What makes a jihadist text popular?

Santiago Segarra[a], Richard Nielsen[b c], and Ali Jadbabaie[b d]
[a]Electrical and Computer Engineering, Rice University
[b]Inst. for Data, Systems, and Society, Massachusetts Institute of Technology
[c]Political Science, Massachusetts Institute of Technology
[d]Lab. for Information and Decision Systems, Massachusetts Institute of Technology

August 7, 2019

Abstract

Jihadist militants often spread their ideas to potential followers through online communications. We analyze 50 million page views of a prominent jihadist web library to identify features of these communications that drive their popularity. We find that author identity is a more accurate predictor of a documents popularity than topic, format, or position within the repository, suggesting that jihadists base their reading decisions on authority. We also find that key events in the development of the Islamic State coincide with temporary increases in the popularity of texts by Sayyid Qutb, a founding father of the jihadist movement. Surprisingly, we find no evidence that counterterrorism operations lead to a sustained martyrdom effect, as the increase in popularity of writings by targeted authors is only temporary. Consequently, fears that counterterrorism efforts will inadvertently repopularize jihadists ideas are overblown, but counterterrorism efforts do not appear to make jihadists ideas less popular either.

1 Introduction

On May 1, 2011, Usama Bin Laden was a struggling author. He was still the most famous living transnational jihadist and his ideas had shaped the jihadist movement but now his relevance was in question. Forced into hiding by a global manhunt, Bin Laden had failed to produce major new works for years. His statements and ideas leaked out in a trickle, with questionable influence on other jihadists. Twenty-four hours later, Bin Ladens face was on the front page of most major newspapers and his writings were being accessed at a furious rate online. True, he was dead – killed by a team of Navy SEALs in his secret Abbottabad compound – but his writings were again popular, at least momentarily. And what every author wants, violent jihadists included, is for their ideas to be read.

Although it is impossible to fully quantify the influence of Bin Ladens ideas, their global effects are profound. In other domains, scholars debate whether ideas matter for shaping political outcomes [1, 2, 3, 4], but the ideas preached by Bin Laden and his ilk have been so catastrophically effective that their influence is not generally in doubt. Rather, the pressing issue is how these ideas can be snuffed out, or at least quarantined to the fringes of Islamist discourse.

The content and evolution of jihadist ideology is well understood [5] and surveys give some insights about the cross-national variation in popular support for violent jihad [6]. New research shows that the availability of international funding shifts the content of jihadist messages [7]. However, less is known about why some jihadist ideas gain more popularity than others. Understanding what drives the popularity of jihadist online writing is important for any effective policy to limit the influence of jihadist ideas. The fact that counterterrorism resources are finite elicits the following fundamental question: When any new piece of writing appears, is it worth the effort to try to suppress it?

Counterterrorism strategies may also have unintended consequences. Killing and capturing jihadist leaders like Usama Bin Laden has been a key strategy of US counterterror efforts. Yet, some observers have expressed concern that counterterrorism strengthens the appeal of violent jihadism by elevating its most vociferous proponents to the status of martyrs [8, 9, 10]. Others claim that these fears are overblown, and that the killings of Bin Laden and other jihadist thinkers over the past decade have been a serious blow to Al-Qaeda, with no blowback effect [11]. Prior research suggests that drone strikes against Al Qaeda did not affect their subsequent propaganda production [12], but this research does not have data to test whether existing propaganda became more or less popular. Given these contradictory views and lack of data, researchers and counterterrorism experts are often in the dark about how to best counter jihadists' violent ideas and actions.

This paper quantitatively investigates what drives the viewership of jihadist writings posted online. We posit two logics of viewership. According to a newsworthiness logic of viewership, a document is likely to become popular because of its topical content, timeliness, and sensationalism. This is the logic of popularity that many observers seem to have in mind when discussing the threat of jihadist media, see for example [13]. In contrast, an authority logic of viewership suggests that a document is likely to become popular because of the eminence of its author, rather than its content. We believe that individuals who click on jihadist materials because of an authority logic are much more concerning than individuals following a newsworthiness logic. Someone who clicks on a document because it is newsworthy is merely trying to stay informed about the world. Someone who clicks because of authority is potentially interested in taking the ideas to heart, and in the context of jihadism, that may encourage them to carry out political violence. The average person who accesses a jihadist text because its author is mentioned in the news is not likely to become violent. By contrast, a reader who is intentionally reading jihadists for their ideas might be.

We use page-view data from a large jihadist web-library to show that authority, rather than content, is the most important factor in making a jihadist text popular. We also examine whether killing jihadist thinkers inadvertently increases the popularity of their ideas. We find that it does, but only temporarily. We compare both of these results to a very different domain where authority matters: academic scholarship in physics, engineering, and related fields. We find a remarkable correspondence between patterns of viewership on the jihadist web library and on a large online repository of academic papers. This correspondence suggests that most of the viewership of the jihadist web library follows an authority logic, and that jihadist authority is similar in some ways to academic authority. The exception is that spikes of interest following the death of jihadist writers appears to follow a newsworthiness logic.

2 Popularity in *The Pulpit of Monotheism and Jihad*

Contradictory views on what makes jihadist texts popular exist in part due to a dearth of systematic, data-centric approaches to this question. To shed light on this issue, we analyze the popularity of documents on a prominent jihadist web library, *The Pulpit of Monotheism and Jihad*, which was the premier source on the open web for Arabic-language jihadist material until its removal in 2015. Jihadists regard the website as important – Bin Laden asked about it specifically in documents recovered from his hideout in Pakistan – and scholars have called it Al Qaeda's premier electronic library [14].

The website reported page-view counts for each document in real time, providing a fine-grained measure of the popularity of jihadist texts with their intended audience. Alexa web statistics service reports visitors to *The Pulpit of Monotheism and Jihad* from throughout the Arab world, especially Egypt, Algeria, Tunisia, and Morocco[1]. We are the first to analyze this web traffic, though other scholars have investigated this website qualitatively [14, 15]. These page views are the best available measure of jihadists' popularity, and we find that they correlate with the prominence of jihadist authors as measured by citations in other canonical jihadist collections (see Supplementary Materials for more details).

We used automated web crawlers to collect the cumulative daily page views reported by the website from February 11th 2011 to December 6th 2014, adding to more than 50 million page views. Consequently, associated with each document we have a corresponding time series of cumulative views at the daily scale; see Fig. S1. These time series contain gaps due to either website malfunctions or faults in the data collection software. In Fig. S2 we report the total number of documents for which we have a cumulative view count in each day. The overall increasing trend is due to the addition of new documents to the repository throughout the period of study. By the end of 2014, the website contained 6,101 documents, ranging from long theological treatises to short fatwas, by 865 authors. Two cleaning procedures were applied to the data of cumulative views: i) imputation of missing data, and ii) outlier detection; see

[1] `https://www.alexa.com/siteinfo/tawhed.ws`, accessed 3/17/2014 and archived by the authors because the website is now offline.

Materials and Methods for more details. Fig. 1A shows the cleaned cumulative views. The distribution of average page views is skewed, as are the distributions of documents per author and views per author; see Fig. 1B-D.

In addition to the cumulative views for each document, we also collect document metadata: attributes such as author, length, or location within the repository (see the Supplementary Material for a complete list). At the beginning of our data acquisition procedure, i.e. February 11th 2011, the repository contained 5,236 documents by 776 different authors. Throughout the 1395 days in which data was collected, 879 new documents were added to the repository and 14 were removed. Some additions have a close relation to concurrent political events. For example, one of the four documents added on January 24th 2014 was the letter An Urgent Call to Our People in Syria by Ayman al-Zawahiri. This occurred exactly the day after this letter was made public by Al-Qaeda's leader. This is also a testament of how active the repository was during the data collection period.

New documents added to the repository are often advertised on the website's homepage, but only for a matter of days or weeks (Fig. S3). This website architecture causes rapid accumulation of page views immediately after documents are posted, but view counts stabilize when homepage advertising stops; see Supplementary Materials for more details. To show the difference between these uptake and steady state phases, Fig. 1E depicts the normalized cumulative views for the first 60 days of each document. More precisely, we plot the cumulative views for the first 60 days of each document as a fraction of the total views accumulated by that document in the first 60 days. If the daily views were constant, the normalized cumulative views would align diagonally (coinciding with the dashed straight line). Instead, all new documents are above the diagonal and show decreasing page-view rates as time passes. In contrast, Fig. 1F shows the same documents but 121 to 180 days after posting, with daily page-view rates that are generally stable. Figs. 1E-F point towards the fact that each document undergoes a boost of interest when added and then switches to a more stable and mature steady-state rate of daily views. We use this rate of views as a measure of the intrinsic popularity of each document.

3 Results and Discussion

3.1 Popularity prediction

What makes some documents more popular than others among jihadists? Notice that our goal is not to make claims about the popularity of general online texts but rather to pinpoint the popularity drivers within this specific target group. To begin answering this question, we build a predictive model of the number of daily page views that a newly posted document is likely to obtain. We study whether features of the text and their authors can explain documents' popularity. We consider four classes of features: source (such as author or magazine), topic (such as topic models see Materials and Methods or words in

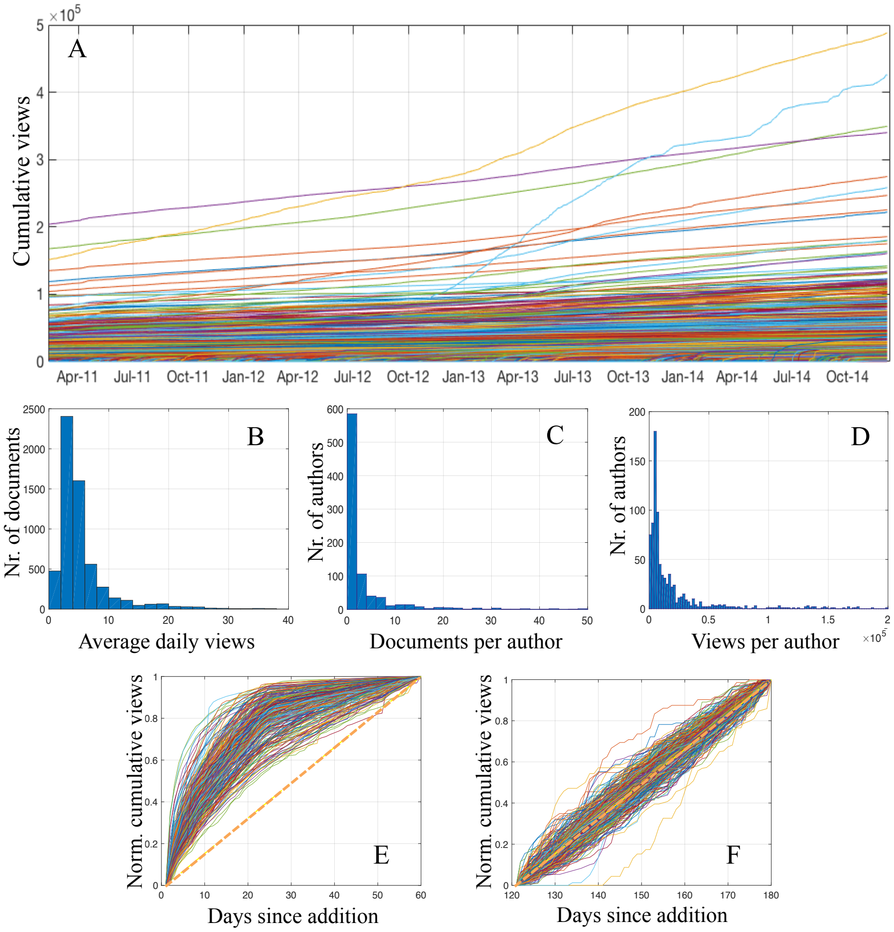

Figure 1: Viewership patterns of jihadist texts. (A) Cleaned cumulative views for every document in the database over the period of study. (B-D) Histograms describing (B) the average daily views per document [mean = 5.9, median = 4.1], (C) the documents per author [mean = 7.1, median = 1], and (D) the total views per author [mean = 57k, median = 8k]. The tails of all histograms have been truncated to facilitate visualization. (E-F) Normalized cumulative views for new documents in (E) the first 60 days and (F) from day 121 to 180.

the title), form (such as the use of poetry or easy-to-read language), and website structure (such as the position of the document within a subpage). All the features considered can be found in Table 1. For simplicity and interpretability, we constrain ourselves to a sparse linear regression model [16] where we predict the popularity of each document as a linear function of a sparse subset of the features; see Materials and Methods. Although alternative nonlinear prediction models might offer modest improvements in prediction accuracy, our goal is to obtain a model that is both sufficiently accurate and interpretable.

To obtain the corpus of documents used for training and testing our regression model, we filter the documents via two procedures that we call rare-source filtering and impulsive denoising (Materials and Methods). Rare sources are authors (or magazines) with so few documents that our out-of-sample prediction is not reliable, so we exclude documents from these sources. Our impulsive denoising procedure discards documents that are particularly popular for idiosyncratic reasons, such as documents that played a unique role in the development of jihadist thought. These are documents that would be better explained in our model by a document-fixed effect, rather than by the predictors (Fig. S4). Each of these two cleaning procedures is controlled by an adjustable parameter and all the results stated here are robust to changes in these parameters (Fig. S5).

In Fig. 2A we plot the predicted popularity of all the documents considered (out-of-sample prediction based on 10-fold cross validation; cf. Fig. S4) against their true daily views. Notice the accumulation along the diagonal, demonstrating a good fit of the model. Indeed, the relative error obtained is $\|\hat{\mathbf{y}} - \mathbf{y}\|_1 / \|\mathbf{y}\|_1 = 0.157$, where $\mathbf{y}$ is a vector containing the real daily views for all documents and $\hat{\mathbf{y}}$ contains the predicted daily views. This means that, on average, we misestimate the steady-state daily view rate of a document by 15.7%. Having established the predictive capabilities when all features are considered, we repeat the prediction only based on one class of features at a time. As expected, the performance markedly decreases, but the source features (authorship) retain the best predictive power with a relative error of 22.7%. The explanatory power of the topic is comparable to that of the website features, achieving respective errors of 27.0% and 26.0%. Finally, features related to format (text length, lines of poetry, use of easy-to-read Arabic words) have very little influence on the popularity enjoyed by jihadist texts with a prediction error of 40.9%; see Fig. 2B. Moreover, the relevance of the source is ratified when checking the regression coefficient associated with the explicit author feature; see Table 1. First, this coefficient is non-zero in the optimal sparse model. Most importantly, this coefficient is positive indicating that stating the author of a document explicitly is associated with 2.4 more daily views.

These results suggest that authorship – more than topic – drives the interest of jihadist readers. In this context, one could argue that counterterrorism operations targeting specific authors could decelerate the spread of violent ideas. However, these operations drive publicity to the targeted authors potentially increasing general interest in their ideas. We study this effect next.

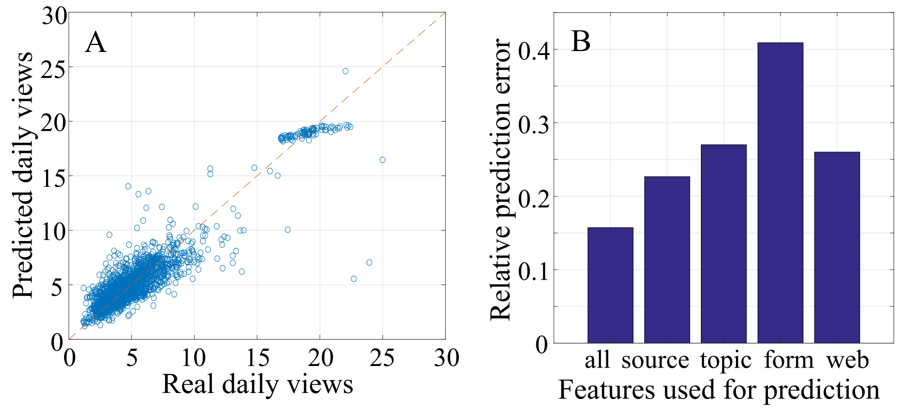

Figure 2: Popularity prediction when different classes of features are considered. (A) Real and predicted daily views for each document when all features are considered in the sparse linear regression model. Out-of-sample prediction based on 10-fold cross validation. The concentration along the diagonal indicates a good fit of the model. (B) Prediction error (10-fold cross validation) when using different sets of features.

3.2 Popularity and counterterrorism

We analyze how the popularity of documents is temporarily affected by counterterrorism operations. Other scholars have debated whether counterterrorism targeting affects the strategic success of terrorist groups [17, 18, 19, 20], but there has been less attention to the impact of targeting on the popularity of terrorists' ideas and propaganda. We identify 11 authors that were killed and 19 that were captured in counterterrorism operations during the data collection window (Table S1). We test whether targeting made the documents of these authors more popular by comparing the page-view trends to other documents on the website using Bayesian structural time-series (BSTS) models [21]. To find appropriate comparison documents, we use Euclidean distance nearest-neighbor matching to identify control documents that have similar page-view trends in the 90 days prior to the targeting event (Materials and Methods).

We find that targeting causes spikes of interest as shown in Fig. 3A. Targeting results in approximately 17 additional page views on the day after targeting, accumulating on average to 37 additional page views by the tenth day after targeting. This effect is driven almost entirely by targeted killings, rather than captures. Each document by a killed jihadist author gets an average of 26 additional page views the day after targeting. This effect quickly fades, with no additional page views attributable to the targeted killing after about 10 days. Cumulatively, targeted killings result in an average of 55 page views. Thus, the 11 authors killed during the period of data collection resulted in approximately 10,700 additional page views (since these 11 authors collectively wrote 195 documents). In contrast, the capture of jihadist authors has, on average, little effect in days following their capture (Fig. 3C). The model predicts 6 additional page views for an author the day after capture, but this result is not statistically significant at conventional levels.

Usama Bin Laden's death is particularly relevant because there were concerns that his death would result in renewed interest in his ideas [8, 9, 10]. Looking only at the 33 documents authored by Bin Laden, we find that his death caused approximately 6,250 additional page views. Among the eleven authors who were killed, seven experience significant increases in page views following their death: Bin Laden (approximately 200 additional page views per document), Abd al-Majeed Abd al-Majid (125 per document), Abu Yahya al-Libi (70 per document), Anwar al-Awlaki (40 per document), Atiyya Allah (35 per document), Abu al-Walid al-Ansary (8 per document), and Khalid Abd al-Rahman al-Husaynan (4 per document). Thus, not every targeted killing results in more page views for the author's writings, but many do. For all targets, the spike in interest is temporary.

Are jihadists more interested in targeted authors because they see them as martyrs? Or are these spikes simply due to news cycle visibility? The evidence suggests that visibility, more than martyrdom, explains these results. All effects we observe are short-lived, and match the duration of news cycle bursts [22]. If social constructions of martyrdom were causing these effects, we expect that targeted killing would result in similar spikes for all authors. If publicity is the cause, then popular authors will have bigger effects because their deaths are more newsworthy. We find that the effects of targeted killing are much larger for previously popular authors (Fig. S6).

False reports of targeting and failed attempts can also differentiate between the effects of visibility and martyrdom, as these events create visibility for authors without making them martyrs. We examine 4 cases of falsely reported targeting and attempted targeting and find detectable spikes in popularity, though they are smaller and shorter in duration (Fig. 3D).

A final way of separating publicity effects from martyrdom effects is to look at natural deaths, which can also create publicity but do not theologically elevate the status of the deceased to that of martyr. In the Figs. 3E-F we report estimates of the effects of natural death. Fig. 3F shows that the natural death of Rifa'i Surur caused his 9 documents to be viewed 1,500 more times per document, given by the cumulative effect over the first ten days after his death. In contrast, Fig. 3E shows that the other natural deaths during this time period have no detectable effect on the popularity of the deceased. The likely cause of the burst in popularity for Rifa'i Surur is that he was eulogized by several prominent jihadists, including Muhammad al-Zawahiri (brother of Ayman al-Zawahiri, current leader of Al-Qaeda). If famous ideologues can have spikes in popularity after dying naturally as well as unnaturally, then fears that targeting will re-popularize jihadists ideas should not be an overwhelming concern for counterterror policy makers.

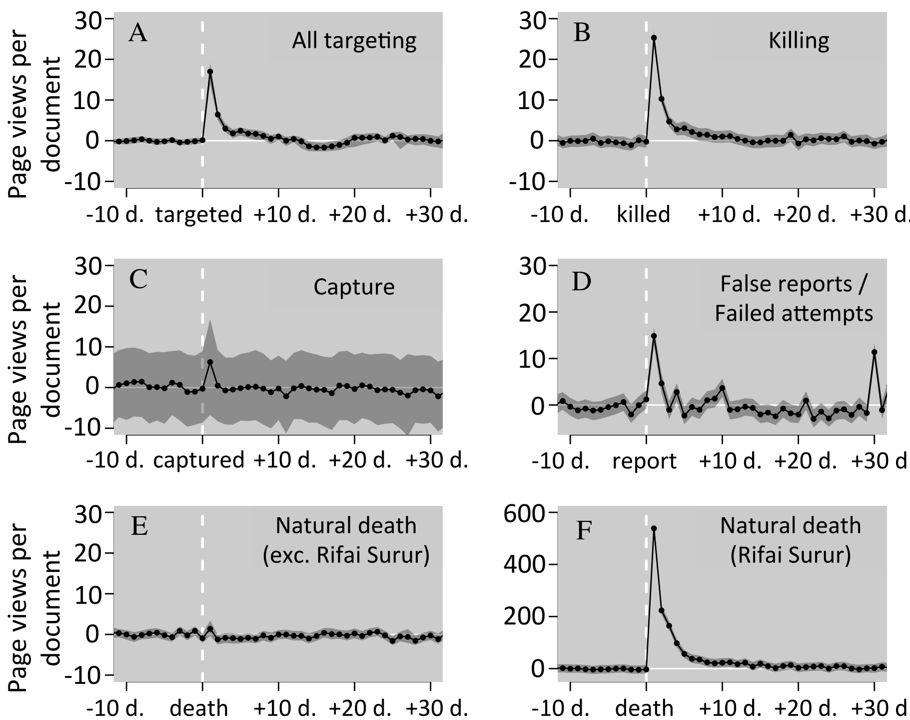

Figure 3: Effects of counterterrorism targeting on the popularity of jihadist texts. Estimated changes in the page views per document for documents of authors that were (A) targeted in any way (30 authors with 362 documents), (B) killed (11 authors with 195 documents), (C) captured (19 authors with 167 documents), (D) falsely reported to be killed or captured (4 authors with 12 documents), (E) died of natural causes, excluding Rifa'i Surur (12 authors with 61 documents), and (F) Rifa'i Surur's natural death (1 author, 9 documents).

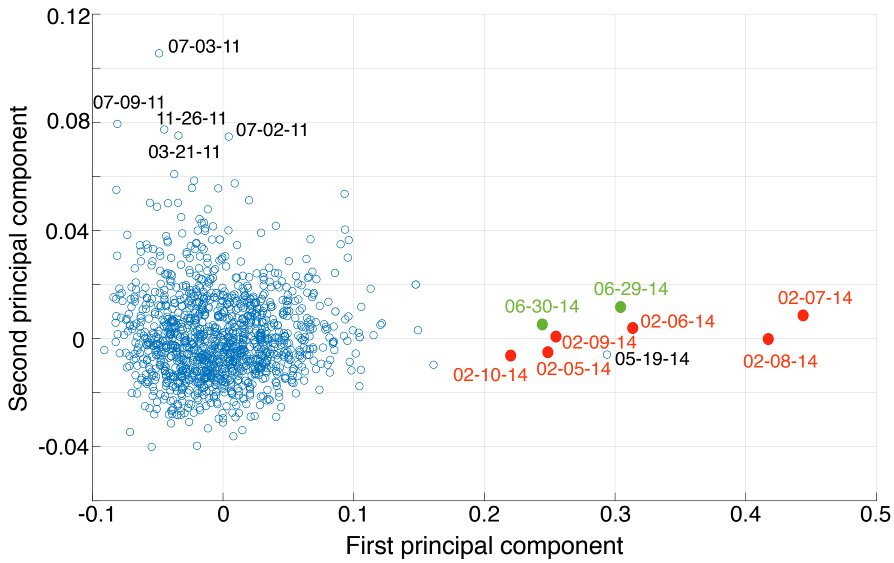

Figure 4: Effect of global events on popularity of jihadist texts. Principal component representation of the dataset. Al-Qaeda's disownment of ISIS (Feb. 3rd 2014) and the declaration of the new Islamic caliphate (June 29th 2014) are possible explanations for the outliers marked in red and green, respectively.

3.3 Popularity and global events

Global events – not directly related to a specific author – also have temporary effects on the popularity of jihadist documents. More specifically, viewership interests can temporarily shift in response to global political events. To detect these effects we run a principal component analysis (PCA) [23] of the daily aggregated viewership of each author. We focus on the cleaned time series of the 5,227 documents that were present in the repository throughout the whole period of study. For each day, we aggregate the views per author in order to obtain a time series of total views for each author. We then perform a transversal normalization where we divide the views of each author in a given day by the total views of that day. The rationale behind this normalization is that we want to capture shifts in relative popularity between authors and not a global increase in popularity of the whole website. We can then think of each date as being a high-dimensional point with 775 features corresponding to the relative popularity of the 775 authors of the documents under study (excluding the empty' author). Using PCA we project every high-dimensional point into the space spanned by the two principal components; see Fig. 4. The first principal component is dominated by the viewership of Sayyid Qutb – concentrating 97.6% of the total energy – whereas the second principal component is dominated by the viewership of Abu Muhammad al-Maqdisi and Hani al-Saba'I – concentrating 80.6% and 12.8% of the total energy, respectively. In this way, an unusually large value of the first principal component for a given date translates into an extraordinary relative interest in Qutb's writings.

Most dates follow a common pattern of popularity concentrated around the origin of the low-dimensional representation. All dates achieving an unusual proportion of the second principal component occurred during 2011, corresponding to key dates in the Arab spring [24, 25]. For dates achieving high values of the first principal component, we find that Al-Qaeda's disownment of ISIS (Feb. 3rd 2014) [26, 27] and the declaration of the new Islamic caliphate (June 29th 2014) [28, 29] are possible explanations for the outliers marked in red and green in Fig. 4, respectively. We found that most of the interest in Sayyid Qutb that spiked during these days was concentrated on his Quran commentary, suggesting that these events might have caused interest in the doctrinally-oriented writings of a prominent jihadist thinker.

4 Correspondence with the Popularity of Academic Manuscripts

We show evidence of a correspondence between the pattern of page views for jihadist documents with the pattern of Twitter mentions and page views for academic documents in an online preprint repository called arXiv[2]. This correspondence suggests that the total views that a document accrues within *The Pulpit of Monotheism and Jihad* come roughly from two sources: a true interest in jihadist authors and their new ideas akin to novel academic papers in arXiv combined with ephemeral spikes of general interest triggered by news cycles – akin to Twitter mentions of academic articles.

The arXiv online repository makes scholarly papers in Physics, Engineering, and other scientific disciplines publicly available, thus promoting a rapid dissemination of ideas. We examine two ways that people interact with papers on arXiv: by downloading them (presumably to read and learn from them) and by posting about them on the social media platform Twitter (presumably because they are newsworthy). These two types of interactions happen on noticeably different timescales. To test this, we rely on arXiv data from an existing study [30] to compute the normalized cumulative views and Twitter mentions of a newly uploaded manuscript. More precisely, for every new arXiv document considered, we plot the cumulative downloads for the first 8 weeks as a fraction of the total downloads accumulated by that document in the first 8 weeks; see Fig. 5A and Materials and Methods for more details. The red curve in Fig. 5A corresponds to the median plot of the cumulative downloads among the 195 documents considered, whereas the confidence intervals are drawn at the 25th and 75th percentiles. We repeat the same procedure but for Twitter mentions (instead of arXiv downloads) to obtain the green curve in Fig. 5A. Twitter mentions follow a news-cycle timescale of approximately one or two weeks. Individuals become aware of a paper when it is posted and tweet about it if it is newsworthy to their social network. This gives rise to a fast-spreading

[2] Available at `https://arxiv.org/`.

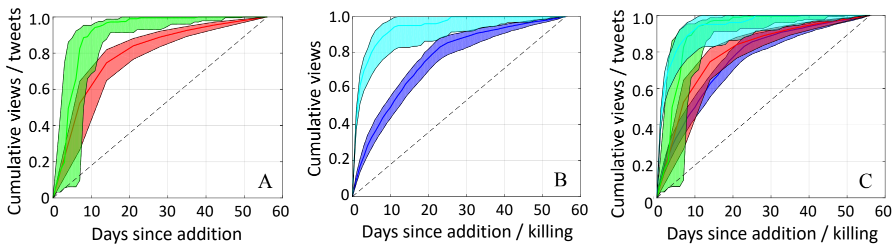

Figure 5: Normalized cumulative views and tweets for arXiv manuscripts and jihadist texts. (A) For a series of manuscripts uploaded to arXiv, we see the normalized cumulative views (red) and tweets (green) during the first 8 weeks after being uploaded. (B) In blue we have the normalized cumulative views of new jihadist texts during the first 8 weeks after being uploaded and in cyan we observe the cumulative views attributable to killings of jihadist authors during the first 8 weeks after the killings. (C) Superposition of the previous two figures revealing a remarkable resemblance between views of new arXiv manuscripts and jihadist texts on one hand and tweets about arXiv manuscripts and additional views attributable to killings of jihadists on the other hand.

and short-lasting presence in Twitter. Indeed, Fig. 5A shows that of all the tweets associated with a paper during its first two months after posting, 90% occur during the first 10 days. In contrast, the download patterns of papers follow a slower timescale. Many people download a paper when it first becomes available, after which there is a natural drop in viewing rates. However, there is a (sometimes sizable) residual activity presumably driven by readers truly interested in the ideas. Consequently, in Fig. 5A we see that only 60% of the paper downloads during the first 8 weeks are concentrated in the first 10 days.

We repeat the above experiment but for the views of new documents posted on *The Pulpit of Monotheism and Jihad* and for the views of jihadist texts attributable to targeted killings. More precisely, the blue curve in Fig. 5B corresponds to the median curve of cumulative views during the first 8 weeks of new jihadist texts with confidence intervals given by the 25th and 75th percentiles. The cyan curve in Fig. 5B represents the normalized cumulative views attributable to a targeted killing for the days directly following the killing. As expected, these latter views fade out more quickly, i.e., a sizable proportion (around 90%) of the views accumulated during the first 8 weeks after killings is concentrated on the first 10 days.

In Fig. 5C we superimpose the viewing patterns of arXiv manuscripts (Fig. 5A) and jihadist texts (Fig. 5B) and find remarkable similarities. Indeed, we find that the timing of page views for new documents posted on *The Pulpit of Monotheism and Jihad* almost perfectly matches the pattern of arXiv downloads. More precisely, there is an initial spike of interest driven by novelty that then reduces to a steady-state viewership of people presumably interested in the ideas. By contrast, the views of jihadist texts attributable to targeted killings closely matches the Twitter mentions of academic papers. This further supports our argument that visibility, more than martyrdom, is what drives the additional views. In this sense, the access to these texts is less ideologically driven and, hence, less concerning.

5 Conclusion

What makes a jihadist text popular? Our study reveals that authorship – more than topic – drives the steady-state popularity of jihadist writings. We also find that external events change viewing patterns. Counterterrorism events temporarily increase the popularity of jihadist authors by increasing publicity, but we find no long-term effects in popularity. There is also an intriguing correlation between key dates in the development of the Islamic State and views of Sayyid Qutb's writings. Together, these findings suggest that a logic of authority best explains the steady-state popularity of jihadist documents. Newsworthiness does matter, especially when a jihadist author becomes newsworthy on account of his death in a counterterrorism activity. But these temporary publicity effects due to targeted killings are relatively small and very short-lived. On one hand, this alleviates concerns by some that capturing and killing jihadist writers inadvertently repopularizes their ideas by making them into martyrs. We find no evidence to support this claim. However, our results suggest that the steady-state page views on the jihadist website we examine are best explained by a logic of authority. This is concerning given that the documents that readers are accessing often advocate political violence. Our findings show that targeting popular authors with counterterrorism action does not induce any detectable decline in their popularity.

On a broader level, we find a correspondence between the time-scales on which texts are accessed in two very different domains: the jihadist web library and the arXiv scientific repository. We conclude from this that the logics of authority and newsworthiness we posit to explain viewership of jihadist texts are also operating in the context of academic scientific discovery. Having a better comprehension of the consumption patterns of jihadist texts opens a window into the jihadist psyche, which may lead to the understanding of how their thinking is shaped.

6 Materials and methods

Data cleaning

Two cleaning procedures were applied to the data of cumulative views: i) imputation of missing data, and ii) outlier detection. For the imputation of missing data, we linearly interpolate the curves of cumulative views for each document. For example, if for a specific document we do not have view data for 20 days and the next available data point records an increase in cumulative views of 1,000 clicks, we distribute these views equally and assume that each of the 20 days contributed 50 views to the total. In terms of outlier detection, the objective is

Table 1: Features considered in the linear regression model

Feature	Class	Type	Description
Author indicator	Source	$\{0,1\}$	One indicator variable per author
Explicit author	Source	$\{0,1\}$	Indicates whether the author is stated explicitly in the website
Magazine indicator	Source	$\{0,1\}$	One indicator variable per magazine
Topic model	Topic	$[0,1]^{50}$	Proportion of topics in text (50 topics; see SI Appendix, Table S1)
Words in title	Topic	$\{0,1\}$	Indicates words on title (Only for words appearing at least 15 times)
Length	Form	$\mathbb{Z}$	Number of words in the document
Log Length	Form	$\mathbb{R}$	Natural logarithm of Length
Nr Title Words	Form	$\mathbb{Z}$	Number of words in the title
Nr Title Chars	Form	$\mathbb{Z}$	Number of characters in the title
In Series	Form	$\{0,1\}$	Indicates if the document belongs to a series of related texts
Common Text	Form	$[0,1]$	Proportion of words used in the text that are common in Arabic
Common Title	Form	$[0,1]$	Proportion of words used in the title that are common in Arabic
Lines Poetry	Form	$\mathbb{Z}$	Number of lines of poetry in the text
Lines Poetry Rel.	Form	$\mathbb{R}$	Lines Poetry divided by Length
Number webpages	Website	$\mathbb{Z}$	Number of subpages in the website in which the document appears
Rank	Website	$\mathbb{Z}$	Position in the list of documents of the same author
Log Rank	Website	$\mathbb{R}$	Natural logarithm of Rank
Rank Relative	Website	$[0,1]$	Rank divided by nr of docs by author
Log Rank Relative	Website	$\mathbb{R}$	Natural logarithm of Rank Relative
Homepage days	Website	$\mathbb{R}$	Number of days spent in homepage
Log Homepage	Website	$\mathbb{R}$	Natural log of (1 + Homepage days)
Homepage binary	Website	$\{0,1\}$	Indicator if Homepage days > 200
Shortest Path	Website	$\mathbb{Z}$	Minimum nr of hyperlinks from the homepage to the document's page
Parent Page	Website	$\{0,1\}$	Indicates parent page from which the doc can be indirectly accessed. Parent pages have shortest path of 1

to discard dates that record spurious, isolated, and out-of-scale view counts for specific documents. These outliers are most likely attributable to either website malfunctions or changes introduced by the webpage administrator rather than actual views of interested readers. In order to detect such outliers, we perform the following procedure: The difference between cumulative clicks of consecutive days is computed in order to obtain a time series of daily clicks for each document. For each of these time series, we slide a moving window of half-width 10 and compare the center value of the window with the median value within the sliding window. If the center value is more than 15 times the median – thus, presumably an outlier – we replace the center value by the median, otherwise we leave it unchanged. In Fig. S7 we illustrate the cleaning procedure on a specific document (The Religion of Abraham and the Call of the Prophets and the Messengers by Abu Muhammad al-Maqdisi), which is among the most popular ones.

Topic modeling

We identify fifty topics and the proportion of each of these topics that a document contains is used as a feature for our sparse linear regression method; see Table 1. The topics are identified using supervised latent Dirichlet allocation (sLDA) [31]. To be more concrete, we use the set of 1220 documents without source no author or magazine features to train the best predictive topics. Notice that for this set of documents we have all the words contained in each document and a rate of daily views. Thus, we run an established sLDA package for Python to infer latent topics predictive of the daily views[3]. Our code making use of this sLDA package is also made available; see Data availability.

The most common words associated with each of the 50 topics recovered are listed in Table S2. The topics are coherent and interpretable; for example, topics 10, 15, 25, and 46 are about geopolitics and war, while topics 35, 37, and 38 are about traditional Islamic legal scholarship. Once the set of 50 topics is determined, these are used to generate the topic features for all the remaining documents. It becomes apparent that different authors are more commonly associated with distinctive topics; see Fig. S8. Finally, from a procedural perspective, notice that the texts without source are not used in our popularity prediction analysis since they lack a source feature. This makes them the perfect candidates to train the sLDA, thus allowing to train the features and test their explanatory power on non-intersecting sets of texts.

Popularity prediction via sparse linear regression

Out of the 5,227 documents that were present in the repository during the whole period of study, we have metadata available – thus, we can construct the features in Table 1 – for 5,191 documents. Given that we want to test, among other things, the effect that authorship has on the popularity of a document,

[3]The package used is available at `https://github.com/Savvysherpa/slda`.

we remove the documents with no source features. In terms of metadata, this means that we discard the documents with empty author that do not belong to a magazine publication, resulting in 4,086 available documents. We further trim this set of documents by two processes that we call rare-source filtering and impulsive denoising. Both processes are controlled by a parameter that we then vary to ensure that the results found are robust with respect to these trimming processes.

In rare-source filtering we discard documents written by authors (or contained in magazines) with less than th_1 documents, where the threshold th_1 controls the aggressiveness of the filtering procedure. For example, if we choose $th_1 = 10$, the corpus of documents is reduced from 4,086 to 2,659. In order to implement our impulsive denoising procedure we solve the following optimization problem

$$\min_{\{\beta_0,\boldsymbol{\beta},\boldsymbol{\alpha}\}} \|\mathbf{y} - \mathbf{1}\beta_0 - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2 + \gamma\|\boldsymbol{\alpha}\|_1, \tag{1}$$

where $\mathbf{y}$ is a vector containing the steady-state daily views of all documents and $\mathbf{X}$ is a matrix containing all features. The objective function of the optimization problem is composed of two terms: The fit of the rate of popularity $\mathbf{y}$ in terms of the linear model plus a free vector variable $\boldsymbol{\alpha}$ and a regularizer that imposes a sparsity structure on $\boldsymbol{\alpha}$. We modify the relative weighting γ to obtain different sparsity levels in $\boldsymbol{\alpha}$, thus resulting in different fits; see Fig. S4A. To constitute our corpus of relevant texts, we discard the documents for which the corresponding elements in $\boldsymbol{\alpha}$ are different from zero. The parameter γ naturally regulates the level of denoising, e.g., if we set $\gamma = 10^{-1.5}$ we have that 12.6% of the documents get filtered, thus obtaining a final corpus of 2,324 texts on which to perform the popularity prediction. Once we have delineated the corpus of documents to be considered, we obtain the regression coefficients $\boldsymbol{\beta}$ via a cross-validated LASSO method [16]. More precisely, denoting by N the total number of documents considered, we seek to minimize the objective

$$\min_{\{\beta_0,\boldsymbol{\beta}\}} \frac{1}{2N}\|\mathbf{y} - \mathbf{1}\beta_0 - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1, \tag{2}$$

composed of a fitting term and a sparsity regularizer on $\boldsymbol{\beta}$ that plays the double role of selecting the most important features and avoiding overfitting. The relative weight λ is selected via 10-fold cross validation; see Fig. S4B. In the Supplementary Material we report the prediction performance of the above sparse linear regression for different values of th_1 and γ, and when considering various subsets of features (Fig. S5).

Effect of counterterrorism on the popularity of texts

Thirty-six authors with 426 documents were targeted (Table S1), but we only consider 33 of these authors (372 documents) because the remaining three authors were targeted too close to the end of the period of study to have sufficient data afterwards. In most cases, authors were targeted with either killing or capture, but we find instances of killing attempts that missed, and mistaken reports of capture and killing where it was later revealed that the reported target was still alive and free. A few individuals were targeted multiple times. We use the page-view trends of the 372 documents produced by these targeted individuals to test whether targeting increases interest in jihadist ideologues.

Inferring whether targeting changed a document's popularity requires a counterfactual estimate of how popular each document would have been in the absence of the targeting event. We construct this counterfactual estimate by identifying documents with similar page-view trends to those that were targeted. Requiring that documents have similar page-view trends prior to the date of targeting allows us to condition on all causes of prior popularity with a single variable. To identify documents with similar page views, we calculate the trend similarity for 90 days prior to the targeting date for every document on the website using Euclidean distance. Although this method of matching trends is simple, it is quite effective; Euclidean distance performs better than many other time-series distance metrics [32, 33]. To select matching documents, we identify the document that minimizes the distance to each treated document. To estimate treatment effects, we use BSTS models, implemented in the CausalImpact package in the R statistical language. The BSTS model is designed to estimate the effects of interventions on a single time series using one or more untreated time series as predictors. We have multiple treated documents in our models, so we combine them into a single time series by taking the mean for each date. We use each of the control documents as the predictors in the BSTS model. Complementary approaches based on a frequentist regression framework as well as on the celebrated difference-in-differences estimator [34] were pursued, confirming the results found via the BSTS model. The number of additional views attributable to killings obtained from the BSTS model were also used in the generation of Fig. 5B. The uncertainty intervals in this figure correspond to 95% Bayesian credible intervals around the estimated cumulative views attributable to counterterrorism targeting. More information can be found in the Supplementary Material.

Cumulative views in arXiv and Twitter

We consider the cohort of 4,606 scientific articles that were analyzed in [30]. These articles were submitted to arXiv between October 2010 and May 2011. We further trim this cohort be focusing on the 500 most popular articles, i.e., the ones that were downloaded the most during the period of study. Of these 500 articles, we kept the 195 articles whose first version was uploaded between October 2010 and February 2011. First, the reason for focusing on first versions is that we want the uploads to be truly new, as opposed to updates to existing manuscripts. Second, the reason for constraining the dates between October 2010 and February 2011 is to ensure that we have 8 weeks of download data after the manuscript was uploaded. We also relied on [30] for the tweet counts of these papers. However, as expected, most papers do not have significant presence in Twitter. For the Twitter plots, we focused on the 12 articles that had more than 25 mentions in Twitter during the period of study and were uploaded between October 2010 and February 2011.

Data availability

All the data and metadata, along with code (R, MATLAB, and Python) to replicate the figures can be accessed[4].

References

[1] P. A. Hall, Ed., *The Political Power of Economic Ideas: Keynesianism across Nations*. Princetion, NJ: Princeton University Press, 1989.

[2] J. Goldstein and R. O. Keohane, Eds., *Ideas and Foreign Policy: An Analyitical Framework*. Ithica, NY: Cornell University Press, 1993.

[3] M. Blyth, "Any more bright ideas? the ideational turn of comparative political economy," *Comparative Politics*, vol. 29(1), pp. 229–250, 1997.

[4] J. M. Chwieroth, "Testing and measuring the role of ideas: The case of neoliberalism in the international monetary fund," *Int. Stud. Q.*, vol. 51, pp. 5–30, 2007.

[5] J. Deol and Z. Kazmi, *Contextualising Jihadi Thought*. Columbia University Press New York, 2011.

[6] Pew Research Center, "The world's muslims: Religion, politics and society," *The Pew Forum on Religion and Public Life*, 2013.

[7] D. Karell and M. Freedman, "Rhetorics of radicalism," *American Sociological Review*, vol. 84, no. 4, pp. 726–753, 2019.

[8] A. K. Cronin, "How al-Qaida ends: The decline and demise of terrorist groups," *Int. Secur.*, vol. 31, no. 1, pp. 7–48, 2006.

[9] A. B. Atwan, "Osama bin Laden's death: A leader's wish fulfilled," *The Guardian*, May 6 2011, available at https://www.theguardian.com/commentisfree/2011/may/03/osama-bin-laden-martyrdom-al-qaida.

[10] E. Husain, "Bin Laden as 'martyr': A call to jihadists," *CNN*, May 4 2011, available at http://www.cnn.com/2011/OPINION/05/02/husain.bin.laden.

[11] R. Simcox, "Killing Al Qaeda's leaders: It works," *Los Angeles Times*, May 24 2011, available at http://articles.latimes.com/2012/may/24/opinion/la-oe-simcox-targeted-killings-work-20120524.

[4] Download `https://drive.google.com/open?id=1F6QuuSsUx2b2fKF-eWdy1UXBityct6HL`. Password: nhb_jihad. Upon a potential publication, the dataset will be made public.

[12] M. Smith and J. I. Walsh, "Do drone strikes degrade al qaeda? evidence from propaganda output," *Terrorism and Political Violence*, vol. 25, no. 2, pp. 311–327, 2013.

[13] M. Bloom and C. Daymon, "Assessing the future threat: Isis's virtual caliphate," *Orbis*, vol. 62, no. 3, pp. 372–388, 2018.

[14] J. M. Brachman and W. F. McCants, "Stealing Al Qaeda's playbook," *Studies in Conflict & Terrorism*, vol. 29, no. 4, pp. 309–321, 2006.

[15] J. Wagemakers, "Protecting jihad: The Sharia council of the Minbar al-Tawhid wa-l-Jihad," *Middle East Policy*, vol. 18, no. 2, pp. 148–162, 2011.

[16] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. R. Stat. Soc. Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[17] J. Jordan, "When heads roll: Assessing the effectiveness of leadership decapitation," *Secur. Stud.*, vol. 18, no. 4, pp. 719–755, 2009.

[18] P. B. Johnston, "Does decapitation work? assessing the effectiveness of leadership targeting in counterinsurgency campaigns," *Int. Secur.*, vol. 36, no. 4, pp. 47–79, 2012.

[19] J. Jordan, "Attacking the leader, missing the mark: Why terrorist groups survive decapitation strikes," *Int. Secur.*, vol. 38, no. 4, pp. 7–38, 2014.

[20] P. B. Johnston and A. K. Sarbahi, "The impact of US drone strikes on terrorism in Pakistan," *Int. Stud. Q.*, vol. 60, no. 2, pp. 203–219, 2016.

[21] K. H. Brodersen, F. Gallusser, J. Koehler, N. Remy, and S. L. Scott, "Inferring causal impact using Bayesian structural time-series models," *Ann. Appl. Stat.*, vol. 9, no. 1, pp. 247–274, 2015.

[22] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the dynamics of the news cycle," in *Proc. Int. Conf. Knowl. Discov. Data Min.*, 2009, pp. 497–506.

[23] I. T. Jolliffe, *Principal Component Analysis*. Springer-Verlag New York, 1986.

[24] M. Lynch, "The arab uprising: The unfinished revolutions of the new middle east," *PublicAffairs, New York*, 2013.

[25] G. Lotan, E. Graeff, M. Ananny, D. Gaffney, and I. Pearce, "The revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions," *Int. J. Commun.*, vol. 5, pp. 1375–1405, 2011.

[26] Al-Qaeda, "Translation of al-Qaeda statement on feb. 3, 2014 acknowledging ISIS officially isn't part of AQ," 2014, available at https://azelin.files.wordpress.com/2014/02/al-qc481_idah-22on-the-relationship-of-qc481idat-al-jihc481d-and-the-islamic-state-of-iraq-and-al-shc481m22-en.pdf.

[27] J. Turner, "Strategic differences: Al Qaeda's split with the Islamic State of Iraq and al-Sham," *Small Wars Insur.*, vol. 26, no. 2, pp. 208–225, 2015.

[28] Islamic State, "This is the promise of Allah," 2014, available at https://ia902505.us.archive.org/28/items/poa_25984/EN.pdf.

[29] W. McCants, *The ISIS Apocalypse: The History, Strategy, and Doomsday Vision of the Islamic State.* Picador, 2016.

[30] X. Shuai, A. Pepe, and J. Bollen, "How the scientific community reacts to newly submitted preprints: Article downloads, twitter mentions, and citations," *PLoS ONE*, vol. 7, no. 11, p. e47523, 2012.

[31] J. D. Mcauliffe and D. M. Blei, "Supervised topic models," in *Advances in Neural Information Processing Systems 20*, J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, Eds. Curran Associates, Inc., 2008, pp. 121–128. [Online]. Available: http://papers.nips.cc/paper/3328-supervised-topic-models.pdf

[32] E. Keogh and J. Lin, "Clustering of time-series subsequences is meaningless: Implications for previous and future research," *Knowl. Inf. Syst.*, vol. 8, no. 2, pp. 154–177, 2005.

[33] E. Keogh and S. Kasetty, "On the need for time series data mining benchmarks: A survey and empirical demonstration," in *Proc. Int. Conf. Knowl. Discov. Data Min.*, 2002, pp. 102–111.

[34] A. Abadie, "Semiparametric difference-in-differences estimators," *Rev. Econ. Stud.*, vol. 72, no. 1, pp. 1–19, 2005.