

---

# Damocles's Switchboard: Information Externalities and the Autocratic Logic of Internet Control

Meicen Sun 

School of Information Sciences, University of Illinois Urbana-Champaign  
Email: [mcsun@illinois.edu](mailto:mcsun@illinois.edu)

---

**Abstract** This paper advances a theory for the autocratic logic of internet control. Politically motivated internet control generates a positive externality for domestic data-intensive firms and a negative externality for domestic knowledge-intensive research entities. Exploiting a major internet control shock in 2014, I find that Chinese data-intensive firms gained 26 percent in revenue over other Chinese firms as the result of internet control. The same shock incurred a 10 percent decline in research quality from Chinese researchers, conditional on the knowledge intensity of their discipline. It also reduced the research quality from Chinese researchers relative to their US counterparts by 22 percent in all disciplines. Due to the positive data externality, internet control enacted to prevent domestic threats challenges the state's competing need for data sovereignty against foreign threats. Meanwhile, the state shields certain foreign knowledge-intensive actors from the negative knowledge externality to avoid the immediate economic costs they might otherwise impose. Qualitative evidence supports both implications, highlighting the centrality of short-term interests and foreign actors in autocratic decision making.

---

## Motivation and Contribution

### *Toward a Framework for the Politics of Internet Control*

The politics of the internet has been studied from a variety of angles. Two, in particular, have proceeded in parallel. First is the burgeoning literature on digital censorship. It has tracked the explosion of censorship technology,<sup>1</sup> and the profusion of citizen responses.<sup>2</sup> Second is the emerging line of inquiry on trade in digital goods and

1. King, Pan, and Roberts 2013, 2014, 2017; Sanovich, Stukal, and Tucker 2018; Stukal et al. 2017.
2. Chen and Yang 2019; Fu, Chan, and Chau 2013; Pan and Siegel 2020.

*International Organization* 78, Summer 2024, pp. 427–59

© The Author(s), 2024. Published by Cambridge University Press on behalf of The IO Foundation. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

doi:10.1017/S0020818324000237

services.<sup>3</sup> It encompasses new forms of trade and trade distortion in the digital age,<sup>4</sup> and new modes of interstate interaction engendered therein.<sup>5</sup>

The parallel is peculiar. Internet control and digital trade are inextricably linked, as observed by numerous practitioners in democratization<sup>6</sup> and in trade liberalization.<sup>7</sup> Internet control, defined here as the restriction of internet traffic via the blocking of web domains,<sup>8</sup> has many a time been decried as digital protectionism that unfairly advantages certain domestic sectors.<sup>9</sup> Disputes over this very issue have occurred on both bilateral and multilateral levels.<sup>10</sup> Nonetheless, there has not been a coherent articulation of how internet control implicates digital trade and how the distributional consequences bear on domestic politics and interstate relations.

In this paper, I advance a framework that connects the dots and, in so doing, traces out the logic of internet control in an autocratic state. It begins by distinguishing between three components of information: (1) *ideas* that propel political action; (2) *data* as a factor of production; and (3) *knowledge* as a driver of innovation. Insofar as all three are bound up in information flow, measures to restrict one also disrupt the others. Because of this, internet control intended to restrict ideas and thus prevent domestic challenges to regime security generates two externalities.

First, controls of this kind benefit domestic data-intensive firms in large economies with a high level of internet connectivity. For sectors that use data as an input factor, internet control not only distorts the quantity of foreign digital products available to domestic consumers. It also boosts the factor endowment for said domestic sectors by forcing domestic consumers to contribute their data to domestic producers. With induced growth, the data-intensive firms become more likely to expand overseas,

3. I follow the definition of a digital product in chapter 19 of the US–Mexico–Canada Agreement and chapter 14 of the Comprehensive and Progressive Agreement for Trans-Pacific Partnership as “a computer program, text, video, image, sound recording, or other product that is digitally encoded, produced for commercial sale or distribution, and that can be transmitted electronically.” Office of the US Trade Representative, “Agreement Between the United States of America, the United Mexican States, and Canada, 7/1/20 Text,” available at <<https://ustr.gov/trade-agreements/free-trade-agreements/united-states-mexico-canada-agreement/agreement-between>>; Australian Government Department of Foreign Affairs and Trade, “CPTPP Text and Associated Documents,” available at <<https://www.dfat.gov.au/trade/agreements/in-force/cptpp/official-documents>>.

4. Aaronson 2019; Rodrik 2018; Weymouth 2017.

5. Farrell and Newman 2019; Liu 2021; Simmons and Kenwick 2022.

6. Freedom House, “Freedom on the Net,” available at <<https://freedomhouse.org/report/freedom-net>>.

7. Office of the United States Trade Representative, “USTR Releases 2023 National Trade Estimate Report on Foreign Trade Barriers,” available at <<https://ustr.gov/about-us/policy-offices/press-office/press-releases/2023/march/ustr-releases-2023-national-trade-estimate-report-foreign-trade-barriers>>.

8. I consult Freedom House’s “Obstacle to Access” (in “Freedom on the Net Research Methodology,” available at <<https://freedomhouse.org/reports/freedom-net/freedom-net-research-methodology>>) and the USTR’s “Key Barriers to Digital Trade” (available at <<https://ustr.gov/about-us/policy-offices/press-office/fact-sheets/2017/march/key-barriers-digital-trade>>) in choosing this operational definition. This distinguishes it from other digital trade barriers such as data localization. As the former report documents, while such blocking is prevalent in autocracies, it is also observed in many democracies.

9. Ferracane, Lee-Makiyama, and Van Der Marel 2018; Wu 2017.

10. See, for example, “United States Tells WTO of Concerns over China’s New Web Access Rules,” Reuters, 23 February 2018.

which increases the likelihood of foreign access to domestic data. This impedes the state's competing objective of preventing foreign challenges to regime security, because it undermines data sovereignty, defined here as the total and absolute control of domestically originated data by the state in question.<sup>11</sup> Second, such controls hurt domestic knowledge-intensive actors who rely on access to knowledge from the outside world in generating innovation. Of these actors, the state will make accommodations for only foreign knowledge-intensive actors in the state who can credibly threaten immediate retaliation otherwise.

To test these two information externalities, I leverage the case of China's system of internet control by exploiting a major internet control shock that occurred in 2014. I discover that internet control gives Chinese data-intensive firms an approximately 26 percent marginal increase in revenue compared to other Chinese firms, and up to 50 percent for the most data-intensive firms. However, this advantage does not translate beyond the domestic context. Despite China's internet control, US data-intensive firms have performed marginally better than their Chinese counterparts. This suggests the presence of countervailing forces, one of which I test through an analysis of China's research sector. There, the same internet control shock is associated with a decline in research quality by 10 percent, and up to 15 percent for the most knowledge-intensive disciplines. An analysis of US and Chinese research output reveals that internet control reduces the research quality of Chinese researchers in *any* discipline by 22 percent compared to their US counterparts. With qualitative evidence, I then explicate how internet control's dual externalities pose not one, but two dilemmas: one between internal and external threats to regime security, and the other between imminent political threats and immediate economic costs. In both instances, foreign actors wield momentous sway over the autocrat's calculus.

### *Contribution to the Literature*

In connecting digital censorship with digital trade, this paper contributes to both strands of the literature. Studies have duly noted the political repercussions of digital censorship,<sup>12</sup> but none have scrutinized its distributional consequences. References to censorship as a "tax" on information access are chiefly confined to the context of political repression.<sup>13</sup> In contrast, this paper shows how information externalities distort market outcomes beyond the political objective of digital censorship. In quantifying the divergent effects of internet control on different actors in the economy, it demonstrates how such control begets dividends for domestic data-intensive sectors but costs for the economy as a whole.

This paper also contributes two novel insights to the growing body of research on digital trade. Prior works have explored the political-economy ramifications of the

11. This operational definition takes stock of related definitions in Chander and Sun 2022; Floridi 2020; Rosenzweig 2012; Woods 2018; K. Xu 2019.

12. Chen and Yang 2019; Guriev, Melnikov, and Zhuravskaya 2021; Roberts 2018.

13. Roberts 2020.

unique properties of informational goods, both quantitatively<sup>14</sup> and qualitatively.<sup>15</sup> My empirical test of the two information externalities in concrete, quantitative terms refines prior conjectures by showing that prevailing trade models underestimate the benefit to domestic data-intensive sectors while overlooking the cost to domestic knowledge-intensive sectors. In so doing, I uncover how, beyond the intended winners and losers,<sup>16</sup> the state's manipulation of information creates unintended winners and losers owing to the structure of information flow, which encapsulates multiple components.

Finally, this paper enriches the debate on the “dictator's dilemma.” Politically motivated control of information flow has been argued to come at an economic cost.<sup>17</sup> Autocrats face a dilemma between political unrest, by allowing in too much information, and economic unrest, by allowing in too little.<sup>18</sup> I challenge this framing in two ways. First, I unpack how it misattributes the source of the incentive for the autocrat to limit control. It is not a general concern about long-term growth but a specific concern about the immediate costs that certain actors may impose. Second, I highlight a new dilemma in the digital age between preventing domestic challenges to regime security through internet control and preventing foreign challenges to regime security through data sovereignty. In empowering domestic firms with droves of data, internet control weakens the autocrat's control over such data when these firms later expand overseas as the result of their growth.

In the next section I present my theory on the two information externalities of internet control and four testable hypotheses. I then introduce my empirical case, China's internet control, and detail my data and methodology. With that, I present my quantitative results. The qualitative section corroborates the implications for state strategy, after which I conclude with reference to future directions and policy relevance.

## Theory and Hypotheses

### *Ideas, Data, Knowledge*

My theory begins by recognizing three distinct components of information—*ideas*, *data*, and *knowledge*—based on earlier conceptualizations of the structure of information.<sup>19</sup> Of particular relevance is the definition of information as consisting of (1) ideas, or bit strings that are “set[s] of instructions for making an economic good”; and (2) data, such as “driving data, medical records, and location data.” Whereas scores of images serve as training data for machine learning algorithms, the resulting algorithm as a set of “forecasting rules” exemplifies an idea.<sup>20</sup>

14. Farboodi and Veldkamp 2021; Farboodi et al. 2019.

15. Liu 2021; Weymouth 2023.

16. Brutger and Strezhnev 2022; Kim 2018.

17. Boas 2000; Saleh 2012.

18. Kedzie 1996; Milner 2006.

19. Jones and Tonetti 2020; Romer 1990.

20. Jones and Tonetti 2020.

While useful for economic analyses, this definition omits a category of information central to civic and political life. Whether it is an ideology deemed threatening to the regime or a rallying call for assembly, information that inspires or facilitates political action has been the prime target for digital censorship.<sup>21</sup> Information of this kind is more like ideas than data in that it requires the interpretation and sense-making of a human actor.<sup>22</sup> Meanwhile, it differs from the foregoing examples of an idea in that the primary objective is to perform a political action rather than produce an economic good. For the purposes of my theory, I term human-actionable information intended for political action *ideas* and that intended for economic production *knowledge*.

One may conceptualize the distinction between data and knowledge with respect to economic production as that between input factor and total factor productivity (TFP). Let the total output,  $Y$ , be a function of TFP,  $A$ ; capital as an input factor,  $K$ ; and labor as an input factor,  $L$ . Whereas knowledge, such as technical know-how, affects total output through TFP by altering the returns to input factors, data does so in a different way. For data-driven firms such as Google and Uber, user data—from search history to driving routes—are used to train algorithms that undergird their core products, from which they derive a major stream of their revenue.<sup>23</sup> Data thus enters the equation as a factor of production that is distinct from capital and labor.

Equation (1) conceptually illustrates how information affects total output via the two components—knowledge and data.<sup>24</sup> The TFP,  $A$ , is a function of knowledge,  $K_n$ , while data,  $D$ , is a factor of production:<sup>25</sup>

$$Y = f(A(K_n), K, L, D) \quad (1)$$

### *Information Externalities and Distributional Consequences*

Given that information contains ideas, data, and knowledge, when a state blocks foreign web domains to restrict the flow of ideas, it also disrupts the flow of both data and knowledge. Domestic consumers now face impeded access to foreign digital products, from search engines to social media platforms. This compels them to switch to domestic substitutes. If Google is blocked, for instance,

21. Roberts 2020.

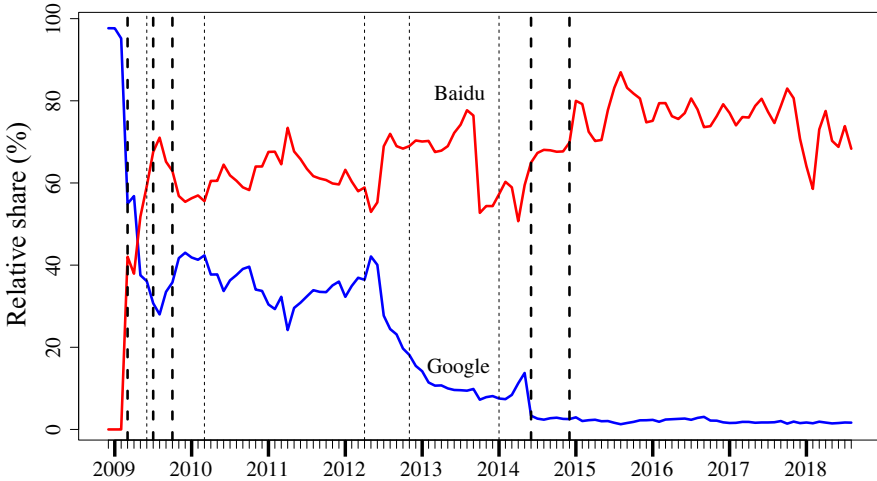
22. The role of human agency in distinguishing between types of information has been widely articulated. Ackoff 1989; Frické 2019; Jørn Nielsen and Hjørland 2014. Even in the age of generative artificial intelligence (AI), humans play a distinctive role in innovation, as recognized by, for instance, the exclusive patentability of humans: “Inventorship Guidance for AI-Assisted Inventions,” *Federal Register*, 13 February 2024, available at <<https://www.federalregister.gov/documents/2024/02/13/2024-02623/inventorship-guidance-for-ai-assisted-inventions>>.

23. *Economist* 2017; Jones and Tonetti 2020.

24. One bit string of information may take multiple forms, as it can be used simultaneously by machines and by humans to various ends. A firm may thus be data-intensive *and* knowledge-intensive if it makes heavy use of both elements.

25. Though it is tempting to express this with a modified Cobb–Douglas production function, such as  $Y = A(K_n) \times K^\alpha \times L^\beta \times D^{1-\alpha-\beta}$ , that would imply more specific relationships between the variables than this paper can formally or empirically demonstrate.

domestic users will resort to an indigenous search engine if one exists. Figure 1 provides a striking visualization of the substitutive relationship between Google and an indigenous search engine when the former's domain experienced disruptions in China.<sup>26</sup>



*Notes:* Vertical lines denote disruptions of varying strength and duration to Google products (data: Google Transparency Report). Google's traffic share fell to nearly zero by 2015 concurrent to the disruptions, as Baidu's rose to around 80%, among the same group of users in China. The tight correlation between dips in Google's traffic share and spikes in Baidu's alludes to the substitution effect that underlies internet control's positive externality for domestic data-intensive sectors.

**FIGURE 1.** *Web traffic to Google versus Baidu from China, December 2008 to August 2018 (data: StatCounter)*

The expanded user base will lead to an increase in both sales revenue and the supply of data. This is due to the prevalence of barter trade, where consumers pay for digital products not with money but with their data.<sup>27</sup> In autocracies, user data collected by domestic producers may be further transacted with the government for the latter's political ends.<sup>28</sup> Treating internet control simply as a tariff or quota without considering these critical features of digital trade would not only overestimate the loss in domestic consumer surplus, given high substitutability between domestic and foreign digital products that are both "free" to use. It would also underestimate

26. The near-perfect substitutability here is not necessary for us to observe at least *some* positive impact on domestic substitutes of foreign digital products as the result of internet control.

27. Farboodi and Veldkamp 2021.

28. Beraja, Yang, and Yuchtman 2023; Beraja et al. 2023.

domestic producer surplus from the supply of data for firms in data-intensive sectors and, in turn, their capacity for growth and expansion.

Concurrently, domestic knowledge-intensive sectors that rely on existing knowledge for their own knowledge production now face impeded access to external knowledge. Anecdotes abound regarding the decline in productivity for researchers when sites such as Google Scholar get blocked. Any or all of three scenarios can occur: (1) Researchers may see a reduction in the amount of external knowledge they can acquire per unit time, such as when network disruptions limit their ability to read articles on Google Scholar (“aware, willing, but unable”).<sup>29</sup> (2) Researchers may be discouraged by such disruptions from trying to acquire external knowledge (“aware but unwilling”).<sup>30</sup> (3) Researchers may be altogether unaware of some external knowledge due to lack of exposure (“unaware”).<sup>31</sup>

Compared to standard trade distortions, welfare transfers to those affected by the negative knowledge externality are complicated by three factors. First, the decline in knowledge production does not immediately translate into a decline in total output. The state must weigh this against more pressing threats to regime security when deciding to impose internet control. Second, the cost to knowledge producers, who are scattered throughout the economy, is more diffuse than the benefit to data-intensive producers, who are fewer in number and better resourced. This presents collective action challenges for the former group.<sup>32</sup> Third, conventional metrics for innovation, discussed later in the empirical analysis, obscure the marginal effect of information access and do not inform precise compensation to those affected by internet control. Attempts at direct welfare transfer through measures such as research-and-development (R&D) spending would thus entail gross inefficiency.<sup>33</sup> These dynamics signify that the “dictator’s dilemma” framing overstates the restraint on the autocrat from the need for innovation.

Figure 2 conceptually illustrates how politically motivated internet control aimed at restricting ideas generates a positive externality for domestic data-intensive sectors and a negative externality for domestic knowledge-intensive sectors. I next spell out the two information externalities as testable hypotheses, before testing them in the sections to follow.

### *Positive Externality for Domestic Data-Intensive Actors*

Different actors in the economy depend on access to data as an input factor to different degrees. Firms that derive most of their revenue from creating data-driven

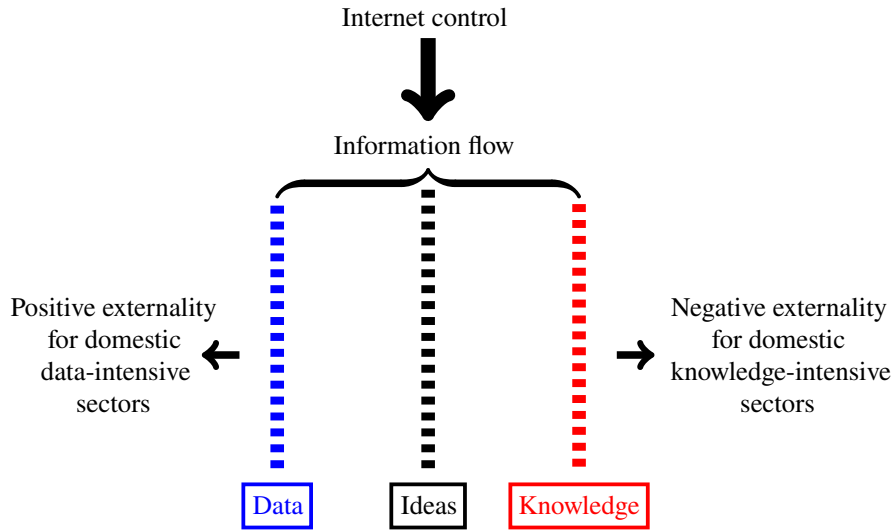
29. Ables 2018; Normile 2017.

30. Fallows 2008; Roberts 2018.

31. Bao 2013; Chen and Yang 2019. All three pathways remain even when circumvention tools such as virtual private networks are used, so long as the cost of internet access in terms of time, resources, and effort exceeds that in the absence of controls. A total inability to access external knowledge is not necessary for us to observe at least some negative impact of internet control on domestic knowledge-intensive sectors.

32. Olson 1971.

33. Doing so would involve calculating a dollar amount for each unit decline in knowledge production resulting from the reduced knowledge access for each knowledge domain.



Notes: Internet control aimed at restricting ideas also disrupts data and knowledge. This generates a positive externality for domestic data-intensive sectors and a negative externality for domestic knowledge-intensive sectors.

FIGURE 2. Two information externalities from internet control

algorithms are more dependent on data than, say, those that profit from producing most physical goods.<sup>34</sup> In the event of internet control, domestic consumers are less able to access foreign digital products and more likely to switch to domestic substitutes, driving up demand for the latter. This leads to an increase in revenue for domestic data-intensive firms, both directly from an increase in sales and indirectly from an increase in the supply of raw materials, or data in this case. Hence,

*Hypothesis 1: Internet control incurs financial gains for domestic data-intensive firms relative to their domestic non-data-intensive counterparts.*

By the same process, foreign data-intensive firms lose out on potential sales and the potential supply of data from consumers in the country under internet control. Hence,

*Corollary Hypothesis 1: Internet control incurs financial gains for domestic data-intensive firms relative to their foreign data-intensive counterparts.*

34. The vast differences in the use of input factors across sectors are evidenced by various input-output tables, such as Bureau of Economic Analysis, "Input-Output Accounts Data," available at <<https://www.bea.gov/industry/input-output-accounts-data>>.



Research on the digital economy indicates a scale effect,<sup>35</sup> which suggests that these hypotheses presuppose a threshold of data endowment in the state. The positive externality therefore applies to states with a large population and a high level of internet connectivity, where a sufficient volume of data can be made available to domestic data-intensive firms that produce substitutes for foreign digital products.<sup>36</sup>

### *Negative Externality for Domestic Knowledge-Intensive Actors*

Similarly, different actors in the economy depend on access to knowledge to different degrees. Researchers who produce knowledge primarily by reviewing the existing literature are more dependent on knowledge than those who do so primarily through other types of activities, such as experiments.<sup>37</sup> In the event of internet control, domestic researchers are less able to access the literature from the outside world. This decrease in knowledge access leads to a steeper decline in the rate of knowledge production for the more knowledge-intensive disciplines, resulting in a greater decline in the quality of research. Hence,

*Hypothesis 2: Internet control incurs a greater decline in research quality for domestic knowledge-intensive disciplines relative to their domestic non-knowledge-intensive counterparts.*

The detriment from internet control affects all domestic researchers, which translates into a decline in research quality for domestic researchers relative to their foreign counterparts across *all* disciplines, regardless of knowledge-intensity. Hence,

*Corollary Hypothesis 2: Internet control incurs a decline in research quality for domestic researchers relative to their foreign counterparts for any given discipline.*

Based on these formulations, I now synthesize the political consequences of the two information externalities and implications for the state's strategy.

### *Implications for State Strategy*

The autocratic state is first concerned with preventing *domestic* challenges to its regime security. Autocracies adept at suppressing and manipulating information are advantaged over overtly violent dictatorships in countering domestic

35. Farboodi et al. 2019; Wilson 1975.

36. While digital substitutes may emerge in response to internet controls, examples including Grab in Southeast Asia demonstrate that they can flourish in the absence of such controls. "Grab Was Already the Uber of Southeast Asia. Now the 'Super-App' Wants to Deliver Financial Equality, Too," *Time*, 1 June 2023.

37. Even a quick survey reveals enormous variation across disciplines in the average volume of literature referenced in a given piece of research. Halevi 2013; Marx and Bornmann 2015; Milojević 2012.

opposition.<sup>38</sup> This incentivizes the autocrat to leverage internet control in restricting the inflow of instigative ideas and domestic communications that facilitate collective action,<sup>39</sup> which causes the two information externalities. Yet the autocrat is also concerned with *foreign* challenges to the regime. One way in which it seeks to prevent such challenges is by pursuing data sovereignty, such as through data localization and cross-border data flow restrictions. As I will explain, the positive data externality creates tension between these two objectives.<sup>40</sup>

### *Political Consequences of Positive Data Externality*

The windfall of data and revenue from the positive data externality makes domestic data-intensive firms more likely to grow and expand globally, such as by listing offshore. Doing so may compel compliance with foreign regulations that curtails the autocratic state's control over the firms' data. This can occur directly, through competing requirements for data localization in foreign territories, or indirectly, through weakened state oversight over these firms. Consequently, one should expect a "one-two punch" from the state to retain control over domestic data held by these firms.

First is a move to reassert data sovereignty with respect to all domestic actors, which may entail stricter and/or more pervasive mandates for state authority over domestic data and prohibitions of foreign access to such data. Second is a move to curb overseas expansion by data-intensive firms which, due to its specificity, may entail targeting individual firms with extensive foreign ownership and/or plans for such expansion. While firm compliance is generally expected in autocracies, signs of noncompliance from data-intensive firms that have benefited from the positive data externality will be met with exceptionally harsh treatment. Being profit-maximizing like all others, the data-intensive firms must now balance growth against the risk of state sanction due to the wealth of domestic data they possess.

### *Political Consequences of Negative Knowledge Externality*

As previously outlined, domestic knowledge-intensive actors are limited in their bargaining power. Direct compensatory welfare transfer by the state would also be inefficient. As a result, the state is not incentivized to offset the negative externality for domestic knowledge-intensive actors beyond limiting the scope of internet control where doing so does not compromise regime security.

One exception is foreign knowledge-intensive actors in the state who are parties to a contract that conditions resource provision to the state on freedom of information access. Typically concentrated in large urban areas, these foreign actors are better

38. Guriev and Treisman 2019; X. Xu 2021.

39. Diamond 2010; King, Pan, and Roberts 2013.

40. This objective function defines the autocratic logic of internet control. A democracy may also adopt this logic, albeit in limited ways, if it follows a similar objective function. This is consistent with the profiles of many democracies in reports such as "Freedom on the Net."

positioned for mobilization than their domestic counterparts. More importantly, they are able to impose *immediate* economic costs on the state, either by invoking legal provisions or by withholding the resources. If the costs are substantial, the state will be incentivized to allow privileged internet access for this specific group of foreign knowledge-intensive actors.

## Data

### *Case Selection: China's System of Internet Control*

I test the two information externalities in my theory through a quantitative analysis of internet control in China. This case uniquely satisfies both the scope and strength requirements for treatment administration. First, my hypotheses on the bifurcated effects on data-intensive versus knowledge-intensive actors require that the internet control in question affects both types of actors—ideally, *all* actors in the economy. In other words, it should be universal or near-universal in scope. China's internet control, popularly dubbed the Great Firewall, offers the closest real-world case to this setting.<sup>41</sup> China's DNS filter blocks hundreds of thousands of domains, with a gamut of subject matter extending far beyond political content.<sup>42</sup>

Second, the internet control must have persisted for a sufficiently long period, with minimal circumvention, to enable meaningful observation of its effects. China's internet control, again, meets this criterion. Unlike censorship shocks elsewhere in the world, which are usually in response to specific events and relatively brief,<sup>43</sup> China's internet control is so entrenched that many in the younger generation have reportedly grown up with little awareness of digital products such as Google and Facebook.<sup>44</sup> As of 2018, only 5 percent of China's urban residents reported attempting to circumvent internet control, and this proportion was presumably much higher than the national average.<sup>45</sup>

### *Treatment Variable: Measuring Internet Control Through Domain Accessibility*

In our case setup, treatment occurred when internet control in China shifted from limited, domain-specific censorship to an across-the-board regime of control. The

41. That is, so long as we can identify the point in time when such near-total control was imposed. I address this in the subsection on treatment measurement. On the scope and scale of China's internet control, see Denyer 2016; Economy 2018; "Internet: Living with the Great Firewall of China," Reuters, 17 October 2017.

42. Hoang et al. 2021.

43. Sun 2019.

44. Yuan 2018.

45. Roberts 2018. Even if circumvention is common among large firms, H1 requires only that *consumers* in China are subject to effective internet control. Total inability to circumvent is not necessary to meet this condition so long as the cost of accessing foreign domains is sufficiently high for these users.

treatment variable must therefore capture both the timing and the degree of this change. In practice, this requires regularly measuring the accessibility of foreign web domains from inside China. Earlier measurements of internet control suffer from various drawbacks, including coder subjectivity, high noise-to-signal ratio, low measurement frequency, narrow scope, sampling bias, and insufficient historical coverage.<sup>46</sup> Given these limitations, I have coded my treatment variable using data from GreatFire, the only known resource of its kind.

GreatFire is an independent group that has used servers in China to test the accessibility of hundreds of thousands of web domains since 2011.<sup>47</sup> The extensive scope is complemented by a high testing frequency—nearly daily for popular domains.<sup>48</sup> I collect accessibility data for the 100 most visited websites in the world.<sup>49</sup> This yields 27,691 observations.<sup>50</sup> Figure 3 depicts the final, interpolated internet control history in China based on the testing data.

I consult previous research and media reports to validate this measurement. Together, they document a massive wave of internet control in 2014,<sup>51</sup> including a major shock around early June when the Chinese state cracked down on foreign websites, allegedly in anticipation of the twenty-fifth anniversary of the Tiananmen Square incident.<sup>52</sup> The anniversary has been nicknamed Internet Maintenance Day in recognition of the state's intensified website blocking around this time each year,<sup>53</sup> making numerous domains inaccessible for days, without explanation.<sup>54</sup>

This wave of internet control is confirmed by the large red segment that begins around early June 2014 marked out in Figure 3. I exploit this shock as my treatment because it uniquely meets the two mentioned conditions: it is near-universal in scope, as it includes almost all of the websites being tested (a “wide” dosage); and it spans a lengthy two years, through mid-2016, including a brief period of relaxation (a “deep” dosage).<sup>55</sup>

46. Still useful for other purposes, these encompass composite measures such as “Freedom on the Net”; application-level measures, including the “Google Transparency Report” (available at <<https://transparencyreport.google.com/traffic/overview>>) and third-party web analytics such as StatCounter (available at <<https://statcounter.com/>>); and technical tools for censorship detection, such as Censored Planet (available at <<https://censoredplanet.org>>).

47. GreatFire Analyzer, available at <<https://en.greatfire.org/analyzer>>. Servers inside China afford a better measurement vantage than proxy servers located overseas.

48. I elaborate on the coding of the treatment variable in the online supplement.

49. *Similarweb*, “Top 100 Websites Ranking on the Web,” available at <<https://www.rankranger.com/top-websites>>, accessed 20 April 2020. Sampling the most popular domains improves measurement precision by reducing the proportion of missing values, as less-visited domains are tested less frequently.

50. I use linear interpolation to address missing values. Stine interpolation yields a similar result.

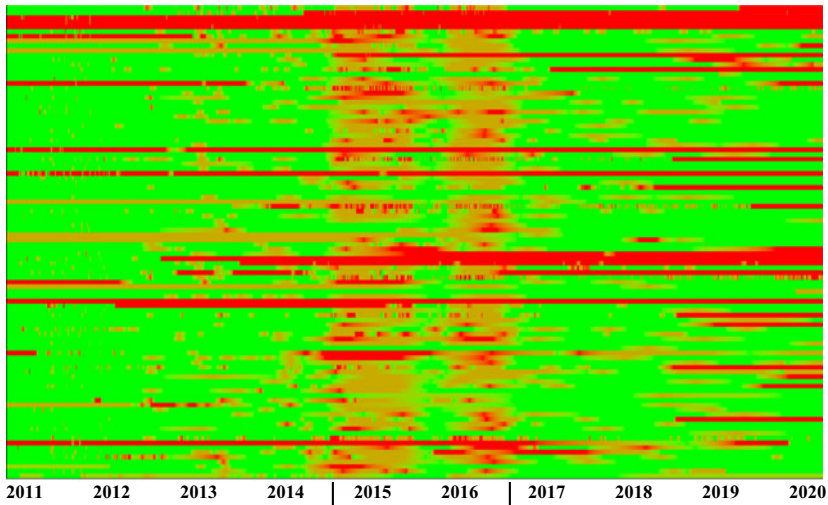
51. Hobbs and Roberts 2018.

52. Levin 2014.

53. Garber 2014; Ng 2014.

54. “What to Expect on June 4, China’s Unofficial and Orwellian ‘Internet Maintenance Day’,” *Tech in Asia*, 3 June 2013. China’s official response to questions about its internet control is that the Chinese internet is “free” and “open” but “manage[d]” (Consulate-General of the People’s Republic of China in Vancouver, “Foreign Ministry Spokesperson Hong Lei’s Regular Press Conference on April 16 2015,” available at <[http://vancouver.china-consulate.gov.cn/eng/fyrth/201504/t20150416\\_4904630.htm](http://vancouver.china-consulate.gov.cn/eng/fyrth/201504/t20150416_4904630.htm)>).

55. The actual impact likely lasted longer, given the chilling effect that often follows the initial shock. Huang 2015.



*Notes:* Each horizontal band corresponds to one of the 100 most-visited websites. Green (light) indicates full accessibility for a given day; red (dark), zero accessibility; and intermediate shades, partial accessibility. The data are from GreatFire.

**FIGURE 3.** *Chinese internet control, 2011–2020*

With internet control as the treatment, I now explain my coding of the two “treatment uptake” variables, which measure how much actors rely on the internet for their productive and innovative activities. These variables measure (1) how much firms in each sector depend on data as an input factor, or “data-intensity”; and (2) how much researchers in each academic discipline depend on knowledge access for research, or “knowledge-intensity.”

#### *Sector-Level Data-Intensity*

There are currently few systematic measurements of sector-level data-intensity. Measurements are either unavailable for Chinese firms, based on outdated data, or too coarse to capture variation across different digital-technology sectors.<sup>56</sup> To address these challenges, I develop two original measures of data intensity tailored for examining the impact of internet disruption on firms across sectors. First, I identify technology classes that contain data-intensive subclasses based on the inclusion

56. These include the US International Trade Commission’s identification of digitally intensive industries (“Digital Trade in the US and Global Economies, Part 2 (Investigation No. 332-540)”), which cannot be replicated for most Chinese firms due to lack of data; and the European Centre for International Political Economy’s measurement, which uses the 2012 BEA classification, where many distinct digital sectors are under the same NAICS code. Bauer, Ferracane, and Marel 2016.

of the keyword “data” in the US Patent and Trademark Office’s patent class list.<sup>57</sup> I then identify patents in the Office’s database that meet this criterion, and the corresponding US and Chinese assignee firms.<sup>58</sup> For each firm, I calculate the percentage of its patents that are data related. This continuous variable of data intensity is subsequently dichotomized and matched with all Chinese firms by NAICS code.

For a second measurement, I assign 1 to sectors that have the word “internet” in their NAICS definition, and 0 otherwise.<sup>59</sup> This assignment is then matched with all Chinese firms by NAICS code. Both of my data-intensity measurements reflect current variation in factor intensity across sectors and discriminate at the five- or six-digit NAICS code level. The second measurement more specifically captures internet-related data intensity. Tables A1 and A2 in the online supplement list the data-intensive sectors identified by these two measurements.

### *Discipline-Level Knowledge-Intensity*

To determine the degree to which researchers in a given discipline rely on the internet, I measure their dependency on the literature to generate research output. This can be proxied by the density of references cited. In bibliometrics, reference density has been quantified using measures such as references per article and references per page to study citation patterns.<sup>60</sup> Of the two, references per page is more suitable for our purposes as it accounts for article length, which varies greatly across disciplines.

I use data from the Web of Science to compile references per page for all disciplines in my sample.<sup>61</sup> To reduce noise in my measurement, I sample the 1 percent most-cited single-discipline research articles in each discipline.<sup>62</sup> For each discipline, I divide the total number of references by the total number of pages. Figure A1 in the online supplement visually summarizes this variable.

### *Dependent Variables and Covariates*

To measure the impact of internet control on firm performance, I use quarterly firm-level revenue data from Compustat Global for Chinese and US listed firms from 2000

57. USPTO, “Classes Arranged by Art Unit,” available at <<https://www.uspto.gov/sites/default/files/documents/caau.pdf>>.

58. Since the period of interest runs from 2011 to 2019, to identify these firms I use data from September 2009, as a midpoint between 2011 and my anti-treatment date, July 2008.

59. North American Industry Classification System, available at <<https://www.census.gov/naics/?58967?yearbck=2012>>. With treatment in 2014, I choose the 2012 NAICS definitions over the 2017 version to minimize post-treatment bias.

60. Halevi 2013; Marx and Bornmann 2015.

61. Web of Science, “Research Areas (Categories/Classification),” available at <[https://images.webofknowledge.com/images/help/WOS/hp\\_research\\_areas\\_easca.html](https://images.webofknowledge.com/images/help/WOS/hp_research_areas_easca.html)>. Since the period of interest runs from 2011 to 2020, I use 2010 data.

62. This percentage tracks the standard given in Web of Science, “Authors / Researchers: What Is Your Impact?” available at <<https://clarivate.libguides.com/authors/impact>>.

to 2019.<sup>63</sup> As covariates, I include firm-level variables that likely correlate with outcome and for which less than a third of observations are missing. These are *total assets*, which proxies for firm size, and *total liabilities*, which proxies for leverage. The online supplement presents summary statistics for Chinese firms in 2013, just before the 2014 shock. Chinese data-intensive firms, many being young technology companies, tended to be smaller in size and leverage than the rest (Tables A3, A4, and A5).

Unlike firms, which are classified by sector, institutions routinely conduct research across disciplines; and research by one institution often involves authors from multiple countries.<sup>64</sup> I therefore measure the impact of internet control on research performance at the research-article level. I collect Web of Science data for all single-discipline research articles produced in mainland China and in the United States from 2011 to 2020.<sup>65</sup> Following earlier approaches,<sup>66</sup> I proxy the quality of an article with the number of forward citations it has received. I include covariates, such as article age, that correlate with outcome. To minimize small-sample bias, I examine only the thirty-one disciplines with at least thirty research articles published from each of the two countries in each year. Table A6 in the online supplement presents summary statistics for the Chinese sample.

## Methodology Overview

In an experimental design, one would randomly assign actors to an environment with internet control or to one without, and compare differences between the two outcomes. In reality, one does not observe the counterfactual performance of Chinese firms or researchers in the absence of internet control. My research design assumes that treatment was exogenous, particularly to the market, or  $T \perp Y(0), Y(1)$ .<sup>67</sup>

We have reasons to believe that the 2014 internet control was not imposed to help the Chinese digital technology companies. The domains of their main competitors, such as Google, Amazon, and Facebook, had either been blocked long before 2014 or were not blocked more than other domains, as my measurement indicates in the previous section. The vast majority of state support did not go to data-intensive sectors.<sup>68</sup> In fact, tension

63. This includes all Chinese firms listed on the Shanghai, Shenzhen, and Hong Kong stock exchanges and in North America, and all US firms listed in North America. Firm nationality is based on headquarters location.

64. Barabási et al. 2002; Newman 2001; Wuchty, Jones, and Uzzi 2007.

65. This is based on author addresses. Because knowledge-intensity is measured at the discipline level, I exclude interdisciplinary articles, which would require weighting each discipline within each article to calculate knowledge-intensity scores.

66. Jaffe, Trajtenberg, and Henderson 1993; Murray et al. 2016.

67. My analysis of research output excludes “politically sensitive” disciplines such as government and law due to insufficient sample sizes. This strengthens the exogeneity assumption, and any survivorship bias would underestimate internet control’s negative impact.

68. Most of China’s USD 5.24 billion of domestic subsidies in the first half of 2014 were from local governments to the steel, cement, and property sectors. Wong 2014. None except one firm, PetroChina, listed under strategic and heavyweight industries by the United States, were in my treatment-uptaking sample. Szamosszegi, Anderson, and Kyle 2009. Chinese data-intensive firms had fared no worse than



between the state and the Chinese tech giants long predates the crackdown that began in 2020, as elaborated later in the qualitative section.<sup>69</sup> Far from being cash cows kept by the government, Chinese tech giants have historically had substantial foreign ownership.<sup>70</sup> China's recent move to rein in its data-intensive sectors through "golden shares" obscures the fact that these shares were first introduced in 2013 to *reduce* the state's role in these sectors.<sup>71</sup> To probe for just what may have prompted the 2014 shock, I interviewed practitioners and industry experts with proximity to the internet policymaking process in China. My interviews suggest that the need for domestic stability has been the principal driver of internet control shocks. Crackdowns typically occur just before anticipated social unrest and major political events, such as the National People's Congress, when protests are more likely than usual.<sup>72</sup>

Yet even with exogeneity in treatment, the treatment uptake variables, data intensity and knowledge intensity, are not randomly assigned. This means that treatment assignment, which is the interaction between treatment and treatment uptake, is not random. Considering this, I leverage a series of empirical strategies to identify the marginal effect of internet control. First, I apply a matching method designed for panel data to identify the effect on Chinese data-intensive firms relative to other Chinese firms. Second, by exploiting the geographical variation in treatment exposure with a triple-difference estimator, I parse out the effect on Chinese data-intensive firms relative to their US counterparts. Third, to identify the effect on Chinese research output, I disentangle the treatment effect from the selection effect using a negative binomial model and a Poisson model with fixed effects. Fourth, I adopt two similar models for a difference-in-differences estimation of research output from China and the United States. The next two sections detail these strategies.

## Empirical Analysis of Positive Data Externality

### *Matching Strategy for Chinese Firms*

Given the quasi-experimental setting, one would match each treatment-uptaking observation with non-uptaking observations to construct the counterfactual.

other Chinese firms in the two years prior to 2014 (see supplemental Figure A4), which further undermines a protectionist motive.

69. See, for example, "Sina Shares Fall After China Strips Its Licence in Web Porn Crackdown," Reuters, 24 April 2014; "China Investigates Search Engine Baidu After Student Dies of Cancer," NPR, 3 May 2016; "China Internet Watchdog to Probe Baidu over Reports It Was Used to Promote Gambling," Reuters, 19 July 2016; "China's Three Internet Giants Being Investigated for Content that 'Endangers National Security'," CNBC, 11 August 2017.

70. As expounded in the qualitative section, foreign ownership had raised suspicion from the state when some of these firms later attempted overseas expansion. Baidu, 2014 Annual Report, available at <<https://ir.baidu.com/static-files/39c9d0ab-4694-4c28-9881-a7989eebf00a>>; US SEC, "Alibaba Group Holdings Limited," available at <<https://www.sec.gov/Archives/edgar/data/1577552/000104746916013400/a2228766z20-f.htm>>; Tencent, 2014 Annual Report, available at <<https://static.www.tencent.com/storage/uploads/2019/11/09/dc4eda2bef30e63399c475accc01824e.pdf>>. All the top shareholders of these firms are presently non-Chinese.

71. "China's New Way to Control Its Biggest Companies: Golden Shares," *Wall Street Journal*, 8 March 2023.

72. Author's interviews.



However, my panel data consist of a different bundle of data-intensive and non-data-intensive firms in each period. The limited number of available covariates also constrains our ability to directly control for potential confounders. To address these, I implement the PanelMatch method, which matches each treated observation with control observations in the same period that have an identical treatment history for up to a specified number of periods. These are refined using matching or weighting methods so that the treated and matched control observations are reasonably balanced on observed confounders. Average treatment effects are then estimated using the difference-in-differences estimator with bootstrapped standard errors.<sup>73</sup> Using PanelMatch, I match each data-intensive Chinese firm with the maximum possible number of non-data-intensive Chinese firms for ten lag periods (calendar quarters) on *total assets*, *total liabilities*, and *revenue*.<sup>74</sup> I then estimate the average treatment effect of the 2014 internet control shock on firm-level revenue for ten lead periods after treatment.

The results strongly comport with my hypothesis of a positive effect of internet control on data-intensive firms (Figure 4). Ten periods after treatment—about two to three years out—the data-intensive firms on average see a 26 percent revenue gain over their non-data-intensive counterparts. The positive effect emerges as early as three quarters after treatment, increases, and plateaus at around nine quarters. Remarkably, the fluctuation from the third to the sixth quarter closely aligns with the noticeable break in treatment in 2015–16, as shown in Figure 3.

I further investigate my hypothesis from the reverse angle. Here, I set the treatment date to July 2008, just before the 2008 Summer Olympics in Beijing. At this time, the Chinese internet underwent an exceptional, brief period of liberalization in anticipation of an influx of foreign visitors.<sup>75</sup> Numerous routinely blocked websites suddenly became accessible. The abrupt removal of the baseline level of control constitutes an “anti-treatment” that should have a *negative* effect on domestic data-intensive firms, and this is indeed the case (see supplemental Figure A2). For a few quarters after the relaxation of internet control, Chinese data-intensive firms saw a decline in revenue relative to other Chinese firms. The brevity of this effect is consistent with the restoration of control right after the Olympics.<sup>76</sup>

### *Robustness Checks and Placebo Tests*

I perform a number of robustness checks, with results presented in supplemental Figure A3. First, I reduce the maximum number of matched observations to twenty and rerun the estimation. Second, I refine my matched set with a variety of

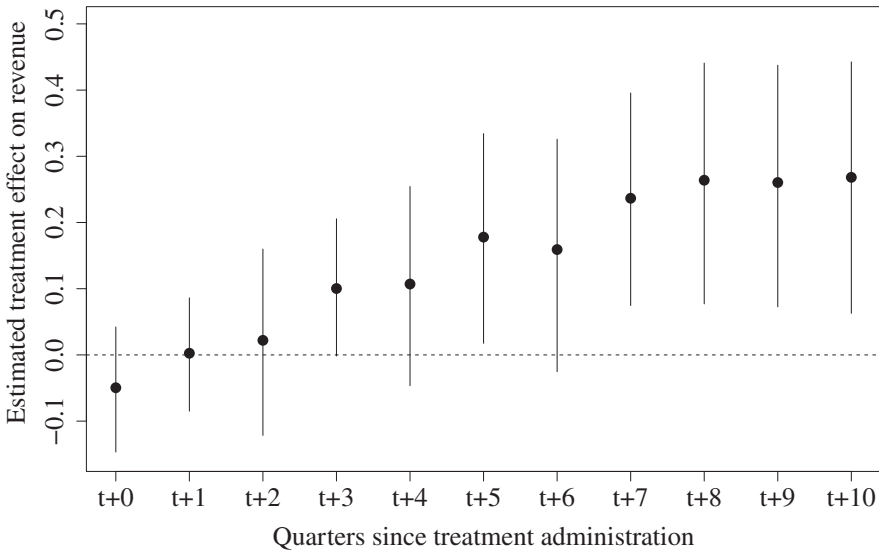
73. Imai, Kim, and Wang 2021.

74. I set the maximum number of matches to 4,003, which is the maximum number of unique non-data-intensive firms in any pretreatment period. This effectively removes the upper limit on the number of matches.

75. Branigan 2008.

76. I do not include this in my main analysis because I rely on media reports for the timing of the anti-treatment, which predates the GreatFire data used to code my treatment variable.

matching and weighting methods, which helps ensure that the result is not driven by any particular method. These estimations return similar results. Third, I include only firms with above-median data-intensity scores in my treatment-uptaking sample to see whether the result is driven by certain stratum of firms. In fact, the effect doubles, to over 50 percent revenue gain for the most data-intensive firms. Among them are those specializing in such products as web search portals (Table A1), as my theory posits. Fourth, I try the alternative data-intensity measurement that uses NAICS keywords, which yields statistically weaker but substantively comparable estimates.



**FIGURE 4.** *Estimated average treatment effect of 2014 internet control shock on Chinese firm-level revenue for ten leads after treatment, with maximum number of observations matched for ten lags before treatment using Mahalanobis distance matching*

Finally, I address concerns with pretreatment trends and spurious treatment effects. Given that the matching strategy relies on the parallel-trend assumption, I conduct a placebo test for two years before treatment. There are no statistically significant differences between the treatment-uptaking and non-uptaking firms throughout this period (Figure A4). For an additional placebo test, I draw a sample of non-data-intensive firms equal in number to the data-intensive firms used in the main analysis, match them with other non-data-intensive firms, and rerun the estimation. I repeat this process for thirty iterations and plot the averaged point estimates with bootstrapped standard errors. As expected, one does not see any treatment effect, which is essentially the difference between two non-uptaking samples (Figure A5).

### *Triple-Difference Estimator for Chinese and US Firms*

My first corollary hypothesis concerns the impact of internet control on Chinese data-intensive firms relative to their foreign counterparts. The US Trade Representative (USTR), for one, views China's internet control as a form of digital protectionism that has cost "billions of dollars in potential US business."<sup>77</sup> Since the internet control shock occurred only in China and not in the United States, I exploit the geographical variation in treatment exposure with a triple-difference estimator.<sup>78</sup> The revenue of firm  $i$  at time  $t$  is given by

$$y_{it} = \alpha_t + \beta_1(D_i \times T_t \times C_i) + \beta_2(D_i \times T_t) + \beta_3(T_t \times C_i) + \beta_4(D_i \times C_i) + \beta_5 D_i + \beta_6 T_t + \beta_7 C_i + \beta_8 Z_{it} + \epsilon_{it} \quad (2)$$

The dummy variable,  $D_i$ , denotes being a data-intensive firm;  $T_t$  denotes being in a treated period; and  $C_i$  denotes being a Chinese firm. This design exploits three sources of variation to account for country-specific confounders, selection into data-intensive sectors by firms in either country, and trends in data-intensive sectors that affect both countries. I add year fixed effects,  $\alpha_t$ , to address time-varying unobserved confounders. In  $Z_{it}$ , I include two salient time-varying firm-level controls, firm size and leverage. Because I hypothesize that internet control in China benefits Chinese data-intensive firms relative to other Chinese firms but not US data-intensive firms relative to other US firms, I expect the coefficient of the triple interaction term,  $\beta_1$ , to be positive and significant. Supplemental Tables A7 and A8 present results for the naive and saturated models. For each model, I use the full sample, the above-median data-intensive sample, and the full sample with the alternative data-intensity measurement. Standard errors are clustered at the sector level, where treatment assignment occurred.

We see that none of the naive estimates are significant, whereas those from the saturated models are significant but *counter* to the expectation. Based on these, one cannot reject the null for Corollary Hypothesis 1. The US data-intensive firms, including many so-called Big Tech firms, appear to have more than offset any data advantage for the Chinese data-intensive firms. A boost in data as an input factor is but one source of revenue growth. That the US data-intensive firms have outperformed their Chinese counterparts despite internet control hints at countervailing forces.

My theory points to one such force: the negative knowledge externality that co-occurs with the positive data externality. In hampering knowledge production, it

77. US Trade Representative 2021; Office of the US Trade Representative, "Fact Sheet on the 2020 National Trade Estimate: Strong, Binding Rules to Advance Digital Trade," available at <<https://ustr.gov/about-us/policy-offices/press-office/fact-sheets/2020/march/fact-sheet-2020-national-trade-estimate-strong-binding-rules-advance-digital-trade>>.

78. The estimator takes the difference between two difference-in-differences, namely the difference between Chinese data-intensive and Chinese non-data-intensive firms before and after treatment, and that between US data-intensive and US non-data-intensive firms before and after treatment.

ultimately undercuts growth for all actors in the economy regardless of the input factor. I now turn to the second set of hypotheses on internet control's detriment to innovation.

## Empirical Analysis of Negative Knowledge Externality

### *Negative Binomial Estimator for Chinese Research Output*

I investigate the impact of internet control on Chinese research output by way of a modified difference-in-differences design. Because citation count data often exhibits high skewness and overdispersion,<sup>79</sup> I adopt a negative binomial model to estimate the 2014 internet control shock's marginal effect on Chinese article-level forward citations:

$$\log(E[Y_i|K_i, T_i]) = \alpha_i + \beta_1(K_i \times T_i) + \beta_2 K_i + \beta_3 T_i + \beta_4 A_i + \beta_5 N_i + \epsilon_i \quad (3)$$

$K_i$  denotes the knowledge intensity of the discipline, and  $T_i$  denotes having been published in a treated period.<sup>80</sup> By exploiting variation in knowledge intensity across disciplines, the model accounts for discipline-specific trends. Because the time dimension collapses in the cross-sectional data set, I control for article age,  $A_i$ , which correlates strongly with citation counts.<sup>81</sup> I also control for number of co-authors,  $N_i$ , which correlates positively with citations.<sup>82</sup> Journal fixed effects,  $\alpha_i$ , are added to all models. Supplemental Figure A6 and Table A9 attest to parallel trends in citations between knowledge-intensive and non-knowledge-intensive disciplines prior to 2014.

Given my hypothesis that internet control engenders a greater decline in research quality for more knowledge-intensive disciplines, I expect the coefficient of the interaction term,  $\beta_1$ , to be negative and significant. Table 1 presents the main results, with incidence-rate ratios in square brackets.<sup>83</sup> Standard errors are clustered at the discipline level, where treatment assignment occurred.

### *Main Results and Robustness Checks*

Across all four models, the coefficients of interest are not only statistically significant (one at  $p < 0.01$ , two at  $p < 0.05$ ) but also substantively large. Models 1 and 2 employ the original, continuous variable of knowledge intensity. Model 2 focuses on the 50 percent most-cited articles published in a given discipline in a given year, which

79. Hausman, Hall, and Griliches 1984; Murray and Stern 2007.

80. I code only years since 2015 as treated. This builds in a seven-month lag after the June 2014 shock, as the effect on research would not be immediate.

81. Furman and Stern 2011.

82. Beaver 2004; Bornmann and Daniel 2008; Freeman and Huang 2015.

83. An incidence-rate ratio of 1 indicates no effect; 1.1 indicates 10 percent more likely; 0.9 indicates 10 percent less likely; and so on.

reduces noise by excluding low-quality articles. The incidence-rate ratios suggest that, on average, internet control in China is associated with a close to 10 percent marginal reduction in research quality, conditional on the knowledge intensity of a discipline.

TABLE 1. *Negative binomial estimates for effect of internet control on research quality*

Dependent variable: Forward citations				
	(1) Cont./full	(2) Cont./top 50%	(3) Dummy 1	(4) Dummy 2
KNOWLEDGE × SHOCK	<b>−0.1015***</b> (0.0376) [0.9035]	<b>−0.0983*</b> (0.0530) [0.9064]	<b>−0.1153**</b> (0.0540) [0.8911]	<b>−0.0945**</b> (0.0417) [0.9098]
KNOWLEDGE	0.8290 (0.8708) [2.2910]	1.875*** (0.4616) [6.5236]	0.2665 (0.4897) [1.3053]	1.694*** (0.0295) [5.4406]
SHOCK	0.7243*** (0.0640) [2.0632]	0.5860*** (0.0763) [1.7968]	0.6306*** (0.0352) [1.8787]	0.5953*** (0.0178) [1.8137]
AGE	0.3004*** (0.0093) [1.3504]	0.2541*** (0.0073) [1.2894]	0.3004*** (0.0093) [1.3504]	0.2875*** (0.0127) [1.3332]
AUTHOR COUNT	0.0204*** (0.0032) [1.0206]	0.0097*** (0.0019) [1.0098]	0.0203*** (0.0032) [1.0205]	0.0213* (0.0129) [1.0215]
Journal FE	Yes	Yes	Yes	Yes
<i>Fit statistics</i>				
Observations	984,993	455,568	984,993	299,811
Squared Correlation	0.06317	0.07946	0.06542	0.29108
Pseudo-R <sup>2</sup>	0.10425	0.09574	0.10428	0.10834
BIC	6,131,291.1	3,461,894.9	6,131,062.0	1,893,563.0
Over-dispersion	1.1943	2.6433	1.1946	1.1413

Note: Clustered (discipline-level) standard errors in parentheses. \**p* < .10; \*\**p* < .05; \*\*\**p* < .01.

I then dichotomize the knowledge intensity variable. In model 3, I assign 1 to disciplines of median knowledge intensity or higher, and 0 otherwise. In model 4, I assign 1 to disciplines of knowledge intensity at least one standard deviation above the mean, and 0 to those at least one standard deviation below the mean. The results largely remain, and the controls for article age and number of co-authors behave as expected across all models.

For an additional robustness check, I repeat the preceding analyses using a Poisson model given only moderate overdispersion in the data.<sup>84</sup> The estimates are even

84. Compared to the negative binomial estimator, the Poisson estimator generally makes less restrictive assumptions about the data-generating process, at the cost of some efficiency. Dupuy 2018; Wooldridge 2010.

greater in significance (three at  $p < 0.01$ , one at  $p < 0.05$ ) and larger in magnitude (Table A10). Based on models 1, 3, and 4, internet control in China is associated with about a 15 percent marginal reduction in research quality, conditional on knowledge-intensity.

#### *Difference-in-Differences Estimator for Chinese and US Research Output*

To examine the impact of internet control on domestic researchers vis-à-vis their foreign counterparts, I again exploit the geographical variation in treatment exposure between China and the US with a difference-in-differences estimator:

$$\log(E[Y_i|T_i, C_i]) = \alpha_{1i} + \alpha_{2i} + \beta_1(T_i \times C_i) + \beta_2T_i + \beta_3C_i + \beta_4A_i + \beta_5N_i + \epsilon_i \quad (4)$$

The dummy  $C_i$  denotes being produced by author(s) in China. I likewise add controls for article age and number of co-authors, and both journal fixed effects and discipline fixed effects,  $\alpha_{2i}$ . Because I hypothesize that internet control hurts Chinese researchers in *any* discipline relative to their US counterparts, I expect the coefficient of the interaction term,  $\beta_1$ , to be negative and significant. Table 2 presents the results for both the negative binomial and Poisson models, with standard errors clustered at the discipline level.

The estimates, significant at  $p < 0.01$  in both models, suggest that internet control has reduced the quality of research by Chinese researchers by more than 22 percent compared to their US counterparts, irrespective of the discipline. While China has caught up with the United States in *aggregate* research quality,<sup>85</sup> such metrics mask the damage from internet control at the *margin*: China would be still more innovative without such controls, even markedly so. This also helps elucidate how the “dictator’s dilemma” exaggerates the autocrat’s concern about internet control’s harm to innovation. Even if sizable, such harm might only manifest when interacted with knowledge intensity or after accounting for confounders.

Based on the foregoing, we can confidently reject the null for both H2 and Corollary Hypothesis 2. In obstructing the flow of knowledge, internet control most acutely hurts domestic knowledge-intensive researchers. But no matter the knowledge domain, it hurts *all* domestic researchers. To the extent that innovation hinges on knowledge creation, internet control inhibits growth regardless of the mix of domains or sectors the state may seek to strategically foster.

### **Evidence for State Strategy**

I conclude this theoretical proposal by presenting preliminary evidence for its implications for state strategy: following internet control, China clamped down on domestic data-intensive sectors that had benefited from the positive data externality. It did so through a

85. Brainard and Normile 2022.

**TABLE 2.** *Difference-in-differences estimates for effect of internet control on research quality (China versus US)*

Model:	Dependent variable: Forward citations	
	(1) Negative binomial	(2) Poisson
CHINA × SHOCK	<b>-0.2497***</b> <b>(0.0249)</b> <b>[0.7790]</b>	<b>-0.2496***</b> <b>(0.0381)</b> <b>[0.7791]</b>
CHINA	-0.2843*** (0.0664) [0.7526]	-0.1171 (0.0995) [0.8895]
SHOCK	0.4939*** (0.0469) [1.6388]	0.2657*** (0.0378) [1.3044]
AGE	0.2292*** (0.0076) [1.2576]	0.1688*** (0.0093) [1.1839]
AUTHOR COUNT	0.0301*** (0.0023) [1.0306]	0.0064*** (0.0011) [1.0064]
Journal FE	Yes	Yes
Discipline FE	Yes	Yes
<i>Fit statistics</i>		
Observations	1,851,164	1,851,164
Squared correlation	$4.22 \times 10^{-6}$	0.06465
Pseudo-R <sup>2</sup>	0.10499	0.48003
BIC	13,151,336.1	49,860,497.5
Over-dispersion	1.1134	

Note: Clustered (discipline-level) standard errors in parentheses. \*\*\* $p < .01$ .

combination of broad-based legislation on data sovereignty and targeted campaigns aimed at curbing individual firms' overseas expansion. Concurrently, the state sought to diffuse discontent from the negative knowledge externality by limiting the scope of internet control generally and by allowing privileged internet access for certain foreign knowledge-intensive actors specifically. These findings underscore the disproportionate influence of short-term interests and foreign actors on the autocrat's decisions.

### *Reasserting Data Sovereignty: Legislation and Crackdown*

In late 2016, shortly after the apparent abatement of internet control (Figure 3), China enacted its Cybersecurity Law.<sup>86</sup> Ambitious in scope but ambiguous in terminology,

86. State Council, Cybersecurity Law, 7 November 2016, available at <[https://www.gov.cn/xinwen/2016-11/07/content\\_5129723.htm](https://www.gov.cn/xinwen/2016-11/07/content_5129723.htm)>.

it set the tone for a succession of laws that would cover all aspects of data sovereignty. These include the National Intelligence Law,<sup>87</sup> the Data Security Law,<sup>88</sup> and the Personal Information Protection Law.<sup>89</sup> Persisting across these legislative efforts is the reassertion of the state's absolute authority over domestic data through localization and handover mandates,<sup>90</sup> and notably through tighter prohibition of access to such data by foreign entities, government or private.<sup>91</sup> The vague definitions grant the state vast discretion in determining the liability of domestic firms and in levying punishment.<sup>92</sup>

However extensive, broad-based legislation could accomplish only part of the state's objective. It could not prevent profit-maximizing firms from seeking opportunities abroad and weakening the state's oversight of their data in doing so.<sup>93</sup> Vague provisions lose potency when challenged by conflicting but better-codified stipulations from another jurisdiction. What became known as China's crackdown on tech was part and parcel of the state's attempt to address this residual concern.<sup>94</sup> State authorities cited anticompetitive behavior, privacy violations, and data security malpractices as bases for the suspension of Ant Group's initial public offering,<sup>95</sup> the investigation leading to DiDi's delisting from the NYSE,<sup>96</sup> and the probe into BOSS Zhipin following its parent company's NASDAQ listing.<sup>97</sup> Beneath these decisions, however, throbbed a pulsating fear of "disorderly capital expansion"—code-speak for when a firm has amassed enough financial clout to pose a political threat to the regime.<sup>98</sup>

Even so, it was the Cyberspace Administration of China, not agencies that oversee offshore listing such as the China Securities Regulatory Commission, that did much

87. National People's Congress, National Intelligence Law, 27 June 2017, available at <[http://www.npc.gov.cn/zgrdw/npc/xinwen/2017-06/27/content\\_2024529.htm](http://www.npc.gov.cn/zgrdw/npc/xinwen/2017-06/27/content_2024529.htm)>.

88. State Council, Data Security Law, 11 June 2021, available at <[https://www.gov.cn/xinwen/2021-06/11/content\\_5616919.htm](https://www.gov.cn/xinwen/2021-06/11/content_5616919.htm)>.

89. State Council, Personal Information Protection Law, 20 August 2021, available at <[https://www.gov.cn/xinwen/2021-08/20/content\\_5632486.htm](https://www.gov.cn/xinwen/2021-08/20/content_5632486.htm)>.

90. For example, Arts. 28, 37, and 50, Cybersecurity Law; Art. 14, National Intelligence Law; Art. 53, Data Security Law; Arts. 36 and 41, Personal Information Protection Law.

91. For example, Arts. 66 and 75, Cybersecurity Law; Art. 36, Data Security Law; Ch. III, Personal Information Protection Law.

92. Maranto 2020; Sacks 2018; Wagner 2017.

93. Auditing requirements, for instance, played a role in China's decision to obstruct its firms' listings in the United States. "China Steps Up Supervision of Overseas-Listed Firms After Didi IPO Drama," Reuters, 6 July 2021.

94. "Xi Jinping's Assault on Tech Will Change China's Trajectory," *The Economist*, 14 August 2021.

95. Feng 2020. Under the substantially foreign-owned Alibaba, Ant Group owns China's largest third-party digital payment platform, Alipay.

96. DiDi is China's largest ride-hailing company and was pivotal in Uber's exit from China. "Uber Looking to Sell Didi, China Market Has Little Transparency, CEO Says," Reuters, 14 December 2021.

97. BOSSZhipin is a large online recruitment platform in China under Kanzhun Ltd. "After Cracking Down on Didi, China Probes Other US-Listed Tech Giants," CNN, 5 July 2021.

98. "China to Strengthen Anti-monopoly Push, Prevent Disorderly Capital Expansion," *Xinhua*, 5 March 2021.



of the disciplining.<sup>99</sup> This hints that data, not just capital, was at stake. With their multitude of domestic data vulnerable to exploitation by foreign actors, these firms, already viewed as a threat from within, now also pose a risk to the regime from without.<sup>100</sup> The apprehension may not be misplaced. The handover of audit working papers, for example, could result in the retention of raw user data and communications between Chinese companies and government agencies for US regulatory inspection for three consecutive years.<sup>101</sup> Even if handover were not mandatory for compliance, it might still pose too great a risk if the data itself were of a particular kind. DiDi, as one of a handpicked group of Chinese entities licensed for detailed surveying and mapping, would present just this type of risk if foreign actors were able to access the company's coveted real-time location data, including data on Chinese defense zones.<sup>102</sup>

The high-flying Chinese data-intensive firms were not simply getting their wings clipped by conflicting compliance requirements. They were being pressed against their primal drive for profit by the regime's insistence on "equal importance to internal and external security."<sup>103</sup> Engorged with a frightful mix of capital and data, even the faintest crack of disobedience could invite crushing force from the state's iron fist.<sup>104</sup> The crackdown cost the Chinese firms trillions and eroded their once-enviable position on a par with their US counterparts.<sup>105</sup> Since then, trade complaints about the Great Firewall and allegations of US Big Tech's "jealousy" of their Chinese rivals have quietly given way to other stressors in bilateral relations.<sup>106</sup> The backlash reset whatever advantage the Chinese firms had won from internet control.<sup>107</sup> The self-same profit motive has sent the Chinese tech giants and the US Big Tech down divergent paths.

### *Minimizing Collateral Damage: AI-Powered Censorship and Selective Accommodation of Foreign Actors*

Due to the inefficiency of directly compensating domestic knowledge-intensive actors for the negative knowledge externality, as previously described, the state will first limit the scope of internet control so long as it does not hinder maintaining

99. R. Lester et al., "China Tightens Control over Overseas Securities Listings in Name of Data Security," *WilmerHale*, 9 July 2021.

100. "What Comes Next as China's Tech Crackdown Winds Down," *Washington Post*, 24 July 2023.

101. "Didi Says It Will Proceed with Delisting from NYSE," *Wall Street Journal*, 23 May 2022. With the aforementioned legislation in China, such requirements would render it practically impossible for Chinese firms to be in compliance in both jurisdictions.

102. "In the New China, Didi's Data Becomes a Problem," *Wall Street Journal*, 18 July 2021.

103. J. Xi, "A Holistic View of National Security," *Qiushi*, 15 April 2014.

104. "Jack Ma Setback Reminds Investors That Beijing Is Still Boss," *Financial Times*, 3 November 2020; "What an Ancient Poem Says About China's Fearful Tech Tycoons," CNN, 12 May 2021.

105. "A Timeline of China's 32-Month Big Tech Crackdown that Killed the World's Largest IPO and Wiped Out Trillions in Value," *South China Morning Post*, 15 July 2023.

106. Yuan 2019.

107. "Instant View: China Halts Ant Group's Mega IPO," Reuters, 3 November 2020.

domestic stability. Figure 3 illustrates such an attempt. Since 2017, across-the-board internet control has eased appreciably. The government has explored tailored measures that target, for example, sensitive segments of a domain while keeping the rest accessible.<sup>108</sup> AI has further fine-tuned censorship, with natural language processing and image recognition now widely embedded in China's popular mobile apps, such as WeChat.<sup>109</sup> Increasingly sophisticated censorship algorithms have driven down both false negatives and false positives.<sup>110</sup> In reducing false negatives, AI detects more anti-regime content faster.<sup>111</sup> In reducing false positives, AI allows through more innocuous content, minimizing the negative knowledge externality without compromising control.

While knowledge-intensive actors in general hold little power over the state, one notable exception is the foreign knowledge-intensive actors in the state. More precisely, they are those with whom the state has entered into various forms of contracts that require the state to ensure them freedom of information access in exchange for their provision of resources. Faced with similar hindrances as their domestic counterparts, these foreign actors have the option to retaliate by imposing an immediate economic cost on the regime. They may do so by invoking provisions for such access in the contract or by withholding the resources. For either to work, however, the threatened cost must be high.

The Sino-Foreign Cooperative University Union is one framework that imparts such de jure leverage to its member institutions, the "joint-venture universities."<sup>112</sup> In principle, these institutions are not subject to the same restrictions on information access as their Chinese counterparts. For US accreditation, the Chinese government must demonstrate that the student experience at these institutions is on a par with that in the United States.<sup>113</sup> In practice, experiences vary. At New York University Shanghai, web domains blocked elsewhere in China are generally accessible via the institution's network. However, at another such institution, Duke Kunshan University, the network follows a different protocol, blocking some domains that are accessible at NYU Shanghai.<sup>114</sup>

108. This was the case with Google Cloud (author's interview). Discontent from those affected by the blocking of websites such as GitHub is one reason for these measures ("Programmers Angry over Blocking of GitHub Code-Sharing Site," *South China Morning Post*, 24 January 2013). Overall, such instances are rare.

109. O'Neill 2019.

110. Author's interview.

111. Knockel et al. 2020.

112. "Secretariat of Sino-Foreign Cooperative University Union," Chinese University of Hong Kong, Shenzhen, available at <<https://tencentlab.cuhk.edu.cn/en/node/1574>>.

113. Q. Yin, "Even as Tensions Grow, US-China Joint Venture Universities Have Room to Develop," Center for Strategic and International Studies, 6 September 2023, available at <<https://www.csis.org/blogs/new-perspectives-asia/even-tensions-grow-us-china-joint-venture-universities-have-room>>.

114. Author's interviews. A 2016 report similarly finds disparity in internet access among US universities operating in China: US Government Accountability Office, "US Universities in China Emphasize Academic Freedom but Face Internet Censorship and Other Challenges," August 2016, available at <<https://www.gao.gov/assets/gao-16-757.pdf>>.

The Schwarzman Scholars program at Tsinghua University represents a different kind of leverage. At over USD 575 million, the program is the “single largest philanthropic effort in China’s history.”<sup>115</sup> An endowment this size enables the founder, Stephen A. Schwarzman, to act as the de facto guarantor of freedom.<sup>116</sup> When asked whether he would “keep things very free” and maintain “total academic freedom” at his college, Schwarzman said, “Yes. Absolutely ... And we’ve made that clear to our friends at Tsinghua and they agree completely.”<sup>117</sup>

Indeed, at the Schwarzman College, virtual private networks are embedded in the network for credentialed users, which affords them a browsing experience similar to that in the United States—unlike their “friends at Tsinghua.” Other students at Tsinghua do not enjoy institution-sponsored unrestricted internet access, nor do those at other elite institutions such as Peking University.<sup>118</sup> Rather than the elevated status or exceptional productivity of the institutions, it is the leverage held by the foreign actors that motivates the state to make accommodations in this peculiarly discriminating manner.

## Concluding Remarks

In this paper I begin with the three distinct components of information: ideas, data, and knowledge. Internet control intended to restrict ideas generates a positive externality for domestic data-intensive sectors and a negative externality for domestic knowledge-intensive sectors. Quantitative analysis of the case of China strongly supports both hypothesized externalities. I then postulate that the positive data externality impedes the state’s competing objective of data sovereignty when domestic data-intensive firms expand overseas. Meanwhile, the state shields certain foreign knowledge-intensive actors from the negative knowledge externality to avoid the immediate costs they might otherwise impose. Qualitative evidence comports with these implications in accentuating the double challenge posed by internet control’s dual externalities.

Many theoretical and empirical extensions can be made, of which I highlight three. First, just as the USTR has accused China of digital protectionism, China has protested the US and the EU sanctions of its firms, such as Huawei, and in some cases threatened retaliation.<sup>119</sup> A fuller assessment of the trade repercussions of

115. Blackstone, “Stephen A. Schwarzman,” available at <<https://www.blackstone.com/people/stephen-a-schwarzman-2/>>.

116. Financial leverage aside, some note the tenuity of such partnerships, which lack the institutional ties to the United States that joint-venture universities embody. B. Allen-Ebrahimian, “The Moral Hazard of Dealing with China,” *The Atlantic*, 11 January 2020.

117. “A Rhodes-Like Scholarship for Study in China,” NPR, 2 May 2013.

118. Author’s interviews.

119. See, for example, “China Asks United States to Stop ‘Unreasonable Suppression’ of Huawei,” Reuters, 16 May 2020; “China Slams EU Ban on Huawei, ZTE Demands Equal Treatment,” Reuters, 16 June 2023.

internet control in a cross-border setting should take into account retaliatory acts and any boomerang effect beyond the initial impact.<sup>120</sup>

Second, a closer look into the negative knowledge externality warrants an investigation into its mechanisms. One hypothesis is that internet control reduces research quality by limiting domestic researchers' exposure to frontier knowledge from the outside world. Text-similarity measures have been used to track idea diffusion, including in scientific innovation.<sup>121</sup> Such methodologies can be applied to test this hypothesis by comparing research from China with that from the rest of the world, where one would expect less similarity between them following internet control.

Third, as internet connectivity continues to rise and indigenous digital products proliferate in the Global South, more states—both autocratic and democratic—will meet the scope conditions of my theory and provide fertile testing ground. It would be worthwhile to explore how information externalities manifest in democracies. The positive data externality may incentivize domestic data-intensive sectors to lobby for the state to block foreign competitors' web domains. The state may likewise be incentivized to pursue such protectionist internet control in return for support from these sectors.<sup>122</sup> Moreover, that the protectionist benefit exists as an externality facilitates the justification of these measures under such guises as national security and privacy concerns. India's increase in internet control concurrent to its stunning increase in internet connectivity typifies a scenario for formulating and testing these hypotheses in a democratic context.<sup>123</sup> My theory also supplies an additional lens for analyzing events, such as the evolving situation of TikTok in the United States, that straddle trade and national security.<sup>124</sup>

One final caveat is that advancements in generative AI may induce heavier reliance on data over knowledge in producing innovation. The positive data externality from internet control may therefore compensate for the negative knowledge externality. However, the resulting innovation may be less novel due to greater data homogeneity.<sup>125</sup> An inquiry into the emergent relationship between politics, information, and innovation in the age of generative AI will illuminate our understanding of state power and of human progress.

120. Anderson 2002; Elliott and Bayard 1994. For an illustration of this dynamic, see “Huawei Ban Timeline: Detained CFO Makes Deal with US Justice Department,” CNET, 30 September 2021.

121. Arts, Cassiman, and Gomez 2018; Dupont and Rachuj 2021.

122. This follows from Ehrlich 2007; Grossman and Helpman 1994. With digital products, consumers contribute both revenue and data to producers, as previously noted. Keener examination of this feature will inform the study of digital trade.

123. See, for example, “The Problem with India’s App Bans,” Atlantic Council, 27 March 2023, available at <<https://www.atlanticcouncil.org/blogs/southasiasource/the-problem-with-indias-app-bans/>>; “India Bans 200-Plus Chinese Mobile Apps in Boon for Paytm,” *Bloomberg*, 6 February 2023; “Amazon Users in India Will Get Less Choice and Pay More Under New Selling Rules,” *New York Times*, 30 January 2019.

124. “Why the US Is Forcing TikTok to Be Sold or Banned,” *New York Times*, 8 May 2024.

125. Bianchini, Müller, and Pelletier 2020; Doshi and Hauser 2023; Yang and Roberts 2023.

## Data Availability Statement

Replication files for this article may be found at <<https://doi.org/10.7910/DVN/OX6G1A>>.

## Supplementary Material

Supplementary material for this article is available at <<https://doi.org/10.1017/S0020818324000237>>.

## References

- Aaronson, Susan Ariel. 2019. What Are We Talking About When We Talk About Digital Protectionism? *World Trade Review* 18 (4):541–77.
- Ables, Kelsey. 2018. China's Rising Tax on Information: The Amount of Economic and Educational Privilege Needed to Jump the Great Firewall Keeps Increasing. *The Diplomat*, 27 February. Available at <<https://thediplomat.com/2018/02/chinas-rising-tax-on-information/>>.
- Ackoff, Russell L. 1989. From Data to Wisdom. *Journal of Applied Systems Analysis* 16 (1):3–9.
- Anderson, Kym. 2002. Peculiarities of Retaliation in WTO Dispute Settlement. *World Trade Review* 1 (2): 123–34.
- Arts, Sam, Bruno Cassiman, and Juan Carlos Gomez. 2018. Text Matching to Measure Patent Similarity. *Strategic Management Journal* 39 (1):62–84.
- Bao, Beibei. 2013. How Internet Censorship is Curbing Innovation in China. *The Atlantic*, 22 April.
- Barabási, Albert-László, Hawoong Jeong, Zoltan Néda, Erzsebet Ravasz, Andras Schubert, and Tamas Vicsek. 2002. Evolution of the Social Network of Scientific Collaborations. *Physica A: Statistical Mechanics and Its Applications* 311 (3–4):590–614.
- Bauer, Matthias, Martina F. Ferracane, and Erik van der Marel. 2016. Tracing the Economic Impact of Regulations on the Free Flow of Data and Data Localization. Centre for International Governance Innovation and Chatham House. Available at <[https://www.cigionline.org/sites/default/files/gcig\\_no30web.pdf](https://www.cigionline.org/sites/default/files/gcig_no30web.pdf)>.
- Beaver, Donald deB. 2004. Does Collaborative Research Have Greater Epistemic Authority? *Scientometrics* 60:399–408.
- Beraja, Martin, Andrew Kao, David Y. Yang, and Noam Yuchtman. 2023. AI-tocracy. *Quarterly Journal of Economics* 138 (3):1349–1402.
- Beraja, Martin, David Y. Yang, and Noam Yuchtman. 2023. Data-Intensive Innovation and the State: Evidence from AI Firms in China. *Review of Economic Studies* 90 (4):1701–1723.
- Bianchini, Stefano, Moritz Müller, and Pierre Pelletier. 2020. Deep Learning in Science. ArXiv preprint 2009.01575.
- Boas, Taylor C. 2000. The Dictator's Dilemma? The Internet and US Policy Toward Cuba. *Washington Quarterly* 23 (3):57–67.
- Bornmann, Lutz, and Hans-Dieter Daniel. 2008. What Do Citation Counts Measure? A Review of Studies on Citing Behavior. *Journal of Documentation* 64 (1):45–80.
- Brainard, Jeffrey, and Dennis Normile. 2022. China Rises to First Place in One Key Metric of Research Impact. *Science* 377 (6608):799.
- Branigan, T. 2008. China Relaxes Internet Censorship for Olympics. *The Guardian*, 1 August.
- Brutger, Ryan, and Anton Strezhnev. 2022. International Investment Disputes, Media Coverage, and Backlash Against International Law. *Journal of Conflict Resolution* 66 (6):983–1009.

- Chander, Anupam, and Haochen Sun. 2022. Sovereignty 2.0. *Vanderbilt Journal of Transnational Law* 55: 283.
- Chen, Yuyu, and David Y. Yang. 2019. The Impact of Media Censorship: 1984 or Brave New World? *American Economic Review* 109 (6):2294–2332.
- Denyer, Simon. 2016. China's Scary Lesson to the World: Censoring the Internet Works. *Washington Post*, 23 May. Available at <[https://www.washingtonpost.com/world/asia\\_pacific/chinas-scary-lesson-to-the-world-censoring-the-internet-works/2016/05/23/413afe78-fff3-11e5-8bb1-f124a43f84dc\\_story.html](https://www.washingtonpost.com/world/asia_pacific/chinas-scary-lesson-to-the-world-censoring-the-internet-works/2016/05/23/413afe78-fff3-11e5-8bb1-f124a43f84dc_story.html)>.
- Diamond, Larry. 2010. Liberation Technology. *Journal of Democracy* 21 (3):69–83.
- Doshi, Anil R., and Oliver P. Hauser. 2023. Generative Artificial Intelligence Enhances Individual Creativity but Reduces the Collective Diversity of Novel Content. SSRN. Available at <<https://dx.doi.org/10.2139/ssrn.4535536>>.
- Düpont, Nils, and Martin Rachuj. 2021. The Ties That Bind: Text Similarities and Conditional Diffusion Among Parties. *British Journal of Political Science* 52 (2):1–18.
- Dupuy, Jean-François. 2018. *Statistical Methods for Overdispersed Count Data*. Elsevier.
- Economist. 2017. Data Is Giving Rise to a New Economy, 6 May.
- Economy, Elizabeth. 2018. The Great Firewall of China: Xi Jinping's Internet Shutdown. *The Guardian*, 29 June. Available at <<https://www.theguardian.com/news/2018/jun/29/the-great-firewall-of-china-xi-jinpings-internet-shutdown>>.
- Ehrlich, Sean D. 2007. Access to Protection: Domestic Institutions and Trade Policy in Democracies. *International Organization* 61 (3):571–605.
- Elliott, Kimberly Ann, and Thomas O. Bayard. 1994. *Reciprocity and Retaliation in US Trade Policy*. Peterson Institute for International Economics.
- Fallows, James. 2008. The Connection Has Been Reset: China's Great Firewall. *Atlantic Monthly*, March.
- Farboodi, Maryam, Roxana Mihet, Thomas Philippon, and Laura Veldkamp. 2019. Big Data and Firm Dynamics. *AEA Papers and Proceedings* 109:38–42.
- Farboodi, Maryam, and Laura Veldkamp. 2021. A Model of the Data Economy. Technical report. National Bureau of Economic Research.
- Farrell, Henry, and Abraham L. Newman. 2019. Weaponized Interdependence: How Global Economic Networks Shape State Coercion. *International Security* 44 (1):42–790.
- Feng, Emily. 2020. Regulators Squash Giant Ant Group IPO. *National Public Radio*, 3 November.
- Ferracane, Martina Francesca, Hosuk Lee-Makiyama, and Erik Van Der Marel. 2018. Digital Trade Restrictiveness Index. European Center for International Political Economy.
- Floridi, Luciano. 2020. The Fight for Digital Sovereignty: What It Is, and Why It Matters, Especially for the EU. *Philosophy and Technology* 33:369–78.
- Freeman, Richard B., and Wei Huang. 2015. Collaborating with People Like Me: Ethnic Coauthorship Within the United States. *Journal of Labor Economics* 33 (S1):S289–S318.
- Frické, Martin. 2019. The Knowledge Pyramid: The DIKW Hierarchy. *Knowledge Organization* 46 (1): 33–46.
- Fu, King-wa, Chung-hong Chan, and Michael Chau. 2013. Assessing Censorship on Microblogs in China: Discriminatory Keyword Analysis and the Real-Name Registration Policy. *IEEE Internet Computing* 17 (3):42–50.
- Furman, Jeffrey L., and Scott Stern. 2011. Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research. *American Economic Review* 101 (5):1933–63.
- Garber, Megan. 2014. There Are 64 Tiananmen Terms Censored on China's Internet Today: and Counting. *The Atlantic*, 4 June.
- Grossman, Gene M., and Elhanan Helpman. 1994. Protection for Sale. *American Economic Review* 84 (4): 833–50.
- Gurieiev, Sergei, Nikita Melnikov, and Ekaterina Zhuravskaya. 2021. 3G Internet and Confidence in Government. *Quarterly Journal of Economics* 136 (4):2533–2613.
- Gurieiev, Sergei, and Daniel Treisman. 2019. Informational Autocrats. *Journal of Economic Perspectives* 33 (4):100–127.
- Halevi, Gali. 2013. Citation Characteristics in the Arts and Humanities. *Research Trends* 32:23–25.

- Hausman, Jerry, Bronwyn Hall, and Zvi Griliches. 1984. Econometric Models for Count Data with an Application to the Patents-RD Relationship. *Econometrica* 52 (4):909–38.
- Hoang, Nguyen Phong, Arian Akhavan Niaki, Jakub Dalek, Jeffrey Knockel, Pellaeon Lin, Bill Marczak, Masashi Crete-Nishihata, Phillipa Gill, and Michalis Polychronakis. 2021. How Great is the Great Firewall? Measuring China's DNS Censorship. In 30th USENIX Security Symposium (USENIX Security 21), 3381–98.
- Hobbs, William R., and Margaret E. Roberts. 2018. How Sudden Censorship Can Increase Access to Information. *American Political Science Review* 112 (3):621–36.
- Huang, Haifeng. 2015. Propaganda as Signaling. *Comparative Politics* 47 (4):419–44.
- Imai, Kosuke, In Song Kim, and Erik Wang. 2021. Matching Methods for Causal Inference with Time-Series Cross-Section Data. *American Journal of Political Science* 67 (3):587–605.
- Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics* 108 (3): 577–98.
- Jones, Charles I., and Christopher Tonetti. 2020. Nonrivalry and the Economics of Data. *American Economic Review* 110 (9):2819–58.
- Jørn Nielsen, Hans, and Birger Hjørland. 2014. Curating Research Data: The Potential Roles of Libraries and Information Professionals. *Journal of Documentation* 70 (2):221–40.
- Kedzie, Christopher Robert. 1996. Communication and Democracy: Coincident Revolutions and the Emergent Dictator's Dilemma. RAND Graduate School.
- Kim, Sung Eun. 2018. Media Bias Against Foreign Firms as a Veiled Trade Barrier: Evidence from Chinese Newspapers. *American Political Science Review* 112 (4):954–70.
- King, Gary, Jennifer Pan, and Margaret E. Roberts. 2013. How Censorship in China Allows Government Criticism but Silences Collective Expression. *American Political Science Review* 107 (2):326–43.
- King, Gary, Jennifer Pan, and Margaret E. Roberts. 2014. Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation. *Science* 345 (6199).
- King, Gary, Jennifer Pan, and Margaret E. Roberts. 2017. How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument. *American Political Science Review* 111 (3):484–501.
- Knockel, Jeffrey, Christopher Parsons, Lotus Ruan, Ruohan Xiong, Jedidiah Crandall, and Ron Deibert. 2020. We Chat, They Watch: How International Users Unwittingly Build Up WeChat's Chinese Censorship Apparatus. Citizen Lab Research Report No. 127, University of Toronto. Available at <<https://tspace.library.utoronto.ca/bitstream/1807/101395/1/Report%23127--wechattheywatch-web.pdf>>.
- Levin, Dan. 2014. China Escalating Attack on Google. *New York Times*, 2 June.
- Liu, Lizhi. 2021. The Rise of Data Politics: Digital China and the World. *Studies in Comparative International Development* 56 (1):45–67.
- Maranto, Lauren. 2020. Who Benefits from China's Cybersecurity Laws? Center for Strategic and International Studies, June.
- Marx, Werner, and Lutz Bornmann. 2015. On the Causes of Subject-Specific Citation Rates in Web of Science. *Scientometrics* 102 (2):1823–27.
- Milner, Helen V. 2006. The Digital Divide: The Role of Political Institutions in Technology Diffusion. *Comparative Political Studies* 39 (2):176–99.
- Milojević, Staša. 2012. How Are Academic Age, Productivity and Collaboration Related to Citing Behavior of Researchers? *PloS One* 7 (11):e49176.
- Murray, Fiona, Philippe Aghion, Mathias Dewatripont, Julian Kolev, and Scott Stern. 2016. Of Mice and Academics: Examining the Effect of Openness on Innovation. *American Economic Journal: Economic Policy* 8 (1):212–52.
- Murray, Fiona, and Scott Stern. 2007. Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the Anti-commons Hypothesis. *Journal of Economic Behavior & Organization* 63 (4):648–87.



- Newman, Mark E.J. 2001. The Structure of Scientific Collaboration Networks. *Proceedings of the National Academy of Sciences* 98 (2):404–409.
- Ng, Jason Q. 2014. 64 Tiananmen-Related Words China Is Blocking Online Today. *Wall Street Journal*, 4 June.
- Normile, Dennis. 2017. Science Suffers as China's Internet Censors Plug Holes in Great Firewall. *Science* 357 (6354):856.
- O'Neill, Patrick Howell. 2019. How WeChat Censors Private Conversations, Automatically in Real Time. *MIT Technology Review*, 15 July.
- Olson, Mancur, Jr. 1971. *The Logic of Collective Action: Public Goods and the Theory of Groups*, with a New Preface and Appendix. Harvard University Press.
- Pan, Jennifer, and Alexandra A. Siegel. 2020. How Saudi Crackdowns Fail to Silence Online Dissent. *American Political Science Review* 114 (1):109–125.
- Roberts, Margaret E. 2018. *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton University Press.
- Roberts, Margaret E. 2020. Resilience to Online Censorship. *Annual Review of Political Science* 23: 401–419.
- Rodrik, Dani. 2018. What Do Trade Agreements Really Do? *Journal of Economic Perspectives* 32 (2): 73–90.
- Romer, Paul M. 1990. Endogenous Technological Change. *Journal of Political Economy* 98 (5, pt. 2): S71–S102.
- Rosenzweig, Paul. 2012. The International Governance Framework for Cybersecurity. *Canada-United States Law Journal* 37:405.
- Sacks, Samm. 2018. China's Emerging Data Privacy System and GDPR. Center for Strategic and International Studies.
- Saleh, Nivien. 2012. Egypt's Digital Activism and the Dictator's Dilemma: An Evaluation. *Telecommunications Policy* 36 (6):476–83.
- Sanovich, Sergey, Denis Stukal, and Joshua A. Tucker. 2018. Turning the Virtual Tables: Government Strategies for Addressing Online Opposition with an Application to Russia. *Comparative Politics* 50 (3):435–82.
- Simmons, Beth A., and Michael R. Kenwick. 2022. Border Orientation in a Globalizing World. *American Journal of Political Science* 66 (4):853–70.
- Stukal, Denis, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker. 2017. Detecting Bots on Russian Political Twitter. *Big Data* 5 (4):310–24.
- Sun, Meicen. 2019. National Borders Don't Stop in the Physical World—They're in Cyberspace Too. World Economic Forum, 16 January. Available at <<https://www.weforum.org/agenda/2019/01/virtual-borders/>>.
- Szamosszegi, Andrew, Charles Anderson, and Cole Kyle. 2009. An Assessment of China's Subsidies to Strategic and Heavyweight Industries. United States-China Economic and Security Review Commission, Washington, DC.
- US Trade Representative. 2021. 2021 National Trade Estimate Report on Foreign Trade Barriers.
- Wagner, Jack. 2017. China's Cybersecurity Law: What You Need to Know. *The Diplomat*, 1 June.
- Weymouth, Stephen. 2017. Service Firms in the Politics of US Trade Policy. *International Studies Quarterly* 61 (4):935–47.
- Weymouth, Stephen. 2023. *Digital Globalization: Politics, Policy, and a Governance Paradox*. Cambridge University Press.
- Wilson, Robert. 1975. Informational Economies of Scale. *Bell Journal of Economics* 6 (1):184–95.
- Wong, Fayen. 2014. Steel Industry on Subsidy Life-Support as China Economy Slows. Reuters, 18 September.
- Woods, Andrew Keane. 2018. Litigating Data Sovereignty. *Yale Law Journal* 128 (2):328–406.
- Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.
- Wu, Mark. 2017. Digital Trade-Related Provisions in Regional Trade Agreements: Existing Models and Lessons for the Multilateral Trade System. Geneva, Switzerland: ICTSD. Available at



- [https://www.zbw.eu/econis-archiv/bitstream/11159/1643/1/rta\\_exchange-digital\\_trade-mark\\_wu-final-1.pdf](https://www.zbw.eu/econis-archiv/bitstream/11159/1643/1/rta_exchange-digital_trade-mark_wu-final-1.pdf).
- Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. 2007. The Increasing Dominance of Teams in Production of Knowledge. *Science* 316 (5827):1036–39.
- Xu, Ke. 2019. Data Security Law: Location, Position and Institution Construction. *Business and Economics Law Review* 3:52–57.
- Xu, Xu. 2021. To Repress or to Co-opt? Authoritarian Control in the Age of Digital Surveillance. *American Journal of Political Science* 65 (2):309–325.
- Yang, Eddie, and Margaret E. Roberts. 2023. The Authoritarian Data Problem. *Journal of Democracy* 34 (4):141–50.
- Yuan, Li. 2018. A Generation Grows Up in China Without Google, Facebook or Twitter. *New York Times*, 6 August.
- Yuan, Li. 2019. Mark Zuckerberg Wants Facebook to Emulate WeChat. Can It? *New York Times*, 7 March.

## Author

**Meicen Sun** is Assistant Professor in the School of Information Sciences at the University of Illinois Urbana-Champaign, and Affiliated Faculty at MIT FutureTech. She can be reached at [mcsun@illinois.edu](mailto:mcsun@illinois.edu).

## Acknowledgments

For extensive feedback I thank Yasheng Huang, In Song Kim, Kenneth Oye, and members of the Kim Research Group. For helpful comments I thank Pablo Beramendi, Daniel Drezner, Richard Freeman, Kathleen McNamara, Abraham Newman, Elan Pavlov, Nathaniel Persily, James Prieger, Robert Reich, Tuan-Hwee Sng, Anton Sobolev, Neil Thompson, Paul Vaaler, Josephine Wolff, and meeting participants at MIT, Stanford University, Georgetown University, Carnegie Mellon University, University of California San Diego, University of Pennsylvania, TPRC, New Faces in Chinese Politics Conference, Cybersecurity Law and Policy Scholars Conference, Politics and Computational Social Science conference, National Bureau of Economic Research, International Political Economy Society, and the American Political Science Association's annual meeting. I am indebted to the editors and the anonymous reviewers for their thoughtful input.

## Funding

Research for this paper received financial support from MIT, Stanford University, Georgetown University, the Smith Richardson Foundation, and the Horowitz Foundation for Social Policy.

## Key Words

Information; data; trade; innovation; China

Date received: January 18, 2023; Date accepted: April 29, 2024