiment on Black label used for entity.



Note: The figures present the average year-semester difference in the probability of a Black entity being represented as being a combined high agency and positive sentiment between when the entity is described using "Black" vs. "Negro" compared to 1952h2. Regression includes newspaper-year-day fixed effects. Standard errors are clustered at the firm level.

5 Media Incentives

In this section, I provide a conceptual framework for studying the effects of firm level incentives on label adoption. I then empirically test for the variation in diffusion rates by demand, supply and market competition.

5.1 Conceptual Framework

I conceptualise firms decisions in a framework that draws on the spatial competition literature, the literature on two-sided markets, and game theoretic herd behaviour, to predict newspaper behaviour. I focus on the commercial incentives of newspaper outlets rather than the political, as there is evidence that commercial considerations likely dominated political factors at the time of my study Baldasty (1992).

Firms are primarily supported by advertising revenue and compete for readers within the demographics most desired by advertisers Armstrong (2002); Ambrus and Reisinger (2006); Anderson and Waldfogel (2011). Large income inequalities between Black and White citizens as well as racial imbalances in business ownership put White readers at the centre of demand.[22] Consumers, heterogeneous in their label preferences, choose the firm that aligns most closely with their prior beliefs, as seen in models by Gentzkow and Shapiro (2006, 2007) and Mullainathan and Shleifer (2005).

Firms, as members of the wider media institution, have an incentive to align themselves with industrial standards (Banerjee, 1992). This could be explained by supply side agents wanting to maintain the credibility of the industry, maintain their own credibility within the industry, to simplify editorial decision-making by relying on industry norms, and/or relying on the wider industrial resources to better predict the preferences of future agents. I assume that this industrial trend is exogenous to any single local newspaper's marginal choice.

Firms must balance the incentive to compete for consumers and the incentive to remain aligned with the institution. I assume that competition strengthens the incentive for newspapers to align with consumer preferences in order to protect their revenue streams, as consumers are more likely to switch to competitors that better reflect their social and political preferences. However, as institutional norms evolve, newspapers may face increasing pressure to conform to industry standards to avoid missing the benefits associated with herd behaviour.

Lastly, I assume that owners, editors, and journalists' non-pecuniary incentives can influence output choices, but are unlikely (Gentzkow and Shapiro, 2010).

Predictions of the Framework

Heterogeneous Adoption Rates: Newspapers targeting audiences with varying racial attitudes are expected to adopt the label "Black" at different rates. In areas with higher racial conservatism, adoption may be slower to avoid alienating traditional readers, with newspapers strategically maintaining the term "Negro" longer. Conversely, newspapers in less conservative regions are likely to adopt "Black" sooner, reflecting the preferences of their local readership. Competitive pressures may amplify this effect, particularly where readers might switch to competitors that better match their racial attitudes.

Gradual Industry Convergence: Whichever label becomes entrenched as the industry norm, newspapers will converge towards that label, driven by institutional pressures, e.g. professional standards, or the reputational risks of clinging to outdated language.

Although this framework does not fully capture the complexities of cultural hegemony (Gramsci, 1971), the concept that firms not only respond to consumer prefer-

[22]The median income of White citizens in the South was more than twice that of non-white citizens at the 1960 census. Median annual income for White citizens was USD17,088, while for non-white citizens median income was USD 8,686 Bureau (1960)

ences but also actively shape and reinforce them, it provides a structured approach to analysing adoption patterns. Historical evidence suggests that racially conservative attitudes in the South have deep roots, with a long-standing history of white racial conservatism shaping public opinion and institutional practices (Woodward, 1955; Key, 1949). These attitudes persist today, as evidenced by higher rates of racial prejudice and segregationist sentiment in these regions compared to other parts of the country (Acharya et al., 2016; Stephens-Davidowitz, 2014). While early adoption in traditionally conservative regions could theoretically accelerate acceptance of progressive norms among the dominant group, my data constraints limit this analysis to pre-treatment measures, sidestepping potential feedback effects from norm reinforcement. By exploiting the competition channel, I partially mitigate concerns about feedback effects, as competition intensifies the focus on responding to existing consumer preferences rather than actively shaping them. This helps isolate the adoption patterns from broader normative shifts influenced by firms themselves.

5.2 Empirical Strategy

Empirical Strategy To estimate the effect of White racial conservatism on the adaptation of racial labels, I implement a generalised difference-in-differences strategy:

$$\widetilde{black}_{nt} = \beta(WRC_n \times Post_t) + \mathbf{\Omega}^{'}\mathbf{X}_c + \mathbf{\Phi}^{'}\mathbf{Z}_n + \gamma_n + \delta_t + \epsilon_{nt} \tag{3}$$

Where $\widetilde{black}_{nt}$ is a measure of the frequency of the "Black" racial label, relative to the "Negro" label, in newspaper n in year-month t; WRC_n represents newspaper exposure to pre-treatment White racial conservatism; $Post_t$ is a dummy variable = 1 in the post treatment period beginning in the second semester of 1966, and 0 otherwise; X_c and Z_n are pre-treatment period county and newspaper level controls, respectively, interacted with year-month fixed effects; γ_n and δ_t are newspaper and year-month fixed effects respectively; and ϵ_{nt} is an idiosyncratic error. The identifying assumption is that the evolution of racial labelling by newspapers in counties with high racial conservatism and counties with low racial conservatism would have followed conditional parallel trends in the absence of a preference shift.

To measure the conditional effect of market power on racial label adoption, I apply a triple differences estimation:

$$\begin{aligned}\widetilde{black}_{nct} = \beta_1(WRC_n \times Post_t \times Comp_n) \\ + \beta_2(WRC_n \times Post_t) + \beta_4(Comp_n \times Post_t) \\ + \mathbf{\Omega}'\mathbf{X}_c + \mathbf{\Phi}'\mathbf{Z}_n + \gamma_n + \delta_t + \epsilon_{nct}\end{aligned} \tag{4}$$

Where $Comp_n$ denotes pre-treatment measures of competition and market power that a newspaper is exposed to. The identifying assumption is that, conditional on controls and fixed effects, pre-1966h2 county newspaper market structure is exogenous

to subsequent changes in the relative use of the racial label "Black" versus "Negro", and that, in the absence of the 1966h2 shock, trends in label usage would have evolved similarly across levels of market competition, irrespective of pre-treatment levels of racial conservatism.

Table A.1 reports the cross-sectional relationship between Thurmond vote share and a range of historical and socio-demographic variables at the county level. These estimates serve two purposes. First, they help characterise the nature of racially conservative media markets. Counties with higher Thurmond support tend to have larger Black populations, deeper legacies of slavery and racial violence, lower income and education, more rural settlement, and stronger Protestant presence. These patterns align with prior findings on the persistence of racial attitudes in the American South (e.g., Acharya et al., 2016; Kuziemko and Washington, 2018).

Second, this table anticipates a key identification concern: that White Racial Conservatism (WRC) may be correlated with other determinants of media behaviour. The balance regressions show that WRC is not randomly distributed. However, the empirical strategy mitigates this concern through several design choices. First, the main regressions control flexibly for observed covariates such as racial composition, income, education, and religious structure, all interacted with time. Second, I absorb newspaper fixed effects to capture unobserved time-invariant confounders. Third, I include state × month-year or DMA × month-year fixed effects to account for time-varying shocks and spatially correlated trends in media environments.

Importantly, I find that WRC, as proxied by the 'Dixiecrat' vote share, is not systematically correlated with market concentration (HHI) within states, reducing concerns that treatment intensity is mechanically related to variation in newspaper competition. Moreover, by measuring WRC and competition at the level of each newspaper's actual circulation footprint, rather than geography alone, I reduce the influence of unobserved spatial clustering that might confound county-level analyses.

5.3 Results - Demand

Before proceeding to the estimations, I assess the difference in the evolution of the race index for each group at 6-month intervals compared to a base period 6 months prior to the shock. **??** shows the evolution for newspapers whose top three markets of circulation are in the top and bottom quartiles of Dixiecrat vote shares compared to the first semester of 1966. Results indicate large differences in label adoption with newspapers exposed to White racial conservatism adopting more slowly. I repeat the exercise for the 1964 voter registration device proxy for WRC with similar results in **??**.

Table 5.1 presents the regression results corresponding to Equation 3, using, as a measure of treatment exposure, (i) the newspaper exposure to 'Dixiecrat' vote share in columns 1 to 4, and (ii) the presence of a voting registration device in the county

in columns 5 and 6. Estimates for the full sample of daily and non-daily outlets in the left-hand panel, and for the subset of daily newspapers for which we have circulation data in the right-hand panel for each exposure measure. In both the full and restricted samples there is a negative relationship between White racial conservatism and the adoption of the label by newspaper outlets. In the preferred specification which controls for state×month-year FE, a one standard deviation increase in the vote share for the 'Dixiecrat' candidate (21%) is associated with a 1.2 p.p decline in the label adoption rate on average in the full sample and a 1.5 p.p decline in the restricted daily sample. This represents a ∼6.5p.p reduction between the highest and lowest vote shares for the 'Dixiecrat' candidate or about 10% lower compared to the post-treatment period mean. Estimates for the alternative measure of WRC in columns (5) and (6) are negative and statistically significant. The effect is larger in magnitude in the full sample compared to the sub-sample of daily newspapers.

All columns include the controls described in the previous section interacted with year-month FE, as well as newspaper outlet fixed effects and year-month fixed effects. Columns 2 and 4 control for year-month × state fixed effects, which absorb the effects of any newsworthy events that vary by state. Estimates for the whole sample and daily sample are stronger and more precise when looking with state×year-month. Voter registration estimates are limited to higher geographical region × time fixed effects since policies largely intersect with state lines.

Table 5.1: Difference in Southern local newspaper racial label adjustment by 'Dixiecrat' vote share

	All		Dailies		All	Dailies
	(1)	(2)	(3)	(4)	(5)	(6)
WRC × Post	-0.004	-0.012***	-0.006**	-0.015***	-0.066***	-0.038***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.015)	(0.012)
Exposure	C-Vote	C-Vote	C-Vote	C-Vote	D-Device	D-Device
N	46,505	46,505	22,783	22,783	46,505	22,783
# of Newspapers	426	426	165	165	426	165
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Newspaper FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
State × Month FE	No	Yes	No	Yes	No	No

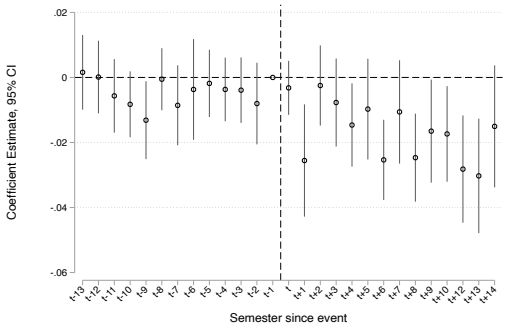
Note: *** p<0.01, ** p<0.05, * p<0.10. Regressions estimate the average difference in the adjustment of racial labels to increasing proportions of vote shares for the 'Dixiecrat' candidate in the 1948 Presidential election after the shock to preferences in 1966h2. The level of observation is media outlet × year-semester. Full sample of outlets in the left-hand panel and sub-sample of daily newspapers in the right-hand panel. Standard errors clustered by county.

In Figure 5.1a and **??**, I estimate dynamic difference in differences to assess the timing of the change and to look for evidence of violations to the parallel trends assumption. I interact the White racial conservatism exposure variables with a set of

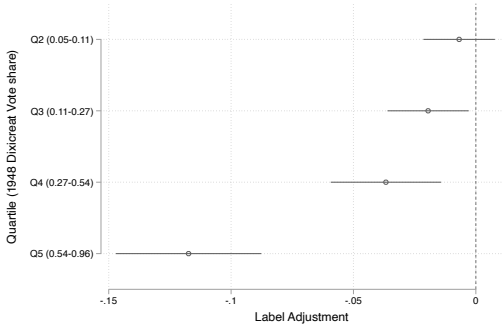
indicators for semester pre-/ post- shock. The omitted category is the semester before the shock. Neither estimations provide any concrete evidence of pre-trends. For both measures, the marginal effect of White racial conservatism on label adoption appears to occur with a few semesters delay after the shock and with more precision in the dailies sub-sample (see **??**). In Figure 5.1b, I estimate a flexible specification that interacts each quintile of the distribution of vote share for the 'Dixiecrat' party candidate with the post-shock period, leaving outlets in the lowest quintile as the reference category. The results suggest that the magnitude of the shock is monotonically increasing in White racial conservatism relatively linearly until the highest quintile, where we see a noticeable jump.

Figure 5.1: The difference in the use of racial labels in Southern local newspapers (1960 - 1973)

(a) Dixiecrat Vote Share



(b) Dixiecrat - Quintiles



Note: Each panel shows the relationship between local newspaper adoption of the term "Black" (relative to "Negro") and pre-treatment measures of racial conservatism. Panel (a) plots coefficients from an event-study regression of the share of "Black" usage on time relative to 1966h1 interacted with continuous pre-treatment county-level Dixiecrat vote share (1948). Panel (b) estimates the post-treatment effect separately for each quintile of the Dixiecrat vote share distribution. Both specifications include newspaper and year-month fixed effects as well as interactions of year-month with pre-treatment county and newspaper level controls. Standard errors are clustered at the county level.

5.4 Results - Competition

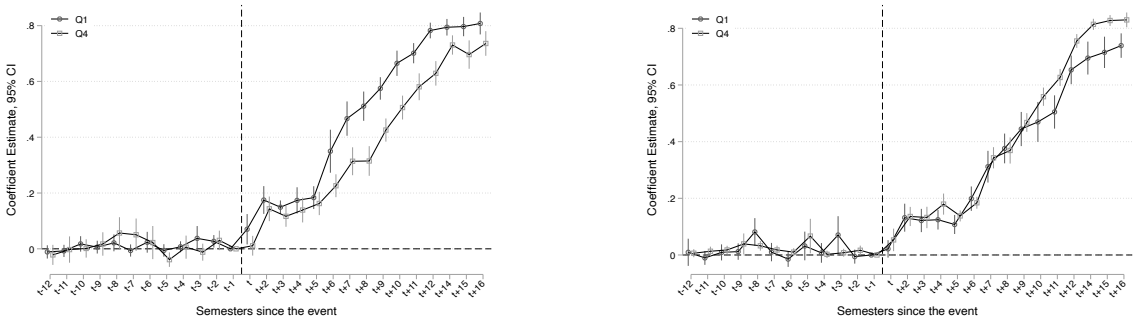
I now empirically analyse the effect of White racial conservatism conditional on local newspaper market competition and provide estimates for the specification in equation 3. I limit the sample to local daily newspapers in counties for which there is circulation data available.

In similar fashion to Section 5.3, I begin with a preliminary descriptive analysis of the evolution of the race indices compared to a base period 6 months before the shock. Figure 5.2 presents the evolution for newspapers operating in counties in the top and bottom quartiles of Dixiecrat vote shares. Panel (a) is newspapers operating in markets with the lowest quartile concentration, i.e. most competitive, and panel (b) the highest quartile. While descriptive in nature, these estimates provide clear preliminary evidence in support of the hypothesis that competition drives newspapers

to appease the dominant consumer group. Descriptive figures are similar utilising the voter registration device dummy (see **??**).

Figure 5.2: Effect of competition and racial conservatism on the evolution of racial labels in Southern local newspapers (1960 - 1973)

(a) Low Concentration (b) High Concentration



Note: Regressions of outlets' semester use of the term "Black" relative to the term "Negro" over time split by the top and bottom quartiles of vote share for Strom Thurmond in the 1948 Presidential Election. The omitted category is the semester before the shock. Panel (a)/(b): is a sub-sample of daily newspapers which circulate in a low/high concentration market. Regressions do not control for any variables and so are descriptive in nature. Standard errors clustered by county.

Estimation. I begin analysis by estimating a generalised dynamic difference in differences version of Equation 3 separately for firms exposed to newspaper market concentration below and above the median exposure. Dixiecrat vote share shows a clear negative relationship between the racial conservatism exposure measure starting in the semester following the shock for firms in low concentration markets, while newspapers in less competitive markets appear to show a slightly positive but insignificant correlation (see Figure 5.3a). Results for the dummy measure of racial conservatism show a similar pattern (Figure 5.3b), with a slightly negative but insignificant correlation over time for firms operating high concentration markets.

Regression estimates are presented in Table 5.2. Columns (1) - (8) provide estimates for the Equation 3 limited to above and below the median of market concentration, while columns (9) - (11) explicitly test for a difference in the effect of concentration on newspapers alignment with White racial conservatism estimating Equation 4.

Columns (1) - (8) indicate that in less concentrated markets, there is a significant negative relationship between WRC_n and label adoption and a marginally negative but sensitive and insignificant relationship in the upper median. These results are robust to year-month $\times$ state fixed effects, year-month $\times$ DMA fixed effects, and a more restrictive specification where I include an outlet trend. This final specification relaxes the assumption that outlets with different exposures to White Racial conservatism follow parallel trends, and identifies the effect of the shock from any sharp change in label adoption at the time of the shock. Estimates are expectedly smaller in magnitude to the baseline specification and significant.

Turning to estimates of Equation 4, we see a significant and robust positive third difference estimate. That is to say, the marginal effect of an increase in White racial

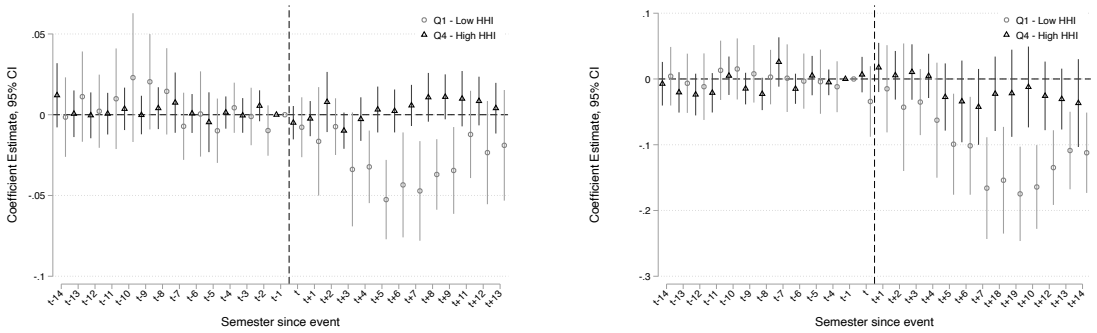
Table 5.2: The effect of local daily newspaper competition on newspaper response to White racial conservatism

	HHI (<50%)			HHI (>50%)			DDD		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
WRC ×	-0.012**	-0.013***	-0.007**	-0.001	-0.001	-0.000	-0.021**	-0.025***	-0.041**
Post	(0.006)	(0.004)	(0.003)	(0.003)	(0.009)	(0.002)	(0.009)	(0.008)	(0.016)
HHI ×							-0.021***	-0.021***	-0.033**
Post							(0.007)	(0.006)	(0.013)
WRC × HHI							0.003**	0.004**	0.009***
× Post							(0.001)	(0.002)	(0.003)
Exposure	C-Vote	C-Vote	C-Vote	C-Vote	C-Vote	C-Vote	C-Vote	C-Vote	C-Vote
N	11,452	11,452	11,452	11,377	11,215	11,377	22,829	22,829	20,128
# of Newspapers	88	88	88	79	78	79	167	167	150
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Newspaper FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Month FE	No	Yes	Yes	No	Yes	Yes	No	Yes	No
Outlet FE × Linear Trend	No	No	Yes	No	No	Yes	No	No	No
DMA × Month FE	No	No	No	No	No	No	No	No	Yes

Note: *** p<0.01, ** p<0.05, * p<0.10. Regressions estimates Equation 3 in columns (1) to (8), split by exposure to market concentration and the triple differences specification of Equation 4 in columns (9) to (11). Each group of regressions increasingly adds stricter fixed effects. The level of observation is media outlet × year-semester. The sample is limited to local daily newspapers which operate in markets for which I have circulation data. Estimates for (9) to (12) should be interpreted as follows: The third line is the coefficient of interest, estimating the difference in the average marginal effect of increasing White racial conservatism when competition is decreased on outlet label adoption rates after the shock. The first line estimates the difference in outlet label adjustment to an increase in White racial conservatism if the newspaper faces perfect competition. The second line estimates the average marginal difference in the outlet label adjustment between decreasing competition when the White racial conservatism variable is set to zero. Standard errors in all regressions clustered by county.

Figure 5.3: Dynamic difference in differences of effect racial conservatism on the evolution of racial labels in Southern local newspapers (1960 - 1973) by HHI quartiles.

(a) Dixiecrat Vote Share (b) Registration Device



Note: This figure presents dynamic difference-in-differences estimates of the effect of racial conservatism on newspaper adoption of the term "Black" (relative to "Negro"), stratified by media market concentration. Panel (a) estimates event-study coefficients for the interaction of Dixiecrat vote share in 1948 with semester dummies, separately for low- and high-concentration counties (HHI quartiles 1 and 4). Panel (b) estimates post-period effects of Dixiecrat vote share by quintile, again separately by market concentration. All regressions include fixed effects for outlet and semester, and interactions of semester with pre-treatment county-level controls (e.g., racial and age composition). The omitted category is the semester immediately prior to the event. The sample includes all daily newspapers. Standard errors are clustered at the county level.

conservatism on racial label adjustment is increasing in magnitude the more competitive is the local daily newspaper market.

Dynamic triple difference-in-differences estimates are provided in **??**. Estimates support the hypothesis that the presence of daily competition increases the magnitude of the effect of White racial conservatism on newspaper outlet label adjustment at the time of the shock and is persistent throughout the period. This is true using both the 'Dixiecrat' vote share variable and voter registration device dummy variable, as well as both the extensive and intensive measures of market competition.

The results suggest that competition drives firms to compete for target consumers and that absence of competition allows firms to target consumers with less popular preferences in the knowledge that larger audiences have no substitutes. This would suggest that competition would both slow adjustment for newspapers facing high WRC and accelerate adjustment for newspapers facing low WRC. **??** supports this narrative. The adjustment towards "Black" is slower in the presence of competition for newspaper facing high levels of WRC. The effect is then reversed for newspapers facing the lowest proportions of White racially conservative readers, i.e. the label adjustment is slower with increasing market power, for firms facing low levels of WRC. **??** highlights this effect is statistically significantly driven by larger firms in regions with high WRC and more by smaller firms more than larger firms in regions with low WRC.

Robustness. Results are robust to many controls and alternative specifications. In all estimates I control for country controls × year-semester FE, year-month and newspaper fixed effects. In column 10, I control for year-month × state fixed effects, which compares firms facing varying levels of concentration and White racial conservatism

within a state absorbing any shocks occurring within the region. I further control for finer fixed effects at the level of year-month × DMA in column 11. This restricts the comparison within areas assumed to have relatively homogeneous news consumption profiles and absorbs any shocks that occur at this finer level. Results appear to be even more pronounced at this more granular level. I repeat the exercise with a dummy indicator for whether there is more than one daily newspaper operating in the same county of commercial residence in Table A.2 columns (1) - (4), and results are substantively similar. This estimate can be thought of as the extensive margin of competition in the county, comparing the effect of White racial conservatism conditional on any competition in a market vs. monopolists in the county of commercial residence. There could be a concern that the same local factors influencing language use in newspapers might have also influenced the Thurmond vote share in 1948. For instance, if media outlets were already more conservative or had more conservative readerships, this could have influenced both the support for Thurmond and their language choices later on. Table A.2 columns (5) - (8) repeats the exercise with the voter registration restriction device measure of White Racial conservatism and shows strong and similar results. I also repeat the exercise with two other alternative and more historical measures of exposure to White Racial conservatism which would be free of any of the potential reverse causality captured in the Strom Thurmond vote: the proportional county slave population from 1850 and, from Acharya et al. (2016), local measures of the agricultural suitability to grow cotton[23] in Table A.2 columns (9) - (10). Both measure have been previously used as proxies for historical racial conservatism and are significant and substantively similar predictors of newspaper label adoption. Estimates are also robust when using alternative transformations of the continuous vote variables, i.e. inverse hyperbolic sine and log transformations, when restricting the sample to exclude large cities and counties with more than 3 local daily newspapers, when using alternative weighting of continuous vote share exposure using all circulated markets, when restricting the definition of voter registration test or device to the original VRA definition from 1964 and using measurements of concentration from an alternative election year prior to 1964 (see Table A.3).

Non-Daily Newspapers Placebo. As a placebo test, I repeat the exercise for non-daily newspapers with commercial residence in the counties where there are daily newspapers with commercial residence. I impute concentration faced by the non-daily newspapers with the concentration faced by daily newspapers in those counties. If county market concentration, or the interaction there of with WRC and/or $Post$, is correlated with some other predictor of racial label choice, estimates for non-daily newspaper operating in the same markets should not be substantively different. Table A.4 columns (1) to (3) show no statistically significant relationships using daily market concentration as a placebo treatment for non-daily newspapers. This suggests that the

[23]The estimates are based on climate averages from 1961 to 1990. The authors averaged "intermediate" and "high" levels of inputs, which refer to the effort required to extract the resource

main effects are not merely artefacts of underlying county characteristics, but instead reflect the competitive environment actually faced by the newspapers making editorial decisions.

In aggregate, these results indicate that the marginal effect of increasing White racial conservatism on slowing the adoption rate of outlet use of the liberal racial label is exacerbated in the face of and intensity of daily newspaper market competition. It suggests that when daily newspapers are competing in a racially conservative region, they tend to, on average, skew their bias towards the preferences of the dominant out-group.

Competitive Differentiation. Majority consumers tastes appear to dominate output decisions in this setting, however, it is possible that there is still some product differentiation occurring, although insufficiently to offset the incentive to please the majority reader on average. The incentive to differentiate from competitors in order to attract more readers and soften price and advertising competition, as in Mullainathan and Shleifer (2005), has been shown to increase political ideological diversity significantly, offsetting the incentive to cater to the the tastes of majority consumers (Gentzkow and Shapiro, 2010). If competition leads to greater diversity in label adoption, one would expect to see increasing variance of label choice across markets with heterogeneous levels of competition, while holding proportions of White racial conservatism constant. I repeat the estimation of equations 3 and 4 at the county level, replacing the outcome variable with the county×year-month variance in label adoption.

Estimates in **??** indicate that there is evidence of statistically significant competitive differentiation in the fae of competition between counties, in the lowest quintile of 'Dixiecrat' vote share (i.e. low WRC). The significance of the effect disappears in all other quintiles. This suggests that firms in more progressive regions, in the face of competition, may differentiate to compete for minority conservative consumers, while in more racially conservative regions, the incentive to compete for minority progressive consumers disappears. This may suggest, non-exhaustively, a higher threat of punishment from more conservative readers, or that supply-side decision makers are slower to produce content in conflict with their own priors on highly salient and controversial matters.

Agency and Sentiment. It is possible that economic incentives may impact the representation of Black entities with respect to adoption. To test this explicitly, I repeat equation 4 for quartiles of firm exposure to the 'Dixiecrat' vote share and market concentration. Results in **??** suggest that in more competitive environments, increasing WRC results in a lower change in probability of a combined high-agency and positive-sentiment representation of Black" entities, compared to Negro". There appears to be no discernible effect in less competitive environments. However, the difference between these effects is not statistically significant.

These findings suggest that, unlike the selective adoption of new group labels, rep-

resentational content, measured via linguistic agency and sentiment, adjusts in a more uniform manner across market environments. While point estimates indicate a smaller shift in high-agency, positive representation in high-WRC, competitive markets, these differences are not statistically significant. This relative uniformity may reflect the influence of broader discursive currents, such as the reach of national newswire content, shared professional norms among journalists, or editorial coordination, that facilitated the widespread uptake of new representational styles. Rather than being tightly tethered to local economic incentives, the framing of Black political actors appears to have been reshaped through a more centralised diffusion of narrative conventions, consistent with the symbolic authority and cultural salience of the Black Power movement.

5.5 Results - Supply Side

While the main estimates, based on variation in newspaper exposure to both racial conservatism and market competition, are strongly suggestive of a demand-driven mechanism, it is still possible that supply-side factors are responsible for the observed adoption patterns. One possibility is that labelling decisions are shaped by centralised editorial policies within multi-paper publishing firms, rather than by market-specific demand. In particular, if publishers own multiple newspapers circulating in different markets and editorial decisions are coordinated across outlets, then observed differences in label use may reflect internal group strategy rather than local responsiveness. A second possibility is that journalists and editors themselves drive content choices, especially if their ideological preferences are aligned with the dominant views of the communities they serve, producing supply-side sorting that mimics demand-side responsiveness. Finally, there are incentives for publishers to cluster in geographically and politically similar markets due to economies of scale in news gathering and editorial production (Lacy and Simon, 1997; Martin and Sunley, 2003). In equilibrium, this could induce correlations between competition, population density, and racial ideology, even if labelling is primarily set at the group or staff level. Together, these concerns raise the possibility that observed adoption patterns reflect internal editorial preferences or firm location strategies, rather than causal responses to external market incentives.

To address these concerns, I exploit a feature of my ABC circulation data that allows me to identify a subset of newspapers that are jointly owned, produced in the same building, and written by the same editorial teams. These outlets, 48 newspapers in total, allow for a clean test of whether variation in adoption reflects market-level exposure or shared supply-side characteristics. Although these newspapers share editorial inputs, they differ in where and how they circulate, creating variation in exposure to White racial conservatism and market competition. I estimate Equation 4 with news-team-by-time fixed effects, isolating within-team variation across markets while controlling for common editorial bias. I also include controls for time-of-day of publi-

cation, as co-owned newspapers often stagger morning and evening editions.

Estimates using this sample of newspapers are presented in Table A.5. Estimates replicating Equation 4 are significant and similar to the whole sample. When I control for news-team×month-year fixed effects, estimates are similar in magnitude, although less precise given that there are 24 pairs being exploited for variation, but still marginally significant. These results are quite suggestive that it is not within-ownership group policy, nor editors or journalists, driving newspaper label adoption rates, external to through the economic incentive channel discussed throughout this paper. Given the number of newspaper pairs is relatively small for these estimates, results should be taken with some level of caution. Further, these results represent supply side effects for newspapers which fit the sample restrictions, i.e. newspapers written by the same teams in varying markets.

5.6 Alternative Explanations

In this section I attempt to rule out the possibility that estimates are being driving by other correlates. On the supply side, I explore whether newspaper outlet political party affiliation or subscription to newswires are driving the results. On the demand side, I rule out that results are being driven by the political alignment of consumers or heterogeneity in Black preferences. Results are in Table A.5 to Table A.9.

Supplier political ideology. Since media outlets can decide how to use racial labels in their editorial production and race was clearly critically entwined in political ideologies, a natural question is how adjustment of the labels differs by outlet political ideology. One might predict that, given the recently updated racially liberal position of the Democrats, firms with a political affiliation with the Democratic party might adopt the term Black more rapidly. If it is the case that there is a correlation between the political ideologies of newspapers, the proportion of White racially conservative citizens, and or the level of competition in a local market, it may in fact be newspaper ideology which is explaining the results. There is, however, some evidence that commercial considerations likely dominated political incentives during the period covered (Baldasty, 2001). Gentzkow et al. (2019) find that partisanship in language choice is not commonplace until after 1994, following political consultancy led innovations in political persuasion. Regardless, the adjustment of racial labels may represent something distinct from the more modern partisan language choices previously studied and so I test it explicitly.

I analyse a sub-sample of 108 newspapers for which there is information about whether the outlet made a public endorsement of a political party or were publicly independent in the lead up to the 1964 Presidential election in USNP. Assuming that a newspaper's affiliation serves as a good proxy for the ideological tilt of the newspaper's content (Gentzkow et al., 2014), I introduce two dummy variables for public endorsement of each of the major parties with publicly independent as the base group

and conduct a statistical horse race replicating both equations 2 and 3. Results in Table A.7 indicate that there is no significant effect of announced party ideology on the outcome or estimates of WRC or the interaction of WRC with market concentration.

Newswire rules. By the 1960's, many newspapers shared stories within larger groups, either within group ownership or via subscription to national and international newswires. It is well documented that locally distributed newspapers who subscribe to newswires can be impacted by sharp adjustments in newswire standards (Djourelova (2023)). What is unclear is whether, in the absence of a sharp change in rules, newswires evolution of language drives adoption on average. Associated Press (AP) style guides[24] from 1960 and 1970 show no difference in the recommendations for describing Black people (see **??**)[25]. The next AP style guide I have access to is from 1977 where there are some more recommendations: (1) that "Negro" and "black" are interchangeable, (2) to avoid the use of "colored" when describing black people, and (3) to avoid the use of "African" as a synonym for "Negro". Unlike the setting in Djourelova (2023), there is no distinct rule against using "Negro" during the sample period but a clear increase in the freedom of label choice when describing Black people. It is also not clear whether competition might affect firms proportional use of newswires in output. Heterogeneity in consumer preferences for locally vs. nationally and internationally relevant news may encourage firms to differentiate in the face of competition manifest in variation in newswire proportional use. Perhaps more convincingly, competition may erode profits forcing firms to save on labour costs by substituting with newswires. Regardless, if the proportional use of newswires correlates with the competitive environment or WRC, then this may bias estimates.

To test this, I count the proportion of news articles[26] written by local and AP journalists in each newspaper from 1965 for a sub-sample of newspapers. I then explicitly introduce this into my regressions as another treatment arm and conduct a statistical horse race. Results in Table A.6 indicate that, in the full sample, the proportion of articles written by local journalists compared to newswires significantly slows the adoption process in the face of WRC. This effect is not present in the daily newspapers subsample, on average. In the Triple differences specification using the daily newspaper sub-sample, the magnitude of the newswire effect is increasing in newspaper market competition, which may suggest strategic choice of newswire articles for daily newspapers, and/or adjustment of language in exisiting newswires. I provide suggestive evidence of the latter in **??**. Regardless, these estimates have no effect on the main estimates, indicating that newswires are not driving the results.

[24]Styleguides can be accessed at https://www.apstylebook.com/archived_stylebooks

[25]Both guides recommend capitalising Negro and using lowercase black, with the recommendation to only describe any race when it is pertinent.

[26]I omit the following categories of newspaper text to isolate 'news' articles: Deaths, Funerals, Obituaries, Marriages, Weather Forecasts, Horoscopes, Real Estate listings, Personal Ads, Commercial Ads, Comics, Games, TV Guides, Racing Cards, Stock Market Readings, Notices (i.e. Church notices, local council notices, and blurbs for images.

Consumer politcal ideology. Given the discussion regarding the alignment of political groups to racial liberalism, it is possible that racially conservative consumers correlate with one of the political parties. To rule this out I conduct another statistical horse race explicitly including vote share Republican Barry Goldwater in the 1964 Presidential election in estimates in Table A.8. In equation 2 estimates, White racial conservatism has a unique and independent effect on local newspaper label adjustment. In regressions including market concentration, Republican vote share are insignificant until the inclusion of the Strom Thurmond vote share, which remains unchanged in significance and magnitude. It is possible that Republican vote share estimates with the multiple treatment arms suffer from contamination bias (Goldsmith-Pinkham et al., 2022).

Black label preferences. While there is evidence that Southern newspapers are unlikely to support the demand of local Black citizens (Torres, 2003; Acham, 2005; Boroghkozy, 2012), and despite that I have explicitly controlled for Black demographics and Black political activism, I test the theory explicitly to rule out the possibility that my estimates capture correlations between the local in-group and out-group. I introduce Black treatment variables capturing pre-treatment Black population proportions by age buckets as well as an NAACP chapter dummy variable capturing Black political activism. These variables are regressed linearly and interacted with one another, with and without the inclusion of my measure for WRC to conduct a statistical horse race of local newspaper label adjustment. In all regression estimates in Table A.9, replicating the main output and heterogeneity by competition, the estimated effects of exposure to the Black preference proxies are much smaller in magnitude and statistically insignificant, while the estimates for exposure to White racial conservatism and interactions with competition variables remain nearly identical in magnitude and significant. The statistical horse race between White racial conservatism and the Black population suggests that White racial conservatism had a unique and independent effect on local newspaper label adjustment.

6 Local Black Governments

Having established that local newspapers varied widely in their adoption of the term "Black" over "Negro," and that this linguistic shift was shaped by structural features of the media market, I now assess whether such symbolic changes in public discourse were associated with concrete shifts in Black political incorporation. The outcome of interest is the share of Black elected officials at the local level. I focus on this measure for two key reasons. First, local officeholding provides a direct and consequential indicator of descriptive representation, capturing the extent to which institutional power was extended to Black communities in the aftermath of enfranchisement. Second, by concentrating on elected officials rather than individual behaviours such as turnout or

registration, I target a political outcome that cannot be mechanically driven by Black media exposure, especially given that Southern newspapers were predominantly read by white audiences. This approach allows for a more stringent test of whether shifts in symbolic language, targeted at white readers, shaped the broader political environment in ways that ultimately facilitated Black representation in local governance.

While previous research has shown that the Voting Rights Act (VRA) dramatically expanded Black officeholding by removing institutional barriers to voting (Bernini et al., 2023b), I ask whether variation in media-based cultural signalling, specifically, the use of more affirming racial terminology, provided an additional, complementary channel through which political incorporation occurred. Conditional on formal access to the ballot and demographic composition, did counties where newspapers adopted the term "Black" earlier or more consistently see faster or larger increases in Black political representation? To answer this, I link newspaper-level measures of label adoption to county-level data on Black elected officials from Bernini et al. (2023b), covering Southern jurisdictions between 1962 and 1974.

This section is not based on an experimental design, but exploits plausibly exogenous variation in symbolic language use across counties after flexibly conditioning on a saturated set of structural, demographic, and political controls. The key identifying assumption is that, conditional on these controls, residual differences in the racial index capture variation in cultural discourse orthogonal to other county-level determinants of Black officeholding. The triple-difference design further isolates variation linked to institutional enfranchisement by interacting the racial index with year and VRA coverage, identifying from divergence in political incorporation between otherwise similar counties that were more or less exposed to affirming media language. While this strategy cannot definitively rule out all forms of unobserved confounding, it substantially narrows the set of alternative explanations and provides a disciplined test of whether symbolic changes in language, net of institutional reform, contributed to shifts in representation.

6.1 Data

County Level Label Diffusion. The first step in this analysis is to aggregate measures of newspapers' label diffusion to county level. I compute, for each county-year, a circulation-weighted average of newspapers' racial conservatism, where weights come from the share of each newspaper's circulation in the county in 1964.

$$\widetilde{black}_{ct} = \frac{\sum_n^N (\widetilde{black}_{nt} \times circ_{nc})}{\sum_n^N circ_{nc}} \tag{5}$$

Where $\widetilde{black}_{nt}$ is the racial index of newspaper n in year t, and $circ_{nc}$ is the circulation of newspaper n in county c. The weighted average $\widetilde{black}_{ct}$ represents the average racial index in county c in year t, weighted by the share of each newspaper's circulation

in that county. A lower value of $\widetilde{black}_{ct}$ indicates that the newspapers circulating most heavily in that county tended to use more racially conservative language in that year.

Black Elected Officials. Data on Black local elected officials are drawn from the digitised county-level series constructed by Bernini et al. (2023b), based on the National Roster of Black Elected Officials (NRBEO). Compiled initially by the Southern Regional Council's Voter Education Project and later maintained by the Joint Center for Political Studies, the NRBEO offers the most comprehensive record of Black officeholding in the U.S. South from the mid-1960s onward.

These data combine multiple sources, including annual surveys of known officeholders, local administrative records, press reports, and government directories, with systematic verification by phone or mail. While undercounts are more likely in the early years due to evolving methods, the resulting panel provides consistent coverage of Black elected officials in counties across the eleven former Confederate states from 1962 to 1980.

I use annual counts of Black officeholders in three main domains, county governments, municipalities, and school districts, to construct measures of descriptive representation. These are merged with data from the Census of Governments on the total number of elected officials by type and year to compute the share of Black officials in each county-sector-year. This yields four main outcomes, capturing representation at different tiers of local governance.

6.2 Empirical Strategy

I assess whether variation in local newspaper language, specifically, the extent to which newspapers adopted the label "Black" over "Negro", predicts changes in Black elected officeholding. The key independent variable is the county-year racial index, $\widetilde{black}_{ct}$, constructed as a circulation-weighted average of newspaper-level language use. Because circulation weights are fixed at their 1964 values, within-county changes in $\widetilde{black}_{ct}$ reflect shifts in newspaper terminology rather than changes in population composition or readership.

To estimate the relationship between media language and Black officeholding, I implement a triple-difference specification that interacts the racial index with VRA coverage and event time. The core estimating equation is:

$$\begin{aligned} Y_{ct} = &\sum_{\tau \neq 1962} \beta_\tau \mathbb{1}[t = \tau] \times \text{VRA}_c \times \widetilde{black}_{ct} + \sum_{\tau \neq 1962} \phi_\tau \mathbb{1}[t = \tau] \times \text{VRA}_c \\ &+ \sum_{\tau \neq 1962} \psi_\tau \mathbb{1}[t = \tau] \times \widetilde{black}_{ct} + \delta_c + \lambda_t + \theta_{s,t} + X'_{ct}\kappa + \varepsilon_{ct} \end{aligned} \tag{6}$$

where Y_{ct} is the share of Black elected officials in county c and year t, VRA_c is an indicator for coverage under Section 5 of the Voting Rights Act (VRA), $\widetilde{black}_{ct}$ is the racial index, δ_c and λ_t are county and year fixed effects, and $\theta_{s,t}$ denotes state-by-year

fixed effects to absorb time-varying shocks at the state level. The vector X_{ct} includes a rich set of baseline covariates, fully interacted with both year and VRA status.

To isolate the effect of media language from structural confounds, I flexibly control for a wide range of county-level characteristics, including Black population share, education, income, urbanisation, white racial conservatism, and media market concentration, two factors that predict the racial index and may also directly shape political incorporation. I additionally control for the presence of pro-Black protest, anti-Black protest, and NAACP chapters, each interacted with year and VRA coverage, to account for political mobilisation or racial conflict that could jointly influence both newspaper language and Black officeholding.

These saturated controls absorb secular trends in demographic and political characteristics while allowing for heterogeneous responses to enfranchisement by county type. In particular, conditioning on white racial conservatism and media competition ensures that the effect of newspaper language is identified from variation orthogonal to these structural drivers. The identifying variation thus comes from differential trends in Black representation across counties with similar structural profiles but diverging media language.

The coefficients β_τ measure the extent to which, in each year $\tau > 1962$, counties exposed to more affirming racial terminology experienced larger increases in Black officeholding, conditional on formal enfranchisement, structural factors, and local political context. This approach separates the institutional channel (formal voting rights and demographic composition) from the symbolic channel (media language).

Nonetheless, several threats to identification remain. First, time-varying unobserved shocks, such as elite strategy or local organising, could bias the estimates if correlated with both language use and political incorporation. Second, reverse causality is possible: increases in Black political representation may themselves prompt more affirming coverage. To address these concerns, I implement a series of robustness checks. I first re-estimate the main specification using lagged values of the racial index, allowing me to test whether the observed effects could reflect contemporaneous feedback from rising Black officeholding. I then residualise the racial index with respect to all structural covariates known to predict media language: white racial conservatism, media market concentration, protest activity, NAACP presence, and Black population share, interacted flexibly with year. This approach isolates variation in newspaper terminology that is orthogonal to structural drivers.

6.3 Results

I begin by replicating estimates from Equation 4 at the county level for the 925 counties for which we have circulation data in Table A.10. Results are similar in magnitude to Section 5.4 and are robustly estimated.

I then estimate a series of cross-sectional regressions at the county level in each

election year between 1962 and 1974. In each year, I regress the share of Black elected officials on the racial index, separately for VRA-covered and non-covered counties, including state fixed effects and all covariates. These models allow for flexible year-by-year heterogeneity in the relationship between local newspaper language and Black representation. **??** displays marginal effects of the racial index by VRA status in each election year.

The results indicate that the relationship between racially affirming newspaper language and Black political representation is consistently stronger in VRA-covered counties than in non-covered counties. From 1970 onward, counties covered by the VRA exhibit increasingly large and positive effects of the racial index across all domains of local officeholding, especially among county and education officials. These effects are statistically significant in several years, particularly in the early 1970s, and are substantially larger than any estimated effects in non-covered counties, where the relationship remains flat and indistinguishable from zero. This pattern is consistent with the interpretation that media-driven cultural signalling through label choice amplified the political gains associated with institutional enfranchisement, particularly in contexts where formal voting protections were in place.

Next, I turn to a fully saturated triple-difference event study design, in which the racial index is interacted with both year and VRA coverage. Figure 6.1 plots the marginal effect of the racial index on Black representation in VRA-covered counties, relative to 1962, netting out trends in non-covered counties. Across all four categories of officeholding, local, county, school district, and municipal, the estimated effects of racially affirming newspaper language grow over time following the implementation of the VRA.

The clearest and most persistent increases appear in county offices, where the coefficient becomes statistically distinguishable from zero by 1968 and remains stable through 1974. For all local officials, the pattern is similar but slightly less precise. Municipal offices show weaker and more variable effects, though still with a consistent upward trend. These patterns suggest that shifts in racial terminology in VRA-covered counties were not merely symbolic: they coincided with tangible increases in Black political incorporation, above and beyond the institutional effects of enfranchisement alone. The results are consistent with the hypothesis that media-based cultural cues helped reinforce or amplify the political activation made possible by formal voting rights. See Table A.11 for full regression results by office type.

These findings are robust to the progressive inclusion of increasingly comprehensive sets of controls. Table A.12 reports coefficient estimates from the baseline triple-difference framework, showing that the interaction of newspaper terminology and the post-1965 VRA treatment remains positive and statistically significant across specifications. Column (1) begins with a simplified model omitting three-way interactions between controls, year, and VRA coverage. In this less saturated specification, the esti-

mate is smaller in magnitude and not statistically significant, suggesting that structural factors correlated with both media language and Black officeholding may bias the naïve model. Columns (2) through (5) sequentially introduce controls for Black population share, racial protest and conflict, and local media market structure, all fully interacted with VRA coverage and year. The estimated effects remain remarkably stable and statistically significant throughout.

In this context, the fully saturated model helps address a plausible concern: that counties experiencing changes in media language also differ in other ways that shape political incorporation. While saturation reduces degrees of freedom and may suppress statistical precision, it provides a more conservative estimate by isolating variation orthogonal to structural drivers of both media behaviour and political outcomes. Column (6) offers a complementary test using a residualised version of the racial terminology index, constructed by regressing the index on pre-treatment county characteristics (white racial conservatism, market concentration, Black share, NAACP presence, and protest activity), fully interacted with year. The triple-difference estimate using these residuals remains positive and statistically significant, reinforcing the interpretation that editorial shifts in racial terminology, net of structural factors, contributed independently to Black political incorporation.

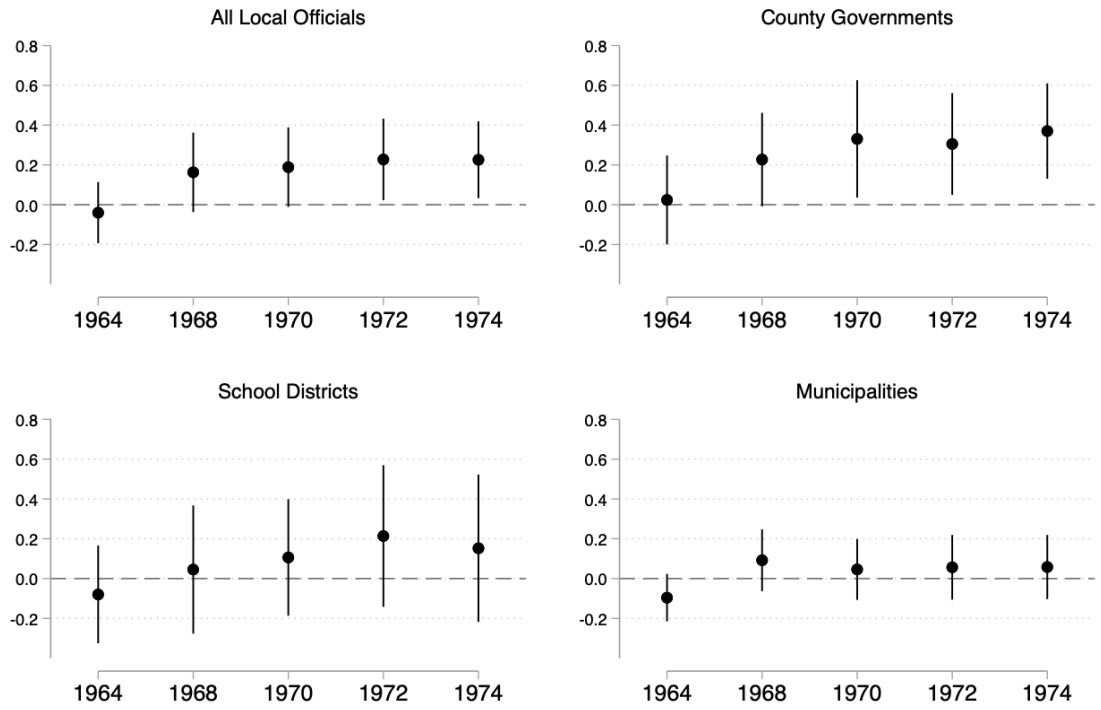
To test whether the estimated effects might reflect media coverage that responds to, rather than drives, Black political empowerment, I re-estimate the main specification using lagged values of the racial index. Table A.13 and Table A.14 report results using one- and two-year lags, respectively. The estimates remain similar in magnitude and precision, indicating that contemporaneous feedback from rising Black officeholding is unlikely to explain the observed patterns.

6.4 Mechanisms: Racial Terminology and White Political Behaviour

The preceding results establish a robust positive relationship between local newspaper language and Black political incorporation, even after conditioning on Black population share and a rich set of structural covariates. But what mechanism plausibly links symbolic changes in white-oriented media to downstream shifts in descriptive representation?

Throughout the paper, I have emphasised that Southern newspapers during this period were overwhelmingly written by, for, and about white actors. They primarily targeted white readers, reflected white elite preferences, and played a central role in shaping white public opinion. As such, the observed relationship between racial terminology and Black officeholding is unlikely to reflect direct effects on Black political mobilisation through media exposure. Instead, a more plausible channel is that shifts in racial language, particularly the adoption of "Black" over "Negro", served to reshape the norms, expectations, and behaviours of white political actors and voters. By legit-

Figure 6.1: Difference in the gradient of Black elected officials to county-level newspaper Black label diffusion between covered and non-covered counties, by type of office.



Note: Estimates are from a fully saturated triple-difference specification interacting year, VRA coverage, and the racial index. See Table A.11 for regression coefficients.

imising Black civic participation, signalling elite acceptance, or reducing coordination barriers in white-dominated political environments, newspapers may have facilitated a broader symbolic inclusion of Black Americans into the local political order.

Three pieces of evidence support this interpretation. First, the effect of the racial index persists after flexibly controlling for the size of the local Black population and its interaction with VRA coverage, ruling out a simple demographic or turnout-based explanation (see again Table A.12).

Second, as shown in earlier sections, the effect of newspaper language is strongest in competitive media environments, precisely where newspapers are most responsive to consumer sentiment and where symbolic shifts are likely to reflect broader social demand.

Third, and most directly, I test whether the relationship between Black population share and Black officeholding depends on the prevailing language of local media. If newspaper language spurred Black mobilisation, we would expect the effect of Black share to be magnified in areas with affirming terminology. However, Table A.15 shows no such interaction. Instead, in VRA-covered counties, newspaper language has a strong independent effect on representation even in areas with small Black populations, and the Black share has no marginal effect conditional on language use. In non-covered counties, where enfranchisement was not institutionally secured, neither

variable predicts political incorporation.

Additional evidence from voter registration data reinforces this interpretation. Using newly digitised county-level records from (Bernini et al., 2023a), I measure the change in voter registrations by county-race between 1964 and 1974 in Table A.16 and find no relationship between newspaper language and registration rates among either Black or white voters in presidential election years. This null result suggests that the observed effects are not driven by changes in direct political participation induced by symbolic media shifts. Instead, they are more consistent with a mechanism in which racial terminology shapes elite behaviour and gatekeeping, altering the perceived legitimacy of Black officeholding in the eyes of white political actors, rather than mobilising Black voters themselves.

These findings suggest that newspaper language shaped the behaviour of white political actors, not by informing Black voters, but by signalling the legitimacy and inevitability of Black political inclusion to a white audience. The symbolic power of media, in this case, operated through norm change: affirming racial terminology helped redefine who could legitimately hold office. This process was catalysed in jurisdictions where institutional enfranchisement had already removed formal barriers to participation, allowing symbolic shifts in media to affect political expectations and outcomes. In sum, newspapers appear to have contributed to the broader process of Black incorporation not simply by reflecting social change, but by actively enabling it.

7 Conclusion

This paper documents a large degree of diffusion of labels and identity redefinition from in-group preferences to media outlets. The fact that this critical juncture in label preference was driven by a small number of in-group agents highlights this potential source of variation to estimate the effects of language in the mass media. By applying this strategy and leveraging a particularly salient change in in-group label preferences, the shift from "Negro" to "Black", I find strong evidence that out-group preferences and market concentration significantly predict media adoption of the in-group desired label. This effect is more pronounced in areas with a higher proportion of out-group members likely to oppose the adjustment, and it increases in magnitude under more competitive market conditions.

This evidence supports existing findings that media producers act as profit maximisers, even in situations where profit-maximising decisions may proliferate prejudice and discrimination. The results are specific to a context where the in-group is undergoing significant political and social change, gaining a stronger voice in the political sphere and securing equal rights to the out-group after a long history of repression. More work is needed to explore the external validity of these findings across different contexts and identity groups.

Beyond documenting the slow adoption of the "Black" label, this paper highlights the broader implications of linguistic agency in media portrayals. I show that the use of "Negro" is associated with lower linguistic agency, which may have contributed to a diminished portrayal of Black agency and activism in areas where the label change occurred more slowly. This suggests that linguistic choices in the media not only reflect societal attitudes but also shape public perceptions of marginalised groups. The consequences of this dynamic may influence political mobilisation, economic outcomes, and social cohesion in racially divided communities.

In addition, by linking media language shifts to downstream political outcomes, I show that earlier adoption of affirming racial terminology facilitated greater Black political incorporation following the Voting Rights Act. Counties whose newspapers embraced "Black" earlier experienced larger increases in Black local elected officials, particularly in contexts where formal barriers to voting had been removed. These results demonstrate that symbolic changes in public discourse, amplified through media outlets, can help accelerate material political inclusion.

Further research should explore the long-term impacts of linguistic agency on social and economic outcomes for minority groups, as well as the role of media in shaping public discourse around other evolving societal norms. By quantifying the effects of media choices at a granular level, this paper provides a framework for analysing how commercial pressures intersect with societal attitudes to either promote or hinder social progress.

References

Acemoglu, D. and Jackson, M. O. (2015). History, expectations, and leadership in the evolution of social norms. *Review of Economic Studies*, 82(2):423–456.

Acham, C. (2005). *Revolution Televised: Prime Time and the Struggle for Black Power*. University of Minnesota Press.

Acharya, A., Blackwell, M., and Sen, M. (2016). The political legacy of american slavery. *The Journal of Politics*, 78(3):621–641.

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

Ambrus, A. and Reisinger, M. (2006). Exclusive vs. overlapping viewership in media markets. *RAND Journal of Economics*, 37(4):859–888.

Anderson, S. P. and Waldfogel, J. (2011). Competition, product diversity and welfare: A study of the us television market. *Review of Economic Studies*, 78(3):923–952.

Armstrong, M. (2002). The theory of access pricing and interconnection. *Handbook of Telecommunications Economics*, 1:295–384.

Baldasty, G. J. (1992). *The Commercialization of News in the Nineteenth Century*. University of Wisconsin Press.

Baldasty, G. J. (2001). Centralizing control in newspaper chains: E. w. scripps and the newspaper enterprise association, 1902–1908. *American Journalism*, 18(2):13–38.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817.

Becker, H. (1963). *Outsiders: Studies in the Sociology of Deviance*. Free Press.

Bennett, L. (1970). What's in a name? *Ebony*, pages 88–96.

Bernini, A., Facchini, G., Tabellini, M., and Testa, C. (2023a). Black empowerment and white mobilization: the effects of the voting rights act. Technical report, National Bureau of Economic Research.

Bernini, A., Facchini, G., Tabellini, M., and Testa, C. (2024). Sixty years of the voting rights act: progress and pitfalls. *Oxford Review of Economic Policy*, 40(3):486–497.

Bernini, A., Facchini, G., and Testa, C. (2023b). Race, representation, and local governments in the us south: the effect of the voting rights act. *Journal of Political Economy*, 131(4):994–1056.

Boroghkozy, A. (2012). *Equal Time: Television and the Civil Rights Movement*. University of Illinois Press.

Bureau, U. C. (1960). *Census of Population: 1960*. U.S. Census Bureau.

Caplan, N. S. and Paige, J. M. (1968). A study of ghetto rioters. *Scientific American*, 219(2):15–21.

Carmines, E. G. and Stimson, J. A. (1981). Issue evolution, population replacement, and normal partisan change. *American Political Science Review*, 75(1):107–118.

Cascio, E. U. and Washington, E. (2014). Valuing the vote: The redistribution of voting rights and state funds following the voting rights act of 1965. *The Quarterly Journal of Economics*, 129(1):379–433.

Choudhary, T. (2024). Political bias in ai-language models: A comparative analysis of chatgpt-4, perplexity, google gemini, and claude.

Chwe, M. S.-Y. (2013). *Rational ritual: Culture, coordination, and common knowledge*. Princeton University Press.

Dell, M., Carlson, J., Bryan, T., Silcock, E., Arora, A., Shen, Z., D'Amico-Wong, L., Le, Q., Querubin, P., and Heldring, L. (2024). American stories: A large-scale structured text dataset of historical us newspapers. *Advances in Neural Information Processing Systems*, 36.

Djourelova, M. (2023). Persuasion through slanted language: Evidence from the media coverage of immigration. *American Economic Review*, 113(3):800–835.

Fairclough, N. (1989). *Language and Power*. Longman.

Fairclough, N. (1992). Discourse and text: Linguistic and intertextual analysis within discourse analysis. *Discourse & Society*, 3(2):193–217.

Fischer, R., Luczak-Roesch, M., and Karl, J. A. (2023). What does chatgpt return about human values? exploring value bias in chatgpt using a descriptive value theory.

Garrow, D. J. (1986). *Bearing the Cross: Martin Luther King, Jr., and the Southern Christian Leadership Conference*. William Morrow and Company.

Gentzkow, M. and Shapiro, J. M. (2006). Media bias and reputation. *Journal of Political Economy*, 114(2):280–316.

Gentzkow, M. and Shapiro, J. M. (2007). What drives media slant? evidence from u.s. daily newspapers. *Econometrica*, 78(1):35–71.

Gentzkow, M. and Shapiro, J. M. (2010). What drives media slant? evidence from us daily newspapers. *Econometrica*, 78(1):35–71.

Gentzkow, M., Shapiro, J. M., and Sinkinson, M. (2014). Competition and ideological diversity: Historical evidence from us newspapers. *American Economic Review*, 104(10):3073–3114.

Gentzkow, M., Shapiro, J. M., and Taddy, M. (2019). Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica*, 87(4):1307–1340.

George, L. and Waldfogel, J. (2003). Who affects whom in daily newspaper markets? *Journal of Political Economy*, 111(4):765–784.

Giles, H. and Johnson, P. (1987). Ethnolinguistic identity theory: A social psychological approach to language maintenance. *International Journal of the Sociology of Language*, 68:69–99.

Goldsmith-Pinkham, P., Hull, P., and Kolesár, M. (2022). Contamination bias in linear regressions. Working Paper w30108, National Bureau of Economic Research.

Gramsci, A. (1971). *Selections from the Prison Notebooks*. International Publishers.

Hill, J. H. (2008). *The Everyday Language of White Racism*. Wiley-Blackwell.

Hogg, M. A. and Abrams, D. (1988). *Social Identifications: A Social Psychology of Intergroup Relations and Group Processes*. Routledge.

Hsieh, C., Hurst, E., Jones, C. I., and Klenow, P. J. (2019). The allocation of talent and us economic growth. *Econometrica*, 87(5):1439–1474.

Jeffries, H. K. (2006). *Bloody Lowndes: Civil Rights and Black Power in Alabama's Black Belt*. NYU Press, New York.

Joseph, P. E. (2007). *Waiting'til the midnight hour: A narrative history of Black power in America*. Macmillan.

Key, V. O. (1949). *Southern Politics in State and Nation*. University of Tennessee Press, Knoxville.

Kuziemko, I. and Washington, E. (2018). Why did the democrats lose the south? bringing new data to an old debate. *American Economic Review*, 108(10):2830–2867.

Lacy, S. and Simon, T. F. (1997). Intercounty group ownership of daily newspapers and the decline of competition for readers. *Journalism & Mass Communication Quarterly*, 74(4):814–825.

Logan, T. D. (2018). Do black politicians matter? Technical report, National Bureau of Economic Research.

Martin, R. and Sunley, P. (2003). Deconstructing clusters: chaotic concept or policy panacea? *Journal of Economic Geography*, 3(1):5–35.

Matthews, D. R. and Prothro, J. W. (1966). *Negroes and the New Southern Politics*. Harcourt, Brace and World, New York.

Miotto, M., Rossberg, N., and Kleinberg, B. (2022). Who is gpt-3? an exploration of personality, values and demographics.

Moore, J., Deshpande, T., and Yang, D. (2024). Are large language models consistent over value-laden questions?

Morris, S. and Shin, H. S. (2002). Social value of public information. *American Economic Review*, 92(5):1521–1534.

Mullainathan, S. and Shleifer, A. (2005). The market for news. *American Economic Review*, 95(4):1031–1053.

Omi, M. and Winant, H. (2014). *Racial Formation in the United States*. Routledge.

Ottinger, S. and Winkler, M. (2022). The political economy of propaganda: Evidence from us newspapers. Unpublished manuscript.

Phinney, J. S. (1990). Ethnic identity in adolescents and adults: Review of research. *Psychological Bulletin*, 108(3):499–514.

Poll, G. (1969). 1969 gallup/newsweek poll of black americans. `https://ropercenter.cornell.edu/ipoll`. Published August 20, 2013.

Ren, Y., Ye, H., Fang, H., Zhang, X., and Song, G. (2024). Valuebench: Towards comprehensively evaluating value orientations and understanding of large language models.

Retzlaff, N. (2024). Political biases of chatgpt in different languages.

Rozado, D. (2024). The political preferences of llms. *PLOS ONE*, 19:e0306621.

Röttger, P. et al. (2024). Political compass or spinning arrow? towards more meaningful evaluations for values and opinions in large language models.

Santurkar, S. et al. (2023). Whose opinions do language models reflect? In Krause, A. et al., editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 29971–30004. PMLR.

Smith, T. W. (1992). Changing racial labels: From "colored" to "negro" to "black" to "african american". *Public Opinion Quarterly*, 56(4):496–514.

Staples, B. (2021). How the white press wrote off black america. *The New York Times*, 10.

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black presidential candidate: Evidence using google search data. *Journal of Public Economics*, 118:26–40.

Strömberg, D. (2004). Radio's impact on public spending. *The Quarterly Journal of Economics*, 119(1):189–221.

Tajfel, H. and Turner, J. C. (1979). *An Integrative Theory of Intergroup Conflict*. Brooks-Cole.

Torres, S. (2003). *Black, White, and In Color: Television and Black Civil Rights*. Princeton University Press.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., and Wetherell, M. S. (1987). *Rediscovering the Social Group: A Self-Categorization Theory*. Basil Blackwell.

Woodward, C. V. (1955). *The Strange Career of Jim Crow*. Oxford University Press, New York.

A Appendix

A.1 Tables

Note: Table contains the phrases which were searched for in local US newspapers in the *newspapers.com* and *newspaperarchives.com* online newspaper databases.

Table A.1: Correlates of White Racial Conservatism across Southern Counties

	No State FE	R^2	State FE	R^2	N
Black Population Proportion (1960)	0.002***	0.305	0.004***	0.467	1091
	(0.000)		(0.001)		
Total Population 000's (1960)	-0.023	0.000	-0.146	0.024	1091
	(0.103)		(0.376)		
Republican Vote Share (1964)	-0.048	0.004	0.296***	0.710	1091
	(0.295)		(0.090)		
Newspaper HHI (1964)	-7.159**	0.016	-2.542	0.057	1075
	(2.999)		(11.576)		
Presence of a NAACP Chapter (1960)	0.000	0.000	0.006	0.221	1091
	(0.002)		(0.003)		
% Slave Population (1850)	0.004***	0.261	0.007***	0.437	675
	(0.001)		(0.001)		
Lynchings per 100,000 in the population	0.000***	0.142	0.000***	0.293	1067
	(0.000)		(0.000)		
Suitability of Cotton	0.001	0.045	0.004***	0.268	1046
	(0.001)		(0.001)		
% urban	-0.005	0.004	-0.009	0.104	1047
	(0.007)		(0.010)		
Median family income (1960)	-0.016**	0.096	-0.015*	0.310	1034
	(0.007)		(0.007)		
% Family income $<$ 3000 USD (1960)	0.151**	0.079	0.160**	0.235	1088
	(0.052)		(0.063)		
% Family income $>$ 10,000 USD (1960)	-0.028	0.037	-0.015	0.204	1085
	(0.016)		(0.017)		
% Married (1960)	0.008	0.025	0.003	0.419	1091
	(0.008)		(0.004)		
% population church members (1952)	-0.126*	0.034	-0.296***	0.191	1067
	(0.064)		(0.074)		
% church members Protestant (1952)	-0.108**	0.037	-0.188***	0.149	1067
	(0.042)		(0.035)		
% church members Roman Catholic (1952)	-0.017	0.001	-0.106	0.261	1067
	(0.078)		(0.074)		

(Continued on next page)

(Continued from previous page)

	No State FE	R^2	State FE	R^2	N
% church members Jewish (1952)	-0.001	0.001	-0.002	0.019	1067
	(0.001)		(0.002)		
% church members other denominations (1952)	0.000	0.000	-0.000	0.014	1067
	(0.000)		(0.000)		
Median age (1960)	-0.010**	0.103	-0.005	0.303	1047
	(0.003)		(0.004)		
% population aged 0-4 (1960)	0.025***	0.135	0.030*	0.205	1091
	(0.004)		(0.014)		
% population aged 21+ (1960)	-0.074***	0.189	-0.078**	0.308	1091
	(0.014)		(0.027)		
% population aged 65+ (1960)	-0.016	0.019	-0.011	0.175	1091
	(0.011)		(0.016)		
Median school years 25+ (1960)	-0.013	0.079	-0.015**	0.314	1044
	(0.007)		(0.005)		
% persons 25+ $<$ 4 yrs schooling (1960)	0.093**	0.096	0.154***	0.231	1090
	(0.035)		(0.036)		
% persons 25+ 12+ yrs schooling (1960)	-0.050	0.023	-0.038	0.222	1090
	(0.037)		(0.033)		
Agriculture: % of land area in farms (1959)	-0.139	0.022	0.058	0.326	1067
	(0.179)		(0.163)		

Table A.2: Alternative measure of White Racial Conservatism and Competition.

	Alt. Competition (1964)				Alt. WRC - HHI(1964)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9) Slavery	(10) Cotton
WRC ×		-0.010**	-0.011*	-0.011*		0.025***	0.027***	0.031*	0.048**	0.038**
Comp × Post		(0.005)	(0.005)	(0.007)		(0.008)	(0.009)	(0.018)	(0.019)	(0.016)
Comp ×	0.010	0.019	0.025	0.012	-0.008	-0.030***	-0.031***	-0.031**	-0.022***	-0.030***
Post	(0.012)	(0.018)	(0.018)	(0.024)	(0.005)	(0.007)	(0.007)	(0.015)	(0.006)	(0.008)
WRC ×		-0.001	-0.006	0.005		-0.145***	-0.133***	-0.130	-0.209**	-0.211***
Post		(0.003)	(0.004)	(0.005)		(0.040)	(0.044)	(0.093)	(0.097)	(0.076)
Exposure	C-Vote	C-Vote	C-Vote	C-Vote	D-Device	D-Device	D-Device	D-Device	C-Slavery	C-Cotton
N	22,756	22,756	22,756	20,123	23,198	23,198	23,198	20,495	14,064	22,155
# of Newspapers	166	166	166	150	170	170	170	153	102	163
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Newspaper FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Semester FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Semester FE	No	No	Yes	No	No	No	Yes	No	No	No
DMA × Semester FE	No	No	No	Yes	No	No	No	Yes	No	No

Note: Regressions estimate the differential adjustment in racial label usage to White racial conservatism (WRC), interacted with market competition, following the 1966h2 shift in racial terminology norms. Measures of WRC include: **D-Device**, an indicator for counties that employed voter registration devices or literacy tests in the 1964 election; **C-Vote**, the county-level vote share for the Dixiecrat candidate in the 1948 election; **C-Slavery**, the estimated share of enslaved population in 1860; and **C-Cotton**, county-level agricultural suitability for cotton, a proxy for slave-based production and modern racial attitudes, from Acharya et al. (2016). In the left panel, market competition is measured as a binary indicator equal to 1 if more than one local daily newspaper operated in the county in 1964; in the right panel, market structure is proxied by the Herfindahl-Hirschman Index (HHI). The sample includes local daily newspapers for which circulation data are available. The outcome is the normalised frequency of the label "Black" in outlet-level text, measured at the outlet × semester level. All regressions include newspaper and year-month fixed effects; some also include state or DMA × month fixed effects as indicated. Standard errors are clustered at the county level. *** p<0.01, ** p<0.05, * p<0.10.

Table A.3: Robustness

	Outcome Transfmorations		Exposure Robustness			Alt. HHI
	(1) IHS	(2) Log	(3) Dailies<4	(4) Vote Spread	(5) VRA-Strict	(6) HHI(1960)
WRC ×	0.003**	0.015*	0.004**	0.003*	0.020**	0.024**
Comp × Post	(0.001)	(0.009)	(0.001)	(0.002)	(0.009)	(0.009)
Comp ×	-0.018***	-0.037	-0.023***	-0.020***	-0.018***	-0.560
Post	(0.006)	(0.036)	(0.007)	(0.007)	(0.007)	(0.349)
WRC ×	-0.018**	-0.061	-0.022**	-0.016	-0.124***	-0.002***
Post	(0.008)	(0.053)	(0.009)	(0.010)	(0.047)	(0.001)
Exposure	C-Vote	C-Vote	C-Vote	C-Vote	D-Device	C-Vote
N	22,779	22,779	22,001	22,779	22,779	21,699
# of Newspapers	165	165	159	165	165	158

Note: This table presents robustness checks for the effect of pre-treatment white racial conservatism (WRC) on the relative use of the term "Black" in Southern local newspapers. Each column estimates a difference-in-differences model where WRC is interacted with a post-period indicator and, in some specifications, with media market concentration (Comp). Columns (1) and (2) use a log and IHS-transformed outcome respectively; Column (3) restricts the sample to outlets in less concentrated counties (fewer than three dailies in 1964); Column (4) uses the complete circulation weighted measure of WRC utilising all counties the newspapers circulate in; Column (5) substitutes in a binary indicator for VRA-era strict registration tests; Column (6) interacts WRC with HHI calculated from 1960 circulation data. All regressions include fixed effects for newspaper and semester, as well as interactions of year-month with county-level demographic and political controls measured pre-treatment. Standard errors are clustered at the county level. *** p<0.01, ** p<0.05, * p<0.10.

Table A.4: Placebo DDD with non-daily newspapers

	DDD: Non-Daily Placebo		
	(1)	(2)	(3)
WRC ×	0.005	0.015	-0.009
Post	(0.021)	(0.024)	(0.052)
HHI ×	0.099	0.627	-1.181
Post	(0.692)	(0.779)	(1.683)
WRC × HHI	0.091	0.072	0.384
× Post	(0.233)	(0.221)	(0.523)
Exposure	C-Vote	C-Vote	C-Vote
N	11,409	11,038	8,356
# of Newspapers	132	129	115
Controls	Yes	Yes	Yes
Newspaper FE	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes
State × Month FE	No	Yes	No
DMA × Month FE	No	No	Yes

Note: Each regression estimates the interaction between pre-treatment white racial conservatism (WRC), measured by 1948 Dixiecrat vote share, and the post-period indicator, as well as a triple interaction with media market concentration in each newspaper county of commercial residence. Sample is limited to non-daily newspapers. All models include outlet and month fixed effects, and interactions of month with county-level pre-shock control cross-sections. Standard errors are clustered at the county level. *** p<0.01, ** p<0.05, * p<0.10.

Table A.5: Within-owner estimates

	(1)	(2)
WRC	0.031***	0.041*
× HHI × Post	(0.011)	(0.023)
WRC	-0.152**	-0.156*
× Post	(0.055)	(0.088)
HHI ×	-0.030***	-2.447
Post	(0.009)	(1.525)
WRC	C-Vote	C-Vote
N	6,199	6,199
# of Newspapers	48	48
Controls	Yes	Yes
Newspaper FE	Yes	Yes
Year-Month FE	Yes	Yes
Company × Month FE	No	Yes

Note: This table reports DDD estimates of newspaper proportional use of the Black dictionary on the interaction between white racial conservatism (WRC), post-period status, and media market concentration (HHI), with fixed effects specified at increasingly granular levels. All specifications include interactions between time (month-year) and a set of county-level pre-treatment controls. Column (1) replicates the main DDD results for the subsample of newspapers operating within the same building including state-by-month fixed effects; Column (2) adds company-by-county fixed effects. The outcome variable is the relative usage of the term "Black" versus "Negro". Standard errors are clustered at the county level in Column (1) and company-county level in Column (2). *** p<0.01, ** p<0.05, * p<0.10.

Table A.6: Newswire Horse Race - Proportion of news articles written by local journalists and not newswire services

	All			Daily			DDD		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
WRC × Comp × Post							0.006*** (0.002)		0.006*** (0.002)
WRC × Post		-0.011*** (0.004)	-0.006* (0.003)		-0.007** (0.003)	-0.007** (0.003)	-0.032*** (0.009)		-0.030*** (0.008)
Local Prop × Comp × Post								-0.053* (0.031)	-0.052* (0.030)
Local Prop × Post	-0.189*** (0.032)		-0.182*** (0.032)	-0.022 (0.041)		-0.026 (0.040)		0.171 (0.111)	0.171 (0.110)
Comp × Post							-0.028*** (0.008)	0.011 (0.014)	-0.001 (0.014)
Treatment	C-Local	C-Local	C-Local	C-Local	C-Local	C-Local	C-Vote	C-Vote	C-Vote
N	34,238	34,238	34,238	20,441	20,441	20,441	20,441	20,441	20,441
# of Newspapers	278	278	278	143	143	143	143	143	143
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Newspaper FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Semester FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Semester FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports a statistical horse race of difference-in-differences and triple-differences regressions of the share of articles using the term "Black" (relative to "Negro") on post-period status, white racial conservatism (WRC), market concentration (HHI), and the proportion of articles written by local journalists (not wire services such as AP or UPI). All regressions include newspaper and semester fixed effects and interactions of year-month with pre-treatment county-level controls. Standard errors are clustered at the county level. *** p<0.01, ** p<0.05, * p<0.10.

Table A.7: WRC vs Newspaper Political Alignment (1964) Horse Race

	WRC					DDD-HHI(1964)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
WRC × Comp × Post						0.005*			0.005**
						(0.003)			(0.002)
WRC × Post	-0.011***		-0.011***	-0.010**	-0.009**	-0.032**			-0.031***
	(0.004)		(0.004)	(0.004)	(0.004)	(0.013)			(0.011)
Comp × Post						-0.023***		-0.007	-0.024*
						(0.007)		(0.017)	(0.014)
Dem × Post		0.025	0.002		0.021		0.024	-0.004	-0.009
		(0.026)	(0.016)		(0.027)		(0.026)	(0.091)	(0.067)
Rep × Post		0.037*		0.011	0.027		0.038*	0.121	0.079
		(0.021)		(0.011)	(0.023)		(0.021)	(0.088)	(0.076)
Dem × Comp × Post								0.008	0.008
								(0.019)	(0.014)
Rep × Comp × Post								-0.017	-0.010
								(0.018)	(0.015)
Exposure	C-Vote	None	C-Vote	C-Vote	C-Vote	C-Vote	None	None	C-Vote
N	14,806	14,806	14,806	14,806	14,806	14,806	14,967	14,967	14,806
R^2	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89
# of Newspapers	106	106	106	106	106	106	107	107	106
party	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes

Note: This table reports a statistical horse race of difference-in-differences and triple-differences regressions of the share of articles using the term "Black" (relative to "Negro") on post-period status, white racial conservatism (WRC), market concentration (HHI), and the Political Affiliation of the newspaper taken from USNP (2014). All regressions include newspaper and semester fixed effects and interactions of year-month with pre-treatment county-level controls. Standard errors are clustered at the county level. *** p<0.01, ** p<0.05, * p<0.10.

Table A.8: WRC vs Republican Vote Share (1964) Horse Race

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
WRC × Comp × Post							0.004**		0.005***
							(0.002)		(0.002)
WRC × Post		-0.014***	-0.013***		-0.009**	-0.008**	-0.025***		-0.031***
		(0.005)	(0.005)		(0.004)	(0.004)	(0.008)		(0.009)
Comp × Post							-0.021***	-0.018	-0.046***
							(0.006)	(0.011)	(0.015)
Republican Vote × Comp × Post								0.000	0.001**
								(0.000)	(0.000)
Republican Vote × Post	-0.001		-0.001	-0.001*		-0.001		-0.001	-0.003***
	(0.001)		(0.000)	(0.001)		(0.001)		(0.001)	(0.001)
Treatment	C-Vote	C-Vote	C-Vote	C-Vote	C-Vote	C-Vote	C-Vote	C-Vote	C-Vote
N	49,751	47,926	47,926	22,779	22,779	22,779	22,779	22,779	22,779
# of Newspapers	456	440	440	165	165	165	165	165	165
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Newspaper FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Semester FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Semester FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DMA × Semester FE							No	No	No

Note: This table reports a statistical horse race of difference-in-differences and triple-differences regressions of the share of articles using the term "Black" (relative to "Negro") on post-period status, white racial conservatism (WRC), market concentration (HHI), and the county vote share for teh Republican candidate in the 1964 Presidential election. All regressions include newspaper and semester fixed effects and interactions of year-month with pre-treatment county-level controls. Standard errors are clustered at the county level. *** p<0.01, ** p<0.05, * p<0.10.

Table A.9: White vs Black Preferences Horse Race

	All			Dailies			HHI-DDD		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
WRC × Comp × Post							0.009***		0.010**
							(0.003)		(0.005)
Comp × Post							-0.033**	-0.041**	-0.087***
							(0.013)	(0.018)	(0.018)
WRC × Post	-0.014***		-0.011**	-0.009**		-0.009**	-0.041**		-0.043*
	(0.005)		(0.005)	(0.004)		(0.004)	(0.016)		(0.025)
Black 20 − 30× post		0.560	1.282		4.040	4.293		-3.414	-10.803
		(3.736)	(3.797)		(4.932)	(4.955)		(27.981)	(25.416)
Black 30 − 40× post		-2.179	-3.109		-0.387	-1.217		17.559	53.722
		(5.412)	(5.395)		(9.951)	(10.124)		(91.419)	(73.375)
Black 40 − 50× post		8.691*	9.706**		-1.062	2.952		8.901	29.312
		(4.964)	(4.916)		(8.446)	(9.591)		(82.566)	(65.242)
Black 50 − 60× post		-3.559	-3.859		-1.495	-8.486		-120.861	-271.915***
		(7.196)	(7.039)		(16.700)	(16.786)		(141.408)	(91.011)
Black > 60× post		-0.026	-0.026		-0.008	0.014		0.677	1.350***
		(0.026)	(0.025)		(0.078)	(0.074)		(0.775)	(0.458)
Black NAACP 20 − 30× post		-1.374	-3.034		-1.692	-2.522		12.083	30.193
		(4.015)	(4.055)		(5.127)	(5.116)		(29.581)	(27.425)
Black NAACP 30 − 40× post		3.313	2.418		-4.047	-2.072		-48.346	-106.844
		(6.505)	(6.855)		(10.244)	(10.576)		(95.279)	(80.109)
Black NAACP 40 − 50× post		-9.708	-8.831		5.399	0.332		30.621	29.367
		(7.422)	(7.613)		(8.420)	(9.744)		(87.423)	(68.612)
Black NAACP 50 − 60× post		8.783	7.232		0.855	6.876		108.798	252.365***
		(8.290)	(7.991)		(17.246)	(17.569)		(145.408)	(93.293)
Black NAACP > 60× post		0.003	0.026		-0.004	-0.009		-0.791	-1.347***
		(0.037)	(0.035)		(0.084)	(0.080)		(0.797)	(0.460)
Black 20 − 30× Comp × Post								3.156	4.272
								(6.586)	(5.540)
Black 30 − 40× Comp × Post								-6.998	-12.696
								(19.631)	(15.478)
Black 40 − 50× Comp × Post								-2.793	-8.039
								(15.924)	(12.683)
Black 50 − 60× Comp × Post								30.223	59.005***
								(26.508)	(18.147)
Black > 60× Comp × Post								-0.160	-0.285***
								(0.148)	(0.091)
Black NAACP 20 − 30× Comp × Post								-3.217	-6.610
								(6.981)	(6.109)
Black NAACP 30 − 40× Comp × Post								11.940	21.611
								(20.285)	(16.673)
Black NAACP 40 − 50× Comp × Post								-6.969	-4.781
								(16.945)	(13.466)
Black NAACP 50 − 60× Comp × Post								-25.641	-54.159***
								(27.657)	(18.764)
Black NAACP > 60× Comp × Post								0.183	0.291***
								(0.157)	(0.092)
Exposure	C-Vote	None	C-Vote	C-Vote	None	C-Vote	C-Vote	None	C-Vote
N	47,926	49,751	47,926	22,829	23,201	22,829	20,128	18,925	18,553
# of Newspapers	440	456	440	167	170	167	150	142	139
R^2	0.60	0.58	0.60	0.86	0.85	0.86	0.87	0.85	0.86

Note: This table reports a statistical horse race between WRC and Black Age demographics. Difference-in-differences and triple-differences regressions are of the share of articles using the term "Black" (relative to "Negro") on post-period status, white racial conservatism (WRC), market concentration (HHI), and the county share of Black by age demographics and the presence of an NAACP chapter, including interactions of the two. All regressions include newspaper and semester fixed effects and interactions of year-month with pre-treatment county-level controls. Standard errors are clustered at the county level. *** p<0.01, ** p<0.05, * p<0.10.

Table A.10: WRC and Competition on Label diffusion - County Level

	(1)	(2)	(3)
WRC ×	-0.081***	-0.064**	-0.040**
Post	(0.024)	(0.027)	(0.014)
HHI ×	-0.644***	-0.510**	-0.380*
Post	(0.120)	(0.166)	(0.187)
WRC × HHI	0.011***	0.008**	0.009**
× Post	(0.003)	(0.003)	(0.004)
N	5,530	5,530	5,482
# of Counties	925	925	917
Controls	Yes	Yes	Yes
County FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
State × Year FE	No	Yes	No
DMA × Year FE	No	No	Yes

Note: This table reports triple-differences regression estimates of county level exposure to the term "Black" (relative to "Negro") on post-period status, white racial conservatism (WRC), market concentration (HHI) at the county level. County level exposure to Black dictionary is a circulation weighted with circulation data taken from the Audit Bureau of Circulations 1964 newspaper level audits. All regressions include newspaper and semester fixed effects and interactions of year-month with pre-treatment county-level controls. Standard errors are clustered at the county level. *** p<0.01, ** p<0.05, * p<0.10.

Table A.11: DDD Event Study Estimates: Media Racial Language and Black Elected Officials

	(1) All local officials (%)	(2) County governments (%)	(3) School board (%)	(4) Municipality (%)
1964 × VRA × $\widetilde{black_ct}$	-0.0400	0.0241	-0.0798	-0.0960
	(0.0778)	(0.113)	(0.124)	(0.0603)
1968 × VRA × $\widetilde{black_ct}$	0.163	0.227*	0.0454	0.0921
	(0.101)	(0.119)	(0.163)	(0.0785)
1970 × VRA × $\widetilde{black_ct}$	0.189*	0.331**	0.106	0.0459
	(0.101)	(0.149)	(0.148)	(0.0773)
1972 × VRA × $\widetilde{black_ct}$	0.227**	0.305**	0.214	0.0566
	(0.104)	(0.129)	(0.180)	(0.0823)
1974 × VRA × $\widetilde{black_ct}$	0.226**	0.370***	0.152	0.0580
	(0.0976)	(0.121)	(0.187)	(0.0815)
1964 × VRA	-0.566	-0.934	0.698	-1.168
	(1.426)	(1.022)	(1.458)	(2.263)
1968 × VRA	3.197	2.124	13.74	-0.404
	(4.540)	(6.553)	(9.669)	(3.849)
1970 × VRA	3.206	-1.738	8.210	5.626
	(7.101)	(9.151)	(10.47)	(7.455)
1972 × VRA	-0.0324	-10.58	2.222	-1.925
	(6.287)	(8.486)	(15.55)	(6.784)
1974 × VRA	4.619	-13.11	3.720	7.015
	(7.644)	(8.803)	(14.76)	(7.536)
1964 × $\widetilde{black_ct}$	0.0103	0.00399	-0.0145	0.0218*
	(0.00871)	(0.00649)	(0.0142)	(0.0118)
1968 × $\widetilde{black_ct}$	-0.0153	-0.0173	0.0269*	-0.0271
	(0.0126)	(0.0132)	(0.0138)	(0.0210)
1970 × $\widetilde{black_ct}$	-0.0192	-0.0145	0.0416**	-0.0399*
	(0.0130)	(0.0118)	(0.0166)	(0.0219)
1972 × $\widetilde{black_ct}$	0.00311	-0.0118	0.0342	0.0208
	(0.0155)	(0.0116)	(0.0207)	(0.0272)
1974 × $\widetilde{black_ct}$	0.00227	-0.0179	0.0217	0.0143
	(0.0219)	(0.0162)	(0.0343)	(0.0281)
N	5,369	5,510	4,980	5,384
County FE	Yes	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes
Controls × VRA × Year-FE	Yes	Yes	Yes	Yes
Black Share × VRA × Year-FE	Yes	Yes	Yes	Yes
Conflict Controls × VRA × Year-FE	Yes	Yes	Yes	Yes

Note: This table reports event-study estimates from triple-difference regressions of Black political incorporation outcomes on interactions between year, Voting Rights Act (VRA) coverage, and a newspaper-level racial language index ($\widetilde{black_{ct}}$). All models include county fixed effects, year fixed effects, and interactions between VRA coverage and county-level controls. Standard errors are clustered at the judicial division level. *** p<0.01, ** p<0.05, * p<0.10.

Table A.12: Triple-Difference Estimates: County Government %

	(1) County gov (%)	(2) County gov (%)	(3) County gov (%)	(4) County gov (%)	(5) County gov (%)	(6) County gov (%)
Post-1965 × VRA × $\widetilde{black_ct}$	0.139	0.315***	0.318***	0.297**	0.286**	0.264***
	(0.0991)	(0.116)	(0.118)	(0.123)	(0.120)	(0.0952)
Post-1965 × VRA	-3.164**	3.671	1.446	-1.583	-3.337	7.205*
	(1.428)	(2.428)	(7.559)	(6.657)	(6.924)	(3.667)
Post-1965 × $\widetilde{black_ct}$	0.0628**	-0.0137	-0.0116	-0.0107	-0.0157	-0.00468
	(0.0286)	(0.0122)	(0.0124)	(0.0108)	(0.0125)	(0.0132)
N	5,710	5,710	5,710	5,710	5,510	5,734
County FE	Yes	Yes	Yes	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls × Year	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls × VRA × Year	No	Yes	Yes	Yes	Yes	Yes
Black Share Controls × VRA × Year	No	No	Yes	Yes	Yes	Yes
Conflict Controls	No	No	No	Yes	Yes	Yes
Market Structure Controls	No	No	No	No	Yes	Yes
Residualised	No	No	No	No	No	Yes
Judicial Divisions (Clusters)	134	134	134	134	134	134

Note: Estimates from triple-difference regressions of Black county official share on post-1965 interactions with VRA coverage and racial language index. Columns vary in controls, including county fixed effects, year fixed effects, VRA-by-controls, Black share-by-VRA-by-year interactions, and residualisation of the index. Standard errors are clustered at the judicial division level. *** p<0.01, ** p<0.05, * p<0.10.

Table A.13: Triple-Difference Estimates: County Government % (Lagged - 1 year)

	1-Year Lag				
	(1)	(2)	(3)	(4)	(5)
	County gov (%)	County gov (%)	County gov (%)	County gov (%)	County gov (%)
Post-1965 × VRA × $\widetilde{black}_{ct}$	0.284**	0.289**	0.267*	0.253*	0.250**
	(0.136)	(0.137)	(0.141)	(0.137)	(0.110)
Post-1965 × VRA	3.680	1.097	-1.908	-3.777	6.582**
	(2.567)	(7.280)	(6.548)	(6.857)	(2.767)
Post-1965 × $\widetilde{black}_{ct}$	-0.0143	-0.0130	-0.0126	-0.0153	-0.0116
	(0.0109)	(0.0108)	(0.00934)	(0.0103)	(0.0107)
N	5,708	5,708	5,708	5,510	5,694
County FE	Yes	Yes	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes	Yes
Baseline Controls × VRA × Year	Yes	Yes	Yes	Yes	Yes
Black Share Controls × VRA × Year	No	Yes	Yes	Yes	No
Conflict Controls	No	No	Yes	Yes	No
Market Structure Controls	No	No	No	Yes	No
Residualised	No	No	No	No	No
Judicial Divisions (Clusters)	135	135	135	135	134

Note: As in Table A.12, but using a one-year lag of the racial language index. Columns progressively add VRA-by-controls, Black share-by-VRA-by-year interactions, and residualised index specifications. Standard errors clustered at the judicial division level. *** p<0.01, ** p<0.05, * p<0.10.

Table A.14: Triple-Difference Estimates: County Government % (Lagged - 2 years)

	2-Year Lag				
	(1)	(2)	(3)	(4)	(5)
	County gov (%)	County gov (%)	County gov (%)	County gov (%)	County gov (%)
Post-1965 × VRA × $\widetilde{black}_{ct}$	0.250**	0.256**	0.240**	0.224**	0.264***
	(0.110)	(0.114)	(0.114)	(0.112)	(0.0952)
Post-1965 × VRA	6.582**	4.128	1.352	-0.429	7.205*
	(2.767)	(7.405)	(6.826)	(7.044)	(3.667)
Post-1965 × $\widetilde{black}_{ct}$	-0.0116	-0.00919	-0.00871	-0.0119	-0.00468
	(0.0107)	(0.0106)	(0.00924)	(0.0104)	(0.0132)
N	5,694	5,694	5,694	5,496	5,734
County FE	Yes	Yes	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes	Yes
Baseline Controls × VRA × Year	Yes	Yes	Yes	Yes	Yes
Black Share Controls × VRA × Year	No	Yes	Yes	Yes	Yes
Conflict Controls	No	No	Yes	Yes	Yes
Market Structure Controls	No	No	No	Yes	Yes
Residualised	No	No	No	No	Yes
Judicial Divisions (Clusters)	134	134	134	134	134

Note: As in Table A.12, but using a two-year lag of the racial language index. Specifications are identical. Standard errors clustered at the judicial division level. *** p<0.01, ** p<0.05, * p<0.10.

Table A.15: Event Study Interactions: Black Share and Media Racial Language by VRA Coverage

	Not Covered	**Covered**
	(1)	(2)
	County gov (%)	County gov (%)
1964 × Black Share	0.0365	0.565
	(0.0329)	(0.351)
1968 × Black Share	-0.0848	-0.562
	(0.0558)	(0.383)
1970 × Black Share	-0.0835	-0.693
	(0.0566)	(0.456)
1972 × Black Share	-0.0835	-0.721
	(0.0663)	(0.533)
1974 × Black Share	-0.102	-1.197**
	(0.0676)	(0.597)
1964 × $\widetilde{black}_ct$	-0.00318	-0.284
	(0.00459)	(0.188)
1968 × $\widetilde{black}_ct$	0.00785	0.494**
	(0.0116)	(0.220)
1970 × $\widetilde{black}_ct$	0.00609	0.669**
	(0.0113)	(0.315)
1972 × $\widetilde{black}_ct$	0.00888	0.671**
	(0.0127)	(0.292)
1974 × $\widetilde{black}_ct$	-0.00718	0.492
	(0.0198)	(0.298)
1964 × Black Share × $\widetilde{black}_ct$	0.000904	0.0133
	(0.000824)	(0.00859)
1968 × Black Share × $\widetilde{black}_ct$	-0.00237	-0.0107
	(0.00154)	(0.00772)
1970 × Black Share × $\widetilde{black}_ct$	-0.00200	-0.0132
	(0.00154)	(0.00953)
1972 × Black Share × $\widetilde{black}_ct$	-0.00198	-0.0130
	(0.00133)	(0.0102)
1974 × Black Share × $\widetilde{black}_ct$	-0.00148	-0.00563
	(0.00145)	(0.0102)
N	2,687	2,823
County FE	Yes	Yes
State-Year FE	Yes	Yes
Baseline Controls × Year	Yes	Yes
Baseline Controls × VRA × Year	Yes	Yes
Black Share Controls × VRA × Year	Yes	Yes
Conflict Controls	Yes	Yes
Market Structure Controls	Yes	Yes
Judicial Divisions (Clusters)	71	71

Note: Event study coefficients from separate regressions for VRA-covered and non-covered counties, including year-by-Black share, year-by-racial language index ($\widetilde{black}_{ct}$), and three-way interactions. All models control for county and state-year fixed effects, and include VRA-by-controls and Black share-by-VRA-by-year interactions. Standard errors clustered at the judicial division level. *** p<0.01, ** p<0.05, * p<0.10.

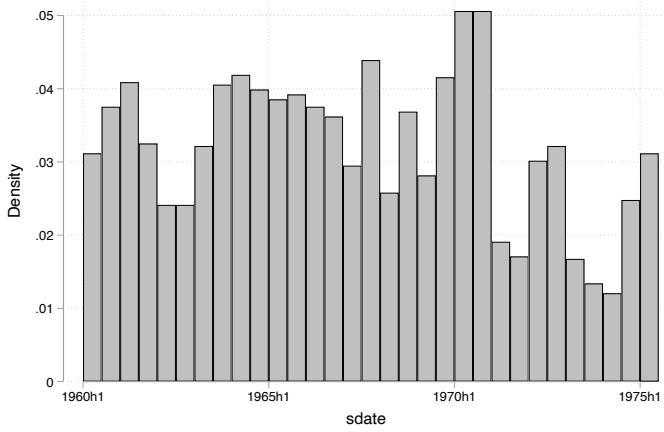
Table A.16: Change in Voter Registration by Race (1964 - 1974)

	(1)	(2)
	Δ ln(Black Reg.)	Δ ln(White Reg.)
Δ $\widetilde{black}_{ct}$	0.491	0.314
	(0.853)	(0.352)
Δ $\widetilde{black}_{ct}$ $\times$ VRA	-1.582	-0.361
	(1.733)	(0.487)
Black Share (1960)	0.0117*	-0.000219
	(0.00602)	(0.00194)
Black Share (1960) $\times$ VRA	0.0175**	0.00625**
	(0.00748)	(0.00274)
N	595	594
State FE	Yes	Yes
Baseline Controls	Yes	Yes
Judicial Divisions (Clusters)	98	98

Note: Regressions of changes in log voter registration rates on change in newspaper racial language index (Δ $\widetilde{black}_{ct}$), its interaction with VRA coverage, and 1960 Black share interacted with VRA. Separate models for Black and White registration. Models include baseline county-level controls and state fixed effects. Observations weighted by 1960 population. Standard errors clustered at the judicial division level. *** p<0.01, ** p<0.05, * p<0.10.

A.2 Figures

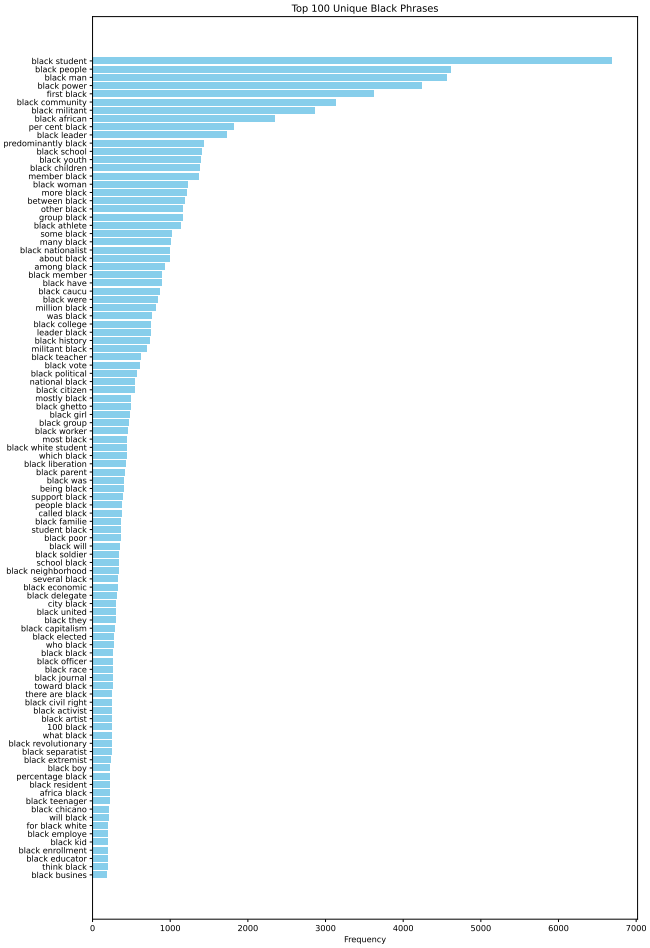
Figure A.1: Training dataset histogram by year×semester.



Note: Histogram of sentence-level training data used in classification tasks, grouped by semester. Bins are six-month intervals based on sentence date ('sdate') in the labeled dataset.

Note: Scatterplot of the relative use of "Black" dictionary compared to "Negro" dictionary by US local newspapers by different methodology of dictionary phrase counting from *newspapers.com*. The y-axis is the year - semester racial index based on the number of pages which make use of each phrase from each dictionary. The x-axis is year - semester racial index based on the total number of mentions of each phrase in each dictionary.

Figure A.2: Frequency of most common unique n-grams in dictionary building corpus.

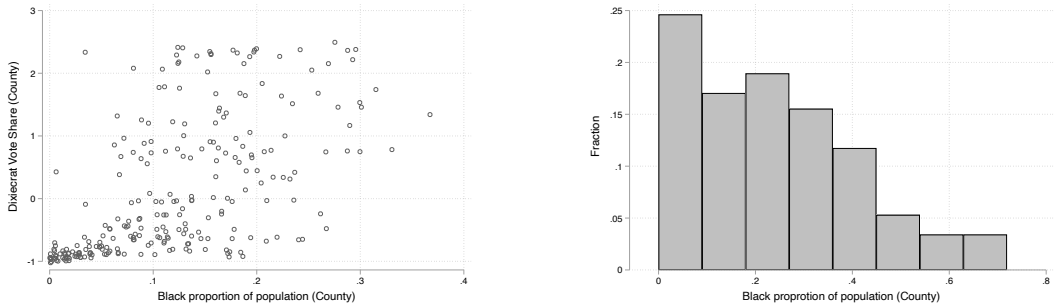


Note: Frequencies of the 100 most common bi-grams and tri-grams containing the word "Black," extracted from a corpus of newspaper articles. Phrases were filtered to exclude stopwords, short words, and known named entities (e.g., "Black Panther"), then de-duplicated using fuzzy string matching.

Figure A.3: Black population (1960) and White racial conservatism (1948) distributions.

(a) Scatter Plot - Dixiecrat vote share and Black population proportion (county)

(b) Histogram - Black population proportion (county)



Note: Panel (a) plots the relationship between the proportion of a county's population that was Black in 1960 and the share of voters supporting the Dixiecrat ticket in 1948. Each dot represents a county. Panel (b) shows the distribution of the Black population share across counties in the sample. The Black population proportion is based on 1960 Census data.

Figure A.4: Scatter Plot - Counts vs Mentions

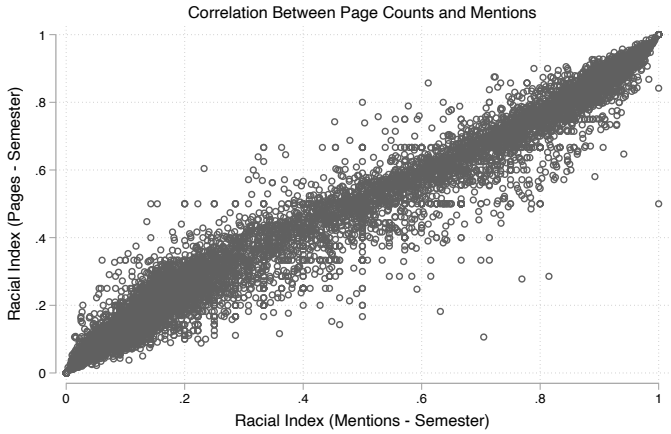


Figure A.5: Change in proportional use of "Black" dictionary by county weighted by circulation (1965 - 1968).

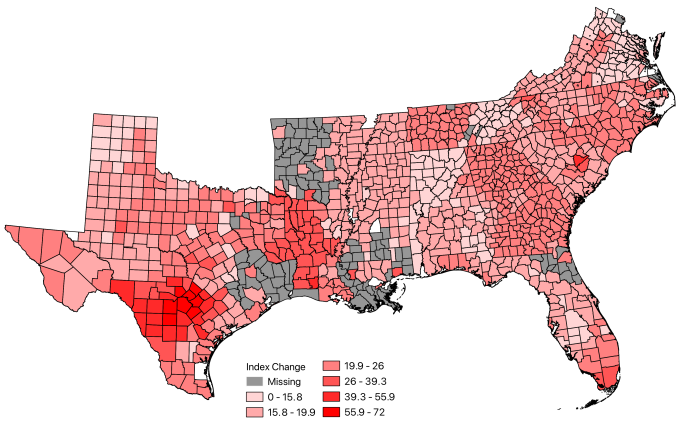
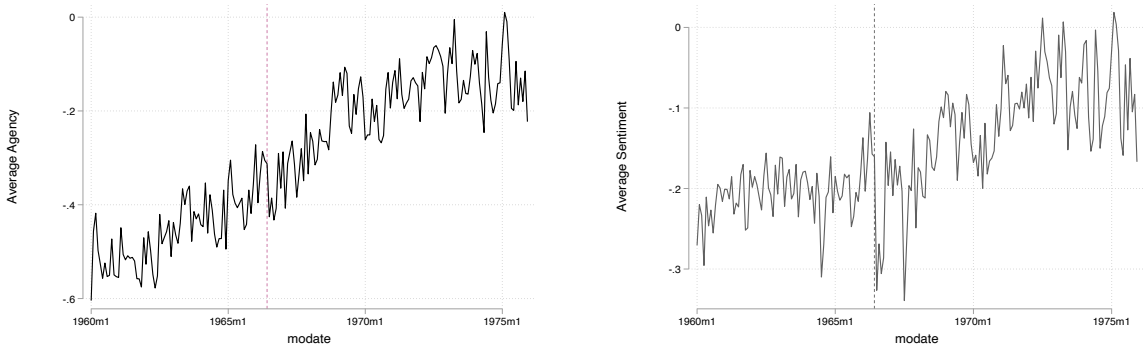


Figure A.6: Evolution of agency and sentiment in Southern Newspapers (1960 - 1975).



(a) Agency (b) Sentiment

Note: The figures present the average agency and sentiment scores of newspaper articles between 1960 and 1975.

Figure A.7: Spatial distribution of voter registration device or test use in the 1964 Presidential election

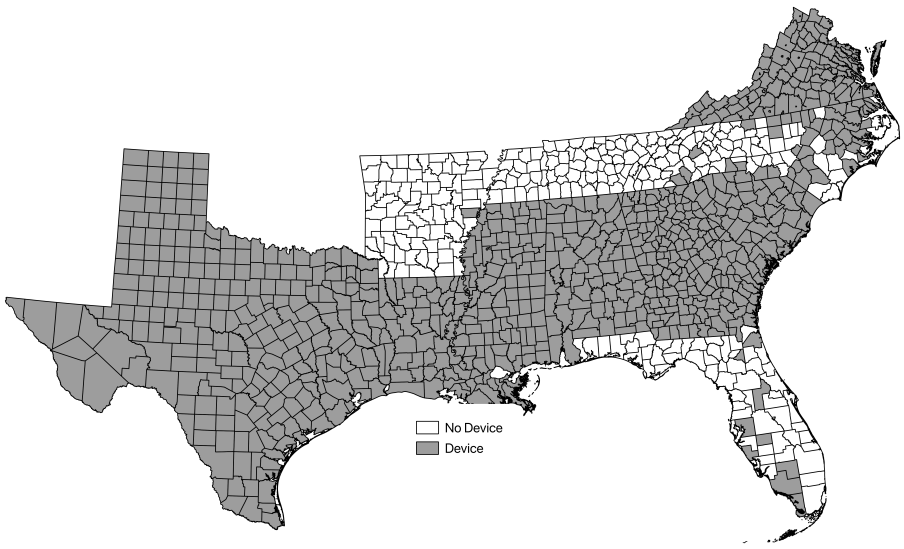
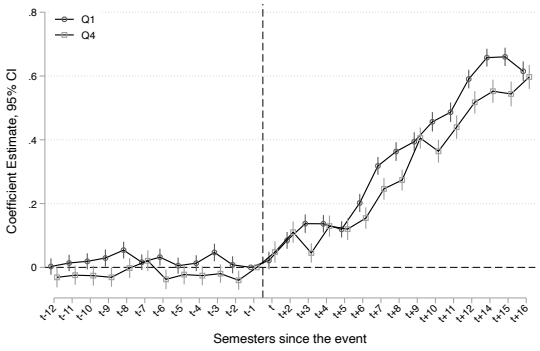
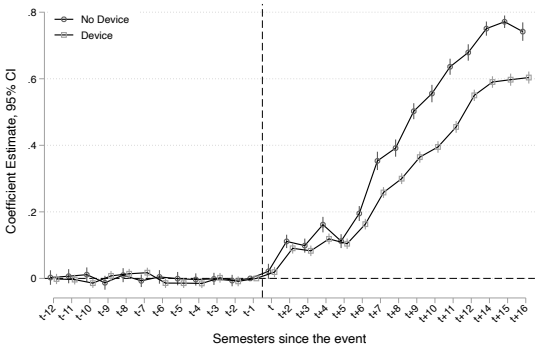


Figure A.8: The difference in the use of racial labels in Southern local newspapers (1960–1973)

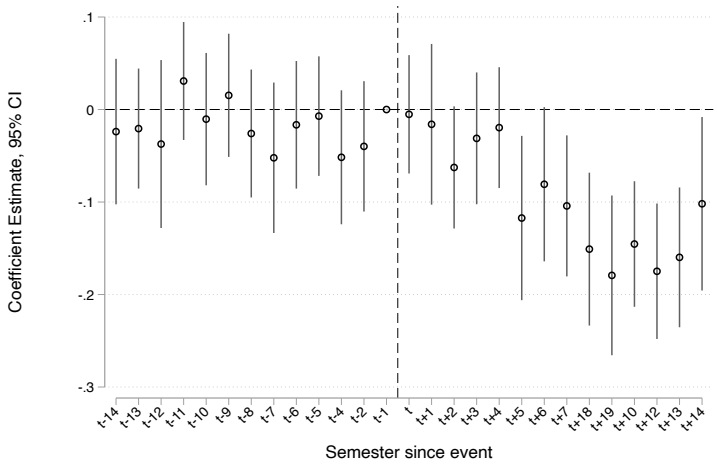
(a) Dixiecrat Vote Share



(b) Registration Device



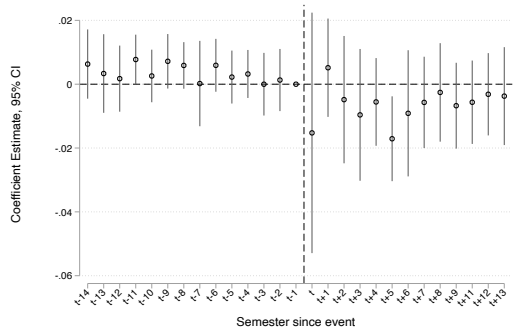
(c) Registration Device



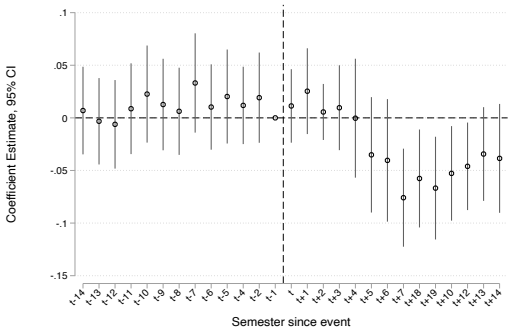
Note: Regressions of outlets' semester use of the term "Black" relative to "Negro" over time, split by (a) quartiles of 1948 Dixiecrat vote share and (b) 1964 use of voter registration devices. Omitted category is the semester before the shock. Regressions are descriptive (no controls). Panel (c) Regressions of outlets' semester use of the term "Black" relative to the term "Negro" on the interaction of county use of a voter registration device or test in the 1964 Presidential election with a set of indicators for semester pre-/ post- shock. The omitted category is the semester before the shock. Regressions include control variables, outlet FE, year-semester FE, State FE × year-semester FE. All regression standard errors are clustered at the county level.

Figure A.9: Dynamic difference in differences - Voter registration tests - US local daily newspapers (1960 - 1973)

(a) 'Dixiecrat' Vote Share



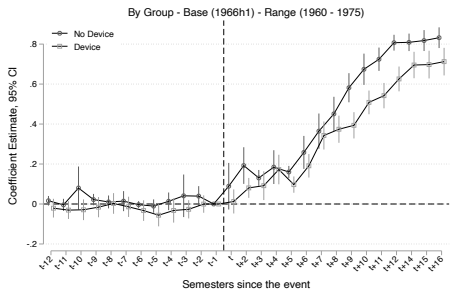
(b) Registration Device



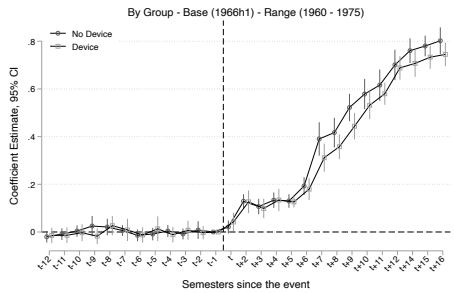
Note: Regressions of outlets' semester use of the term "Black" relative to the term "Negro" on the interaction of (a) newspaper exposure to the 1948 presidential election 'Dixiecrat' vote share; and (b) county use of a voter registration device or test in the 1964 Presidential election with a set of indicators for semester pre-/ post- shock. Sample restricted to daily newspapers. The omitted category is the semester before the shock. Regressions include control variables, outlet FE, year-semester FE, State FE × year-semester FE. Standard errors are clustered at the county level.

Figure A.10: Effect of competition and racial conservatism on the evolution of racial labels in Southern local newspapers (1960 - 1973)

(a) Low Concentration



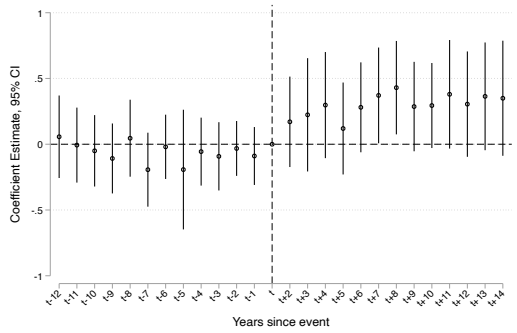
(b) High Concentration



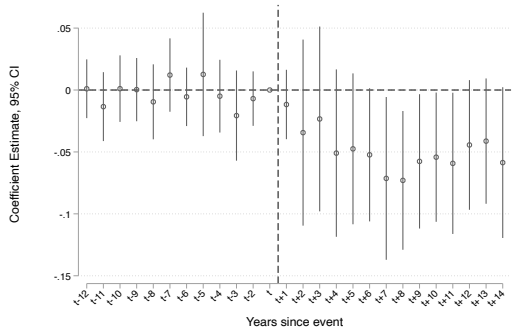
Note: Regressions of outlets' semester use of the term "Black" relative to the term "Negro" over time split by whether the county made use of a voter registration device or test in the 1964 Presidential election. The omitted category is the semester before the shock. Panel (a)/(b): is a sub-sample of daily newspapers which circulate in a low/high concentration market. Regressions do not control for any variables and so are descriptive in nature. Standard errors clustered by county.

Figure A.11: Dynamic Double Interactions - Racial Conservatism - Market Competition (1960 - 1973)

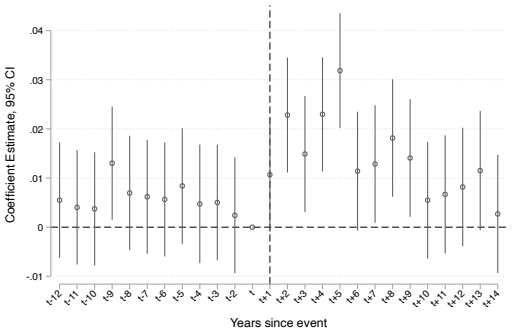
(a) 'Dixiecrat' ×Concentration×Post



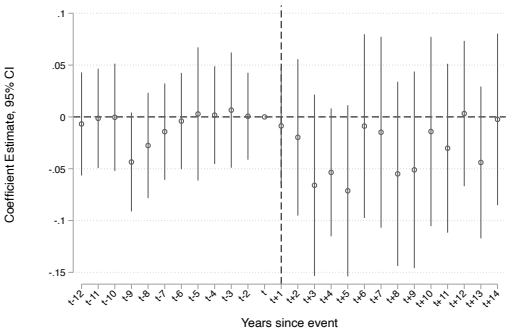
(b) 'Dixiecrat' ×Comp-Dummy×Post



(c) Device×Concentration×Post



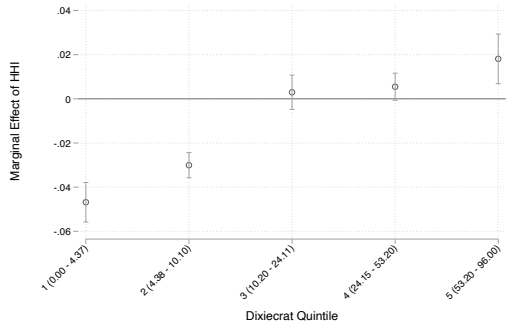
(d) Device ×Comp-Dummy×Post



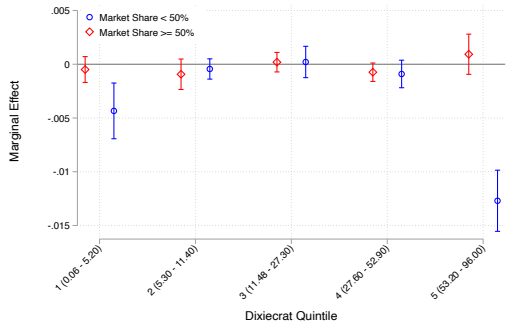
Note: Dynamic regressions of outlets' semester use of the term "Black" relative to the term "Negro" on the double interaction of measures for White racial conservatism with: (1) whether the daily newspaper outlet faces competition from another local daily newspaper outlet, and (2) a set of indicators for semester pre-/ post- shock. The omitted category is the semester before the shock. Regressions include outlet FE, year-month FE, State FE × year-month FE and control variables × year-month FE. Standard errors are clustered at the county level. Panels (a) and (b) use the 'Dixiecrat' vote share to proxy White Racial Conservatism. Panels (c) and (d) use a dummy indicator for the presence of a restrictive voter registration device in the 1964 presidential election as a proxy for White racial conservatism. For measures competition, (a) and (c) use the continuous HHI measure, while panels (b) and (d) use a dummy for the presence of more than one newspaper operating in the county of commercial residence.

Figure A.12: Marginal Effect of HHI by Dixiecrat Quintiles

(a) All

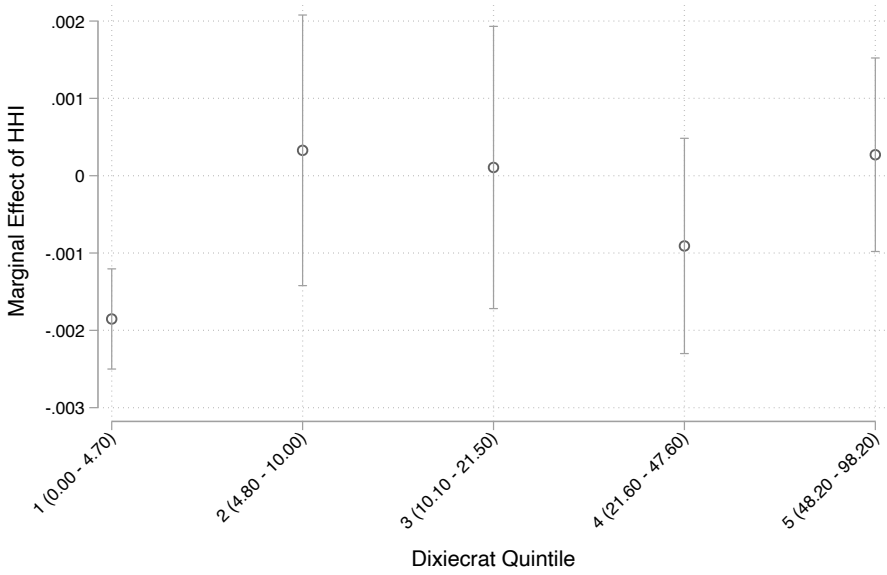


(b) Split by 50% Market Share



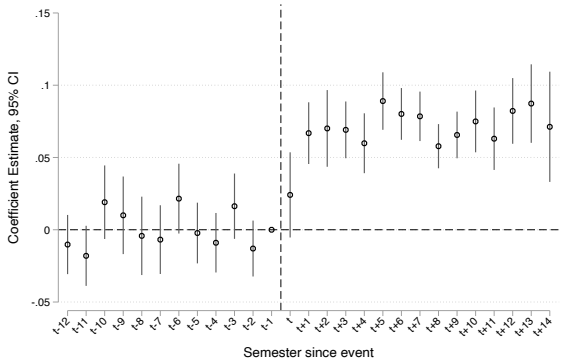
Note: Results are linear combinations from OLS regressions of the effect of HHI×post treatment on outlets' semester use of the term "Black" relative to "Negro," by Dixiecrat Vote Share Quintiles. Standard errors are clustered at the county level.

Figure A.13: Marginal Effect of HHI on County label variance in Racial Index, by 'Dixiecrat' Quintile (Post-Treatment).

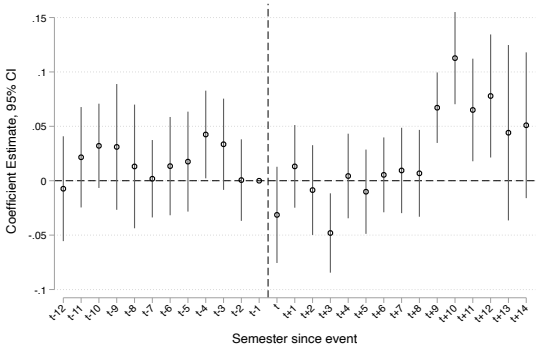


Note: Results are linear combinations from OLS regressions of the effect of HHI×post treatment on the county variance in label adjustment by Dixiecrat Vote Share Quintiles. Standard errors are clustered at the State level.

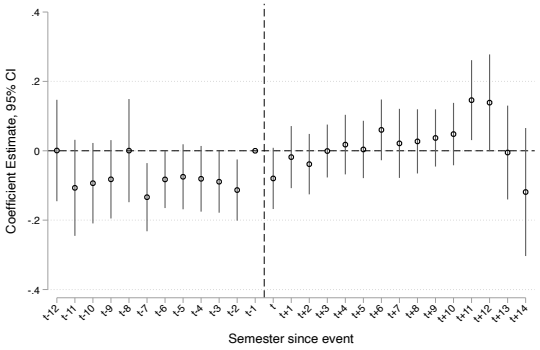
Figure A.14: Dynamic DiD - High agency and positive sentiment on Black label used for entity (by entity type)



(a) Group



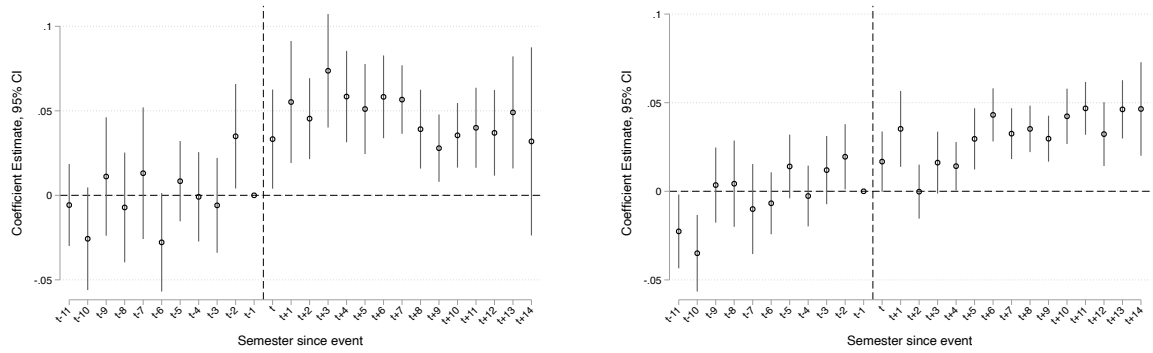
(b) Person



(c) Organisation

Note: The figures present the average year-semester difference in the probability of a Black entity being represented as being a combined high agency and positive sentiment between when the entity is described using "Black" compared to "Negro" over the period 1960 to 1973. The omitted semester is 1965h2. The top figure represents Group entities, while the bottom two represent Person and Organisation entities.

Figure A.15: Dynamic DiD - High agency and positive sentiment on Black label used for entity.

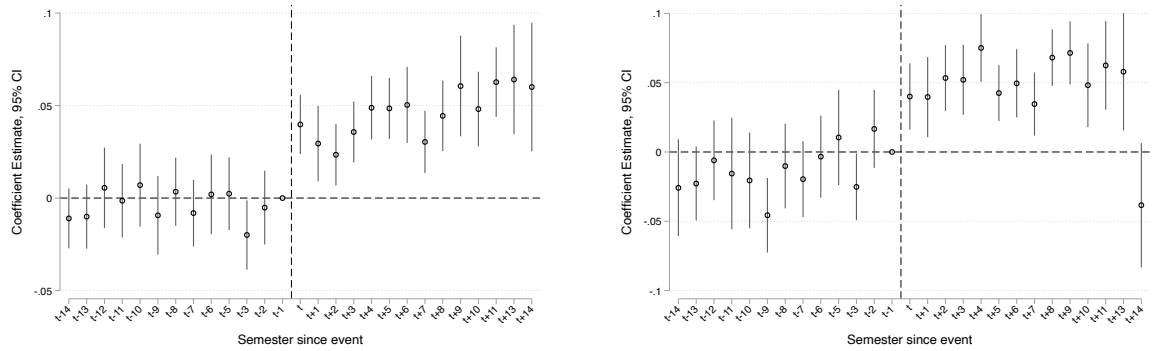


(a) Within-article. Entities held constant.

(b) Within-entity-outlet.

Note: The figures present the average year-semester difference in the probability of a Black entity being represented as being a combined high agency and positive sentiment between when the entity is described using "Black" vs. "Negro" compared to 1952h2. Panel a) includes article-fixed effects and year-day fixed effects and restricts the sample to articles which feature sentences with both "Black" entities and "Negro" entities, and those entities featured both before and after June 16 of 1966. Panel b) includes newspaper-entity-fixed effects and day fixed effects. Standard errors for all regressions are clustered at the firm level.

Figure A.16: Dynamic DiD - High agency and positive sentiment on Black label used for entity by source (newswire vs. local)

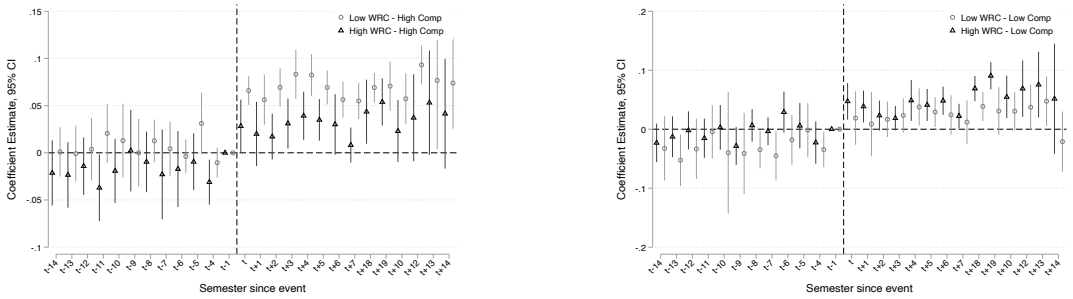


(a) Locally Written

(b) Associated Press

Note: The figures present the average year-semester difference in the probability of a Black entity being represented as being a combined high agency and positive sentiment between when the entity is described using "Black" compared to "Negro" over the period 1960 to 1973. The omitted semester is 1965h2. The left panel uses sentences from articles written by local journalists, while the right uses sentences form articles written by the Associated Press.

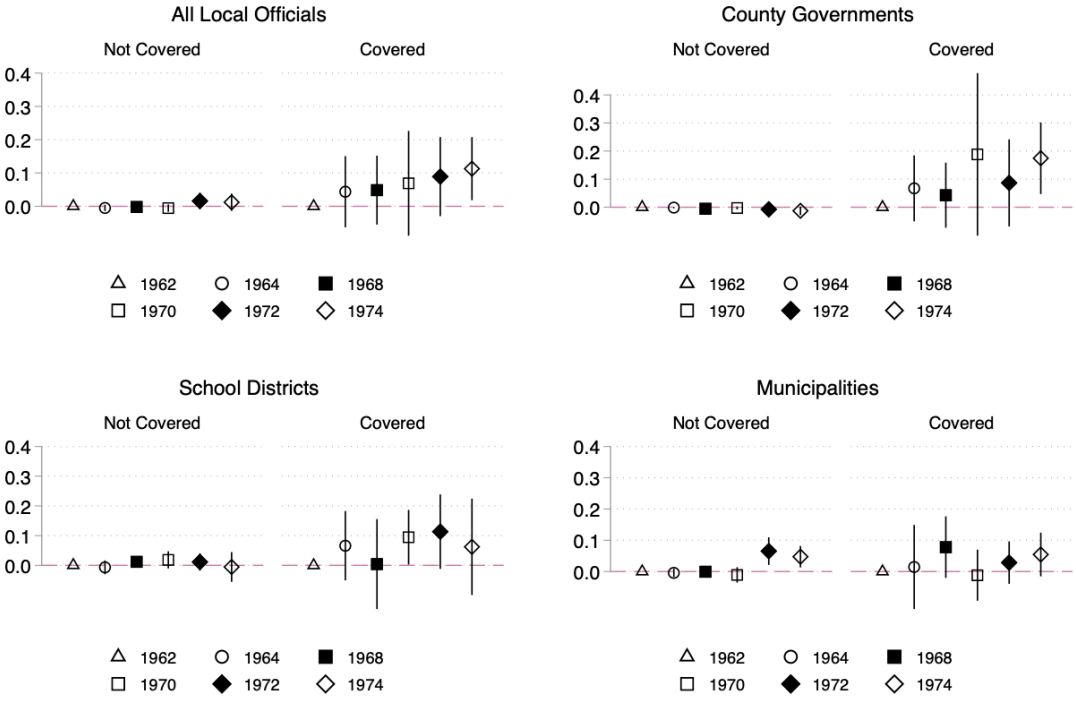
Figure A.17: Dynamic DiD - High agency and positive sentiment on Black label used for entity by WRC and Concentration



(a) Q1 - HHI (b) Q4 - HHI

Note: The figures present the average year-semester difference in the probability of a Black entity being represented as being a combined high agency and positive sentiment between when the entity is described using "Black" compared to "Negro" over the period 1960 to 1973. The omitted semester is 1965h2. Estimates in both panels are split into the first and fourth quartiles of newspaper exposure to White Racial Conservatism. Panel a) restricts the sample to newspapers exposed to the lowest quartile of concentration, i.e. most competitive. The left panel restricts the sample to the highest quartile of concentration, i.e. least competitive. Standard errors clustered at the newspaper level.

Figure A.18

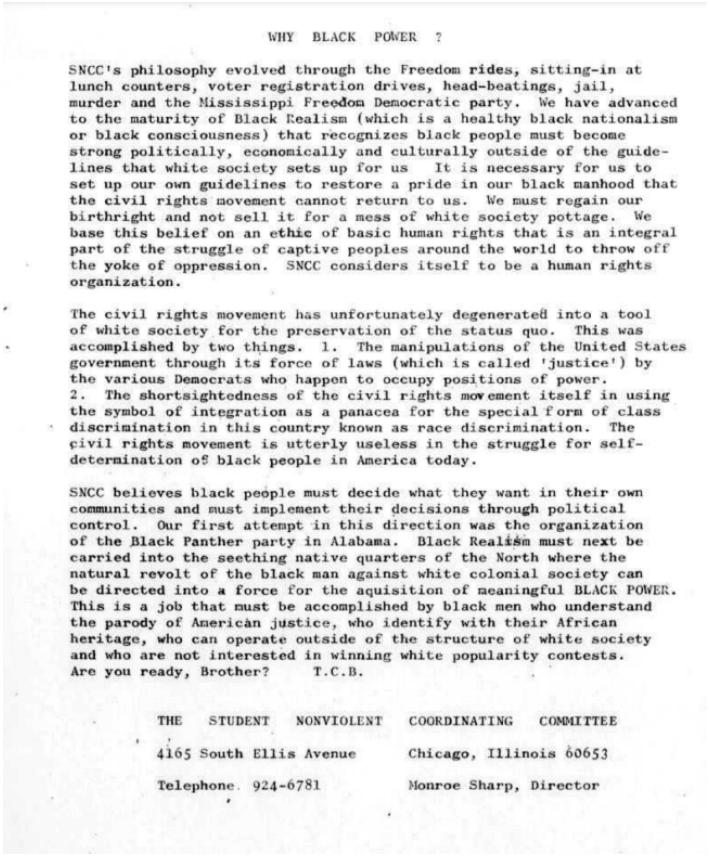


Note: Each panel displays the estimated marginal effect of county-level newspaper language measured by the proportion of total newspaper mentions using the term *Black* (rather than *Negro*), circulation-weighted, on the share of local government officials who are Black. Estimates are from separate cross-sectional regressions for each year (1962, 1964, 1968, 1970, 1972, 1974), with *ri* interacted with an indicator for coverage under Section 5 of the Voting Rights Act. Panels correspond to: (a) all local officials, (b) county governments, (c) school districts, and (d) municipalities. All regressions include controls for baseline demographics, newspaper characteristics, and their interactions with Section 5 coverage, with state-by-year fixed effects absorbed. Standard errors clustered by judicial district.

B Appendix B

B.1 Black Power Marketing Material

Figure B.1: Guidelines and marketing material for the Black Power movement. July, 1966



WHY BLACK POWER ?

SNCC's philosophy evolved through the Freedom rides, sitting-in at lunch counters, voter registration drives, head-beatings, jail, murder and the Mississippi Freedom Democratic party. We have advanced to the maturity of Black Realism (which is a healthy black nationalism or black consciousness) that recognizes black people must become strong politically, economically and culturally outside of the guidelines that white society sets up for us It is necessary for us to set up our own guidelines to restore a pride in our black manhood that the civil rights movement cannot return to us. We must regain our birthright and not sell it for a mess of white society pottage. We base this belief on an ethic of basic human rights that is an integral part of the struggle of captive peoples around the world to throw off the yoke of oppression. SNCC considers itself to be a human rights organization.

The civil rights movement has unfortunately degenerated into a tool of white society for the preservation of the status quo. This was accomplished by two things. 1. The manipulations of the United States government through its force of laws (which is called 'justice') by the various Democrats who happen to occupy positions of power. 2. The shortsightedness of the civil rights movement itself in using the symbol of integration as a panacea for the special form of class discrimination in this country known as race discrimination. The civil rights movement is utterly useless in the struggle for self-determination of black people in America today.

SNCC believes black people must decide what they want in their own communities and must implement their decisions through political control. Our first attempt in this direction was the organization of the Black Panther party in Alabama. Black Realism must next be carried into the seething native quarters of the North where the natural revolt of the black man against white colonial society can be directed into a force for the aquisition of meaningful BLACK POWER. This is a job that must be accomplished by black men who understand the parody of American justice, who identify with their African heritage, who can operate outside of the structure of white society and who are not interested in winning white popularity contests. Are you ready, Brother? T.C.B.

THE STUDENT NONVIOLENT COORDINATING COMMITTEE

4165 South Ellis Avenue Chicago, Illinois 60653

Telephone 924-6781 Monroe Sharp, Director

Table B.1: Example of editorial discretion of label choice in published identical newswire articles. 18 July, 1966

Violence Flares in Mississippi

GREENWOOD, Miss. (UPI) — Dr. Martin Luther King Jr. and Black Nationalist leader Stokely Carmichael led a 600-man march on this delta town's courthouse Friday, their followers screaming for "black power."

The Odessa American (Texas)

–June 18, 1966–

Marchers Shout for 'Power'

GREENWOOD, Miss. (UPI) — Dr. Martin Luther King Jr. and Negro leader Stokely Carmichael led a 600-man march on this delta town's courthouse yesterday with their followers screaming for "black power."

The Tampa Tribune (Florida)

Figure B.2: Newspaper Page Text Extraction Procedure Example



(a) Bounding Boxes

(b) Multi-column article to article and headline matching

Note: Panel a) is an example of newspaper page bounding box identification using a pre-trained model released by Dell et al. (2024). It identifies: articles, headlines, captions, advertisements, and other layout regions. Right: Novel algorithm to i) match articles, authors and headlines spanning multiple columns, and ii) calculate space allocated to the same. Text is then extracted using pre-trained OCR: Tesseract

B.2 Example of Article Corpus Output

The following JSON example illustrates the structure of the article corpus output. This structure captures key metadata, textual content, classification information, and additional space-related data for each article in the dataset:

Example: Article Corpus Output in JSON Format

```
{
    "n_id": "newspaper_id",
    "newspaper": "Newspaper title",
    "date": "YYYY-MM-DD",
    "page": "page number",
    "article_id": "Article_id",
    "headline": "Headline text",
    "author": "Author of article if present",
    "article": "Article text",
    "topic_number": Topic Number,
    "topic": "Topic Name",
    "newswire": Indicator,
    "AP": Indicator,
    "UPI": Indicator,
    "black_story": Indicator,
    "black_person": Indicator,
    "negro_person": Indicator,
    "black_count": "Number of references to ‘‘Black’’ entities",
    "negro_count": "Number of references to ‘‘Negro’’ entities",
    "black_index": "Article level relative reference to ‘‘Black’’ entities",
    "space": {
        "article": "Number of articles",
        "article_size": "Total size of articles",
        "headline": "Number of headlines",
        "headline_size": "Total size of headlines",
        "photograph": "Number of photographs",
        "photograph_size": "Total size of photographs",
        "image_caption": "Number of image captions",
        "image_caption_size": "Total size of image captions",
        "cartoon_or_advertisement": "Number of cartoons or advertisements",
        "cartoon_or_advertisement_size": "Total size of cartoons or advertisements"
    }
}
```

B.3 Linguistic Agency and Sentiment Classification Prompt

The following prompt was used to classify linguistic agency and sentiment in sentences referring to Black individuals, groups, or communities:

GPT Prompt for Linguistic Agency and Sentiment Classification

Classify the entity being assessed, linguistic agency, and sentiment of the following sentence referring to Black individuals, groups, or communities as either **"High Agency"**, **"Neutral Agency"**, or **"Low Agency"**. The terms **"Black"** and **"Negro"** should be treated identically.

Sentence: `"sentence"`

Definitions for Entity Type:

- **Person:** relates to a specific person, e.g., a black man..., a black..., the black minister Thomas Carlton ..., James Jones, a Black Panther Party member, etc.
- **Group:** relates to a group of Black people who are not an official organisation, e.g., the black community, five black men, blacks, negroes, a black group, etc.
- **Organisation:** relates to an organised group, e.g., The Black Panthers, The NAACP, Black United Students (BUS) organization, Black power groups, etc.
- **Object:** relates to an object or set of objects, e.g., the black residences, black-owned houses, etc.

Definitions for Agency:

- **High Agency:** The subject is portrayed as actively initiating impactful actions or demonstrating influence.
- **Neutral Agency:** The subject is mentioned without significant action or impact, neither actively performing actions nor being passive.
- **Low Agency:** The subject is portrayed as the recipient of actions or in a passive manner, having actions done to them.

Definitions for Sentiment:

- **Positive:** The entity is described in a favourable or supportive tone.
- **Neutral:** The entity is described without any clear positive or negative tone.
- **Negative:** The entity is described in a critical or unfavourable tone.

Respond in the following format:

```
- Entity: [Entity name]
- Entity_type: [Person, Group, Organisation, Object]
- Classification: [Agency Level] (High Agency, Neutral Agency, Low Agency)
- Sentiment: [Positive, Neutral, Negative]
- Confidence Level: [High, Medium, Low]
```

B.4 Examples of Output for Linguistic Agency and Sentiment Classification

The following examples illustrate the output of the classification system, showing how linguistic agency, sentiment, and confidence are determined for sentences referring to Black individuals, groups, or communities:

Example 1: Support of the Negro Community

Source: Birmingham Post-Herald, 1967-04-24
Sentence: "Some 3000 persons marched through Harlem Sunday in what the parade grand marshal described as a demonstration of the Negro community's support of their fellow Americans in Vietnam."
Classification:

- **Entity:** 3000 persons in the Negro community
- **Type:** Group
- **Agency:** High
- **Sentiment:** Positive
- **Confidence:** High

Example 2: Demonstration by Blacks

Source: Alabama Journal, 1974-03-06
Sentence: "About 50 blacks hissed and cursed each time Shockley attempted to speak."
Classification:

- **Entity:** 50 blacks
- **Type:** Group
- **Agency:** High
- **Sentiment:** Negative
- **Confidence:** High

Example 3: Striking Black Jurors

Source: Alabama Journal, 1973-03-05
Sentence: "They noted this morning that the state struck all of the black jurors leaving them with an all-white jury."
Classification:

- **Entity:** Black jurors
- **Type:** Group
- **Agency:** Low
- **Sentiment:** Negative
- **Confidence:** High

Example 4: Former Fry Cook Sentenced to Death

Source: El Paso Times, 1968-09-28
Sentence: "Marion, a 26-year-old Negro and former fry cook, was sentenced to death by a Lubbock jury in 1964 for the bludgeoning death of 66-year-old Midland socialite Mrs. J. Fred Turner in her 15-room Midland mansion."
Classification:

- **Entity:** Marion, a 26-year-old Negro
- **Type:** Person
- **Agency:** Low
- **Sentiment:** Negative
- **Confidence:** High

B.5 Examples of Articles mentioning both Black and Negro

Example 1: Mixed Label Usage in Racial Commentary

Source: Birmingham Post-Herald, 1967-05-30, p.8, Birmingham, Alabama
Excerpt: "Guidelines and integration is about all one can read about in the daily papers. Of course there is always the threat that welfare aid will be cut off unless people agree to destroy the white race by sending their children to school with the **Negroes**. I cannot understand why our Southern congressmen do not tell the big bosses in Washington to go jump in the lake and that we will not ruin our race by mixing with **Negroes**. There is something sinister and diabolical about all this mess about integration. Our government is not content for **blacks** and whites to have the same kind of schools and other institutions but they are trying to force the races to mix together, Why? ... Keep up this forced integration and you won't have to get a sun tan, you will have a mulatto race with no return to the white race."

Example 2: Mixed Label Usage in Military Reporting

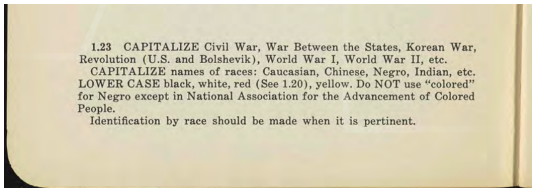
Source: Kingsport Times, 1969-12-01, p.5, Kingsport, Tennessee
Excerpt: "MILLINGTON, Tenn. (UPI) – Four young Marines accused of rioting in connection with a fight between **black** and white servicemen last July go on trial today in a military court-martial. A joint court-martial for the four **Negro** servicemen, all Vietnam veterans, was set Nov. 20 after the trial of one of the men, Pvt. Oscar W. Terry Jr., Paducah, Ky., ended in a mistrial. Others accused are Cpl. Perry Backstrom, 21, Lauderdale, Miss.; Pfc. Arthur McCall, 20, Birmingham, Ala., and Pfc. Memphis. The men are charged with assault and battery, conspiracy to riot and other lesser charges. In Terry's trial, a day of testimony was on the records when Lt. Cmdr. J. W. Erhard said he had misgivings about his impartiality and he had discussed his feelings with other members of the court. Naval authorities have accused the four of invading the casual company barracks and a snack bar at the base hospital, and beating several persons."

Example 3: Mixed Label Usage in Sports Reporting

Source: Longview News-Journal, 1969-11-07, p.11, Longview, Texas
Excerpt: "BLOOMINGTON, Ind. (UPI) – Indiana football coach John Pont said Thursday there was no change in the status of 10 **Negro** athletes dropped from the Rose Bowl contender for missing two straight practice sessions. Pont, who met with the 10 earlier in the day, said their grievances were not racial. The meeting between Pont and the athletes was requested through a **black** faculty member to discuss the players' status. The meeting resulted in a complete review of the boycott situation and the presentation of grievances by the athletes. Pont said the grievances were personal and not racial, but refused to release the list. He told the five underclassmen of the group that those who still wanted to play football at Indiana could report for spring practice or during the fall. The status of the scholarships that are to be renewed each year was still uncertain.
The underclass players are Greg Harvey, Larry Highbaugh, Gordon May, Bob Pernell, Don Silas and Greg Thaxton. The other four **blacks** who refused to return to the IU practices were seniors Benn Norman, Clarence Price, Mike Adams and Charlie Murphy. Four other **Negro** players who missed practice Tuesday returned to the Hoosier squad without penalty. Indiana is in the midst of the battle to determine the Big Ten representative to the Rose Bowl on New Year's Day. Those players who returned were starting tight end John Andrews, Steve Porter, Tim Roberson and Bob Jones."

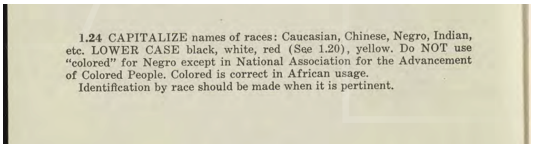
Figure B.3: AP style guide for Black and Negro

(a) 1960



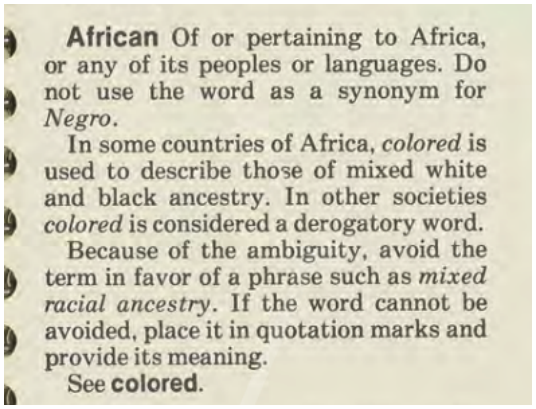
1.23 CAPITALIZE Civil War, War Between the States, Korean War, Revolution (U.S. and Bolshevik), World War I, World War II, etc.
CAPITALIZE names of races: Caucasian, Chinese, Negro, Indian, etc. LOWER CASE black, white, red (See 1.20), yellow. Do NOT use "colored" for Negro except in National Association for the Advancement of Colored People.
Identification by race should be made when it is pertinent.

(b) 1970



1.24 CAPITALIZE names of races: Caucasian, Chinese, Negro, Indian, etc. LOWER CASE black, white, red (See 1.20), yellow. Do NOT use "colored" for Negro except in National Association for the Advancement of Colored People. Colored is correct in African usage.
Identification by race should be made when it is pertinent.

(c) 1977 (page 7)



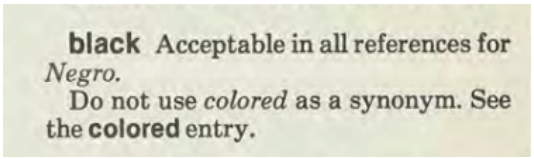
African Of or pertaining to Africa, or any of its peoples or languages. Do not use the word as a synonym for *Negro*.

In some countries of Africa, *colored* is used to describe those of mixed white and black ancestry. In other societies *colored* is considered a derogatory word.

Because of the ambiguity, avoid the term in favor of a phrase such as *mixed racial ancestry*. If the word cannot be avoided, place it in quotation marks and provide its meaning.

See **colored**.

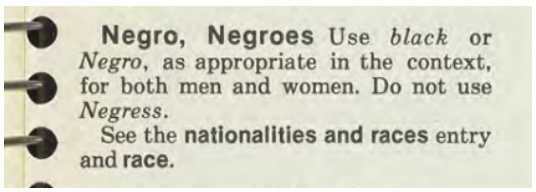
(d) 1977 (page 28)



black Acceptable in all references for *Negro*.

Do not use *colored* as a synonym. See the **colored** entry.

(e) 1977 (page 149)



Negro, Negroes Use *black* or *Negro*, as appropriate in the context, for both men and women. Do not use *Negress*.

See the **nationalities and races** entry and **race**.

Chapter III

From the Pulpit to the Polls: The Electoral Consequences of Christian Talk Radio

Abstract

This paper studies the political effects of Christian-conservative radio during a period of rapid media deregulation in the United States. I focus on the national expansion of Salem Communications, a major religious broadcaster that began systematically integrating conservative political talk with Christian programming in the early 1990s. Exploiting the staggered rollout of Salem-owned stations across U.S. media markets, I examine how exposure to partisan religious broadcasting affected voter turnout and Republican vote share in presidential elections. I find that increased exposure to Salem's broadcasts led to a meaningful increase in Republican vote share, with effects concentrated in counties with larger evangelical populations. These results suggest that religious media can serve as an effective vehicle for political mobilisation when ideological content is embedded within faith-based programming. More broadly, the findings highlight how changes in media regulation facilitated the emergence of partisan religious broadcasting and shaped political behaviour during a critical period of partisan realignment.

1 Introduction

The entanglement of religion and politics in the United States has grown increasingly pronounced over the past five decades. While religious leaders have long intervened in public debates, the late twentieth century witnessed a marked transformation: a growing segment of evangelical Christians began to align with the Republican Party, culminating in the emergence of a self-conscious and institutionally embedded "religious right". This shift, initially catalysed by figures such as Jerry Falwell and the Moral Majority in the 1980s (Buccione and Knight, 2024), has since evolved into a durable feature of American electoral politics. Its influence was vividly on display during the January 6 Capitol riot, where Christian nationalist symbols and rhetoric were prominently featured among the rioters[1]. Recent scholarship has explored how elite religious endorsements, voter identity, and single-issue activism contributed to this realignment. Yet, the role of religiously infused media institutions in sustaining and expanding the political power of conservative Christians remains incompletely understood.

In this paper, I examine the political impact of exposure to Christian-conservative radio programming by addressing the following question: how and to what extent can partisan religious media shape electoral behaviour? I provide evidence from a quasi-natural experiment generated by the staggered geographic expansion of Salem Media Group's radio stations during the early 1990s, when the company began systematically integrating conservative political talk with Christian teaching. Using fine-grained variation in predicted radio signal strength across counties and over time, I estimate the effect of exposure to Salem's broadcasts on county-level voting outcomes in U.S. presidential elections. To shed light on the underlying mechanism, I examine whether the electoral effects of exposure vary systematically with local religious composition, consistent with persuasion operating through evangelical audiences.

Salem Media Group is a major U.S. Christian-conservative radio network and one of the largest owners of talk-format stations in the country. Founded in 1974, Salem grew rapidly in the decades following Reagan's presidency by constructing a vertically integrated national media apparatus that fused Christian teaching with partisan political commentary, rather than relying on televangelism or denominational outreach. Between 1994 and 2005, the company expanded from 18 to over 100 AM and FM stations, capitalising on deregulatory changes in broadcasting, most notably the repeal of the Fairness Doctrine and the Telecommunications Act of 1996. This radio empire expansion provides an ideal setting for understanding the potential for religious broadcasting persuasion of religious voters for three key reasons. Firstly, Salem specifically aimed to persuade evangelical listeners to engage in political life by voting for conservative politicians. Company documents and interviews with senior management highlight

[1] Elana Schor, "Christianity on Display at Capitol Riot Sparks New Debate," *Associated Press*, January 28, 2021. Available at: `https://apnews.com/article/christianity-capitol-riot-6f13ef0030ad7b5a6f37a1e3b7b4c898`.

that the company strategic goals are driven by ideological impact rather than commercial appeal. Branded channels systematically co-located religious programming with conservative talk radio, targeting audiences primed by Christian content and delivering coordinated political messaging to a cohesive evangelical listenership.

Second, the staggered expansion provides a source of quasi-natural variation in exposure to its message across time and space. Beginning after 1992, the company launched an expansion strategy, which targeted multiple US radio markets over the following 25 years.

Third, Salem's programming model was institutionally stable over the period studied. Following its mid-1990s strategic shift, the company explicitly committed to integrating Christian teaching with conservative political talk across its stations, and internal reports emphasise consistency in format rather than experimentation or frequent rebranding. Because station acquisitions were typically followed by rapid reprogramming into standardised Salem formats, variation in exposure primarily reflects differences in signal reach rather than changes in message content. As a result, the estimated effects can be interpreted as arising from exposure to a coherent and persistent ideological broadcast strategy, rather than from transitory programming shocks.

My empirical strategy exploits the fact that Salem Media Group expanded its radio network into different U.S. media markets at different times to assess the political effects of exposure to Christian-conservative broadcasting. Specifically, I employ a difference-in-differences approach that compares county-level voting outcomes before and after local exposure to Salem's signal. The central identifying assumption is that the timing of Salem's market entry is unrelated to other factors driving changes in partisan voting behaviour. Internal company documents and contemporaneous reports indicate that Salem's expansion strategy prioritised broad audience reach and rapid network growth rather than fine-grained political targeting, and that new acquisitions were frequently clustered around existing infrastructure, reflecting path dependence and economies of scale rather than local political conditions. Consistent with this account, I show that pre-expansion political, demographic, and religious characteristics are balanced with respect to subsequent exposure, and that pre-trends in voting outcomes are flat in event-study analyses. As a final safeguard against residual selection, I condition on potential exposure based on transmitter placement and use variation in effective signal reception generated by local topography to isolate differences in realised exposure that are unlikely to reflect strategic targeting.

The analysis combines several novel data sources. I digitise monthly station ownership records from *Broadcasting and Cable Magazine* and merge them with engineering records from the Federal Communications Commission that report station location, effective radiated power, and antenna height. I use these parameters to estimate county-level signal coverage using the Irregular Terrain Model, a professional radio propagation algorithm validated in prior economic research (Wang, 2021). These broadcast data are

linked to county-level presidential election returns from 1980 to 2020 and to measures of religious adherence from the Longitudinal Religious Congregations and Membership File. The resulting panel enables a granular analysis of both short-run and long-run effects across counties with differing demographic and religious profiles.

I find that increased exposure to Salem's broadcasts led to a statistically significant and persistent rise in Republican presidential vote share. In the initial expansion period from 1992 to 1996, a one standard deviation increase in terrain-adjusted signal strength raised Republican vote share by approximately 0.8 percentage points, with effects concentrated in counties with high evangelical Protestant populations. Event-study estimates reveal that these shifts emerged immediately following first exposure and persisted over subsequent presidential elections, with no evidence of pre-trends or anticipatory effects. Extending the panel through 2020, I show that early-treated counties exhibit long-run partisan realignment, with elevated Republican support persisting through at least five election cycles. While aggregate effects attenuate modestly over time, they remain especially pronounced in ideologically receptive counties. These findings suggest that exposure to Salem's religious-conservative programming not only shaped short-run vote choices but produced durable changes in political preferences. The estimated persuasion rate, approximately 3 percent of listeners, places the effect within the lower range of influential media interventions,[2] yet its persistence and demographic specificity suggests lasting political consequences of ideologically targeted religious broadcasting.

This paper makes several contributions to the literatures on media influence, religious mobilisation, and electoral behaviour. First, I contribute to a growing body of work examining how religious media institutions influence political preferences and behaviour. While previous studies have shown that individual religious leaders or organisations can shift voter allegiances, such as the Moral Majority's endorsement of Ronald Reagan over Jimmy Carter in 1980 (Buccione and Knight, 2024), this paper shifts focus to the institutional media infrastructure that helped sustain and scale such mobilisation in the decades that followed. In contrast to earlier work emphasising the role of elite endorsements, televangelism, or candidate religiosity, I study a deregulation-enabled, profit-driven radio network that fused conservative theology with partisan messaging over a much longer period.

Second, I add to the literature on media effects in political settings by introducing new evidence on religiously affiliated broadcast infrastructure. Whereas much of the existing work focuses on state-sponsored or secular commercial media, such as Nazi radio in Germany (Adena et al., 2015), Mediaset in Italy (Durante et al., 2019), or Fox News in the U.S. (DellaVigna and Kaplan, 2007), I show that religious commercial broadcasters can operate with parallel persuasive power. Relatedly, I extend insights

[2]For comparison, DellaVigna and Kaplan (2007) estimate that exposure to Fox News persuaded 11.6% of viewers in newly treated markets. Other studies report persuasion rates ranging from 6% to 20% (e.g., Enikolopov et al., 2011; DellaVigna et al., 2014; Adena et al., 2015).

from Wang (2021), who studies the influence of a charismatic Catholic radio demagogue in the 1930s, by focusing on the institutionalisation of that model: a corporate media network with national reach, formal syndication agreements, and programmatic alignment with electoral goals.

Third, this paper contributes to the study of how deregulation and market structure shape political outcomes. Salem's expansion was facilitated by the repeal of the Fairness Doctrine in 1987 and the Telecommunications Act of 1996, both of which loosened content neutrality requirements and ownership caps. I provide new evidence on how these regulatory changes allowed ideologically motivated actors to reshape local media environments and voter behaviour over time. In doing so, the paper complements research on the long-run consequences of media deregulation (Prat and Strömberg, 2013), while emphasising the distinct role of religious content and religious audiences in these dynamics. While not the central focus of the paper, the findings suggest that structural shifts in media policy can have broader implications for the political information environment.

Finally, this paper contributes to the broader literature on single-issue voters and political mobilisation. Like previous work showing how intense preferences on issues like abortion or gun rights can shape turnout and candidate support (Bouton et al., 2021), I show that religiously grounded media messaging can catalyse political engagement in ways responsive to theological identity. This dynamic is especially salient in areas with high evangelical adherence, where listeners were primed to respond to narratives that link politics with spiritual stakes.

Taken together, these contributions advance our understanding of how institutional media networks can fuse religious and political messaging, mobilise partisan constituencies, and reshape the U.S. electoral landscape in the post-Fairness Doctrine era. The remainder of the paper proceeds as follows. Section 2 provides historical context on evangelical mobilisation and the rise of Salem Communications. Section 3 describes the construction of the dataset. Section 4 outlines the empirical strategy, including the radio signal model and identification approach. Section 5 presents the main findings and robustness checks and Section 6 concludes.

2 Background

2.1 Salem: Christian Radio and Conservative Politics

Salem Media Group is a publicly traded media conglomerate specialising in Christian and conservative content. Founded in 1974 by Edward Atsinger III and Stuart Epperson Sr., the company grew from a small religious broadcaster into one of the largest owners of AM radio stations in the United States. As of 2023, Salem owned 103 stations in 36 markets, including 63 in 23 of the top 25 U.S. media markets, comprising 33 FM

and 70 AM stations.[3] These stations operate under three principal formats: Christian Teach and Talk (CTT), Conservative News Talk, and Christian Contemporary Music.

From its inception, Salem's founders have operated at the nexus of ministry and politics. Both men served on the Council for National Policy (CNP), a private network of Christian conservatives and Republican donors, where they championed a vision of the United States as a "Christian nation" and opposed secular influences they deemed morally corrosive.[4] Epperson, named one of Time magazine's 25 most influential evangelicals in 2005, played a central role in articulating this worldview.[5] In the mid-1980s, he twice ran unsuccessfully for Congress as the Republican nominee in North Carolina's fifth district. Meanwhile, Atsinger co-founded the Allied Business PAC in the early 1990s, which spent over $9 million supporting conservative candidates and was instrumental in the Republican capture of the California State Assembly in 1994.[6]

It was during this period that Salem began pioneering a strategy of co-locating Christian teaching with conservative political talk, using religious programming as a gateway to promote partisan content and build a cohesive, ideologically aligned audience. In 1995, the company formally announced its shift from an exclusively religious format to one that explicitly incorporated conservative political radio.

This transformation was not only ideological but also operational. According to its 2004 Annual Report, Salem reprogrammed nearly all acquired stations with explicitly Christian and conservative content, and required several years of local promotion to build new listener bases. Though Salem publicly described its audience as "Christian and family-themed," its programming consistently aligned with the theological and political worldview of white evangelical Protestants, a group central to the company's revenue model and ideological strategy.[7] To deepen market penetration, Salem pursued a cluster strategy in major cities, operating multiple stations per market, each aimed at a different segment of the evangelical audience, whether through sermons, news talk, or contemporary Christian music. This clustering allowed Salem to bundle audiences for advertisers, cross-promote ideologically aligned content, and embed itself in local evangelical communities. It also partnered with churches and religious organisations, hosted pastor appreciation events and Christian music festivals, and maintained block programming renewal rates exceeding 90 percent annually.

This capacity to mobilise evangelical audiences was evident even prior to Salem's national expansion. In 1988, KKLA-FM, a Salem-owned Christian talk station in Los Angeles, played a role in coordinating evangelical protest against the film *The Last Temptation of Christ*. Host John Stewart used his broadcast to organise rallies that

[3]See Ballotpedia, "Salem Media Group," https://ballotpedia.org/Salem_Media_Group.

[4]SPLC, "Council for National Policy," 2016 - https://www.splcenter.org/resources/hatewatch/council-national-policy-behind-curtain/

[5]See https://content.time.com/time/specials/packages/completelist/0,29569,1993235,00.html

[6]Los Angeles Times, 1997.https://www.latimes.com/archives/la-xpm-1993-07-08-mn-11197-story.html

[7]Salem Communications, 2004 Annual Report, SEC Form 10-K.

ultimately brought more than 25,000 protesters to the gates of Universal Studios, a moment he later described as the first time many evangelicals realised they could act collectively in the public square.[8] This episode shows how Salem stations functioned not only as ideological platforms, but as infrastructure for religious mobilisation capable of activating coordinated protest around a Christian nationalist worldview.

By the 2010s, Salem's political interventions extended beyond broadcasting. It co-sponsored and helped moderate Republican primary debates in 2015, 2016, and 2023, and internal reports reveal that company executives pressured radio hosts to portray Donald Trump more favourably during his presidential campaigns.[9] In one high-profile incident, longtime conservative radio host Craig Silverman was fired mid-broadcast after criticising Trump, an act he described as politically motivated censorship.[10] These episodes reflect Salem's commitment to shaping partisan loyalties, not just through ideology but through direct editorial control.

Salem's national expansion was catalysed by two regulatory turning points. First, the repeal of the Fairness Doctrine in 1987 lifted the requirement for broadcasters to present opposing views on controversial issues. Second, the Telecommunications Act of 1996 eliminated limits on station ownership, paving the way for rapid consolidation. Between 1994 and 2005, Salem expanded from 18 stations to more than 100.[11] Unlike commercial broadcasters reliant on advertisers, Salem's revenue model has long been sustained by paid religious programming, which insulates its operations from commercial backlash and enables long-term ideological consistency.[12]

Regarding content, Salem promotes a distinctly Christian nationalist worldview. Internal strategy documents highlight the coherence of this programming vision. According to its 2003 Annual Report, the company viewed Christian teaching and conservative news talk as complementary vehicles for advancing "conservative views and family values".[13] The report emphasised that Christian teaching served as both "a learning resource and personal support," helping listeners navigate issues ranging from parenting to their "religious legal rights in education and the workplace". In this framing, Salem's stations function not merely as broadcasters but as platforms for religious and political instruction. News and talk formats were selected for their ideological alignment with Christian teaching, reinforcing a shared worldview across different program genres. As the company put it, this strategy positioned Salem to "improve upon its leadership position in religious and family-themes radio".

[8]Orange County Register, "Recalling the Time Local Christians Took on Hollywood," August 15, 2008.

[9]New York Times, 2023https://www.nytimes.com/2023/11/08/us/politics/moderators-nbc-republican-debate.html; CNN, 2018 https://money.cnn.com/2018/05/09/media/salem-radio-executives-trump/index.html

[10]NPR, 2019 https://www.npr.org/2019/11/23/782255191/radio-host-says-he-was-fired-for-criticizing-trump

[11]See FCC ownership filings; Divided Dial, Episode 2.

[12]See Salem Communications SEC FORM 10-K filings.

[13]Salem Communications, 2003 Annual Report, SEC Form 10-K.

Salem's existing syndicated hosts, including Eric Metaxas, Charlie Kirk, and Sebastian Gorka, have used their national reach to promote narratives of cultural decline, elite betrayal, and Christian persecution, frequently advancing conspiracy theories about election fraud and government corruption. As of 2025, these hosts had a combined following exceeding 10 million across the popular political social media platforms X and Truth Social. In 2020, Salem served as executive producer of *2000 Mules*, a widely discredited film alleging widespread voter fraud in key swing states.[14] Lawsuits stemming from the film prompted Salem to cease its distribution in 2024.[15] Yet as late as 2022, its nationally syndicated hosts continued to promote claims of a "stolen election," placing Salem among the few major networks that refused to moderate their messaging in the aftermath of the January 6 Capitol Riots.

According to senior executives, this model is intentional. Phil Boyce, Salem's senior vice president of spoken word, explained that the company aims not for commercial appeal but for ideological impact: "We're in it to save America". Hosts are not journalists, he noted, but "storytellers, opinion makers, thought leaders".[16] The result is a broadcasting ecosystem that operates largely out of sight of national media scrutiny, but which plays a critical role in shaping political discourse among conservative religious communities.

This transformation is central to the identification strategy in this paper. In 1995, Salem formally announced that it was transitioning from exclusively religious broadcasting to a format that explicitly combined Christian teaching with conservative political talk. This marked a strategic inflection point: the company began integrating ideological content across its stations, systematically co-locating religious and right-wing political programming to cultivate a cohesive, mobilised listenership. The years immediately surrounding this shift, particularly the 1992 to 1996 period, therefore represent a natural window in which to study the political consequences of exposure to partisan religious broadcasting. My empirical strategy leverages this timing to isolate the effect of ideological content delivery, not merely radio expansion, by focusing on Salem's initial market entries during the moment it redefined its programming model.

Religious Realignment and the Clinton Presidency The early 1990s were a period of political uncertainty among religious voters, particularly white evangelical Protestants. In the 1992 presidential election, Bill Clinton attracted notable support

[14]See Reuters (2022)https://www.reuters.com/article/factcheck-usa-mules/fact-check-does-2000-mules-provide-evidence-of-voter-fraud-in-the-2020-u-s-presidential-election-idUSL2N2XJ0OQ/, CNN "'2000 Mules' creator admits some of film's claims are flawed", Dec 2024 https://edition.cnn.com/2024/12/02/politics/2000-mules-creator-admits-some-of-films-claims-are-flawed, and Washington Post, "2000 Mules, a key piece of election misinformation, has its day in court", Oct 2023 for investigations into the film's claims and Salem's distribution.

[15]CNN, "Right-wing media company Salem apologizes, stops distributing 2020 election conspiracy film '2000 Mules' after lawsuit," May 2024 https://edition.cnn.com/2024/05/31/media/salem-will-stop-distributing-2000-mules

[16]See "The Divided Dial," Episodes 2 and 5 available at https://www.wnycstudios.org/podcasts/otm/divided-dial.

from this group, receiving approximately 35 percent of the white evangelical vote, an increase from the 24 percent received by Democratic candidate Michael Dukakis in 1988.[17] Clinton's Southern background and emphasis on family values were often cited as factors contributing to his initial appeal. However, by the 1996 election, this support had declined, with Clinton receiving closer to 26 percent of the white evangelical vote.[18]

The decline in support coincided with broader debates within religious communities about the role of moral character in public leadership, as well as growing concern over social issues such as abortion, LGBTQ rights, and church-state separation. These discussions took place alongside a broader shift in partisan alignment among evangelical voters, which had been underway since the 1980s and continued through the 1990s.

Salem's decision to integrate conservative political talk with Christian teaching occurred during this period of reorientation within the religious electorate. Its programming shift followed closely after the 1994 midterm elections, in which Republicans gained control of Congress, and emerged in a media environment increasingly oriented toward ideological targeting. In this context, Salem's programming strategy was less a departure than an adaptation, channeling existing religious anxieties and partisan realignments into a commercially viable broadcast formula.

2.2 Deregulation

Salem's transformation from a niche religious broadcaster into a national network of politically conservative radio stations was made possible by two major deregulatory shifts in U.S. media policy: the repeal of the Fairness Doctrine in 1987 and the passage of the Telecommunications Act of 1996.

The Fairness Doctrine, implemented by the Federal Communications Commission (FCC) in 1949, required licensed broadcasters to present contrasting viewpoints on controversial issues of public importance. Its aim was to ensure ideological diversity on the public airwaves, given their limited bandwidth and public licensing. While initially seen as a safeguard against propaganda and demagoguery, critics, including many conservatives, argued that it suppressed political expression by incentivising self-censorship among broadcasters. In 1987, under the Reagan administration, the FCC formally abolished the doctrine, concluding that it was no longer necessary in an increasingly diverse media environment.[19] The repeal enabled stations to adopt ideologically consistent programming without legal obligation to provide balance, catalysing the rise of single-perspective formats such as conservative talk radio.

The second, and arguably more consequential, shift came with the Telecommunications Act of 1996. Signed into law by President Bill Clinton, the Act marked the most

[17] ABC News exit polls, 1992; Green et al. (1996).

[18] Pew Research Center, "Religion and the 1996 Vote," 1996.

[19] See FCC Report, "Fairness Doctrine Complaint of Syracuse Peace Council against Telelevision Station WTVH. Syracuse. New York," 1987 https://quello.msu.edu/wp-content/uploads/pdfs/OCR_James_H_Quello_498.pdf.

comprehensive overhaul of American media regulation in over six decades. Among its many provisions, it eliminated national caps on the number of radio stations a single firm could own and significantly relaxed local ownership limits.[20] Prior to 1996, a firm could own no more than 40 radio stations nationwide. After the Act, no national limits applied, and local ownership caps were doubled or eliminated, depending on market size.

The late 1990s and early 2000s witnessed a wave of consolidation, with large conglomerates such as Clear Channel (now iHeartMedia) acquiring hundreds of stations nationwide. Salem Communications followed suit, expanding from just 18 stations in 1994 to over 100 by 2005.[21] Crucially, Salem was one of the few broadcasters that simultaneously scaled religious and political programming, often colocating "Christian Teach and Talk" (CTT) stations with their conservative talk brands, such as "The Patriot" and "The Answer". These stations shared audiences, infrastructure, and cross-promotional airtime.

3 Data

This section describes the construction of three core datasets: the expansion of Salem radio stations, the geographic reach of their broadcasts, and county-level voting outcomes. Together, these data allow me to estimate the causal effect of conservative Christian radio exposure on electoral behaviour.

3.1 Radio Expansion

To document Salem Media Group's expansion, I compile a novel dataset on the full history of Salem-owned AM and FM radio stations between 1980 and 2020. The primary source is *Broadcasting & Cable Magazine*, a monthly trade publication that regularly reports changes in station ownership, format, and market entry. I digitise and manually code Salem's acquisition and launch dates, as well as the city of license, station frequency, and call sign for each listed property[22] (see Figure A.1 for an example). For stations with broadcast call signs which change over time, I track these changes using the address of the broadcast tower and the ownership details to ensure consistency between signals with changing broadcast signs[23]. I cross-reference these station broadcast call signs and locations with Salem SEC filings and Annual reports to confirm

[20] Telecommunications Act of 1996, Pub. L. No. 104-104, 110 Stat. 56.

[21] FCC Ownership Reports; Divided Dial, Episode 2.

[22] I obtain archival pdf versions of each issue from https://www.worldradiohistory.com/Archive-All-BC/Broadcasting-Magazine.htm

[23] For example, KDOW (1220 AM) in Palo Alto, California, has undergone several call sign changes, originally KIBE (1949–1984), then KDFC (1984–1997), KBPA (1997–1999), KBZS (1999–2001), KSFB (2001–2004), and KNTS (2004–2008), before adopting its current identifier. Despite these changes, the signal and ownership lineage remain traceable through FCC tower registration and licensing data.

purchase and transmission launch dates. These data are further cross-referenced with license and engineering records from the Federal Communications Commission (FCC), which include precise tower coordinates, transmitter power (in kilowatts), and antenna height (in meters).

For each station, I confirm transmission parameters using the FCC's Consolidated Database System (CDBS) and Licensing and Management System (LMS), ensuring consistent technical metadata across the full sample.[24] These data form the basis for the subsequent estimation of broadcast signal coverage and intensity. Between 1992 and 2005, Salem expanded from 18 to over 100 stations, with a notable clustering of new entries in major urban centres and counties with large evangelical populations. I focus on stations operating on the contiguous US, i.e. excluding Hawaii (see Table 3.1 below.)

Table 3.1: Number of Salem-Owned Stations (Contiguous US)

Year	1980	1984	1988	1992	1996	2000	2004	2008	2012	2016
Stations	5	8	15	18	34	71	93	92	92	93

Political Switch

3.2 Radio Coverage

A central challenge in estimating the political effects of Salem's broadcasting expansion is that direct measures of radio exposure are unavailable at a fine geographic scale. To overcome this, I construct a measure of signal strength by modelling the effective reach of each Salem transmitter across U.S. counties. Radio signal propagation follows well-established physical laws: in free space, signal strength diminishes with distance, while terrain features such as hills and elevation changes can significantly attenuate transmission. I capture this variation using the Irregular Terrain Model (ITM), a professional engineering algorithm developed by the U.S. Department of Commerce. The ITM accounts for both transmitter characteristics and local topography, and has been validated in recent empirical work on media effects (Olken, 2009; Yanagizawa-Drott, 2014; Durante et al., 2019; Wang, 2021).[25]

For each Salem transmitter, I collect technical specifications including frequency (MHz), effective radiated power (ERP), and antenna height above average terrain (HAAT) from Federal Communications Commission (FCC) license records. I combine these with 90-meter resolution elevation data to calculate the ITM-predicted signal loss

[24] FCC CDBS and LMS Public Access Databases. See: `https://www.fcc.gov/media/radio/cdbs-database-public-files`

[25] I am grateful to Benjamin Olken for generously sharing the ITM software, which underpins the signal propagation model used in this study.

between each transmitter and the centroid of every U.S. county. The resulting path loss estimates allow me to compute predicted signal strength for each transmitter–county pair. I assign each county the maximum predicted signal strength across all transmitters active in a given year. This continuous measure of signal exposure forms the basis of my empirical analysis (see Appendix C for complete calculations). Following the approach introduced by Olken (2009), I also construct the free-space path loss. Holding all technical parameters constant, I recalculate path loss for each county as if the earth were smooth and unobstructed by terrain. This hypothetical measure captures variation due to transmitter proximity and power but excludes topographic attenuation.

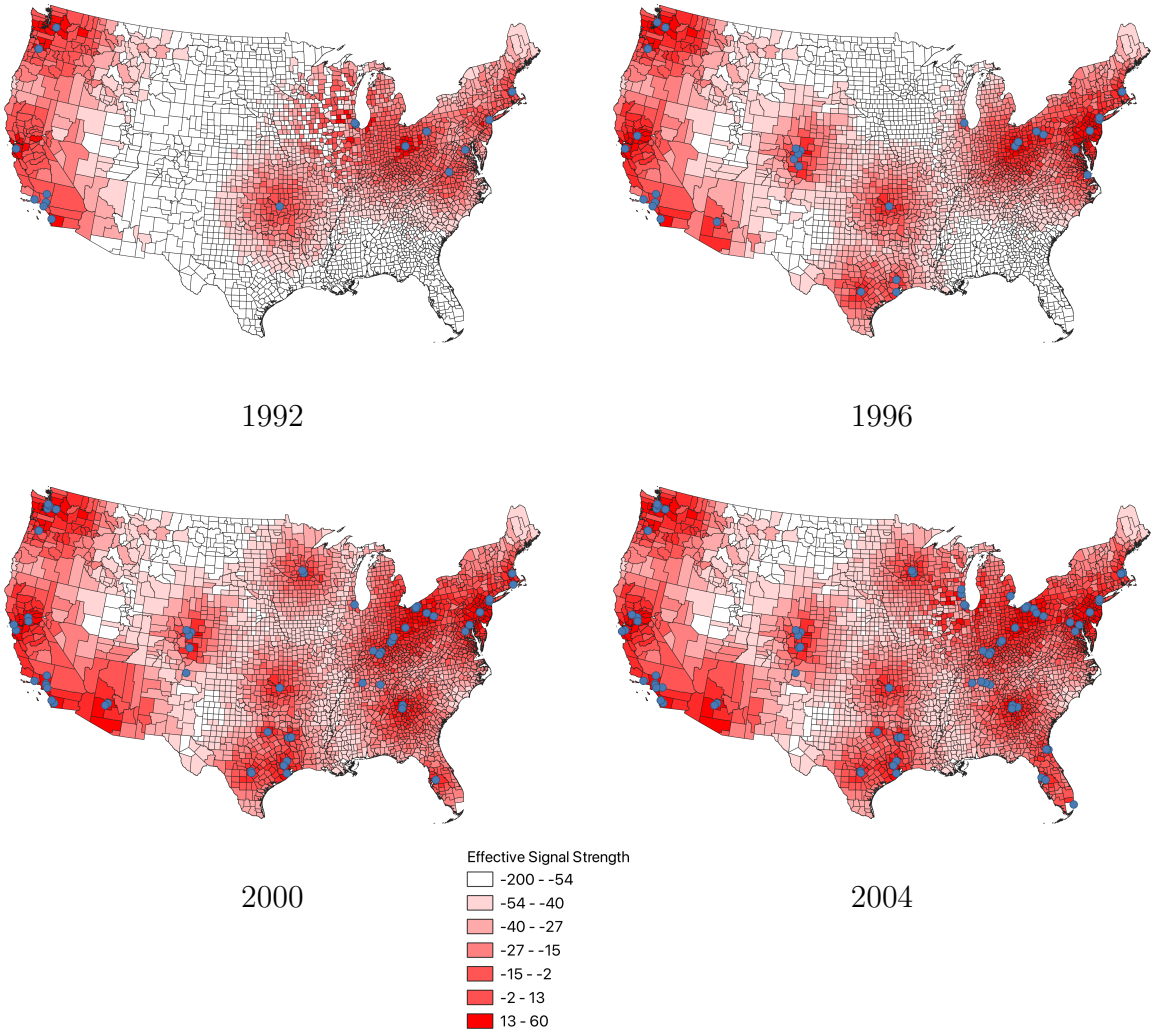


Figure 3.1: Geographic expansion of Salem Media Group and modelled signal strength across U.S. counties (1992–2004). Signal strength is calculated using the Irregular Terrain Model (ITM), with coverage thresholds based on prior engineering standards.

Figure 3.1 visualises the geographic expansion of Salem's broadcast footprint between 1992 and 2004, based on ITM-modeled signal coverage. Counties with signal strength above the reception threshold (50 dBµV/m) are increasingly clustered in urban areas as Salem expands its network. While Salem continued to add stations after

2004, most of the major expansion in both station count and geographic signal coverage occurred during this initial period. Appendix Figure A.2 shows the broadcast footprint between 2004 and 2016, illustrating the relative stability of Salem's coverage after the early 2000s. This motivates my focus on the 1992–2004 period as the primary treatment window.

3.3 Voting, Religious and Demographic Covariates

County-level presidential election returns are obtained from Algara and Amlani (2021), who provide harmonised vote shares for Democratic, Republican, and third-party candidates across all counties from 1872 onward. The data are cleaned and merged by county FIPS codes county boundaries are harmonised to ensure consistency over time.

To measure county-level religious composition, I use the Longitudinal Religious Congregations and Membership File (1980–2010), which provides counts of religious adherents and congregations by denomination and tradition (Grammich et al., 2018). These data are drawn from the Church Membership Surveys and aggregated by the Association of Statisticians of American Religious Bodies (ASARB). I focus on evangelical Protestant adherence as a proxy for potential sensitivity to Salem's Christian-conservative content.

Historical demographic characteristics, such as population size, racial composition, urbanisation, and income, are drawn from Haines et al. (2010). For time-invariant county-level covariates not covered in this series, I supplement with data from ICPSR 20660 (2008). Measures of county ruggedness and land area are taken from the replication package for Wang (2021).

These covariates allow me to flexibly control for both long-run geographic and demographic characteristics, as well as potential moderators of media influence. The final dataset is an unbalanced panel of U.S. counties between 1980 and 2020, with variation in Salem signal coverage, political outcomes, and county-level characteristics across time. For the main analysis, I focus on the largest and most rapid expansion between 1992 and 2004.

4 Empirical Strategy

This section outlines the empirical strategy used to estimate the causal effect of exposure to Salem's Christian-conservative broadcasts on county-level voting outcomes. The baseline analysis focuses on the period between the 1992 and 1996 U.S. presidential elections, which coincides with Salem's initial large-scale expansion into new radio markets and its strategic shift toward conservative political talk radio. During this period, Salem began co-locating religious programming with explicitly partisan political content, integrating Christian teaching with right-leaning commentary.

Figure 3.1 shows that this expansion aligns with contemporaneous accounts from senior management indicating that early acquisitions prioritised rapid network growth and presence in large media markets rather than fine-grained audience targeting.[26] Station acquisitions were made at the Designated Market Area (DMA) level rather than the county level, and there is no indication that Salem systematically targeted markets based on local political alignment or ideological persuadability. While larger markets may differ systematically from smaller ones, available evidence suggests that Salem's early expansion was driven primarily by economies of scale in management, content production, and advertising sales, as well as by the goal of establishing a visible national footprint.[27]

Further, Salem's expansion strategy exhibits substantial path dependence. The firm frequently co-located new stations in markets where it already operated, clustering acquisitions around legacy infrastructure established prior to its shift toward explicitly partisan programming. To the extent that new entries were shaped by inherited infrastructure and organisational scale rather than contemporaneous political conditions, selection based on local political trends is unlikely to drive observed exposure patterns. Subsequent sections formally assess this claim using balance tests and event-study analyses of pre-expansion political trends.

4.1 Econometric model

To estimate the effect of exposure to Salem's Christian-conservative broadcasts on electoral outcomes, I employ a difference-in-differences strategy that exploits the rollout of Salem-owned radio stations across U.S. media markets during the early 1990s. Exposure is a continuous measure of the intensity of reception, allowing the design to compare changes in voting outcomes across counties experiencing differential increases in effective exposure following market entry. Specifically, I estimate the following equation on a county-level panel of presidential election outcomes, using variation in exposure before and after local broadcast availability.

I estimate the following specification:

$$\text{VoteShare}_{ct} = \beta \, \text{Signal}_{ct} + \gamma \; + \mathbf{X}_c \cdot \delta_t + \eta_c + \lambda_t + \epsilon_{ct}, \tag{1}$$

The outcome of interest is Republican presidential vote share in county c and year t. The primary regressor is standardised effective signal strength from Salem transmitters, denoted Signal_{ct}. η_c are county fixed effects, λ_t are year fixed effects, and $\mathbf{X}_c \cdot \delta_t$ denotes a vector of pre-period cross-sectional county characteristics interacted with year dummies to allow for flexible differential trends. These controls include: demographic composition, educational attainment, income, unemployment, poverty, population density, age structure, industrial composition, religious adherence, prior voting outcomes,

[26] Salem Annual Report, 1999.
[27] Salem Annual Report, 2004.

and geographic characteristics such as elevation, area, and terrain ruggedness. In additional specifications, I further include state×year fixed effects which effectively restricts comparisons to within the same state. Standard errors are clustered at the state–county level to account for spatial correlation and grouped treatment timing.

In most specifications, I also include the free-space signal strength as a control for potential exposure based on transmitter location and technical characteristics. Conditioning on this measure holds fixed the set of counties that lie within the mechanical reach of Salem's broadcast infrastructure in a given period. Remaining variation in effective signal strength reflects differences in realised reception across counties within the same potential coverage area, arising from local topography.

The coefficient β captures the reduced-form effect of exposure to Salem's broadcasts on Republican vote share. Identification comes from differential changes in effective signal strength across counties during the early phase of Salem's expansion, when a subset of counties that were not covered in 1992 became reachable by 1996. The identifying assumption is that, conditional on county and year fixed effects, within-county changes in effective signal strength following Salem's expansion are uncorrelated with time-varying unobserved determinants of voting behaviour.

To empirically evaluate the validity of the identifying assumption, I first conduct a series of placebo tests using county-level Republican vote shares in presidential elections prior to Salem's expansion. Specifically, I estimate equation 1 on the intervals 1980–1984, 1984–1988, and 1988–1992, restricting attention to counties that would later fall within Salem's potential broadcast range. Table 4.1 reports the results. In all three placebo windows, change in the signal strength between 1992 and 1996 is not significantly correlated with prior changes in Republican support. Point estimates are small and imprecise, and R^2 values remain low despite the inclusion of baseline controls and state-by-year fixed effects. This pattern reinforces the interpretation that pre-existing political trends do not predict subsequent changes in effective signal strength following Salem's expansion, supporting the conditional exogeneity of signal exposure in the post-expansion period.

To further explore the validity of the identifying assumption, I assess the balance of observed covariates across counties with differing levels of predicted signal strength. Table 4.2 reports the coefficients from univariate regressions of signal strength on pre-treatment county characteristics. Consistent with the possibility of strategic market entry, signal strength is significantly correlated with several demographic and economic variables in the bivariate case. However, once I include controls for counterfactual free-space signal strength, geographic characteristics (elevation, terrain ruggedness, and county land area), and state fixed effects, these correlations largely disappear. In the fully controlled specification, signal strength is not systematically associated with population composition, industrial structure, or pre-1992 turnout. The exceptions are a small set of variables, median income, farm size, and educational attainment,

Table 4.1: Exposure to Salem Communications signals and voting in past Presidential Elections (Placebo Tests)

	1980-1984	1984-1988	1988-1992
	(1)	(2)	(3)
Signal	0.653*	-0.099	0.392
	(0.362)	(0.203)	(0.470)
N	5,568	5,568	5,568
All Baseline Controls	Yes	Yes	Yes
State × Time FE	Yes	Yes	Yes
R^2	0.005	0.000	0.001
Dependent Variable Mean	42.242	42.242	42.242

Notes: This table presents placebo estimates of the effect of Salem signal strength on Republican presidential vote share during three pre-treatment periods: 1980–1984, 1984–1988, and 1988–1992. Each column reports coefficients from a separate two-year panel regression using lagged signal values prior to Salem's actual entry. All regressions include county fixed effects, state-by-year fixed effects, and the full set of baseline controls. The absence of consistent, significant effects across these placebo windows supports the validity of the identification strategy. Standard errors are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

for which coefficients remain statistically significant at conventional levels, but effect sizes are modest. These results indicate that, conditional on baseline geography and potential signal exposure, cross-county differences in realised signal strength are not systematically related to pre-expansion political outcomes or observable characteristics that predict subsequent partisan change.

Table 4.2: Exposure to Salem Communications broadcasting signal and pre-expansion County Characteristics (Balance Tests)

	Univariate		SignalFree, State FE and Terrain		
	Coefficient	R^2	Coefficient	R^2	N
Average Republican Vote Share (1980 - 1988)	0.041	0.008	-0.015	0.319	3084
	0.031		0.017		
Average turnout (1980 - 1988)	-0.001	0.000	0.000	0.031	2919
	0.000		0.000		
Republican vote share (1980)	0.042	0.006	-0.001	0.502	3084
	0.033		0.015		
Republican vote share (1984)	0.055	0.014	-0.020	0.303	3084
	0.030		0.019		
Republican vote share (1988)	0.024	0.003	-0.025	0.253	3084

Continued on next page

	0.037		0.019		
% Presidential Election Turnout (1980)	-0.001	0.001	0.000	0.067	2996
	0.000		0.000		
% Presidential Election Turnout (1984)	-0.001	0.000	0.001	0.022	2950
	0.001		0.001		
% Presidential Election Turnout (1988)	-0.001	0.002	-0.000	0.144	2754
	0.000		0.000		
Population Total (1990)	510.951	0.002	74.811	0.139	3084
	259.847		465.561		
% White (1990)	-0.032	0.002	0.012	0.560	3084
	0.048		0.046		
% Black (1990)	0.019	0.001	-0.023	0.639	3084
	0.043		0.038		
% Native (1990)	0.011	0.002	0.011	0.215	3084
	0.013		0.015		
% Asian/Pacific Islander (1990)	0.002	0.001	-0.000	0.353	3084
	0.001		0.003		
% Hispanic (1990)	0.226	0.175	0.187	0.562	3084
	0.098		0.113		
Births per 1000 in pop. (1990)	0.015	0.014	0.009	0.206	3084
	0.005		0.007		
Deaths per 1000 in pop. (1990)	-0.018	0.019	-0.012	0.177	3084
	0.006		0.010		
Infant deaths per 1000 in pop. (1990)	-0.013	0.001	-0.046	0.023	3084
	0.009		0.025		
% High school graduates (age 25+, 1990)	-0.005	0.000	-0.033	0.508	3084
	0.051		0.028		
% College graduates (age 25+, 1990)	0.039	0.017	-0.007	0.235	3084
	0.021		0.010		
% unemployed (1990)	0.006	0.002	0.012	0.253	3084
	0.006		0.008		
Median HH income (1990)	-7.652	0.001	-28.347	0.316	3084
	20.360		15.773		
% below poverty line (1990))	0.043	0.018	0.046	0.420	3084
	0.027		0.016		
% nonfarm services (1990)	0.011	0.002	0.010	0.145	3084
	0.008		0.011		
Median age 1990 (100%)	-0.011	0.005	-0.005	0.195	3084
	0.006		0.009		
Average size of farm (Acres, 1992)	17.439	0.030	7.989	0.273	3084
	3.787		4.378		

Continued on next page

Federal expenditures per capita (1990)	6.686	0.003	2.617	0.121	3084
	3.954		4.887		
Evangelical Protestants per 1000 in pop.	-0.653	0.007	-1.878	0.621	3027
	0.537		0.909		
% Farm operators (1992)	0.000	0.000	0.000	0.162	2947
	0.000		0.000		
% Manufacturing (1990)	-0.001	0.004	-0.000	0.085	2947
	0.000		0.000		

Notes: This table reports balance regressions of pre-expansion county-level covariates on baseline exposure to Salem Communications' signal strength and market concentration, including their interaction. The left panel reports univariate correlations. The right panel includes controls for signal-free power, state fixed effects, and county-level terrain. Robust standard errors are in parentheses.

In addition to ruling out observable sources of endogeneity, I also address the possibility that Salem's terrain-induced signal variation might be correlated with that of other ideologically similar broadcasters. Section 5.3 presents placebo tests using signal exposure from Bott Radio Network and American Family Radio, two evangelical networks active during the same period, which reveal no comparable effects on vote share. To the extent that terrain-related attenuation also reduced exposure to stations with opposing ideological content, the estimates reported may be interpreted as conservative.

5 Results

5.1 Baseline Results

Table 5.1 presents the estimated effect of exposure to Salem's Christian-conservative radio broadcasts on Republican vote share and voter turnout between the 1992 and 1996 presidential elections. The dependent variable in columns 1–7 is the Republican vote share at the county level, while columns 8 and 9 report estimates for total and Republican-attributed turnout, respectively.

Column 1 begins with a parsimonious specification that includes only county and year fixed effects. In this baseline model, a one standard deviation increase in received signal strength is associated with a 1.25 percentage point increase in Republican vote share. The estimate is statistically significant at the 1 percent level. Column 2 adds the free-space signal strength measure, which accounts for the mechanical correlation between signal strength and proximity to transmitters. The coefficient on the effective signal remains large and significant, increasing to 2.79 percentage points.

Subsequent columns introduce a series of progressively richer control sets. Column 3 adds state-by-year fixed effects and time-varying controls for county demographics

and socioeconomic conditions. Column 4 adds controls for population density, area, and terrain ruggedness. Column 5 further adjusts for past presidential voting patterns (1980–1988), and column 6 includes religious adherence rates across denominations. The estimated effect remains remarkably stable throughout, hovering between 0.81 and 0.83 percentage points per standard deviation increase in signal strength.

Column 7 reports results from the most saturated and preferred specification, which includes all control variables and fixed effects. The coefficient estimate of 0.81 remains statistically significant at the 1 percent level. This effect corresponds to an increase of roughly 2.1 percent relative to the baseline mean. The stability of the point estimate across specifications lends strong support to the identifying assumption that topography-driven variation in signal strength is orthogonal to unobserved county-level political trends.

Columns 8 and 9 assess the effect of signal exposure on voter turnout. In both specifications, the coefficient on signal strength is statistically indistinguishable from zero and substantively negligible, indicating that the primary margin of response was vote switching rather than mobilisation.

I next implement an alternative specification to assess whether the political effects of Salem exposure were driven by changes in content rather than coverage alone. This analysis restricts the sample to counties with stable signal strength between 1992 and 1996, excluding areas where Salem's signal was newly introduced or dropped, and compares changes in Republican vote share before and after Salem's 1995 pivot to explicitly partisan Christian-conservative programming. The estimated coefficient on the interaction between signal exposure and the post-treatment period in Table B.1 is positive and similar in magnitude to the baseline effect, suggesting that the ideological shift may have amplified the political consequences of broadcast exposure. However, the estimates are imprecise and not statistically significant, and should be interpreted with caution. Still, the direction and size of the effect provide suggestive evidence that it was the content of Salem's programming, not solely the presence of a signal, that shaped local electoral outcomes.

Evangelical Population Heterogeneity: To assess whether the political effects of religious broadcasting were stronger in communities aligned with Salem's target audience, I next examine heterogeneity by religious composition. Salem Media's programming explicitly caters to evangelical Christian audiences, and contemporary accounts indicate that its flagship talk shows and pastoral content were designed to resonate with socially conservative Protestant communities. If the network was more persuasive in areas with a higher baseline demand for Christian-conservative content, then we should expect the political influence of exposure to Salem's signal to be larger in counties with higher concentrations of evangelical Protestants.

To test this hypothesis, I interact signal strength with an indicator for counties in the top quartile of evangelical Protestant adherence as of 1990. Column 1 of Ta-

Table 5.1: Exposure to Salem Communications signals and voting in Presidential Elections (1992–1996)

	Republican Vote Share							Turnout (*Total*)	Turnout (*Republican*)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Signal	1.247***	2.787***	0.819***	0.819***	0.832***	0.815***	0.814***	-0.004	0.007
	(0.460)	(0.640)	(0.284)	(0.284)	(0.282)	(0.277)	(0.290)	(0.035)	(0.014)
N	5,568	5,568	5,568	5,568	5,568	5,568	5,568	5,568	5,568
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Signal Free	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE × Time FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
County Terrain Controls	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Past Electoral Controls	No	No	No	No	No	Yes	Yes	Yes	Yes
Religious Controls	No	No	No	No	No	No	Yes	Yes	Yes
R^2	0.019	0.107	0.008	0.008	0.008	0.008	0.008	0.001	0.002
DV Mean (1992)	39.781	39.781	39.781	39.781	39.781	39.781	39.781	0.582	0.229
DV SD (1992)	8.591	8.591	8.591	8.591	8.591	8.591	8.591	1.952	0.720
Signal SD (1992)	33.005	33.005	33.005	33.005	33.005	33.005	33.005	33.005	33.005

Notes: This table reports coefficient estimates from a series of nested two-way fixed effects models of Republican vote share and turnout between 1992 and 1996. The main explanatory variable is received signal strength from Salem affiliates. All models include county and year fixed effects, with controls incrementally added across specifications (see row labels). Turnout columns report total and Republican turnout, respectively. Standard errors are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

ble 5.2 shows that the effect of a one standard deviation increase in exposure to Salem broadcasts on Republican vote share is significantly larger in evangelical-rich counties. The estimated effect in counties outside the top quartile of the evangelical population is 0.81 percentage points, while the interaction term for the top quartile is 0.44 percentage points higher, implying a total effect of approximately 1.25 percentage points, approximately 3 percent of the 1992 mean Republican vote share.

In columns 2 - 4, I repeat this exercise for counties in the top quartile of Catholic, non-Evangelical Protestant and other Protestant religious adherence, respectively. The estimated interaction effects are smaller and statistically insignificant in all cases. These results are consistent with the view that Salem's programming exerted greater persuasive influence in counties demographically aligned with its intended religious audience.[28]

Columns 4 and 5 show no statistically significant effects on turnout, either overall or for Republican voters specifically, suggesting that Salem's influence operated primarily through persuasion rather than mobilisation. Taken together, these results indicate that the persuasive effect of Salem's broadcasts was strongest in religious constituencies aligned with the network's message.

AM/FM and Format-Type Signal Heterogeneity: To better understand the mechanisms through which Salem's broadcasts influenced political behaviour, Table B.3 disaggregates signal exposure by both frequency band and content format. The first set of columns distinguishes between AM and FM signals. As expected, the AM co-

[28]In each regression, the interacted religious group is excluded from the set of baseline religious controls, while the remaining religious group shares remain included.

Table 5.2: Exposure to Salem Communications signals and voting in Presidential Elections (1992–1996) – DDD

	Republican Vote Share				Turnout (*Total*)	Turnout (*Republican*)
	(1)	(2)	(3)	(4)	(5)	(6)
Signal	0.813***	0.774**	0.838***	0.773**	0.003	0.011
	(0.269)	(0.317)	(0.303)	(0.301)	(0.039)	(0.014)
Signal × Evangelical	0.437**				-0.085	-0.031
	(0.202)				(0.081)	(0.028)
Signal × Catholic		0.118				
		(0.202)				
Signal × Other Protestant			-0.009			
			(0.287)			
Signal × Other Religion				0.204		
				(0.255)		
N	5,568	5,568	5,568	5,568	5,568	5,568
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
State × Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Baseline_controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.009	0.008	0.008	0.007	0.001	0.002
DV Mean (1992)	42.242	42.242	42.242	42.242	0.582	0.229
DV SD (1992)	9.948	9.948	9.948	9.948	1.952	0.720
Signal SD (1992)	32.677	32.677	32.677	32.677	32.677	32.677

Notes: This table reports difference-in-differences-in-differences (DDD) estimates of the effect of Salem signal exposure on Republican presidential vote share and turnout between 1992 and 1996. Each column includes county fixed effects, time fixed effects, state-by-year fixed effects, and a full set of baseline covariates (demographic, geographic, and religious). Signal strength is interacted with indicators for counties in the top quartile of 1990 religious group share to test for heterogeneity in treatment effects. Columns 1–4 report vote share outcomes; Columns 5–6 report effects on total and Republican turnout, respectively. Standard errors are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

efficients are larger and statistically significant, while FM estimates are smaller and imprecisely estimated. This aligns with the programming structure of Salem Media Group, which reserved AM frequencies for politically oriented talk shows, such as The Michael Medved Show and The Dennis Prager Show, while FM stations primarily delivered Christian music and general faith-based lifestyle content. Given AM radio's central role in the rise of conservative political broadcasting, this pattern is consistent with an interpretation in which persuasion, not religious ambience, drives political outcomes.

The second set of columns directly tests this mechanism by classifying stations by format: Christian Teach & Talk (CTT), News Talk (NT), Top 40 (contemporary music), and Other. Results show that CTT and NT formats, both of which emphasise spoken-word content with ideological and moral themes, are strongly associated with

increased Republican vote share. The interaction effects with the evangelical population are also positive and significant, indicating that persuasive impacts were amplified in counties aligned with Salem's religious messaging. In contrast, Top 40 and Other formats have smaller or insignificant effects, reinforcing the view that content, not just signal presence, determined political influence. These findings show that Salem's political effects were not generic to Christian broadcasting, but instead operated through intentional format targeting of ideologically resonant content to receptive audiences.

5.2 Robustness

I conduct several robustness checks to validate the main findings. Table B.2 presents estimates from alternative specifications of Equation 1. Columns (1) to (3) replace the continuous signal strength measure with indicator variables equal to one if signal strength exceeds varying thresholds. This approach follows the discrete treatment specification common in ITM-based research (Yanagizawa-Drott, 2014; Chopra, 2021). In all cases, the estimated effect on Republican vote share remains positive and statistically significant.

In Columns (4) to (7), I assess whether the main results are driven by reductions in signal strength rather than the introduction of Salem's broadcast coverage. Specifically, in columns (4) and (5) I hold signal strength constant in counties where coverage declined between 1992 and 1996. In columns (6) and (7) I restrict the sample by omitting counties for which the signal strength reduced between 1992 and 1996. The coefficients in each specification remain similar in magnitude and significance, suggesting that the results are not driven by exits from coverage. Online Appendix Table E1, I explore spatial heterogeneity and find consistent effects across different geographic regions newly exposed to Salem's broadcasts.

In Column (8), I test for heterogeneity in treatment effects by replacing the top-quartile Evangelical dummy with county-level quintiles of Evangelical Protestant share. Interactions between signal strength and each quintile (excluding the lowest, which serves as the reference group) reveal a monotonic increase in the estimated effect, with the top quintile displaying the largest and most precisely estimated coefficient. The fourth quintile is positive but not statistically distinguishable from zero. Overall, these results provide support for the baseline findings and suggest stronger effects in more religious counties.

5.3 Alternative Religious Radio Signal Placebo

To validate that the political effects identified for Salem Media Group are not simply artefacts of religious broadcasting or conservative Christian radio more broadly, I conduct placebo tests using two other religious radio networks: Bott Radio Network and American Family Radio (AFR). Both networks operated nationally during the study

period, had ties to the Council for National Policy, and shared a broadly evangelical orientation, but differed in programming style and political engagement.

Bott Radio Network, founded in 1962, focused exclusively on non-partisan religious teaching and avoided political commentary. AFR, launched by the American Family Association in the late 1980s, promoted a Christian nationalist worldview but maintained a primary emphasis on moral activism rather than electoral influence during the early 1990s.

For both networks, I digitise the historical station expansion using *Broadcasting & Cable Magazine* and calculate county-level signal strength using the same terrain-adjusted methodology applied to Salem. Replicating the main analysis with these alternative exposure measures yields no significant effects on Republican vote share or turnout between 1992 and 1996 (see Appendix Tables B.4 and B.5).

The absence of comparable effects for Bott and AFR supports the interpretation that it was Salem's distinctive integration of religious identity and partisan political messaging, not religious broadcasting alone, that drove the observed shifts in voting behaviour.

5.4 Persuasion Rate

To better understand the magnitude of the political effects associated with Salem's Christian-conservative broadcasts, I calculate a persuasion rate following the method introduced by DellaVigna and Kaplan (2007) and formalised in DellaVigna and Gentzkow (2010). The persuasion rate measures the share of persuadable individuals who switched to voting Republican as a result of exposure to Salem's programming. That is, those who would not have voted Republican in the absence of exposure but did so because of it.

Using the interaction estimate from Section 5, along with evangelical population shares and baseline Republican vote support, I estimate a persuasion rate of approximately 3.0% (standard error: 1.4%). The calculation is detailed in Appendix D.

This estimate falls within the lower range of media persuasion effects reported in the economics literature (Enikolopov et al., 2011; DellaVigna et al., 2014; Adena et al., 2015). For instance, DellaVigna and Kaplan (2007) estimate that Fox News persuaded 11.6% of viewers, while other studies report effects between 6% and 20%. Given Salem's emphasis on ideological reinforcement rather than mass-audience news framing, a more modest persuasion rate is plausible. Yet even a small average effect, when applied across counties with tens of thousands of voters, can generate substantively meaningful aggregate shifts in electoral outcomes.

5.5 Staggered Entry and Persistence of Effects

Up to this point, the analysis has focused on the initial wave of Salem's expansion between 1992 and 1996, when the network grew from 18 to 34 owned-and-operated stations. I now extend the analysis in two stages. First, I examine the full staggered rollout of Salem-owned stations through 2004, a period during which the network expanded to 103 towers nationwide. This staggered entry permits estimation of the dynamic political effects of increased signal exposure using the continuous-treatment event-study framework developed by de Chaisemartin et al. (2024), which accommodates both staggered timing and continuous variation in treatment intensity. I estimate placebo effects for the three presidential elections prior to treatment and trace the evolution of effects over the three elections following initial exposure.

In the second stage, I extend the panel through to 2020 to assess the long-run persistence of early exposure effects. It is important to note, however, that the measure of exposure used throughout the analysis is based solely on Salem-owned broadcast infrastructure. It does not capture the rapid post-2004 growth in syndicated programming delivered via the Salem Radio Network (SRN), which allowed thousands of unaffiliated AM and FM stations to carry Salem-produced content. Because I do not observe where or when unaffiliated stations adopted Salem programs, exposure after 2004 reflects only a subset of total potential ideological influence. As a result, signal strength from Salem-owned towers in the post-2004 period should be interpreted as a lower bound on exposure to Salem's content, in contrast to the pre-2004 period, which more cleanly identifies exposure to exclusively Salem-controlled programming.

Nevertheless, the extended panel provides a valuable test of durability: whether early exposure to Salem's distinctive religious-conservative programming produced persistent shifts in political behaviour, even as the broader media landscape fragmented with the rise of cable news, digital platforms, and alternative ideological sources. This long-run analysis allows for an assessment of both persistence and potential interaction, whether the effects of early exposure were amplified or eroded as ideological media competition increased.

$$Y_{ct} = \sum_{\tau \neq -1} \beta_\tau \cdot \mathbb{1}\{\text{event time}_c = \tau\} + \sum_{s=1}^{S} \sum_{t} \delta_{st} \cdot T_{c0}^{s} \cdot \mathbb{1}\{t = t\} + \eta_c + \lambda_t + \varepsilon_{ct} \quad (2)$$

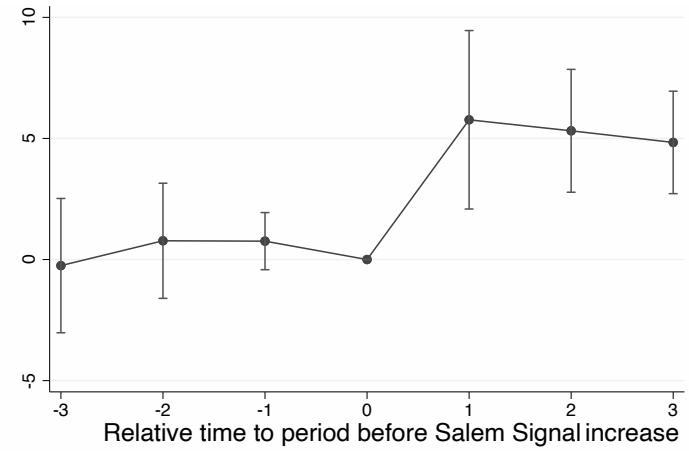
In this specification, Y_{ct} denotes the outcome of interest for county c in year t, such as Republican vote share. The variable event time$_c$ indexes time relative to the change in signal strength for county c, and β_τ captures the dynamic treatment effect at event time τ, relative to the reference period $\tau = -1$, which is omitted for normalisation. The term T_{c0} represents the baseline (e.g., 1992) level of the continuous treatment variable (signal strength), and the interaction $T_{c0}^{s} \cdot \mathbb{1}\{t = t\}$ flexibly controls for differential time trends by allowing year-specific coefficients δ_{st} on polynomial functions of T_{c0} up

to degree S. County fixed effects η_c control for time-invariant heterogeneity, and year fixed effects λ_t absorb common shocks. The error term ε_{ct} captures residual variation, clustered at the state level.

Results

Results in Figure 5.1, which plots dynamic effects from 1992 to 2004, show an immediate and persistent increase in Republican vote share following Salem's entry. There is no evidence of pre-trends in the three election cycles preceding exposure, supporting the identifying assumption of parallel trends. The treatment effect emerges sharply in the first post-treatment election, reaching around 5 percentage points, and remains elevated in the next two election cycles.

Figure 5.1: Dynamic exposure to Salem signal across time (1992–2004)

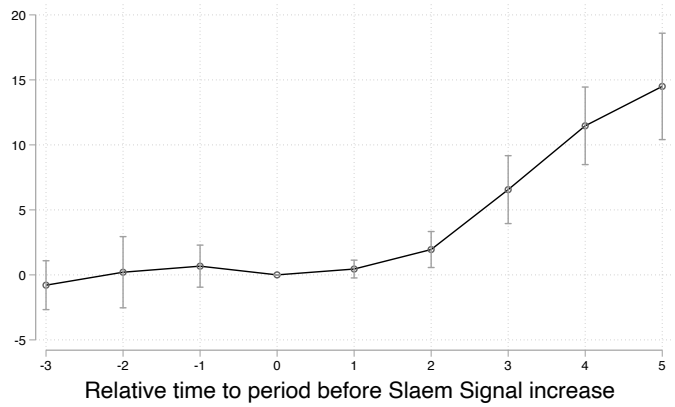


Notes: This figure plots event-study estimates of the effect of Salem signal exposure on Republican presidential vote share from 1992 to 2004. Counties are grouped by time relative to their first significant increase in Salem signal strength. Bars indicate 95% confidence intervals. The reference period is the election immediately prior to treatment.

The extended event-study estimates in Figure 5.2 show that these partisan shifts persisted well beyond the initial expansion period. Because the estimator aligns counties by event time, the longest-run effects are identified from counties first exposed in the early 1990s, while shorter-run dynamics are informed by the larger set of counties treated later in the rollout. The figure demonstrates that effects remain statistically and substantively significant through at least five election cycles after initial exposure.

Importantly, there is no evidence of anticipatory behaviour or differential trends in the pre-treatment periods, reinforcing the credibility of the identification strategy. The persistence of treatment effects in the long run, despite increasing competition from digital and ideological media sources, suggests that Salem's religious-conservative programming generated durable changes in political preferences, particularly in counties exposed early in the network's growth.

Figure 5.2: Dynamic exposure to Salem signal across time (1992–2020)



Notes: This figure extends the event-study analysis through the 2020 election. Counties are indexed by time relative to their first increase in Salem signal strength. Coefficients reflect dynamic treatment effects relative to the election immediately before treatment. Bars represent 95% confidence intervals. Longer-run effects reflect early-treated counties only.

To complement the dynamic event-study approach, I estimate a series of two-way fixed effects regressions that pool counties over sequential three-election windows. These specifications compare counties exposed to increases in signal strength to counties who never and have not yet been exposed to increases in signal strength, isolating the average effect of increased signal strength within each period. The baseline model in Table 5.3 columns 1, 3, and 5 shows that exposure to Salem's signal is associated with a statistically significant increase in Republican presidential vote share in the earlier periods, ranging from 0.81 percentage points in 1992–1996 to 0.61 percentage points in 1992–2000, although the point estimates decline and become statistically indistinguishable from zero by the 1992–2004 window.

In Columns 2, 4, and 6, I interact signal strength with evangelical population shares. The interaction terms are positive, statistically significant, and grow in magnitude over time. By 2004, the signal effect in high-evangelical counties reaches 1.68 percentage points, suggesting that Salem's influence was especially pronounced in counties with greater baseline receptivity to religious-conservative content. The results highlight the importance of local ideological context in conditioning the political effects of partisan media exposure. They also reinforce the interpretation from the event-study that early effects were strongest where Salem's messaging resonated most and that these effects attenuate only modestly over time in aggregate, while remaining persistent in receptive environments.

Table 5.3: Exposure to Salem Communications signals and voting in Presidential Elections (Staggered DiD Matrix, 1992–2004)

	(1)	(2)	(3)	(4)	(5)	(6)
Signal	0.814***	0.780***	0.613**	0.311	0.417	-0.063
	(0.290)	(0.269)	(0.297)	(0.295)	(0.322)	(0.338)
Signal ×		0.511**		1.254***		1.675***
Evang Pop		(0.212)		(0.250)		(0.322)
Years	1992 - 1996	1992 - 1996	1992 - 2000	1992 - 2000	1992 - 2004	1992 - 2004
N	5,568	5,568	8,352	8,352	11,136	11,136
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE × Time FE	Yes	Yes	Yes	Yes	Yes	Yes
All BL controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.008	0.008	0.004	0.012	0.002	0.016
Signal SD	0.841	0.841	0.821	0.821	0.804	0.804

Notes: This table reports estimates from two-way fixed effects regressions of Republican presidential vote share on Salem signal strength. Each column corresponds to a different three-election window. All specifications include county fixed effects, state-by-year fixed effects, and a full set of baseline controls (demographic, geographic, and religious). Interaction terms are included in even-numbered columns to capture heterogeneity by 1990 evangelical population. Standard errors are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

6 Conclusion

This paper provides new evidence on the political consequences of religiously affiliated media infrastructure in the post-deregulation era. Focusing on Salem Communications, a Christian-conservative radio network that rapidly expanded its national footprint during the 1990s and early 2000s, I show that increased exposure to partisan religious broadcasting measurably influenced U.S. presidential voting behaviour. Using plausibly exogenous variation in signal strength induced by local topography, combined with the staggered entry of Salem stations across counties and years, I identify a persistent and demographically targeted increase in Republican vote share attributable to Salem's broadcasts. These effects are strongest in counties with high concentrations of evangelical Protestants and are driven primarily by politically oriented AM radio formats, rather than religious music or apolitical programming.

The findings speak to several important literatures. First, they extend a growing body of work on media effects by demonstrating that ideologically motivated religious broadcasters can exert durable influence on electoral outcomes, comparable in magnitude and persistence to well-studied secular outlets such as Fox News. Second, they highlight the role of deregulation, particularly the repeal of the Fairness Doctrine and the relaxation of ownership caps under the Telecommunications Act of 1996, in facilitating the rise of ideologically coherent media networks that fuse religious identity with

political messaging. Third, the results contribute to our understanding of political persuasion by showing that religiously framed partisan content can move vote choices even in the absence of overt mobilisation efforts, with effects concentrated in ideologically aligned constituencies.

Importantly, the effects of exposure persist well beyond initial contact. Counties first exposed to Salem's broadcasts in the early 1990s continue to exhibit elevated Republican support through at least five subsequent presidential elections. This suggests that partisan religious media can generate not only immediate shifts in electoral behaviour, but also longer-run realignments in political preferences, particularly when delivered through trusted theological frameworks. At the same time, the heterogeneity in treatment effects highlights the importance of local ideological context: Salem's influence is strongest where its message resonates with preexisting religious worldviews.

More broadly, this paper illustrates how transformations in media market structure, enabled by regulatory change and shaped by religious institutions, can contribute to durable shifts in political coalitions. As new forms of ideological communication continue to emerge across digital platforms, understanding the interplay between media content, identity, and political behaviour remains an urgent task. The case of Salem Communications provides a historical analogue to contemporary dynamics, showing how private actors operating within deregulated communication environments can build powerful infrastructure for political persuasion, often outside the scope of conventional media scrutiny.

These findings also speak to the broader institutional consequences of media deregulation. Salem's expansion occurred in the wake of two major policy shifts: the repeal of the Fairness Doctrine and the relaxation of ownership limits under the Telecommunications Act of 1996, which enabled more consolidated and ideologically defined media structures to emerge. While these reforms were primarily justified on economic and technological grounds, the results presented here indicate that such changes may also have downstream political implications. More broadly, the findings highlight the potential for structural media policy to shape the information environments in which voters form political preferences, particularly in settings where content is closely aligned with group-based identity. These patterns merit further investigation, particularly in the context of evolving digital and broadcast media regulation.

References

Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., and Zhuravskaya, E. (2015). Radio and the rise of the nazis in prewar germany. *Quarterly Journal of Economics*, 130(4):1885–1939.

Algara, C. and Amlani, S. (2021). Replication data for: Partisanship & nationalization in american elections: Evidence from presidential, senatorial, & gubernatorial elections in the u.s. counties, 1872–2020.

Bouton, L., Castanheira, M., and Llorente-Saguer, A. (2021). The political economy of single-issue politics. *Review of Economic Studies*, 88(5):2226–2261.

Buccione, G. and Knight, B. G. (2024). The rise of the religious right: Evidence from the moral majority and the jimmy carter presidency. Technical report, National Bureau of Economic Research.

Chopra, F. (2021). Media persuasion and consumption: Evidence from the dave ramsey show. *Available at SSRN 3992358*.

de Chaisemartin, C., d'Haultfœuille, X., and Vazquez-Bare, G. (2024). Difference-in-difference estimators with continuous treatments and no stayers. In *AEA Papers and Proceedings*, volume 114, pages 610–613. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

DellaVigna, S., Enikolopov, R., Mironova, V., Petrova, M., and Zhuravskaya, E. (2014). Cross-border media and nationalism: Evidence from serbian radio in croatia. *American Economic Journal: Applied Economics*, 6(3):103–132.

DellaVigna, S. and Gentzkow, M. (2010). Persuasion: empirical evidence. *Annu. Rev. Econ.*, 2(1):643–669.

DellaVigna, S. and Kaplan, E. (2007). The fox news effect: Media bias and voting. *Quarterly Journal of Economics*, 122(3):1187–1234.

Durante, R., Pinotti, P., and Tesei, A. (2019). The political legacy of entertainment tv. *American Economic Review*, 109(7):2497–2530.

Enikolopov, R., Petrova, M., and Zhuravskaya, E. (2011). Media and political persuasion: Evidence from russia. *American economic review*, 101(7):3253–3285.

Grammich, C., Hadaway, K., Houseal, R., Jones, D. E., Krindatch, A., Stanley, R., and Taylor, R. H. (2018). Longitudinal religious congregations and membership file, 1980–2010 (county level). Open Science Framework.

Haines, M. R., university Consortium for Political, I., and Research, S. (2010). Historical, demographic, economic, and social data: The united states, 1790–2002.

ICPSR 20660 (2008). County characteristics, 2000–2007 [united states].

Olken, B. A. (2009). Do television and radio destroy social capital? evidence from indonesian villages. *American Economic Journal: Applied Economics*, 1(4):1–33.

Prat, A. and Strömberg, D. (2013). The political economy of mass media. *Advances in Economics and Econometrics*, 2:135–187.

Wang, T. (2021). Media, pulpit, and populist persuasion: Evidence from father coughlin. *American Economic Review*, 111(9):3064–3092.

Yanagizawa-Drott, D. (2014). Propaganda and conflict: Evidence from the rwandan genocide. *The Quarterly Journal of Economics*, 129(4):1947–1994.

Appendix

A Figures

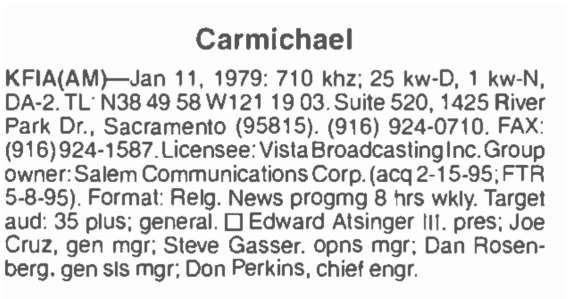
Figure A.1: Excerpts from *Broadcast & Cable Yearbook* 2000



Salem Communications Corp., 4880 Santa Rosa Rd., Suite 300, Camarillo, CA 93012. (805) 987-0400. FAX: (805) 384-4511. Executives: Edward G. Atsinger III, pres/CEO; Stuart W. Epperson, chmn; Eric H. Halvorson, exec VP/COO.

Stns: 34 AM, 15 FM. KPXQ(AM) Phoenix, AZ; KFIA(AM) Carmichael (Sacramento), KIEV(AM) Glendale, KLTX(AM) Long Beach (Los Angeles), KKLA-FM Los Angeles, KDAR(FM) Oxnard (Ventura), KTKZ(AM) Sacramento, KLTH(AM) San Bernardino (Riverside), KFAX(AM) San Francisco and KPRZ(AM) San Marcos-Poway (San Diego), all CA; KRKS-FM Boulder (Denver), KBJD(AM) Denver. KNUS(AM) Denver, KRKS(AM) Denver, KPRZ-FM Fountain, KBIQ(FM) Manitou Springs and KGFT(FM) Pueblo, all CO; WYLL(FM) Des Plaines (Chicago), IL; WLSY(FM) New Albany, IN; WRVI(FM) Valley Station, KY; WITH(AM) Baltimore, MD; WEZE(AM) Boston, MA; KYCR(AM) Golden Valley and KKMS(AM) Richfield, both MN; WWDJ(AM) Hackensack, NJ (New York, NY); WMCA(AM) New York, NY;

Salem Owned Stations (2000)



Carmichael

KFIA(AM)—Jan 11, 1979: 710 khz; 25 kw-D, 1 kw-N, DA-2. TL: N38 49 58 W121 19 03. Suite 520, 1425 River Park Dr., Sacramento (95815). (916) 924-0710. FAX: (916) 924-1587. Licensee: Vista Broadcasting Inc. Group owner: Salem Communications Corp. (acq 2-15-95; FTR 5-8-95). Format: Relg. News progmg 8 hrs wkly. Target aud: 35 plus; general. □ Edward Atsinger III, pres; Joe Cruz, gen mgr; Steve Gasser, opns mgr; Dan Rosenberg, gen sls mgr; Don Perkins, chief engr.

KFIA (AM) - Carmichael, Sacramento: Tower, Programming and Ownership Details

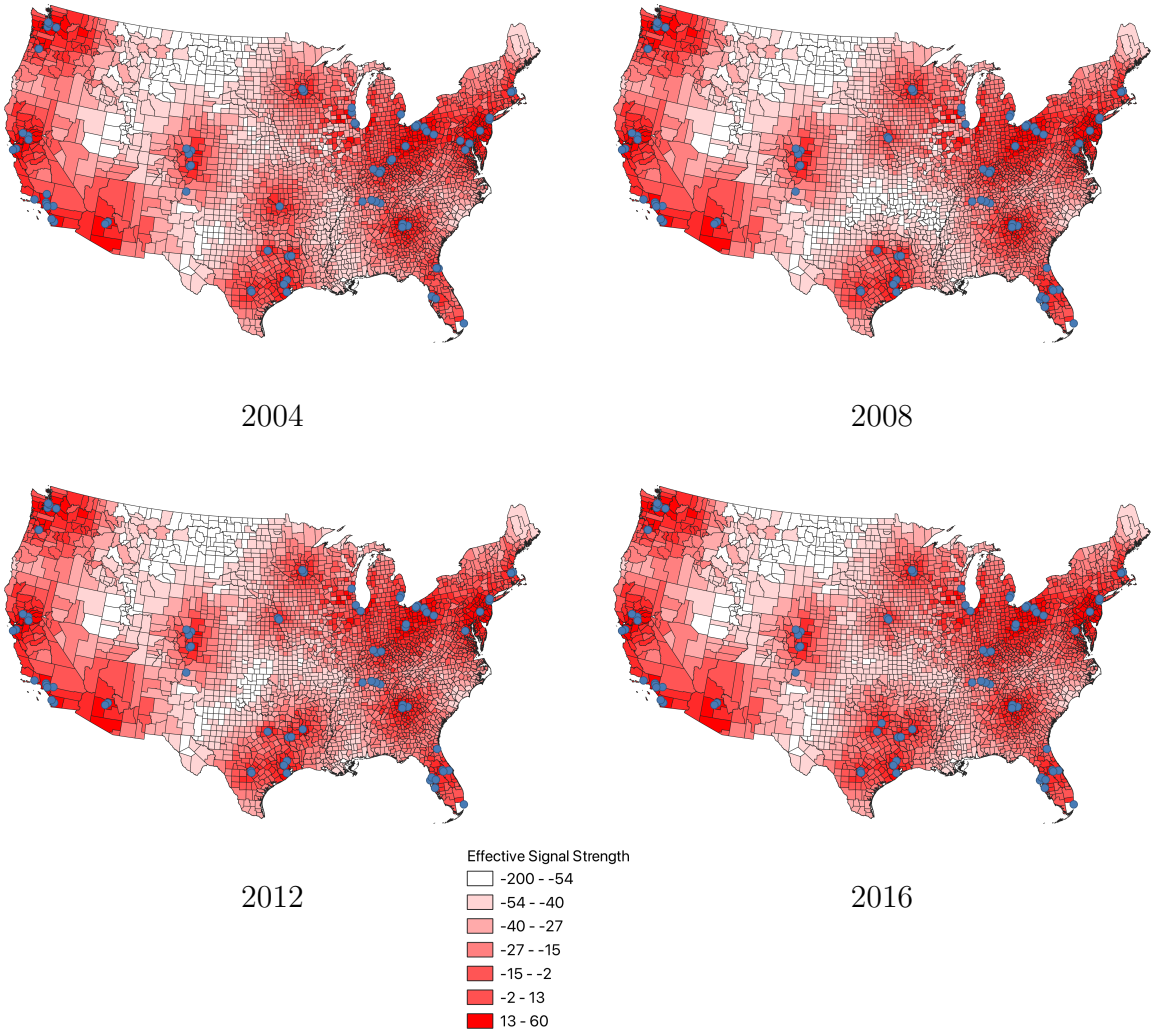


Figure A.2: Post-2004 expansion of Salem Media Group's broadcast coverage based on ITM-modelled signal strength. The marginal increase in signal footprint between 2004 and 2016 justifies restricting the main analysis window to the initial expansion period.

B Tables

Table B.1: Exposure to Salem Communications signals and voting in Presidential Elections (1992–1996)

	Republican Vote Share	
	(1)	(2)
Signal	0.893	0.878*
× Post	(0.606)	(0.464)
Evangelical		-0.798
× Post		(0.529)
Signal		0.061
× Evangelical × Post		(0.735)
N	4,012	4,012
County FE	Yes	Yes
Time FE	Yes	Yes
State FE × Time FE	Yes	Yes
All BL controls	Yes	Yes
R^2	0.008	0.012
Signal SD	33.005	33.005

Notes: This table reports coefficient estimates from two-way fixed effects regressions of Republican vote share between 1992 and 1996. The key independent variable is signal strength from Salem Communications affiliates, interacted with an indicator for the post-treatment year (1996). Column (1) estimates the average effect of exposure for counties with stable signal coverage across both years. Column (2) includes a triple interaction with an indicator for counties in the top decile of white evangelical adherence in 1990. All models include county fixed effects, year fixed effects, state-by-year interactions, and a full set of baseline controls. Standard errors are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B.2: Robustness Main Results (1992 – 1996)

	Republican Vote Share							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Signal (≥ -40)	0.828**							
	(0.344)							
Signal (≥ -20)		1.288***						
		(0.474)						
Signal (≥ 0)			0.517**					
			(0.242)					
Signal				0.825***	0.813***	0.757***	0.747***	0.315
				(0.276)	(0.269)	(0.274)	(0.262)	(0.321)
Signal × Evangelical					0.437**		0.454**	
					(0.202)		(0.216)	
Signal × Q2 Evangelical								0.635*
								(0.375)
Signal × Q3 Evangelical								0.666**
								(0.300)
Signal × Q4 Evangelical								0.337
								(0.478)
Signal × Q5 Evangelical								0.956**
								(0.367)
N	5,568	5,568	5,568	5,596	5,568	4,616	4,588	5,568
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE × Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
All Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.004	0.009	0.001	0.008	0.009	0.006	0.007	0.011
DV Mean (1992)	39.781	39.781	39.781	39.781	39.781	39.781	39.781	39.781

Notes: This table reports robustness checks of the estimated effect of Salem Communications' signal exposure on Republican presidential vote share between 1992 and 1996. Columns 1–3 vary the signal exposure threshold used to define treatment ($\geq$ –40 dBm, –20 cBm, and 0 dBm respectively). Columns 4 and 5 replicate the baseline and DDD specification with signal strength held constant for any counties which witnessed reduced signal strength between 1992 and 1996. Columns 6 and 7 replicate the baseline and DDD specification with the sample restricted to counties facing a constant or increase in signal strength between 1992 and 1996. Column 8 replaces the binary evangelical interaction with quartiles (Q2–Q5) of the county-level evangelical distribution. All specifications include county and year fixed effects, state-by-year fixed effects, and baseline controls. Standard errors are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B.3: Radio Programming Heterogenetity (1992 – 1996)

	Republican Vote Share											
	AM		FM		CTT		NT		Top 40		Other	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Signal	0.967***	0.926***	0.773	0.829	0.727**	0.683**	0.745***	0.717***	0.662**	0.660**	0.970*	0.981
	(0.319)	(0.312)	(0.546)	(0.520)	(0.281)	(0.275)	(0.207)	(0.205)	(0.295)	(0.294)	(0.554)	(0.602)
Signal × Evangelical		0.591**		-0.353		0.610***		0.394***		0.036		-0.089
		(0.266)		(0.331)		(0.170)		(0.144)		(0.149)		(1.127)
N	5,568	5,568	5,568	5,568	5,568	5,568	5,568	5,568	5,568	5,568	5,568	5,568
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE × Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
All BL controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.008	0.009	0.007	0.008	0.980	0.980	0.980	0.980	0.980	0.980	0.980	0.980
Signal SD	33.005	33.005	33.005	33.005								

Notes: This table evaluates the heterogeneity in the effect of radio signal exposure on Republican vote share between 1992 and 1996 by signal frequency and programming format. Columns 1–2 report estimates for AM signals, while columns 3–4 report FM signals. Columns 5–6 correspond to Christian Teach & Talk (CTT) stations; columns 7–8 to News Talk (NT); columns 9–10 to Top 40 music; and columns 11–12 to other formats. Even-numbered columns interact signal strength with an indicator for counties in the top quartile of 1990 evangelical Protestant adherence. All regressions use a two-period panel and restrict the sample to counties that experienced no decline in signal strength between 1992 and 1996. All specifications include county fixed effects, year fixed effects, state-by-year fixed effects, and a full set of baseline demographic, geographic, and religious controls. Standard errors are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B.4: Exposure to Bott Radio Network signals and voting in Presidential Elections (1992–1996)

	Republican Vote Share						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Signal	110.841**	110.841**	7.353	7.353	7.603	7.911	6.918
	(53.869)	(53.869)	(42.405)	(42.405)	(40.136)	(39.383)	(40.006)
N	5,568	5,568	5,568	5,568	5,568	5,568	5,568
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Signal Free	No	Yes	Yes	Yes	Yes	Yes	Yes
State FE × Time FE	No	No	Yes	Yes	Yes	Yes	Yes
County Controls	No	No	No	Yes	Yes	Yes	Yes
County Terrain Controls	No	No	No	No	Yes	Yes	Yes
Past Electoral Controls	No	No	No	No	No	Yes	Yes
Religious Controls	No	No	No	No	No	No	Yes
R^2	0.001	0.001	0.000	0.000	0.000	0.000	0.000
DV Mean (1992)	39.781	39.781	39.781	39.781	39.781	39.781	39.781
DV SD (1992)	8.591	8.591	8.591	8.591	8.591	8.591	8.591
Signal SD (1992)	6.811	6.811	6.811	6.811	6.811	6.811	6.811

Notes: This table reports coefficient estimates from a series of nested two-way fixed effects models of Republican vote share and turnout between 1992 and 1996. The main explanatory variable is received signal strength from Bott Radio Network. All models include county and year fixed effects, with controls incrementally added across specifications (see row labels). Turnout columns report total and Republican turnout, respectively. Standard errors are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B.5: Exposure to American Family Association Radio Network signals and voting in Presidential Elections (1992–1996)

	Republican Vote Share						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Signal	-0.125	-0.079	0.447	0.447	0.423	0.450	0.449
	(0.941)	(1.270)	(0.403)	(0.403)	(0.397)	(0.384)	(0.377)
N	5,156	5,156	5,156	5,156	5,156	5,156	5,156
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Signal Free	No	Yes	Yes	Yes	Yes	Yes	Yes
State FE × Time FE	No	No	Yes	Yes	Yes	Yes	Yes
County Controls	No	No	No	Yes	Yes	Yes	Yes
County Terrain Controls	No	No	No	No	Yes	Yes	Yes
Past Electoral Controls	No	No	No	No	No	Yes	Yes
Religious Controls	No	No	No	No	No	No	Yes
R^2	0.000	0.000	0.002	0.002	0.001	0.002	0.002
DV Mean (1992)	39.781	39.781	39.781	39.781	39.781	39.781	39.781
DV SD (1992)	8.591	8.591	8.591	8.591	8.591	8.591	8.591
Signal SD (1992)	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: This table reports coefficient estimates from a series of nested two-way fixed effects models of Republican vote share and turnout between 1992 and 1996. The main explanatory variable is received signal strength from American Family Association Radio Network. All models include county and year fixed effects, with controls incrementally added across specifications (see row labels). Turnout columns report total and Republican turnout, respectively. Standard errors are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

C Signal Calculations

This appendix describes how effective radio signal strength from Salem transmitters is calculated at the county level using engineering parameters and the Irregular Terrain Model (ITM). The resulting measure serves as the key exposure variable in the main analysis.

Data Inputs and Preprocessing

For each Salem-owned AM and FM station, I collect technical specifications from Federal Communications Commission (FCC) records, including: transmitter coordinates, operating frequency (AM/FM band), effective radiated power (in kilowatts), and antenna height above average terrain (in meters). These parameters are merged with elevation data and county centroid locations. The distance (in kilometres) between each transmitter and each county centroid is computed as a necessary input for propagation modelling.

Transmission Power and Gain Settings

Transmitter power is converted from kilowatts to decibel-watts (dBW):

$$\text{Transmitter Power}_i = 10 \times \log_{10}(\text{ERP}_i \times 1000)$$

I set transmitter antenna gain (TAG) to 2.5 dB and receiver antenna gain (RAG) to 10 dB for all stations. These values are consistent with typical assumptions in engineering and prior work (Olken, 2009).

Signal Loss Estimation via ITM

Two measures of signal loss are calculated:

- **Free-space loss**: estimated assuming no terrain obstacles
- **Effective loss**: estimated using the ITM to account for geographic attenuation (e.g., hills, mountains)

The ITM software calculates terrain-based loss between each transmitter and each county centroid. The following received power calculations are made:

$$\text{Received Power}_{ic}^{\text{Free}} = \text{Transmitter Power}_i - \text{Free-Space Loss}_{ic} + \text{TAG} + \text{RAG}$$
$$\text{Received Power}_{ic}^{\text{Eff}} = \text{Transmitter Power}_i - \text{ITM Loss}_{ic} + \text{TAG} + \text{RAG}$$

Aggregation

Each county c is assigned the maximum effective received power across all transmitters i:

$$\text{Signal Strength}_c = \max_i \left(\text{Received Power}_{ic}^{\text{Eff}}\right)$$

This final signal strength measure is used as the continuous treatment variable in the analysis. A corresponding free-space signal strength measure is also constructed and included as a control..

D Calculation of the Persuasion Rate

Following the methodology of DellaVigna and Kaplan (2007), I calculate the persuasion rate (PR) as:

$$\text{PR} = \frac{\Delta V}{S(1 - V_0)} \tag{3}$$

where:

- $\Delta V = 0.00437$ is the estimated interaction effect of Salem signal strength and high evangelical adherence on Republican vote share (from Table 5.2),
- $S = 0.25$ is the share of the population in counties with top quartile evangelical adherence,
- $V_0 = 0.422$ is the baseline Republican vote share in those counties in 1992.

Substituting in these values:

$$\text{PR} = \frac{0.00437}{0.25 \cdot (1 - 0.422)} = \frac{0.00437}{0.1445} \approx 0.0302$$

which implies a persuasion rate of 3.02 percent. To compute the standard error, I apply the delta method:

$$SE(\text{PR}) = \frac{SE(\Delta V)}{S(1 - V_0)} = \frac{0.00202}{0.1445} \approx 0.0140$$

Hence, the final persuasion rate is estimated at **3.0%** with a standard error of **1.4%**.

Conclusion

This thesis examines how structural features of media markets shape the content and political consequences of communication, focusing on the intermediaries that filter, amplify, and constrain political expression. Across three empirical settings: digital platforms, local newspapers, and religious radio networks, I investigate how media actors respond to economic and institutional incentives, and how those responses affect discourse and democratic participation.

Each paper addresses a distinct but interrelated question: how upstream infrastructure affects content moderation (Paper 1), how market incentives influence symbolic representation (Paper 2), and how ideological media alter partisan alignment (Paper 3). Together, the findings show that media are not passive conduits of information, but strategic actors embedded in commercial and regulatory environments that shape their behavior.

The thesis contributes to political economy, media economics, and institutional analysis by documenting how platform governance, label adoption, and ideological segmentation each operate through market structures to influence political norms and outcomes. In doing so, it highlights the central, but often opaque role of media intermediaries in shaping modern democratic life.