Wikimedia Foundation v. NSA
No. 15-cv-0062-TSE (D. Md.)

Plaintiff's Exhibit 2

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, *Plaintiff*, v. NATIONAL SECURITY AGENCY / CENTRAL SECURITY SERVICE, *et al.*, *Defendants*.	No. 1:15-cv-00662-TSE

<u>REPLY DECLARATION OF JONATHON PENNEY</u>

TABLE OF CONTENTS

I. INTRODUCTION

1. My name is Jonathon Penney. I have been asked by the plaintiff's counsel in *Wikimedia Foundation v. National Security Agency*, No. 1:15-cv-00662-TSE (D. Md.), to provide this reply declaration to address the Defendants' reply to the Plaintiff's brief and to my declaration, both of which were dated December 18, 2019. Results of my additional analysis in support of this reply declaration are included herein in the Appendix.

2. My qualifications and expertise are discussed in detail in my opening Declaration ("Declaration"). *See* ECF No. 168-02. Unless otherwise stated, I have personal knowledge of the facts herein.

II. DR. SALZBERG'S ANALYSIS IS FLAWED

3. In support of their Reply motion, Defendants submit the Declaration of Dr. Alan Salzberg ("Salzberg Declaration"). *See* ECF No. 178-3. Dr. Salzberg's analysis misunderstands my interrupted time series (ITS) design and study in fundamental ways and ignores relevant literature on methodological best practices for ITS studies. Furthermore, in critiquing my study, the Salzberg Declaration relies primarily on a visual inspection of data, which can often be misleading,[1] rather than formal testing mechanisms that can be verified. Salzberg's reliance on a visual inspection of the data causes him to formulate misguided critiques and conclusions as to my study's reliability and validity. In actuality, my methodology, method of analysis, and regression model is superior to any alternatives he suggests.

4. **First**, Salzberg's use of disaggregated line plots (see ¶¶ 12-14 and Figure 1 of the Salzberg Declaration) to analyze the Wikipedia page view data is an inferior method of analysis

---

[1] GENE S. FISCH, *EVALUATING DATA FROM BEHAVIORAL ANALYSIS: VISUAL INSPECTION OR STATISTICAL MODELS?* 54 Behavioural Processes 137, 137 (2001) (quoting Howard Wainer: "A graph is nothing but a visual metaphor. To be truly evocative, it must correspond closely to the phenomena it depicts… If a graphic depiction of data does not faithfully follow this notion it is almost sure to be misleading.").

for ITS studies, as compared to the segmented linear regression trend analysis of aggregated data that I used. Dr. Salzberg's approach ignores relevant literature on methodological best practices. Indeed, the recommended method of analysis for ITS design studies is segmented linear regression analysis, which I employed in my study, as it allows researchers to: (1) control for prior trends in the data; (2) measure the dynamics of change in response to an intervention; (3) tolerate fewer time points than alternative methods; (4) adjust for serial correlation in the data; and (4) apply these methods to aggregate level data.[2] By contrast, Salzberg's disaggregated line-plots method offers none of these strengths or advantages and is neither recommended nor discussed in relevant literature.

5. **Second**, Salzberg's use of disaggregated line plots adds "noise," both visual and statistical, which masks actual overall trends in the data best understood through analysis of *aggregated* monthly page view data. Consistent with my approach, a majority of ITS design studies use aggregated data.[3] This is to reduce or remove noise in the data and to allow for more

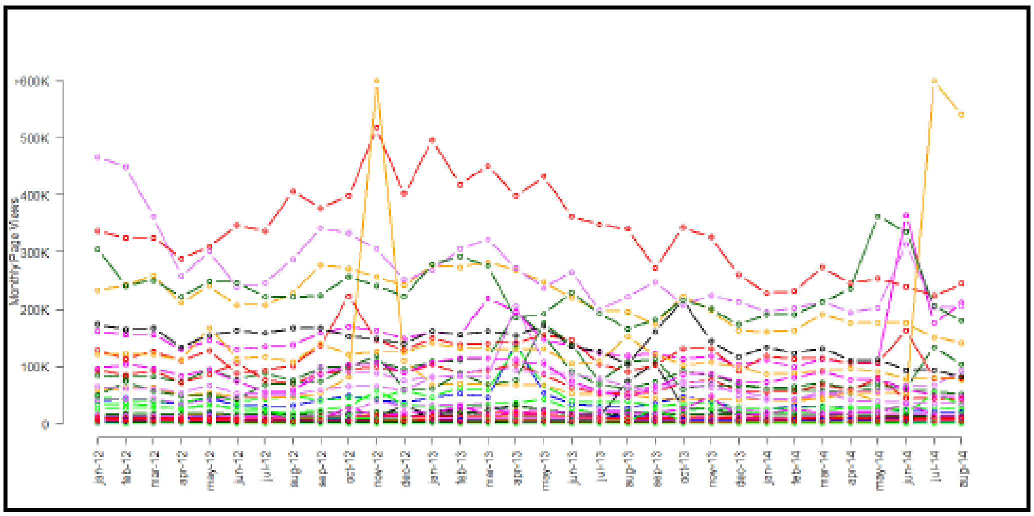
---

[2] The leading peer reviewed works on ITS design, methodology, and analysis recommends segmented linear regression analysis over several alternatives methods and models, including generalized estimating equations (GEE) method and autoregressive integrated moving average (ARIMA) modeling. FANG ZHANG, A.K. WAGNER, ET AL., *METHODS FOR ESTIMATING CONFIDENCE INTERVALS IN INTERRUPTED TIME SERIES ANALYSES OF HEALTH INTERVENTIONS*, 62:2 J. Clinical Epidemiology 143, 143-144 (2009) (discussing the advantages of segmented linear regression analysis for ITS designs compared to alternatives); A.K. WAGNER ET AL., *SEGMENTED REGRESSION ANALYSIS OF AN INTERRUPTED TIME SERIES IN MEDICATION USE RESEARCH*, 27 J. Clinical Pharmacy & Therapeutics 299, 299, 208 (2002) (describing segmented linear regression analysis as a "powerful statistical method or estimating intervention effects" in ITS studies, and describing "strengths"); MYLENE LAGARDE, *HOW TO DO (OR NOT TO DO) … ASSESSING THE IMPACT OF A POLICY CHANGE WITH ROUTINE LONGITUDINAL DATA,* 27:1 Health Policy and Planning 76, 79 (2012) (noting this method "controls for secular trends and can also adjust for potential serial correlation of the data"); ROBERT B. PENFOLD & FANG ZHANG, *USE OF INTERRUPTED TIME SERIES ANALYSIS IN EVALUATING HEALTH CARE QUALITY IMPROVEMENTS*, 13:6 Acad. Pediatrics S38 (2013) (discussing the advantages and limitations of employing time series analysis to understand and explore the impact of health policy changes).

[3] Wagner (2002), *id.*, at 308 ("Segmented regression typically aggregates individual-level data by time point" and noting a leading ITS study where the "unit of analysis" was a monthly aggregated data, as used in this study); JANDOC, ET AL., *INTERRUPTED TIME SERIES ANALYSIS IN DRUG UTILIZATION RESEARCH IS INCREASING: SYSTEMATIC REVIEW AND RECOMMENDATIONS*, 68 J. Clinical Epidemiology 950, 950 (2015) ("Interrupted time series methods use aggregate data collected over equally spaced intervals before and

sophisticated statistical tests and analysis.[4] Salzberg departs from this approach, which is standard in a majority of ITS studies, leading him to incorrect inferences and conclusions in his analysis. For example, Figure 1 of the Salzberg Declaration distorts and hides important trends by plotting individual line plots for the 48 Terrorism articles. Here, line plots for a majority of the 48 Articles cannot be seen as they have page views too small to be visualized with the large page view scale (0 to 600,000 page views) used on the vertical axis of the graph:

Figure 1: Individual Page Views for Each of the Articles Within the Terror 48, Which The Penney Declaration Hypothesized Show an Immediate Decline Beginning in June 2013



6. Salzberg's Figure 1 creates a false impression there are no patterns or trends for overall page views over 32 months. But those trends are easily visible when individual article page views are analyzed as aggregated monthly page views as visualized in **Figure 1a** below. This

---

after an intervention, with the key assumption that data trends before the intervention can be extrapolated to predict trends had the intervention not occurred"); EMMA BEARD, *USING TIME-SERIES ANALYSIS TO EXAMINE THE EFFECTS OF ADDING OR REMOVING COMPONENTS OF DIGITAL BEHAVIOURAL INTERVENTIONS AND ASSOCIATIONS BETWEEN OUTCOMES AND PATTERNS OF USAGE*, Centre for Behaviour Change (CBC) Conference, University College of London 15 (2017), https://www.ucl.ac.uk/behaviour-change/events/presentations-17/beard.pdf (noting that a "[m]ajority of studies use aggregated data". She also specifically notes that linear regression may be used for "interrupted time series design" if autocorrelation is controlled).

[4] BEARD, *id.*, at 15 (noting that a "[m]ajority of studies use aggregated data" as this "removes noise and allows for more sophisticated tests which require continuous or rate type data").

figure plots the median aggregated page views for the 48 Terrorism Articles with trend lines included to understand the shift in trend.

Figure 1a: Aggregated Median Page Views for 48 Terrorism Wikipedia Articles

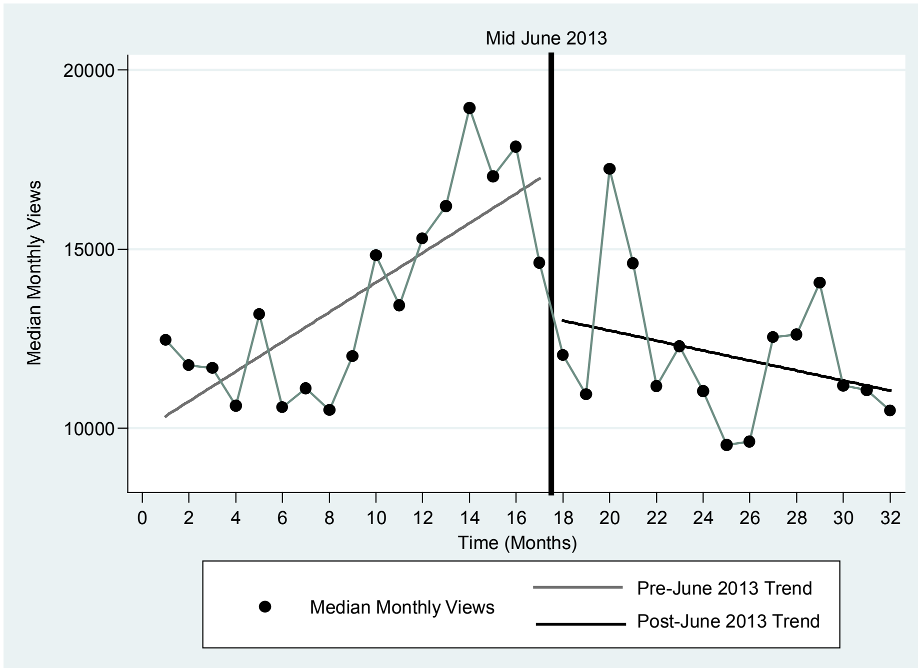
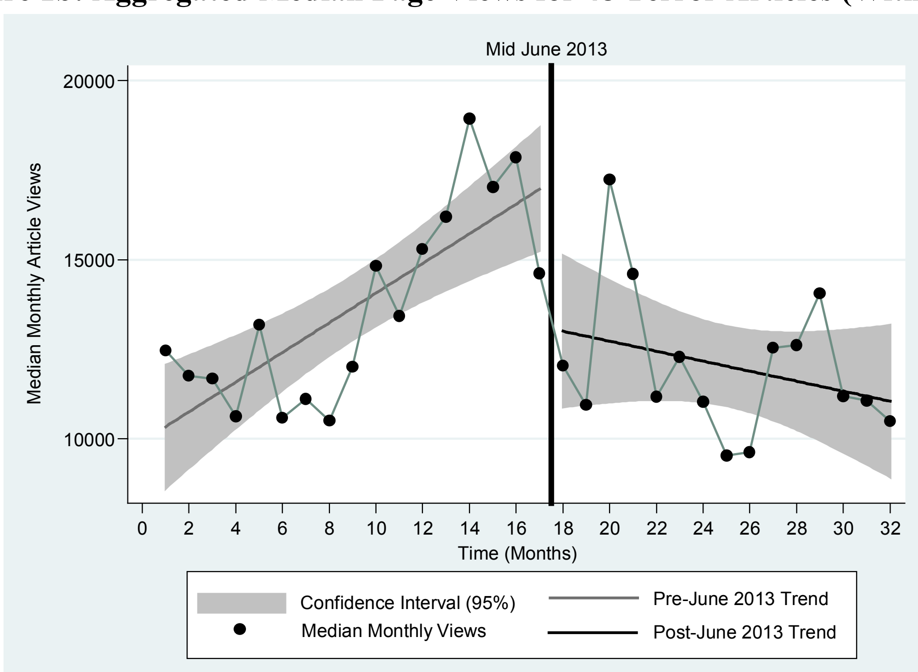


Figure 1b: Aggregated Median Page Views for 48 Terror Articles (With C.I.)

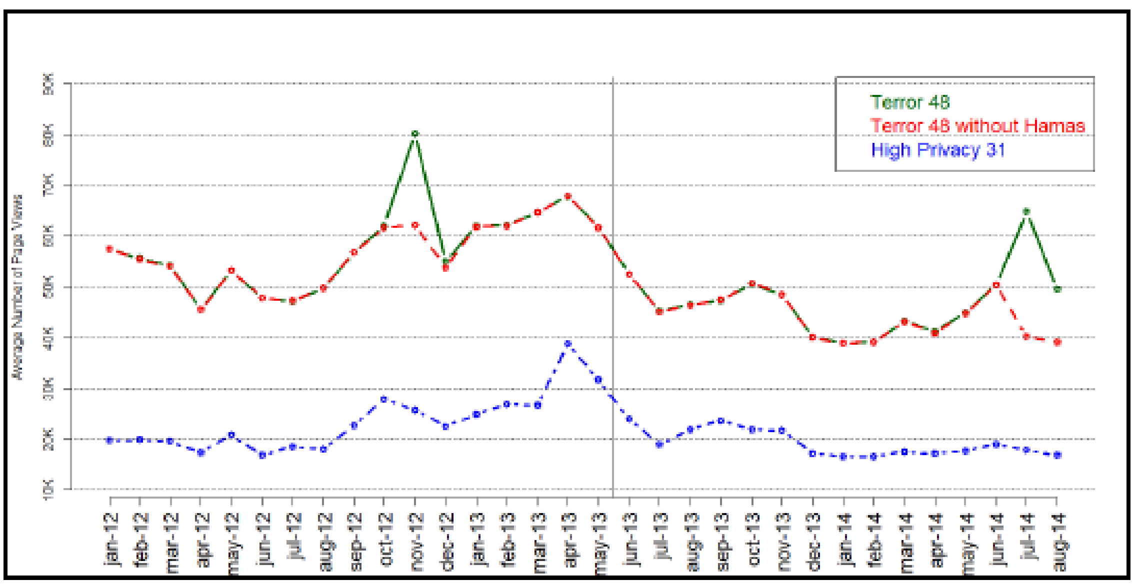


7. **Figure 1b** visualizes the same data with pre/post June 2013 trend lines and

confidence intervals (the gray shaded area), and demonstrates that the trend change before and after June 2013 was statistically significant, as there is no overlap of confidence intervals during these two periods. Specifically, this visualization of the data demonstrates a statistically significant drop in June 2013 and reduction in overall monthly page views. This is a far clearer visualization, with clear trends, compared to Salzberg's Figure 1, which masks these trends in the "noise" of 48 disaggregated line plots.

8. Figures 5 and 6 are of the Salzberg Declaration are similarly distorted. These figures present line-plots for the 48 Terror Articles, 47 Terror Articles, and 31 High Privacy Articles together. For example, Figure 5 visualizes a line plot for the average monthly page views for those three article sets:

Figure 5: Average Page Views Show a Peak in April 2013 or Before



9. Again, by presenting the data associated with these three different sets of articles on the same graph with the same scale on the vertical axis (10,000 to 90,000 average page views), the Figure distort the presentation of the data, creating a false impression that both the 47 Terror Articles and the 31 articles with more privacy sensitive ratings ("31 Higher Privacy Articles") have

a flat trend over the course of the 32 months. These distortions are easy to visualize when the average monthly page views for the 31 Higher Privacy Articles are plotted and presented with an appropriate scale as in **Figures 2a** and **2b** below.

Figure 2a: Average Monthly Page Views for the 31 Higher Privacy Articles Plotted Alone Show Increase Until June 2013 And Then A Drop-Off After That Month

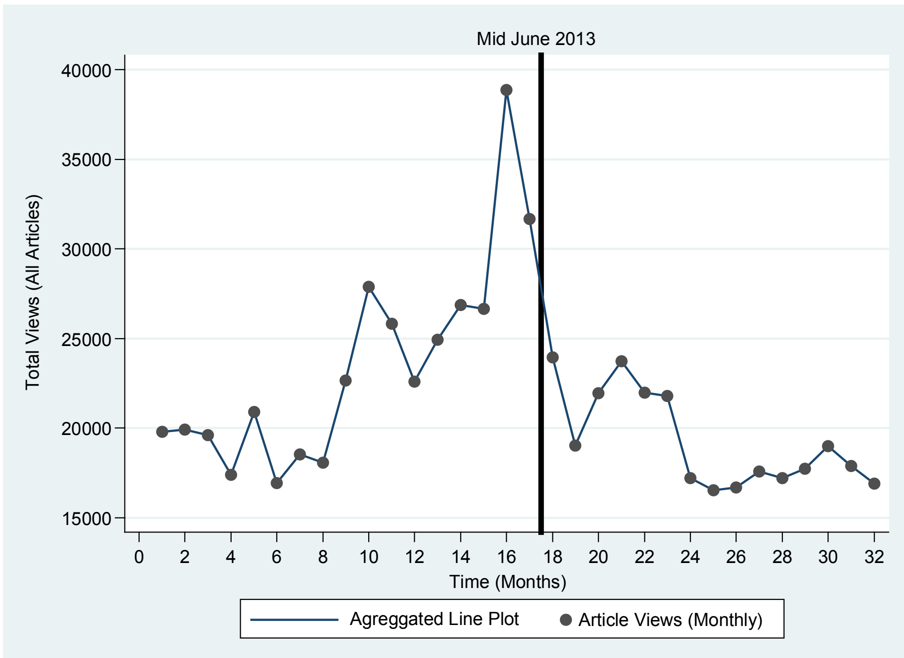
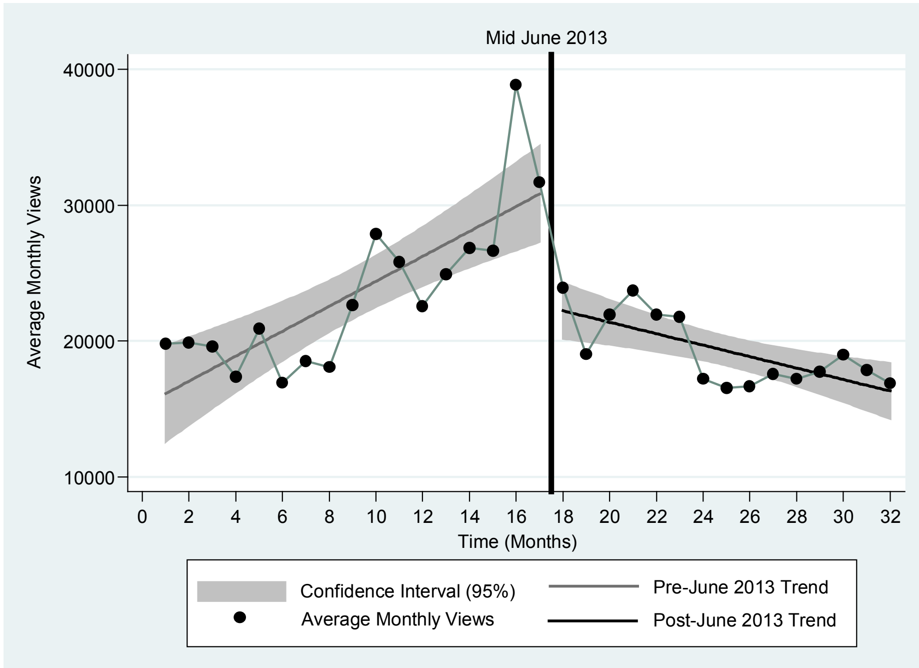


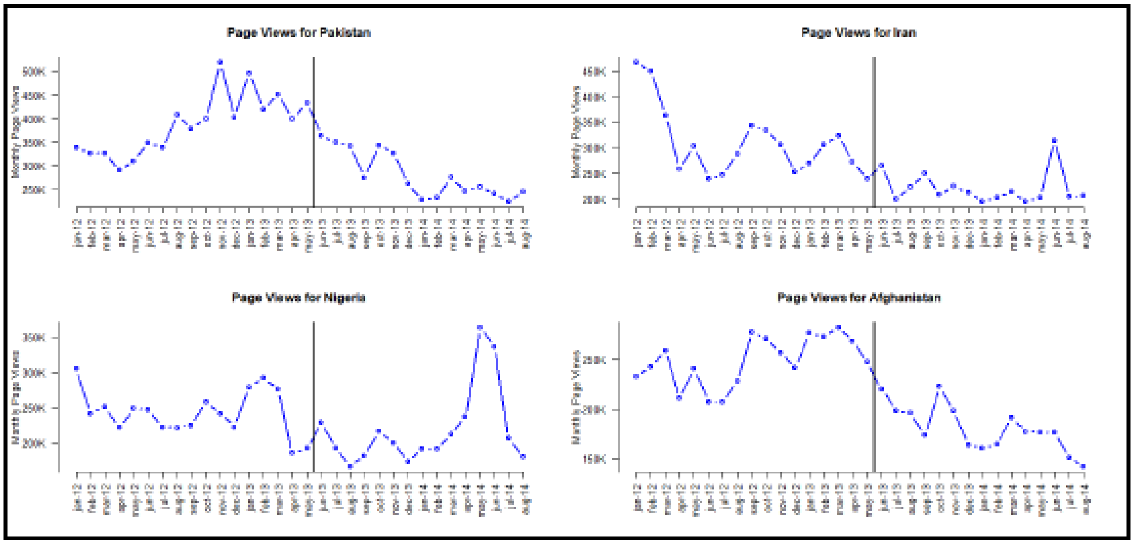
Figure 2b: Average Monthly Views for the 31 Higher Privacy Articles Plotted Alone Show Increase Until June 2013 And Then A Sharp Drop-Off (With Trend Lines)



10. In **Figure 2a** above, the average monthly page views for the 31 Higher Privacy Articles show a clear trend and not the flattened pattern reflected in Figure 5 of the Salzberg Declaration. With an appropriate scale on the vertical axis, page views increase until mid-June 2013 and declined thereafter. This point is even clearer in **Figure 2b**, which plots the very same data but adds trend lines and confidence intervals for clarity.

11. **Third**, Salzberg focuses his analysis on cherry-picked individual articles that obscure and mislead about actual trends in the data. For example, Figure 2 of the Salzberg Declaration visualizes monthly page view line-plots for four articles: Pakistan, Iran, Nigeria, and Afghanistan:

Figure 2: Individual Articles show no Association of June 2013 with a Decline in Page Views



12. While these four Wikipedia articles did form part of the 48 Terror Article set, they have among the lowest privacy-sensitivity scores among all articles in the set (*see* Table 12 of my Declaration). That is, these articles raised few privacy concerns for survey participants. As discussed in my opening declaration, 415 independent Internet users participated in a survey in which they provided feedback on how keywords associated with each of the 48 Terrorism Articles may raise privacy-related concerns ("Privacy Evaluation Survey").[5] In the Privacy Evaluation Survey, the combined average privacy-sensitivity rating for all 48 Terror Articles was 2.15 and the median was 2.07. The articles that Salzberg cherry-picked fell far below that mean and median: Pakistan (1.82), Iran (1.85); Nigeria (1.71); Afghanistan (1.83). Since the hypothesis that I tested in my study concerns a ***privacy***-based chilling effects theory [6]—i.e., that Wikipedia users avoided privacy-sensitive Wikipedia articles due to awareness of NSA Upstream Surveillance—it is inappropriate to rely on trends for these four articles, which fell far below the mean and median privacy-sensitive rating. As such, to the extent Salzberg isolates and relies on these four articles to

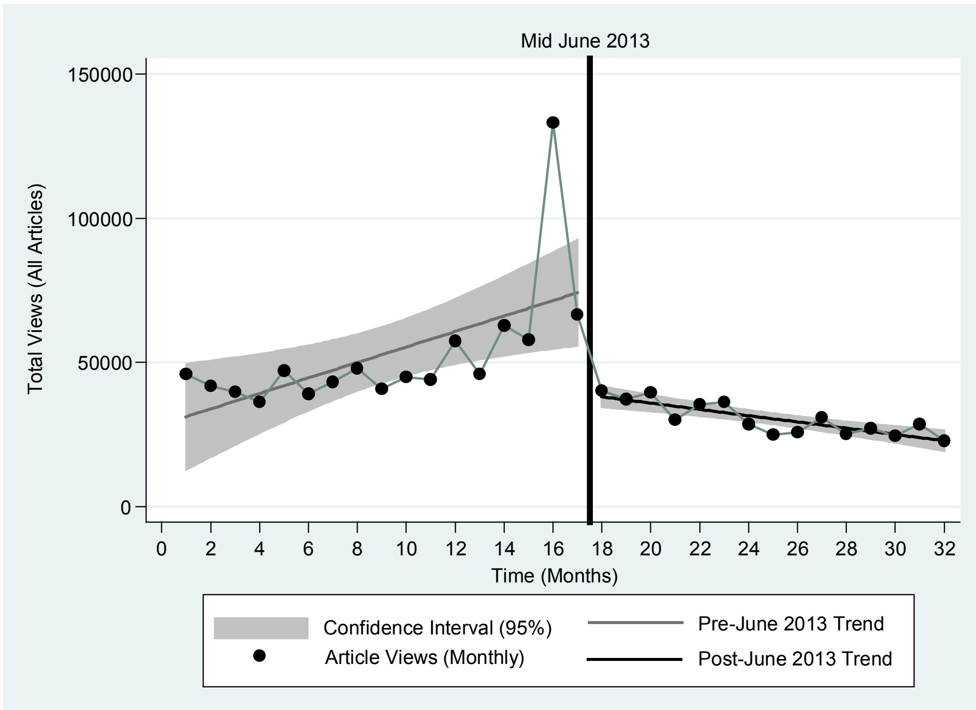
[5] See ¶¶ 32-33 of my Declaration dated December 18, 2018.
[6] See ¶¶12-21 of my Declaration dated December 18, 2018.

reject a chilling effects hypothesis (*see* paragraph 16 of the Salzberg Declaration), his analysis is misleading, unreliable, and masks actual trends found in my study.

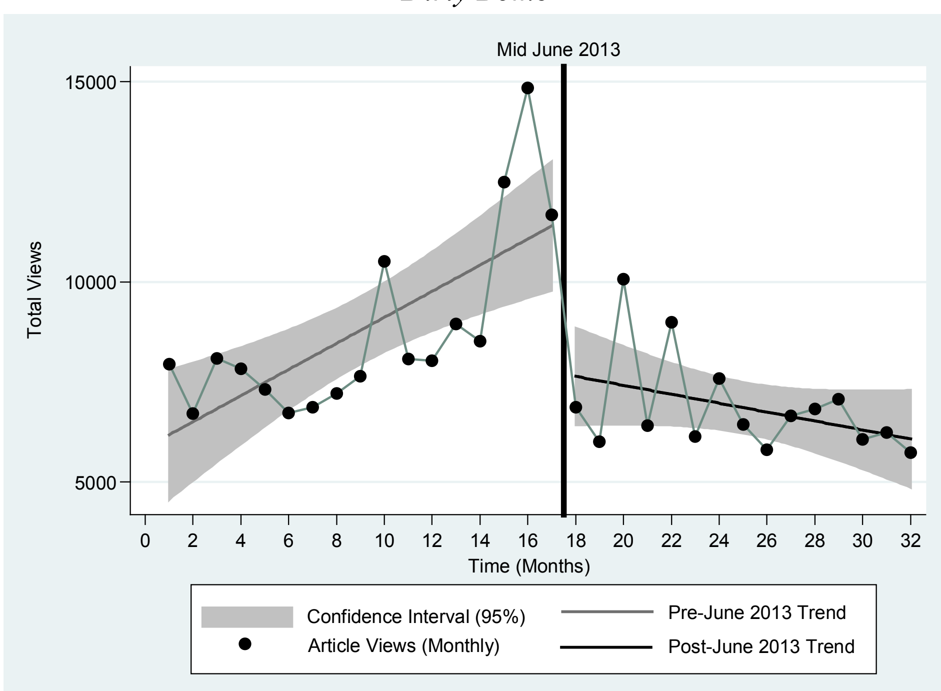
13. **Fourth**, the Salzberg Declaration focuses on disaggregated line-plots that mask aggregated data trends or less privacy-sensitive or privacy-concerning articles like the four articles noted above (Pakistan, Iran, Nigeria, and Afghanistan). This approach is inappropriate given my study tests a chilling effects hypothesis based on a privacy theory. Moreover, Salzberg's approach leads to a flawed analyses and conclusions. In fact, the page views for the four articles (among the 48 Terrorism Articles) with the very highest privacy-sensitivity scores according to the Privacy Evaluation Survey—improvised explosive device (2.86), dirty bomb (2.81), car bomb (2.81), and ammonium nitrate (2.61)—are entirely consistent with a chilling effects hypothesis in June 2013. **Figure 3a** depicts monthly page views for each of these most privacy-sensitive articles. Each figure demonstrates page view trends consistent with a chilling effects hypothesis: a monthly increase in page views leading up to June 2013, an abrupt statistically significant decline and subsequent change in trend to a monthly decrease in page views.

Figure 3a: Page Views For The Four Most Privacy-Sensitive Articles Are Consistent With A Chilling Effects Hypothesis in June 2013

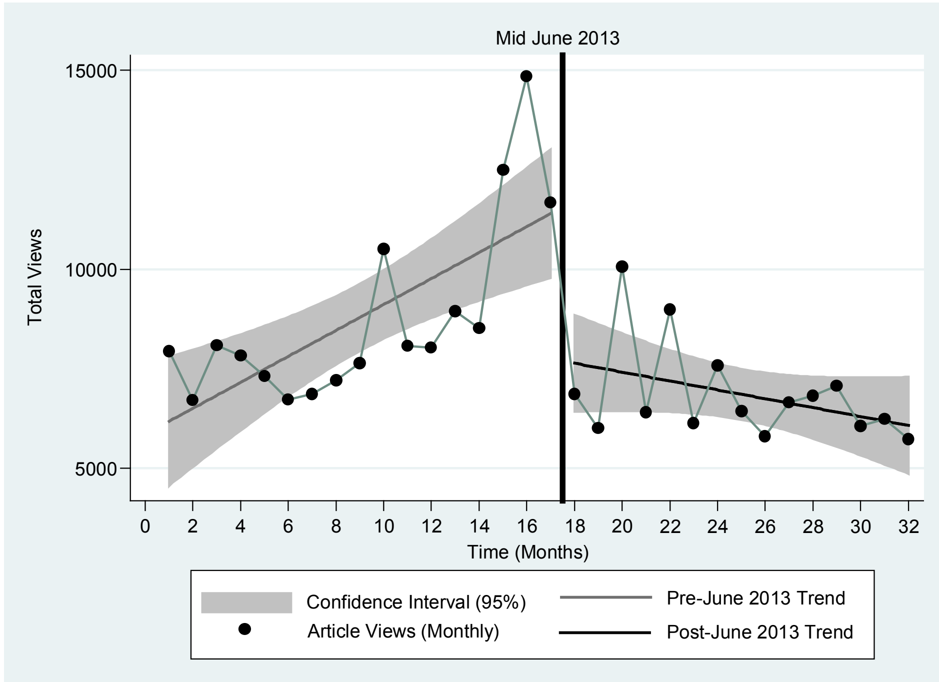
Improvised Explosive Device



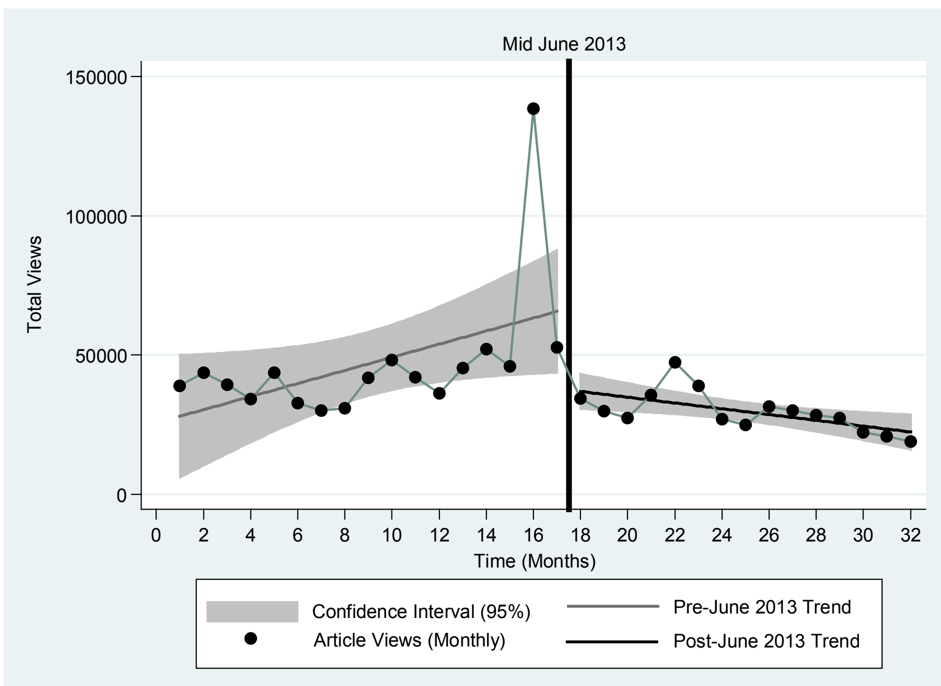
Dirty Bomb



Car Bomb

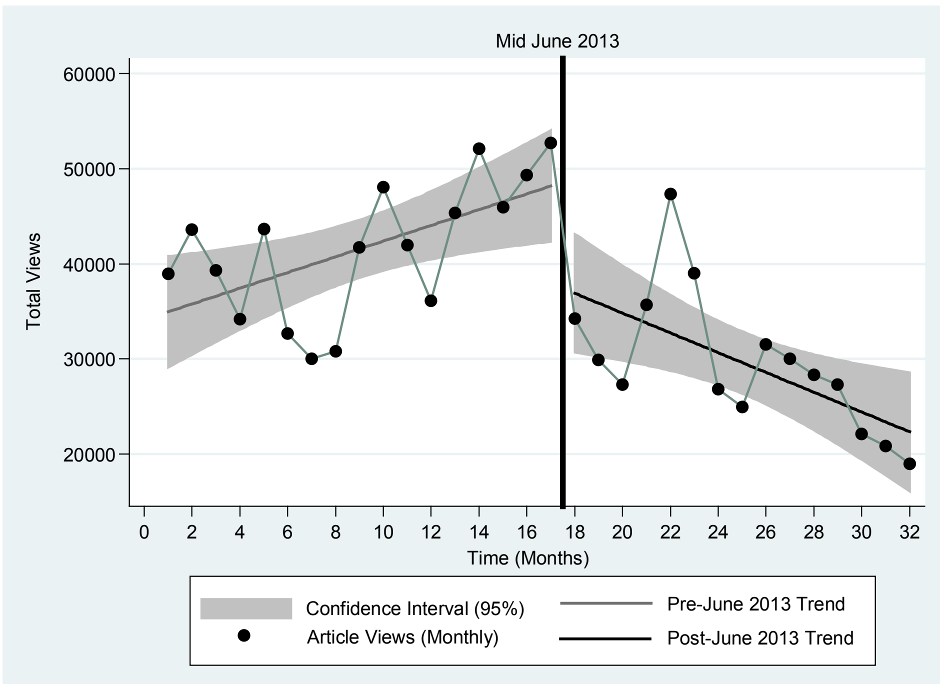


Ammonium Nitrate



14. Salzberg claims that the "ammonium nitrate" article is an outlier in its number of page views in April 2013.[7] Even assuming this is correct, if the article's page views for that month are normalized,[8] the overall trend for the ammonium nitrate article remains consistent with a chilling effect hypothesis. **Figure 3b** provides the ammonium nitrate article's normalized page views over 32 months, which are consistent with a June 2013 chilling effects hypothesis: increasing monthly articles views in the months leading up to June 2013, and then an abrupt statistically significant decline in June and a subsequent monthly reduction in views:

Figure 3b: Page Views for Normalized Ammonium Nitrate Article Are Consistent With June 2013 Chilling Effects Hypothesis



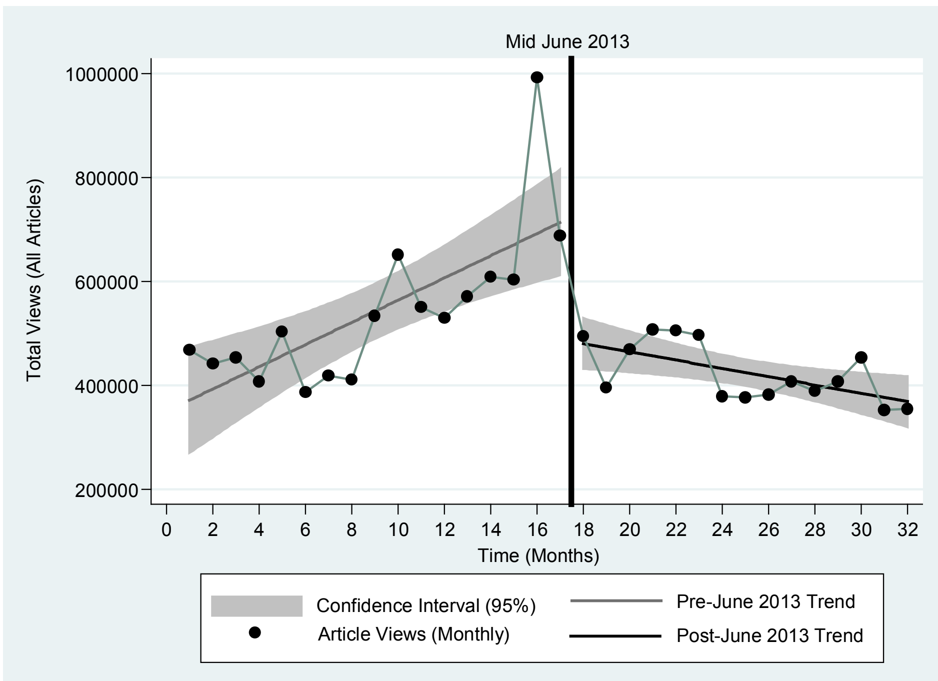
[7] See ¶ 60 of the Salzberg Declaration.

[8] I replaced the outlier value for the article in April 2013 (138363) with an average of the total page views for the article (49316) in the two adjacent months (May and March 2013). Correcting, modifying, or deleting an outlier value or observation in a data set is consistent with best-practices in dealing with outliers: HERMAN AGUINIS, RYAN K. GOTTFREDSON & HARRY JOO, *BEST-PRACTICE RECOMMENDATIONS FOR DEFINING, IDENTIFYING, AND HANDLING OUTLIERS*, Organizational Res. Methods 8, 20–23 (2014), http://orm.sagepub.com/content/early/2013/01/11/1094428112470848.abstract ("Once error outliers have been identified, the correct procedure is to either adjust the data points to their correct values or remove such observations from the dataset").

15. Furthermore, the aggregate total monthly page views, average monthly page views, and median monthly page views for the 23 most privacy-sensitive articles among the set of 48 Terrorism Articles in the study[9] are also consistent with a chilling effect hypothesis in June 2013, when measured over a 32-month period. **Figure 4** visualizes monthly median page views for these 23 most privacy sensitive Wikipedia articles:

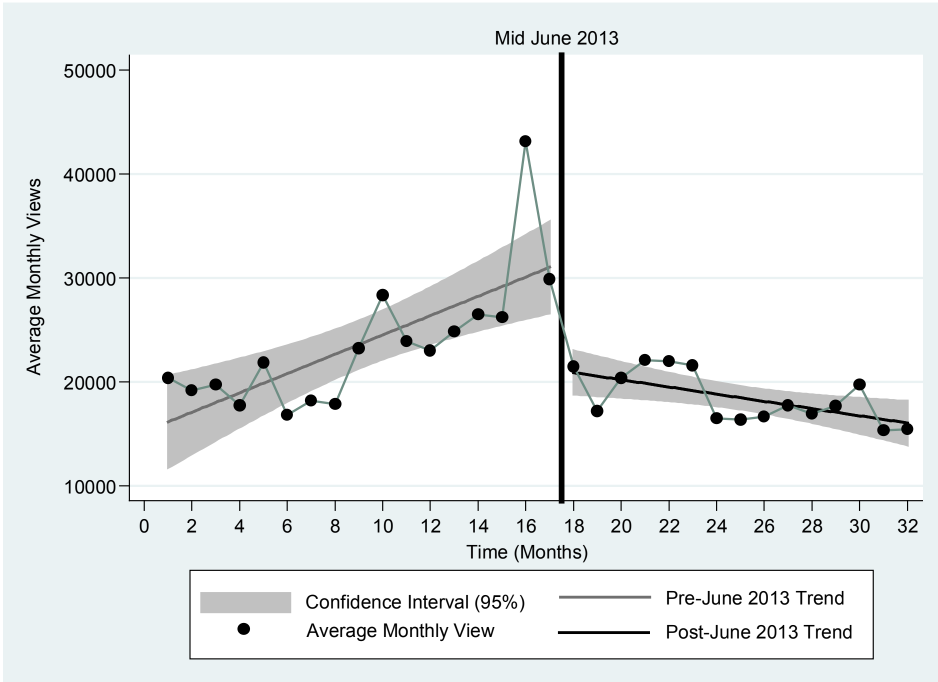
Figure 4: Total Monthly Page Views, Average Monthly Page Views, and Median Monthly Page Views for 23 Most Privacy-Sensitive Wikipedia Articles Over 32 Months Are All Consistent With A Chilling Effects Hypothesis

Raw Total Monthly Page Views: 23 Most Privacy Sensitive Articles

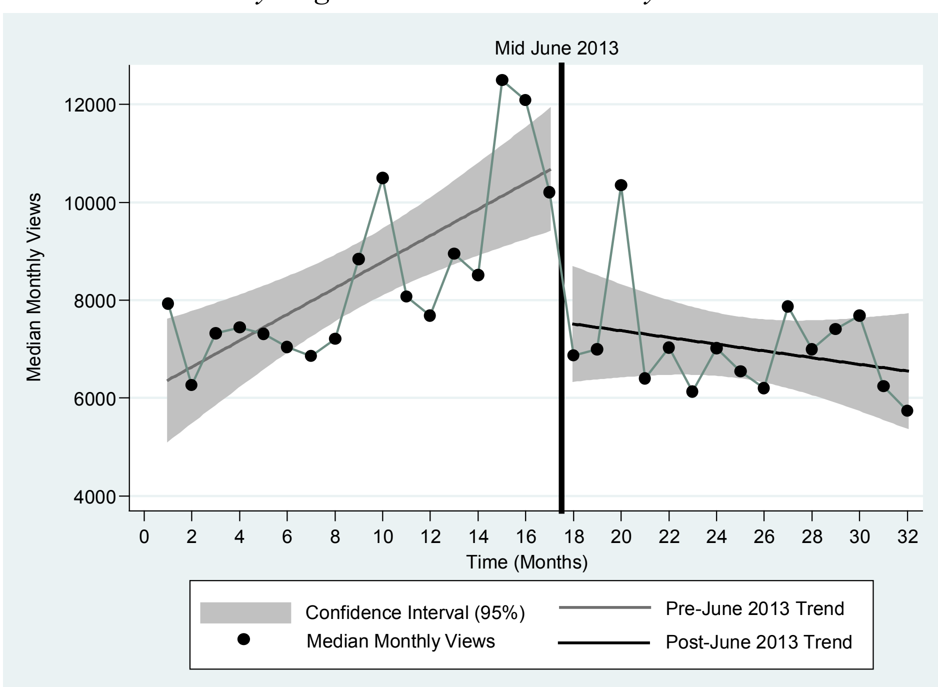


[9] This set of the 23 Most Privacy-Sensitive Article includes all articles in the 48 Terrorism Article group with a combined average privacy-sensitivity score greater than the median of those combined scores (2.07). This group includes: improvised explosive device (2.86), dirty bomb (2.81), car bomb (2.81), ammonium nitrate (2.61), biological weapon (2.60), chemical weapon (2.51), suicide attack (2.50), suicide bomber (2.44), Nuclear Enrichment (2.39), environmental terrorism (2.39), eco terrorism (2.39), weapons grade (2.39), jihad (2.35), Al Qaeda (2.34), terrorism (2.30), conventional weapon (2.27), Taliban (2.22), AL Qaeda in the Arabian Peninsula (2.17), Al Qaeda in the Islamic Maghreb (2.17), terror (2.15), Abu Sayyaf (2.14), Tehrik-i-Taliban Pakistan (2.12), and attack (2.08).

Average Monthly Page Views: 23 Most Privacy Sensitive Articles



Median Monthly Page Views: 23 Most Privacy Sensitive Articles



16. Here, in each graph set out in **Figure 4**, the total raw, average, and median monthly page views for these 23 articles increase in the time period leading up to June 2013. They also

demonstrate a statistically significant decline in June 2013, as well as a statistically significant trend reversal, with monthly page views declining after June 2013. Again, these findings are entirely consistent with a chilling effect hypothesis.

17. In the end, these results—focused on the most privacy-sensitive articles, analyzed both on an individual disaggregated analysis and aggregate monthly analysis—are entirely consistent with a chilling effect hypothesis. By ignoring the privacy theory upon which the chilling effect hypothesis is based, the Salzberg Declaration is deeply flawed.

18. **Fifth,** one of Salzberg's primary critiques of my analysis rests on a false premise: that my study "assumes a single peak in May 2013." This premise is false because my study makes no such assumption. My study hypothesizes a surveillance chilling effect in June 2013. Consistent with other ITS design studies, my study analyzes an outcome variable measured at consistent intervals (monthly privacy-sensitive Wikipedia article view data) to test that hypothesis over 32 months, by examining for statistically significant changes in level and trend in that data both before and after June 2013.

19. **Sixth,** Salzberg claims that my study's model can be altered to "prove" an April 2013 peak or earlier peak (based on a theory that the Boston Marathon bombings caused the page view trend reversal). However, he cites no cross-validation analysis to compare models and results to support his claim.[10] Cross-validation analysis is an established technique for understanding whether the results of a statistical test are robust – i.e., if we leave out datapoints at random from our dataset, do the results still hold? To answer that question and disprove Salzberg's theory, I

[10] LUKE JOHN KEELE, SEMIPARAMETRIC REGRESSION FOR THE SOCIAL SCIENCES 86 (Wiley & Sons, 2008) (describing "cross-validation" as a "general technique for assessing model fit based on resampling that can be applied to most statistical models").

conducted a "leave one out" cross-validation analysis[11] on both the 23 Most Privacy-Sensitive Wikipedia Articles set, as well as the larger 47 Terrorism Article set (the 48 Terrorism Articles without the Hamas article) to compare different statistical models based on a March, April, May, or June 2013 intervention effect. For comprehensive analysis, I used three data sets for each of these article sets—raw total page views, average monthly page views, and median monthly page views. Furthermore, for the 47 Terrorism Article Set, I excluded the "fundamentalism" article, making it a set of 46 total articles. (Salzberg noted that the "fundamentalism" article had too similar values to the "recruitment" article in the broader 48 Article set, so I have excluded it from this supplemental analysis. (Salzberg Decl. ¶ 7.)) I also performed the analysis both including and excluding the "ammonium nitrate" and "jihad" articles from the sets (Salzberg claims these two articles have outlier values).

20. The results of this cross-validation analysis show that for the 23 Article set of the Most Privacy-Sensitive Wikipedia articles, a statistical model based on a June 2013 intervention effect was superior to models based on March, April, and May 2013 interventions ***in every single data set analyzed*** (raw total monthly page views, average monthly page views, median monthly page views). That is, a June 2013 statistical model resulted in fewer estimation errors (lower room mean square errors (RMSE) and mean absolute errors (MAE)) than the other models. These results held even when the "ammonium nitrate" and "jihad" articles were removed from for analysis.

21. For the 46 Terrorism Article Set, a statistical model based on a June 2013

[11] *Id*. at 8 (describing "leave one out" cross validation as "probably the most commonly used method" as it "works well with most any sample size"; also that with "leave-one-out cross-validation, one observation is randomly selected and then omitted from the data set. The analyst then fits one of the possible models to this slightly truncated data set and calculates measure of fit. Next, a new data point is dropped, and the measure of fit is calculated again. This process is repeated as each of the data points is removed from the data set. The cross validation score is the averaged measure of model fit and can be used to compare different model specifications.")

intervention effect was also superior to models based on March, April, and May interventions for both the raw monthly page views set and the average monthly page views data set. These results held even when the "ammonium nitrate" and "jihad" articles were removed from for analysis. For the median monthly page view set, a June 2013 model was also superior to models based on both April and March 2013 interventions, but not May, where the RMSE and MAE scores for both models were very close (only a 2.5% difference in the MAE and 5% difference in the RMSE scores). These results all held or without "ammonium nitrate" and "jihad" in the sets.

22. In short, these results demonstrate the strength and robustness of my June 2013 model and its findings: it proved superior to comparable models in 46 of 48 total tests, and even in those two remaining tests, the difference in results were minimal. Moreover, when focused on the most privacy-sensitive Wikipedia articles, whether including or excluding the "jihad" and "ammonium nitrate" articles that Salzberg claims were outliers due to the Boston Marathon Bombing-- my model was a better "fit" to the data than the alternative models that Salzberg proposes based on earlier interventions in every single data set and scenario.

23. **Seventh,** Salzberg's comparative analysis of recent page view data is fundamentally undermined by the fact that Wikimedia's "page view" definition has changed over time. (*See* Salzberg Decl. ¶¶ 27-32.) Wikimedia has publicly published explanations on recent changes in the page view definition. (*See, e.g.*, Wikimedia Downloads: Analytics Datasets, https://dumps.wikimedia.org/other/analytics/; Research:Page View, https://meta.wikimedia.org/wiki/Research:Page_view.) The page view data that Salzberg relies on from July 2015 through November 2018 includes data on *mobile* page views, and therefore is incomparable to the data from the time period that I studied. As I explained in my Declaration, my study "used data for English language article view counts from stats.grok.se, an online portal

that provided access to *non-mobile* Wikipedia article view count data on a daily and monthly basis." (Decl. ¶ 34 (emphasis added); *see id.* at Table 3, 8, 9 (expressly indicating non-mobile data used.) Salzberg provides an example link to the Pageviews tool that he used to gather the more recent data, which shows that he selected all "Platform" types, including mobile. (*See* Salzberg Decl. ¶ 27, n. 17.) The difference between page views with non-mobile vs. mobile data included is often very significant, and therefore Salzberg's "extended data" comparison analysis is deeply flawed at the source and should be ignored. For example, using the "Hamas" example Salzberg offers, the difference between the "All" platforms and the "Desktop" (non-mobile) data for the month of May 2018 is over 100,000 views.

III. RESPONSES TO DR. SALZBERG'S SIX METHODOLOGICAL CRITIQUES

24. In his Declaration, Salzberg presents a series of purported critiques regarding my analysis. (Salzberg Decl. ¶¶ 47-66.) I respond to these issues below.

25. **Salzberg's first critique**: Aggregation "masks the differences in the changes over time by article" and was "performed without any analysis of the individual datasets" to determine whether it was the appropriate method. Standard methods for analyzing this kind of "panel data" were ignored. (Salzberg Decl. ¶¶ 48–50.)

26. **My Response:**

(a) As noted earlier, my method of analysis to test the June 2013 surveillance chilling effect hypothesis was an ITS design using aggregated data with segmented regression trend analysis. I chose an ITS design because it is an "ideal design" for assessing the impact of a "population-wide" intervention—like the effects of mass online government surveillance—that "affects the whole population and where randomization or a control group is impossible."[12] ITS

[12] N. BRUCE BASKERVILLE, ET AL., *IMPACT OF CANADIAN TOBACCO PACKAGING POLICY ON USE OF A TOLL-FREE QUIT-SMOKING LINE: AN INTERRUPTED TIME-SERIES ANALYSIS*, 6(1) CMAJ Open E59, E64 (2016)

design studies have also been commonly used in contexts like this one, to study information systems context (e.g., computing context)[13] and the impact of media coverage.[14] Within ITS design studies, use of segmented regression to analyze aggregated data to understand pre/post intervention trends in the data is not only "standard," but the ***recommended*** method and approach.[15]

(b) Second, there is no single determinative method or factor to decide whether an aggregated or disaggregated analysis of data is appropriate. Most ITS design studies use aggregated data,[16] because such time series designs "examine aggregate effects"[17] and are "strong designs for estimating the effects of instituting uniform, full-coverage programs or the effects of

---

(discussing "interrupted time-series design" as an "ideal design for assessing the effects of a population-wide intervention"; a "robust method for the evaluation of a policy that affects the whole population and where randomization or a control group is impossible"); RICHARD MCCLEAR ET AL., DESIGN AND ANALYSIS OF TIMES SERIES EXPERIMENTS 7, 297 (2017) (describing interrupted time series designs as "the major application of time series data for causal inference" and as a "strong quasi-experimental design… when random assignment was unfeasible"); CHESTER L. BRITT, DAVID J. BORDUA, & GARY KLECK, *A REASSESSMENT OF THE D.C. GUN LAW: SOME CAUTIONARY NOTES ON THE USE OF INTERRUPTED TIME SERIES DESIGNS FOR POLICY IMPACT ASSESSMENT*, 30 Law & Soc'y Rev. 361, 361 (1996) ("Interrupted time series designs provide one of the most common means for assessing the impact of a change in law or in social policy"); D.T. CAMPBELL, *REFORMS AS EXPERIMENTS*, 24(4) American Psychologist 409 (1969) (this seminal article by Campbell was among the first to advocate for interrupted time series designs in cases where natural experiments are not possible); WAGNER ET AL., *supra* note 1, at 308 (describing ITS designs as the "strongest, quasi-experimental designs" to estimate intervention effects in "non randomized settings").

[13] *See e.g.*, S. ASGARI & NUNES BAPTISTA, *EXPERIMENTAL AND QUASI-EXPERIMENTAL RESEARCH IN INFORMATION SYSTEMS,* IADIS International Workshop Information Systems Research Trends: approaches and methodologies (ISRTAM 2011), 20-26 July (noting ITS designs have been "used often in the field of [Information Systems]").

[14] *See e.g.,* MELANIE A WAKEFIELD, BARBARA LOKEN,& ROBERT C HORNIK, *USE OF MASS MEDIA CAMPAIGNS TO CHANGE HEALTH BEHAVIOR*, 376 The Lancet 1261, 1262-1263, (2010) (discussing interrupted time series analyses studies in the health context); RANDY ELDER ET AL., *EFFECTIVENESS OF MASS MEDIA CAMPAIGNS FOR REDUCING DRINKING AND DRIVING AND ALCOHOL-INVOLVED CRASHES*, 27(1) Am. J. Prev. Med. 57 (2004) (ROBERTO GRILLI ET AL., *MASS MEDIA INTERVENTIONS: EFFECTS ON HEALTH SERVICES UTILIZATION*, 1 Cochrane Database of Systematic Reviews 1, 1 (2002) (providing a comprehensive review of research studying the impact of media coverage on health service use— and noting that among the "twenty studies" reviewed in the work, all used interrupted time series designs).

[15] See works cited at *supra* note 2 and accompanying text.

[16] See also works cited at *supra* note 3 and accompanying text.

[17] JEFFREY M. WOOLDRIDGE, INTRODUCTORY ECONOMETRICS A MODERN APPROACH 15 (5th ed., 2012) ("…time series data are often used to look at aggregate effects. An example of a time series data set on unemployment rates and minimum wages…").

making changes in such programs."[18] NSA surveillance, including Upstream surveillance, was "uniform" and "full coverage," within the meaning of this guidance. It also has aggregate effects, as the entire U.S. Internet-using population is subject to their reach. It was therefore appropriate that my study employed aggregate page view analysis, since I sought to make aggregate level inferences about large scale NSA surveillance effects.[19]

(c) Third, aggregated data and analysis is further appropriate in ITS studies where the aim is to explore national or major regional rates and trends;[20] to reduce or remove "noise" in the data;[21] and to allow for more sophisticated statistical tests and analysis.[22] All of these circumstances apply to this study. First, my study was focused on evaluating the large-scale national aggregated effects or impact of mass awareness of NSA surveillance in June 2013 and after. Thus, examining the Wikipedia article page view data in aggregate is consistent with that aim. Second, analyzing the Wikipedia articles in aggregate helped reduce "noise" in the data given that, inevitably, individual Wikipedia article page views would like fluctuate and vary widely over 32 months; if analyzed in aggregate, broader overall trends or patterns could be discerned. Analyzing aggregated data makes particular sense in this ITS study, as Wikipedia page view data has a particularly high signal-to-noise ratio—that is, where the signal or "true" patterns in data (like longer term trends due by chilling effects) may be obscured by "noise," that is, more

---

[18] PETER ROSSI ET AL., EVALUATION: A SYSTEMATIC APPROACH 352-354 (6th ed., 1999) (noting "[m]ost existing times series involve aggregated data" as they often aim to study national, state, or large regional subjects).

[19] L. LEYDESDORFF, *THE SCIENCE CITATION INDEX AND THE MEASUREMENT OF NATIONAL PERFORMANCE IN TERMS OF NUMBERS OF SCIENTIFIC PUBLICATIONS*, 1-2 Scientometrics 111, 113 (1989) ("In general, one should prefer aggregated data for inferences at the aggregated level, since otherwise methodological problems of inference may emerge.").

[20] ROSSI ET AL., *supra* note 18, at 354.

[21] BEARD, *supra* note 4, at 2 (noting that a "[m]ajority of studies use aggregated data" as this "removes noise and allows for more sophisticated tests which require continuous or rate type data").

[22] *Id*.

temporary variations for individual article page views for other reasons.[23] Several prior studies have observed that Wikipedia page view data has such "noise" and those studies likewise used aggregated page views for analysis.[24] Third, analyzing the Wikipedia article page view data in aggregate also allowed for the more sophisticated statistical tests and analysis. In this case, that was segmented regression trend analysis pre/post June 2013, which is the recommended method of analysis for ITS design studies. In short, an aggregated analysis of the Wikipedia article page view data was both a "standard" method and entirely justifiable.

(d) Fourth, my opinions *are* supported by an analysis of individual article page views. Specifically, I examined the page view trends for individual and smaller groups of articles with higher privacy-sensitivity scores to verify the results of my aggregate data analysis showing: (1) a statistically significant drop in June 2013, and (2) trend change in monthly page views from increasing views before that month to a monthly decline after that month. As the chilling effect hypothesis I was testing is based on privacy sensitivity, then page view trends for the most privacy-sensitive Wikipedia articles would reveal page view trends consistent with the aggregate data results. As seen from **Figures 3a, 3b, and 4**, those page views were consistent with a June 2013 chilling effects hypothesis. Furthermore, I also examined and analyzed individual article page views to identify and investigate outliers. For example, this was done with the "Hamas" article, among others. Examining individual and aggregate level data allowed me to identify overly

---

[23] N. GENEROUS, ET AL., *GLOBAL DISEASE MONITORING AND FORECASTING WITH WIKIPEDIA*, 10(11) PLoS Comput Biol 1, 8, 12 (2014)(study on using Wikipedia page view data to track and predict global diseases using aggregated data as "signal-to-noise" ratio in Wikipedia page view data may mean "[t]rue patterns in data may be obscured by noise", in this case, the "noise" being variations in page views of health information on Wikipedia unrelated to the personal diagnosis); J.D. SHARPE, ET AL., *EVALUATING GOOGLE, TWITTER, AND WIKIPEDIA AS TOOLS FOR INFLUENZA SURVEILLANCE USING BAYESIAN CHANGE POINT ANALYSIS: A COMPARATIVE ANALYSIS*, 2(2) JMIR public health and surveillance 1, 2, 4 (2016) (noting the "signal-to-noise" ratio in Wikipedia page view data can be "problematic" and also aggregating).

[24] *Id*.

influential articles. This is a standard aspect of conducting regression model diagnostics. Additionally, I used individual articles as part of my cross-validation analysis.

(e) Fifth, Salzberg describes my data as "panel data." This is accurate in the broad definition of the term—my data is "longitudinal," involving repeated measurements from a group to study the impact of an intervention (monthly measurements of individual Wikipedia article page views aggregated into the larger sets).[25] However, using more precise definitions informed by the aim and design of my study, the data is more accurately described as time series data. This is because the subject of my study is the aggregate Wikipedia page view trends both before and after June 2013 in order to test the aggregate effects of NSA surveillance on Wikipedia and its users. In short, *time* is a central unit of analysis in my study.[26] By contrast, panel data and panel studies typically follow *individuals* over time, and include multiple-dimensional observations from each individual (e.g., schooling, employment, marital status, training, child rearing, health, etc.).[27] This conclusion is supported by Woolbridge in *Introductory Econometrics: A Modern Approach*, the same text Salzberg cites. It states that a "time series data set consists of observations on a variable or several variables over time," with time an "important dimension in a time series data set."[28] As I was interested in examining the aggregate effects of media coverage of NSA surveillance introduced at a specific point in time (June 2013), and trends before and after that time, the "variable" in my study was aggregate Wikipedia page views, and a central dimension was time. Woolbridge also describes panel data as longitudinal data that follows or "attempts to follow" individuals over time, collecting multiple observations on a range of different data points,

---

[25] CHENG HSIAO, PANEL DATA ANALYSIS 1 (CUP, 2017); ROSSI ET AL., *supra* note 18, at 352.
[26] GREGORY B. MARKUS, ANALYZING PANEL DATA 7-8 (sage, 1979) (describing the difference between time series data and panel data, with the former having "time" as the central unit of analysis, while panel studies take individuals as the central unity of analysis).
[27] HSIAO, *supra* note 25, at 1; MARKUS, *id* 7-8.
[28] WOOLBRIDGE, *supra* note 17, at 8.

with repeated measures not necessarily uniform or taken at regular intervals.[29] This creates additional layers of complexity for analysis suited for panel studies.[30] The "standard" methods for that Salzberg recommends at paragraph 50 of his Declaration, in fact, concern panel data analysis, specifically data collected at only two or three points in time.[31] These are not applicable to my time series data set collected at regular intervals over 32 months.

27. **<u>Salzberg's second critique</u>**: Penney's model "assumes a single peak in May 2013" rather than "letting the data reveal where, if anywhere a peak in the data exists." Penney's model can be altered to "prove" an April 2013 peak and support the theory that the Boston Marathon bombings caused the page view trend reversal. A "polynomial model" further shows that Penney's hypothesized peaks in page views were incorrect. (Salzberg Decl. ¶¶ 51–54.)

28. **<u>My Response:</u>**

(a) First, my ITS design does not "assume a single peak"—it tests for the effects of real-world events happening at a particular time—June 2013. In order to test a surveillance chilling effect hypothesis in that month, I examined page view trends before an intervention point. This is a standard approach in naturalistic studies like this, where the aim is to test the impact of an intervention at a given point in time (here, NSA surveillance revelations and media coverage in June 2013). Any peak before or after June 2013 arises from the data itself, and is not any assumption or requirement in my model.

(b) Second, Salzberg's approach of "letting the data reveal where, if anywhere, a peak in the data exists," is not a sound or reliable social scientific approach and can lead to substantial bias in results. My ITS study, designed to examine a June 2013 impact due to media

---

[29] WOOLBRIDGE, *supra* note 17, at 448.
[30] WOOLBRIDGE, *supra* note 17, at 448.
[31] WOOLBRIDGE, *supra* note 17, at 459-474.

coverage of NSA surveillance programs, was based on a hypothesis grounded in existing empirical and theoretical research on privacy, surveillance studies, and chilling effects. By contrast, Salzberg's approach of visually inspecting data and running various statistical models, including a "polynomial model," until a "fit" showing "earlier peaks in 2013" is found, is a biased approach.[32] Any earlier "peak" or polynomial model is not grounded on any *a priori* hypothesis, theory or research. Nevertheless, even assuming a "polynomial model" estimating earlier peaks in 2013 is not biased. It also does not discount or disprove a chilling effects hypothesis in June 2013. My statistical model and results did not require or assume any such "peak" in May. A statistically significant drop in June 2013 based on prior trends, or a reduction in monthly views thereafter, would each be consistent with a chilling effect hypothesis, notwithstanding any earlier peaks in the page view data in the 32 months. My analysis found both, but neither required any "peak" in May or April 2013.

(c) Third, my statistical model based on a June 2013 intervention is a superior fit for the page view data compared to models based on March, April, and May 2013 interventions in every single data set analyzed. This is demonstrated by my cross-validation analysis on both the 23 Article Set (raw monthly page views, average monthly page views, median monthly page views) and the 28 Article Set (raw monthly page views, average monthly page views, median monthly page views).[33] My June 2013 model was a better "fit" when focused on the page view data for the most privacy-sensitive Wikipedia articles compared to Salzberg's alternative models,

---

[32] ANDREW GELMAN & ADAM ZELIZER, *EVIDENCE ON THE DELETERIOUS IMPACT OF SUSTAINED USE OF POLYNOMIAL REGRESSION ON CAUSAL INFERENCE*, Research & Politics (2015) (describing reported effects, based on a "curve fitting" polynomial model that is "statistically significant but substantively dubious, and are sensitive to model choice"); MEGAN L. HEAD ET AL., *THE EXTENT AND CONSEQUENCES OF P-HACKING IN SCIENCE*, 13(3) PLoS Biol 1 (2015) ("Inflation bias… occurs when researchers try out several statistical analyses and/or data eligibility specifications and then selectively report those that produce significant results.").

[33] *See supra* ¶¶ 16-17.

notwithstanding any impact of the Boston Marathon media coverage in April or May 2013. In short, my ITS approach, statistical regression model, and method of analysis is the best way to understand the page views trends for the most privacy-sensitive Wikipedia articles.

29. **Salzberg's third critique:** Penney's model is "oversimplified, leaving out virtually all factors that could affect page views of terror-related articles from the model." For example, the model fails to account for seasonality or major news events. The model tacitly acknowledges this failure in how it handles the Hamas outlier data, which is ultimately manipulated in a way that is favorable to the hypothesis. Penney does not consider other real world variables that may not be favorable to the hypothesis, like the Boston Marathon bombings, which happened shortly before the NSA disclosures. (Salzberg Decl. ¶¶ 55–60.)

30. **My Response:**

(a) My analysis ***does*** account for various external factors that may affect page views in the model. First, while seasonality is a confounding concern in ITS designs, there is no basis to expect large seasonal effects with these page views—that, for example, Internet users tend to view terrorism-related content in the spring but not in the summer. In any event, to account for possible seasonality and seasonal effects in the data in the ITS design, I went beyond the "general recommendation" for 12 data points before and after the hypothesized intervening chilling effect (June 2013),[34] and instead collecting data for 17 data points before (January 2012 through May 2013) and 14 data points following (July 2013 through August 2014). This longer study period allows for better assessment of overall page view trends by identifying and accounting for seasonal trends. This is particularly challenging with Wikipedia page view data because of its high signal-

---

[34] WAGNER ET AL., *supra* note 2, at 301 ("A general recommendation is for 12 data points before and 12 data points after the intervention (8), although this number is not based on estimates of power. Rather, with 24 monthly measures, the analyst can adequately evaluate seasonal variation.").

to-noise ratio, that is, variability of page views over the course of a study period due to a range of daily factors.[35]

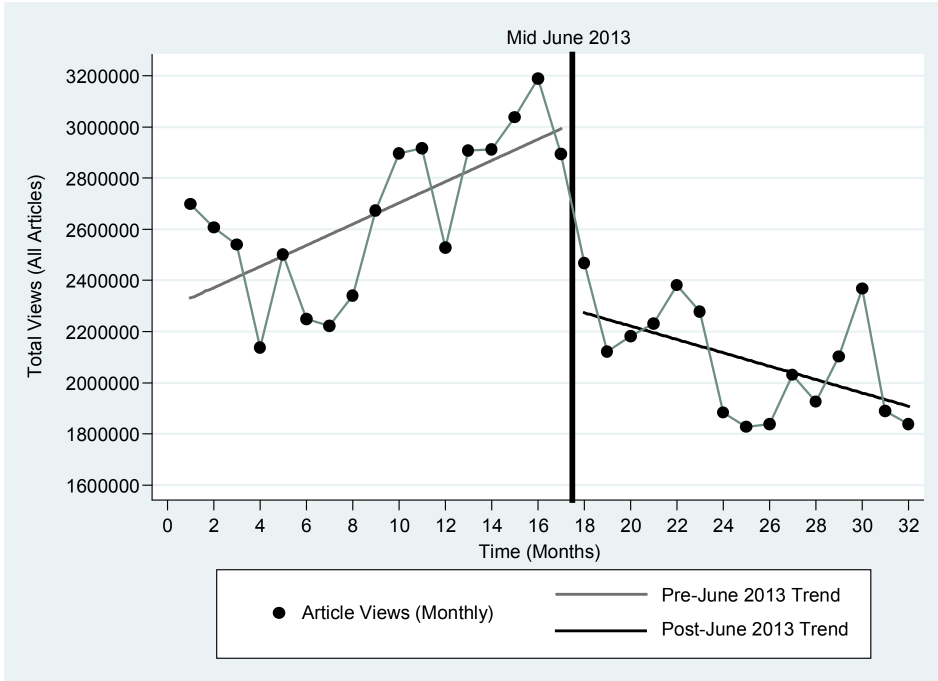
(b) Second, the data ***was*** also analyzed for seasonal trends as well for real-world events to determine whether any such events would have an outsized effect on page views. This is precisely how I identified the "Hamas" article as an outlier and related it to real world events. This was not done by cherry-picking or manipulation, but through standard regression model diagnostics, as well as best practices for identifying and dealing with outlier observations.[36]

(c) Third, no seasonal or "real world" event-related variations identified by Salzberg explain the actual trends apparent in the aggregated data before and after June 2013. Salzberg points to a seasonal "trough" in the summer of 2012 and a "peak" due to the Boston Marathon bombing in April 2013 for page views in the 47 Terrorism articles. These are visible in **Figure 5** below:

[35] See works cited at *supra* note 23 and accompanying text.

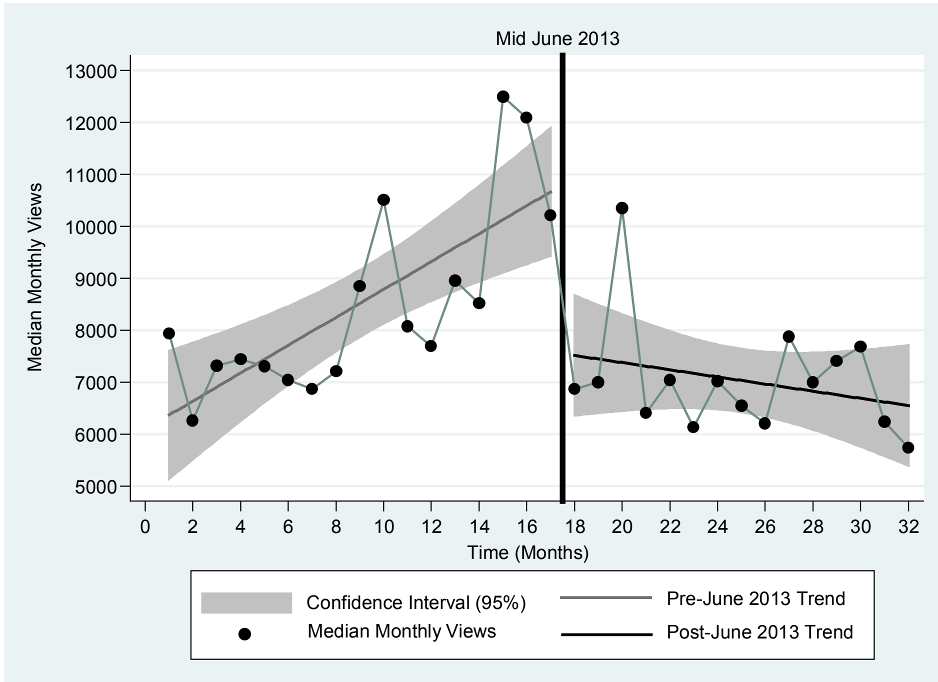
[36] The Hamas article was identified using standard regression model diagnostics, including examining cook's distance, leverage, residual values, z-scoes, among others. Handling of the influential outlier was also done according to best practices, with the outlier removed but results before and after removal conveyed. See AGUINAS ET AL., *supra* note 8, at 20-23.

Figure 5: Aggregated Total Monthly Page Views for 48 Terror Articles (Without Hamas)



However, the added trend analysis of page views in **Figure 5** demonstrates that despite these possibly seasonal and "real world" event-related variations over the course of the 17 months prior to June 2013 and 14 mounts after that period, there is no variation or "trough" comparable to the one that occurs between the total page views as of the beginning of June 2013 (2,893,553 page views) and that at the end of July 2013 (2,121,335). Indeed, during time period there was a decline of 772, 218 page views, or roughly 27%. By contrast, the decline between April and May 2013 that Salzberg highlights (294, 820 page views) is slightly greater than 9%. Nor is there any point in the entire 32-month study period where there are consistently fewer monthly page views than in the months of December 2013 through August 2014. These points are consistent with a June 2013 chilling effects hypothesis. These points are even clearer when examining the median monthly page views for the 23 Most Privacy Sensitive Articles visualized in **Figure 6:**

Figure 6: Aggregated Median Monthly Page Views for 23 Most Privacy-Sensitive Articles



(d) Here, again, there is a significant drop from the median monthly page views as of the beginning of June 2013 (12,090) and the end of July 2018 (6,864) totaling 5,226, which is slightly greater than a 45% drop in median monthly views. Nor, despite some variation, is there any other period in the 32 months where monthly median page views are trending so clearly lower over time as the months after June 2013 onto August 2014. Again, this is consistent with a chilling effects hypothesis. Salzberg offers no alternative explanation beyond identifying "peaks" and "troughs" in 2012 and 2013, which do not account or explain these findings.

(e) Fourth, in a naturalistic study outside the experimental context, it is not possible to control for all confounding factors, like the impact of all real world events on page views over 32 months. However, the ITS design was chosen for the very fact that its pre/post design *can* help control for other explanatory factors as any such known or unknown confounding variables would be present in both the pre and post measurements (monthly page views), thus any changes after the intervention (June 2013) can be attributed to intervention itself. This dimension of the ITS design can be further strengthened by adding one or more comparators,[37] which was also done in this study. Comparators help control for confounding factors and seasonality.[38] Here, the comparator groups—security, infrastructure, popular, and Wikipedia English home-page traffic—not expected to be impacted or affected by the ITS design intervention (here, surveillance chilling effects in June 2013) can be compared to the study group (page views for privacy-sensitive terrorism related Wikipedia articles). As my results showed, the privacy-sensitive terrorism Wikipedia were impacted (statistically significant drop in June 2013 and trend reversal after that month) while none of the comparator articles showed the same effects. This is consistent with a chilling effect hypothesis.

(f) Salzberg claims that my comparator groups do not corroborate my findings as they are not "proper" controls groups that "exhibit the trend" shown by the terrorism articles before June 2013.[39] Of course, comparator groups identical to the study group are ideal but are often not feasible.[40] In fact, Campbell and Stanley, in their leading 1966 text *Experimental and Quasi-Experiment Designs for Research* that Salzberg cites, recommend that for an ITS study of a "major administrative change" that a researcher use a "similar institution" as a comparator not expected to undergo the change and upon which the same ITS design can be tested.[41] Wagner et al.'s (2002) leading article on ITS design recommends that where an identical control group is not

---

[37] R. BARKER BAUSELL, *THE DESIGN AND CONDUCT OF MEANINGFUL EXPERIMENTS INVOLVING HUMAN PARTICIPANTS 199* (OUP, 2015) (noting that an ITS design can be "significantly buttressed" by adding one or more comparator group); MARY A. M. ROGERS, *COMPARATIVE EFFECTIVENESS RESEARCH 94* (OUP, 2014).
[38] BAUSELL, *id.* at 200; Wagner et al., *supra* note 2, at 306.
[39] Salzberg Decl. at ¶¶ 33-46.
[40] BAUSELL, *id.* at 199; Wagner et al., *supra* note 2, at 306.
[41] DONALD CAMPBELL & JULIAN STANLEY, *EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR RESEARCH 55* (Houghton-Mifflin, 1966).

possible, a study may examine a "different but related group . . . not expected to change following the intervention, in the same group of subjects."[42] This is precisely the methodology that I adopted. Since identical or randomly sampled control groups were not possible, I used comparator groups drawn from a "different but related group" of Wikipedia articles drawn from a "different but related" group of DHS key words (security and infrastructure), as well as a set of popular Wikipedia articles. Given that security, infrastructure, and popular Wikipedia articles do not raise privacy concerns, they would not be expected to change after a surveillance chilling effect intervention in June 2013. We would also expect similar viewer audiences for these articles (e.g. someone with an interest in national security would be just as likely to view Wikipedia articles on terrorism as domestic or infrastructure security articles, the comparators in this study). My results showed that they were not impacted, while the privacy-sensitive terrorism-related articles were. This, as noted, was consistent with a surveillance chilling effect hypothesis.

31. **Salzberg's fourth critique:** The model did not take into account that the 48 terror articles chosen based on 2011 DHS list would naturally rise and decline in interest over time. In other words, the 2011 terrorism-related key words would "undoubtedly become stale over time." The same is true for the trends in the comparator article groups. (Salzberg Decl. ¶¶ 61–64.)

32. **My Response:**

(a) The articles in my study were chosen based on keywords associated with "terrorism" that DHS uses to track and monitor social media. Since the media coverage relating to the Snowden revelations framed the issue of NSA surveillance as a matter of national security and terrorism threats, it was logical for me to use DHS keywords associated with "terrorism" to create the Wikipedia articles that represent the sort of articles that users may be chilled from accessing

[42] WAGNER ET AL., *supra* note 2, at 306.

in light of privacy concerns about government surveillance. I used this approach for pragmatic methodological reasons, as there is no pre-determined list reflecting all privacy-sensitive Wikipedia articles from which to draw a random sample. Furthermore, cherry-picking a list of articles relating to certain sensitive topics (like Syria or ISIL, as Salzberg suggests) would be subject to serious selection biases. Using government keyword lists to study government surveillance is not a novel or unprecedented approach;[43] indeed, my methodology is similar to the methodology of an earlier, peer-reviewed study exploring the chilling effects associated with the NSA surveillance in Google search data, which also used these DHS keyword lists.[44]

(b) The fact that some likely privacy-sensitive articles (like Syria and ISIL) that were not included in the DHS keyword list may have recorded higher page views for some period of time during my study does not in any way undermine the overall conclusions for the privacy-sensitive Wikipedia articles examined over the entire 32-month period that I studied.

(c) Salzberg's Declaration offers no evidence for his claim that the "many of the 2011 terrorism-related keywords undoubtedly became stale over time" and thus "page views dropped."

33. **Salzberg's fifth critique:** The page view data examined only extends for 32 months through August 2014, therefore the "results do not and cannot imply that an effect of the June 2013 disclosures persists today, or did so even in 2015." (Salzberg Decl. ¶ 65.)

---

[43] JEDIDIAH R. CRANDALL & MASASHI CRETE-NISHIHATA ET AL., *CHAT PROGRAM CENSORSHIP AND SURVEILLANCE IN CHINA: TRACKING TOM-SKYPE AND SINA UC*, First Monday, July 1, 2013, http://firstmonday.org/ojs/index.php/fm/article/view/4628/3727 [https://perma.cc/M5FJ-T4D5]; JEFFREY KNOCKEL, JEDIDAH CRANDALL & JARED SAIA, *THREE RESEARCHERS, FIVE CONJECTURES: AN EMPIRICAL ANALYSIS OF TOM-SKYPE CENSORSHIP AND SURVEILLANCE*, 16:4 FOCI '11: USENIX Workshop on Free & Open Comm. on Internet (2011), https://www.cs.unm.edu/~crandall/foci11knockel.pdf [https://perma.cc/FH8H-JUBA].

[44] ALEX MARTHEWS & CATHERINE TUCKER, *GOVERNMENT SURVEILLANCE AND INTERNET SEARCH BEHAVIOR*, *IN* CAMBRIDGE UNIVERSITY HANDBOOK ON SURVEILLANCE LAW (David Gray et al. eds., 2017).

34. **My Response:**

(a) The statistically significant trend reversal from increasing monthly views prior to June 2013 to a downward trend, with a monthly reduction in page views afterwards, is indicative of a lasting chilling effect. This is supported by other research on long-term online chilling effects due to public awareness of surveillance.

(b) First, an Massachusetts Institute of Technology (MIT) study on Google search data later published a peer reviewed chapter in the *Cambridge University Handbook on Surveillance Law*, found a statistically significant reductions in privacy-sensitive Google searches after the June 2013 Snowden disclosures about NSA surveillance.[45] The findings, the authors concluded, provided "substantial empirical documentation of a chilling effect," both in the "shorter term" and "in the longer term", that "appeared to be related to increased awareness of government surveillance online."[46]

(c) Second, a 2017 peer reviewed study on Wikipedia editors found evidence, based on qualitative interviews in the spring and summer of 2015, that editors were chilled from certain activities on Wikipedia due to awareness of government surveillance.[47] For example, one Wikipedia editor stated, "for the Edward Snowden page, I have pulled myself away from adding sensitive contributions, like different references, because I thought the name may be traced back to me in some way." The fact that some Wikipedia users have avoided sensitive or controversial topics in 2015, two years after the Snowden revelations in 2013, is consistent with my findings.

---

[45] ALEX MARTHEWS & CATHERINE TUCKER, *GOVERNMENT SURVEILLANCE AND INTERNET SEARCH BEHAVIOR* 1, 3-4 (MIT Sloane Working Paper No. 14380, 2015); MARTHEWS & TUCKER, *supra* note 44.
[46] ALEX MARTHEWS & CATHERINE TUCKER, *GOVERNMENT SURVEILLANCE AND INTERNET SEARCH BEHAVIOR* 1, 3-4 (MIT Sloane Working Paper No. 14380, 2015); MARTHEWS & TUCKER, *supra* note 44.
[47] ANDREA FORTE, NAZANIN ANDALIBI, AND RACHEL GREENSTADT, *PRIVACY, ANONYMITY, AND PERCEIVED RISK IN OPEN COLLABORATION: A STUDY OF TOR USERS AND WIKIPEDIANS*, *in* CSCW 1800 (2017).

(d) Third, a recent 2018 study exploring how journalists have been impacted by "potential surveillance by government," which involved qualitative interviews with American journalists in 2015, found all seven journalists in the study indicated that "their work and lives have changed under a real or perceived threat of mass government surveillance."[48] The author concluded that "participants reported an increased awareness of mass government surveillance" and "[i]n every case, they reported adjusting their behavior to some degree."[49] This is also consistent with my findings.

(e) Fourth, a Pew Research Center survey of 475 adult Americans conducted between November 26, 2014 and January 3, 2015 found that, among the 87% of respondents aware of "government surveillance programs" due to the Snowden revelations, 34% had taken "at least one step to hide or shield their information from the government," including avoiding using "certain terms in online communications."[50] The survey also found 25% changed the patterns of their own use of various online platforms "a great deal" or "somewhat" since the Snowden revelations. These findings from a survey administered in late 2014 and early 2015 are also consistent with my conclusions.

(f) Fifth, a PEN America survey of 520 American writers in October 2013[51] found that 28% of the writers surveyed had "curtailed or avoided" certain online activities due to "fear of surveillance" and another 12% "seriously considered" doing so; 24% "deliberately avoided certain topics in phone or email conversations," and another 9% have "seriously

---

[48] STEPHENSON WATERS, *THE EFFECTS OF MASS SURVEILLANCE ON JOURNALISTS' RELATIONS WITH CONFIDENTIAL SOURCES, DIGITAL JOURNALISM,* 6:10 Digital Journalism 1294, 1310 (2018).
[49] *Id*. at 1310.
[50] LEE RAINIE ET AL., *PEW RES. INTERNET PROJECT*, Americans' Privacy Strategies Post-Snowden 4 (Mar. 16, 2015), http://www.pewinternet.org/files/2015/03/PI_AmericansPrivacyStrategies_0316151.pdf.
[51] FDR GROUP & PEN, AMERICAN CENTER, *CHILLING EFFECTS: NSA SURVEILLANCE DRIVES U.S. WRITERS TO SELF-CENSOR 3–4* (2013), http://www.pen.org/sites/default/files/ Chilling%20Effects_PEN%20American.pdf.

considered it"; and 16% have refrained from "conducting Internet searches or visiting websites on topics that may be considered controversial or suspicious," and another 12% have "seriously considered it." These results are consistent with my conclusions as to a long term chilling effect.

(g) Also, as explained above, Salzberg's extended comparison analysis that relies on more recent page view data is fundamentally invalid because it compares across page view definitions—the more recent data includes "mobile" page views, while my study relied on "non-mobile" data.[52]

35. **<u>Salzberg's sixth critique:</u>** Penney's model fails to isolate the "particular effect of public 'awareness' about the NSA Upstream program" from potential other effects of the Snowden disclosures, including increased awareness about other NSA surveillance activities. (Salzberg Decl. ¶ 66.)

36. **<u>My Response:</u>**

(a) In any study of naturalistic changes in human behavior, it will not be possible to isolate the source of all causes and effects on behavior. It is enough for purposes of establishing whether Upstream likely had a chilling effect on Wikipedia users that the reporting on NSA surveillance in June 2013 included multiple references to international Internet communications monitoring, and that general public awareness of NSA surveillance grew due to media coverage after June 2013.[53] As earlier noted, a Pew Research Center survey of adult Americans conducted between November 26, 2014 and January 3, 2015 found 87% of respondents were aware of NSA surveillance programs due to the Snowden revelations.[54]

(b) Furthermore, in-line with the empirical conclusions of my Declaration,

---

[52] *See supra* ¶ 23.
[53] See ¶¶ 32-33 of my Declaration dated December 18, 2018.
[54] *See supra* ¶ 30(e).

Wikimedia has introduced other evidence establishing the particular chilling effect that awareness of Upstream surveillance had on Wikimedia's readers and contributors. The Declarations of Michelle Paulson and James Alexander describe the chilling effect that Upstream surveillance had on the Wikimedia community at large, particularly among users abroad who engage with the platform concerning privacy-sensitive topics. (*See* Pl.'s Ex. 3, ¶¶ 41, 45, 46; Pl.'s Ex. 4, ¶¶ 4-11.) The Declaration of Emily Temple-Wood, an active Wikimedia community member, further describes first-hand the chilling effect that awareness of Upstream surveillance has had among the community of readers and contributors. (*See* Pl.'s Ex. 6, ¶¶ 20-21.)

(c) Salzberg's critique that the "particular effect" of Upstream cannot be entirely isolated is not actually a methodological critique, but rather, a general observation about a naturalistic studies. However, courts have rejected such challenges when ruling on *Daubert* motions. *See A Woman's Choice-East Side Women's Clinic v. Newman,* 904 F. Supp. 1434 (S.D. Ind. 1995) (upholding naturalistic study against *Daubert* challenge when ruling on preliminary injunction motion).

I declare under penalty of perjury under the laws of the United States that the foregoing is true and correct to the best of my knowledge and belief.

Executed on March 8, 2019 in Halifax, Canada.

Jonathon Penney

APPENDIX

23 Most Privacy Sensitive Article Set
Cross Validation Analysis

RAW total monthly views

```
.
. reg HP23RawViews time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.82
       Model |  3.1487e+11     3  1.0496e+11           Prob > F      =  0.0000
    Residual |  1.8571e+11    28  6.6324e+09           R-squared     =  0.6290
-------------+------------------------------           Adj R-squared =  0.5893
       Total |  5.0057e+11    31  1.6148e+10           Root MSE      =   81439

------------------------------------------------------------------------------
HP23RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   21383.58   4031.855     5.30   0.000      13124.7    29642.46
intervention |  -224931.2   58212.11    -3.86   0.001    -344173.3   -105689.1
   postslope |  -29367.59   6320.044    -4.65   0.000    -42313.61   -16421.56
       _cons |   349787.6   41314.22     8.47   0.000     265159.3      434416
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   89506.354
Mean Absolute Errors     |   63503.274
Pseudo-R2                |     0.49622
-----------------------------------------

. reg HP23RawViews time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.34
       Model |  2.8498e+11     3  9.4993e+10           Prob > F      =  0.0000
    Residual |  2.1560e+11    28  7.6998e+09           R-squared     =  0.5693
-------------+------------------------------           Adj R-squared =  0.5232
       Total |  5.0057e+11    31  1.6148e+10           Root MSE      =   87749

---------------------------------------------------------------------------------
   HP23RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
           time |    22020.4   4758.837     4.63   0.000     12272.36    31768.43
interventionMAY |  -152067.1   62229.88    -2.44   0.021    -279539.2   -24594.96
   postslopeMAY |  -34404.22   6730.013    -5.11   0.000    -48190.03   -20618.42
          _cons |   345966.8   46015.77     7.52   0.000     251707.7    440225.8
---------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   100292.59
Mean Absolute Errors     |   71401.914
Pseudo-R2                |     0.37718
-----------------------------------------
```

```
. reg HP23RawViews time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.66
       Model |  2.4087e+11     3  8.0289e+10           Prob > F      =  0.0003
    Residual |  2.5971e+11    28  9.2752e+09           R-squared     =  0.4812
-------------+------------------------------           Adj R-squared =  0.4256
       Total |  5.0057e+11    31  1.6148e+10           Root MSE      =   96308

------------------------------------------------------------------------------------
     HP23RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |    13623.8   5755.506     2.37   0.025     1834.181    25413.42
interventionAPRIL |   65634.32   68033.27     0.96   0.343    -73725.51    204994.1
   postslopeAPRIL |   -34751.7   7473.906    -4.65   0.000    -50061.31    -19442.1
            _cons |   393547.5   52329.76     7.52   0.000     286354.8    500740.1
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   108143.33
Mean Absolute Errors     |   69579.447
Pseudo-R2                |      0.28104
-----------------------------------------

. reg HP23RawViews time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.46
       Model |  2.3801e+11     3  7.9335e+10           Prob > F      =  0.0004
    Residual |  2.6257e+11    28  9.3775e+09           R-squared     =  0.4755
-------------+------------------------------           Adj R-squared =  0.4193
       Total |  5.0057e+11    31  1.6148e+10           Root MSE      =   96837

------------------------------------------------------------------------------------
     HP23RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   13450.41   6420.254     2.09   0.045     299.1154     26601.7
interventionMARCH |   88472.13   68403.98     1.29   0.206    -51647.07    228591.3
   postslopeMARCH |  -33519.11   7782.969    -4.31   0.000     -49461.8   -17576.42
            _cons |   394472.2   54666.49     7.22   0.000       282493    506451.4
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   105572.3
Mean Absolute Errors     |   69332.589
Pseudo-R2                |      0.30842
-----------------------------------------
```

Average monthly views

```
.
. reg HP23AvgViews time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.83
       Model |  595220203     3  198406734           Prob > F      =  0.0000
    Residual |  351044760    28  12537312.9           R-squared     =  0.6290
-------------+------------------------------           Adj R-squared =  0.5893
       Total |  946264964    31  30524676.2           Root MSE      =  3540.8

------------------------------------------------------------------------------
HP23AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   929.7206   175.2961     5.30   0.000     570.6429    1288.798
intervention |  -9779.688   2530.932    -3.86   0.001    -14964.07   -4595.308
   postslope |  -1276.853   274.7814    -4.65   0.000    -1839.717   -713.9885
       _cons |   15208.16    1796.25     8.47   0.000     11528.71    18887.61
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |    3891.5395
Mean Absolute Errors     |    2760.9415
Pseudo-R2                |      0.49623
-----------------------------------------

. reg HP23AvgViews time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.34
       Model |  538718524     3  179572841           Prob > F      =  0.0000
    Residual |  407546439    28   14555230           R-squared     =  0.5693
-------------+------------------------------           Adj R-squared =  0.5232
       Total |  946264964    31  30524676.2           Root MSE      =  3815.1

---------------------------------------------------------------------------------
   HP23AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
           time |   957.4059   206.9046     4.63   0.000     533.5811    1381.231
interventionMAY |  -6611.619   2705.628    -2.44   0.021    -12153.85   -1069.391
   postslopeMAY |  -1495.838   292.6072    -5.11   0.000    -2095.217   -896.4595
          _cons |   15042.05   2000.672     7.52   0.000     10943.86    19140.24
---------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |    4360.523
Mean Absolute Errors     |    3104.3749
Pseudo-R2                |      0.37718
-----------------------------------------

. reg HP23AvgViews time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.66
       Model |  455331527     3  151777176           Prob > F      =  0.0003
    Residual |  490933436    28   17533337           R-squared     =  0.4812
-------------+------------------------------           Adj R-squared =  0.4256
```

```
       Total |   946264964    31  30524676.2           Root MSE      =  4187.3

------------------------------------------------------------------------------------
     HP23AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   592.3357    250.238     2.37   0.025     79.74639    1104.925
interventionAPRIL |   2853.676   2957.952     0.96   0.343    -3205.414    8912.766
   postslopeAPRIL |  -1510.946   324.9507    -4.65   0.000    -2176.577   -845.3148
            _cons |   17110.78   2275.194     7.52   0.000     12450.26    21771.31
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   4701.8614
Mean Absolute Errors     |   3025.1117
Pseudo-R2                |     0.28105
-----------------------------------------

. reg HP23AvgViews time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.46
       Model |   449922646     3   149974215           Prob > F      =  0.0004
    Residual |   496342317    28  17726511.3           R-squared     =  0.4755
-------------+------------------------------           Adj R-squared =  0.4193
       Total |   946264964    31  30524676.2           Root MSE      =  4210.3

------------------------------------------------------------------------------------
     HP23AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   584.7802   279.1393     2.09   0.045     12.98935    1156.571
interventionMARCH |   3846.882   2974.063     1.29   0.206    -2245.209    9938.974
   postslopeMARCH |  -1457.347   338.3873    -4.31   0.000    -2150.502   -764.1918
            _cons |   17151.08   2376.785     7.22   0.000     12282.45     22019.7
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   4590.0594
Mean Absolute Errors     |   3014.3548
Pseudo-R2                |     0.30843
-----------------------------------------
```

Median monthly views

```
. reg HP23Median time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   11.65
       Model |  48268345.7     3  16089448.6           Prob > F      =  0.0000
    Residual |  38654921.8    28  1380532.92           R-squared     =  0.5553
-------------+------------------------------           Adj R-squared =  0.5077
       Total |  86923267.5    31  2803976.37           Root MSE      =    1175

------------------------------------------------------------------------------
  HP23Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   268.5098   58.16923     4.62   0.000     149.3555    387.6641
intervention |  -3080.506   839.8499    -3.67   0.001     -4800.86   -1360.151
   postslope |  -337.3348   91.18187    -3.70   0.001    -524.1124   -150.5572
       _cons |   6097.706   596.0572    10.23   0.000     4876.738    7318.674
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1273.658
Mean Absolute Errors     |   1017.3205
Pseudo-R2                |      0.41399
-----------------------------------------

. reg HP23Median time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    9.23
       Model |  43222658.1     3  14407552.7           Prob > F      =  0.0002
    Residual |  43700609.4    28  1560736.05           R-squared     =  0.4973
-------------+------------------------------           Adj R-squared =  0.4434
       Total |  86923267.5    31  2803976.37           Root MSE      =  1249.3

---------------------------------------------------------------------------------
     HP23Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
           time |   279.9441   67.75249     4.13   0.000     141.1594    418.7288
interventionMAY |  -2201.731   885.9788    -2.49   0.019    -4016.576   -386.8856
   postslopeMAY |  -406.6324   95.81649    -4.24   0.000    -602.9035   -210.3612
          _cons |     6029.1   655.1354     9.20   0.000     4687.116    7371.084
---------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1373.4931
Mean Absolute Errors     |   1147.5357
Pseudo-R2                |      0.32413
-----------------------------------------

. reg HP23Median time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    6.36
       Model |  35235938.7     3  11745312.9           Prob > F      =  0.0020
    Residual |  51687328.8    28  1845976.03           R-squared     =  0.4054
-------------+------------------------------           Adj R-squared =  0.3417
```

```
       Total |  86923267.5    31  2803976.37           Root MSE      =  1358.7

------------------------------------------------------------------------------------
       HP23Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |     234.75   81.19588     2.89   0.007     68.42777    401.0722
interventionAPRIL |  -483.0456   959.7803    -0.50   0.619    -2449.066    1482.975
   postslopeAPRIL |   -435.625   105.4382    -4.13   0.000    -651.6054   -219.6446
            _cons |     6285.2   738.2428     8.51   0.000     4772.978    7797.422
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1508.2152
Mean Absolute Errors     |   1224.239
Pseudo-R2                |      0.20017
-----------------------------------------

. reg HP23Median time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    6.84
       Model |  36767609.2     3  12255869.7           Prob > F      =  0.0013
    Residual |  50155658.3    28  1791273.51           R-squared     =  0.4230
-------------+------------------------------           Adj R-squared =  0.3612
       Total |  86923267.5    31  2803976.37           Root MSE      =  1338.4

------------------------------------------------------------------------------------
       HP23Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   146.1165   88.73404     1.65   0.111    -35.64696    327.8799
interventionMARCH |   1425.536   945.4084     1.51   0.143    -511.0457    3362.117
   postslopeMARCH |  -402.6335   107.5681    -3.74   0.001    -622.9767   -182.2903
            _cons |   6757.912   755.5431     8.94   0.000     5210.252    8305.572
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1433.4557
Mean Absolute Errors     |   1161.2393
Pseudo-R2                |      0.26252
-----------------------------------------
.
```

21 article set (23 set minus Ammonium Nitrate and Jihad articles)

RAW total monthly views

```
. reg HP21RawViews time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   20.38
       Model |  1.5222e+11     3  5.0742e+10           Prob > F      =  0.0000
    Residual |  6.9713e+10    28  2.4898e+09           R-squared     =  0.6859
-------------+------------------------------           Adj R-squared =  0.6522
       Total |  2.2194e+11    31  7.1593e+09           Root MSE      =   49898

------------------------------------------------------------------------------
HP21RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   14449.77   2470.296     5.85   0.000     9389.599    19509.95
intervention |    -134792   35666.25    -3.78   0.001      -207851   -61732.97
   postslope |  -22040.86   3872.258    -5.69   0.000    -29972.82    -14108.9
       _cons |   284385.7   25313.01    11.23   0.000     232534.4      336237
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   54117.269
Mean Absolute Errors     |   43122.186
Pseudo-R2                |      0.58202
-----------------------------------------

. reg HP21RawViews time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.65
       Model |  1.3902e+11     3  4.6342e+10           Prob > F      =  0.0000
    Residual |  8.2913e+10    28  2.9612e+09           R-squared     =  0.6264
-------------+------------------------------           Adj R-squared =  0.5864
       Total |  2.2194e+11    31  7.1593e+09           Root MSE      =   54417

---------------------------------------------------------------------------------
   HP21RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
           time |   14548.22   2951.169     4.93   0.000     8503.022    20593.41
interventionMAY |  -81022.03   38591.54    -2.10   0.045    -160073.2   -1970.848
   postslopeMAY |   -25025.8   4173.583    -6.00   0.000    -33574.99    -16476.6
          _cons |     283795   28536.44     9.95   0.000     225340.8    342249.3
---------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   60680.401
Mean Absolute Errors     |   47288.848
Pseudo-R2                |      0.47884
```

```
-----------------------------------------

. reg HP21RawViews time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.26
       Model |  1.2601e+11     3  4.2004e+10           Prob > F      =  0.0000
    Residual |  9.5927e+10    28  3.4260e+09           R-squared     =  0.5678
-------------+------------------------------           Adj R-squared =  0.5215
       Total |  2.2194e+11    31  7.1593e+09           Root MSE      =   58532

------------------------------------------------------------------------------------
     HP21RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   10747.09   3497.949     3.07   0.005     3581.866    17912.31
interventionAPRIL |   29848.38   41347.69     0.72   0.476    -54848.53    114545.3
   postslopeAPRIL |  -25421.95   4542.319    -5.60   0.000    -34726.47   -16117.43
            _cons |   305334.8   31803.78     9.60   0.000     240187.7    370481.8
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   64314.986
Mean Absolute Errors     |   43954.927
Pseudo-R2                |      0.41567
-----------------------------------------

. reg HP21RawViews time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.20
       Model |  1.2575e+11     3  4.1916e+10           Prob > F      =  0.0000
    Residual |  9.6190e+10    28  3.4354e+09           R-squared     =  0.5666
-------------+------------------------------           Adj R-squared =  0.5202
       Total |  2.2194e+11    31  7.1593e+09           Root MSE      =   58612

------------------------------------------------------------------------------------
     HP21RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   10464.43   3885.943     2.69   0.012     2504.441    18424.43
interventionMARCH |   52992.36    41402.4     1.28   0.211    -31816.62    137801.3
   postslopeMARCH |  -24766.06   4710.744    -5.26   0.000    -34415.58   -15116.54
            _cons |   306842.3   33087.61     9.27   0.000     239065.4    374619.1
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   63048.122
Mean Absolute Errors     |   43494.51
Pseudo-R2                |      0.43631
-----------------------------------------
```

Average monthly views

```
. reg HP21AvgViews time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   20.38
       Model |   345191835     3   115063945           Prob > F      =  0.0000
    Residual |   158083216    28  5645829.13           R-squared     =  0.6859
-------------+------------------------------           Adj R-squared =  0.6522
       Total |   503275051    31  16234679.1           Root MSE      =  2376.1

------------------------------------------------------------------------------
HP21AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   688.1005   117.6342     5.85   0.000     447.1377    929.0633
intervention |   -6418.64   1698.409    -3.78   0.001    -9897.673   -2939.608
   postslope |  -1049.597   184.3949    -5.69   0.000    -1427.313    -671.881
       _cons |   13542.04   1205.392    11.23   0.000      11072.9    16011.17
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |    2577.0343
Mean Absolute Errors     |    2053.4405
Pseudo-R2                |      0.58203
-----------------------------------------

. reg HP21AvgViews time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.65
       Model |   315262322     3   105087441           Prob > F      =  0.0000
    Residual |   188012729    28  6714740.31           R-squared     =  0.6264
-------------+------------------------------           Adj R-squared =  0.5864
       Total |   503275051    31  16234679.1           Root MSE      =  2591.3

---------------------------------------------------------------------------------
   HP21AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
           time |   692.7941    140.532     4.93   0.000     404.9274    980.6608
interventionMAY |  -3858.256   1837.694    -2.10   0.045    -7622.601   -93.91028
   postslopeMAY |  -1191.737   198.7422    -6.00   0.000    -1598.842   -784.6318
          _cons |   13513.87    1358.88     9.94   0.000     10730.34    16297.41
---------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |    2889.5437
Mean Absolute Errors     |    2251.8857
Pseudo-R2                |      0.47886
-----------------------------------------
```

```
. reg HP21AvgViews time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.26
       Model |   285751156     3  95250385.3           Prob > F      =  0.0000
    Residual |   217523895    28  7768710.54           R-squared     =  0.5678
-------------+------------------------------           Adj R-squared =  0.5215
       Total |   503275051    31  16234679.1           Root MSE      =  2787.2

------------------------------------------------------------------------------------
     HP21AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   511.7821   166.5695     3.07   0.005       170.58    852.9843
interventionAPRIL |   1421.438   1968.944     0.72   0.476     -2611.76    5454.636
   postslopeAPRIL |  -1210.601   216.3015    -5.60   0.000    -1653.674   -767.5272
            _cons |   14539.61    1514.47     9.60   0.000     11437.36    17641.86
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   3062.6289
Mean Absolute Errors     |   2093.0978
Pseudo-R2                |      0.41568
-----------------------------------------

. reg HP21AvgViews time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.20
       Model |   285155599     3  95051866.3           Prob > F      =  0.0000
    Residual |   218119452    28  7789980.44           R-squared     =  0.5666
-------------+------------------------------           Adj R-squared =  0.5202
       Total |   503275051    31  16234679.1           Root MSE      =  2791.1

------------------------------------------------------------------------------------
     HP21AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   498.3099   185.0451     2.69   0.012     119.2623    877.3575
interventionMARCH |   2523.736   1971.545     1.28   0.211    -1514.791    6562.263
   postslopeMARCH |   -1179.36   224.3213    -5.26   0.000    -1638.862    -719.859
            _cons |   14611.46   1575.602     9.27   0.000     11383.99    17838.94
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   3002.2907
Mean Absolute Errors     |   2071.122
Pseudo-R2                |      0.43632
-----------------------------------------
```

Median monthly views

```
. . reg HP21Median time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    5.51
       Model |  20811047.6     3  6937015.87           Prob > F      =  0.0042
    Residual |  35281531.9    28  1260054.71           R-squared     =  0.3710
-------------+------------------------------           Adj R-squared =  0.3036
       Total |  56092579.5    31  1809438.05           Root MSE      =  1122.5

------------------------------------------------------------------------------
  HP21Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   206.3873    55.5731     3.71   0.001     92.55092    320.2236
intervention |  -1514.518   802.3669    -1.89   0.069    -3158.092    129.0565
   postslope |  -282.8123   87.11237    -3.25   0.003    -461.2539   -104.3707
       _cons |   5062.868   569.4548     8.89   0.000     3896.392    6229.343
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1218.2317
Mean Absolute Errors     |   886.83799
Pseudo-R2                |     0.18535
-----------------------------------------

. reg HP21Median time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    4.20
       Model |  17413400.8     3  5804466.93           Prob > F      =  0.0142
    Residual |  38679178.7    28  1381399.24           R-squared     =  0.3104
-------------+------------------------------           Adj R-squared =  0.2366
       Total |  56092579.5    31  1809438.05           Root MSE      =  1175.3

---------------------------------------------------------------------------------
     HP21Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
           time |   185.0985   63.74119     2.90   0.007     54.53063    315.6664
interventionMAY |  -427.3015   833.5242    -0.51   0.612    -2134.698    1280.095
   postslopeMAY |  -313.7162   90.14365    -3.48   0.002    -498.3671   -129.0653
          _cons |     5190.6   616.3479     8.42   0.000     3928.068    6453.132
---------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1293.5702
Mean Absolute Errors     |   965.87467
Pseudo-R2                |     0.10277
-----------------------------------------

. reg HP21Median time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    4.84
       Model |  19161806.3     3  6387268.76           Prob > F      =  0.0077
    Residual |  36930773.2    28  1318956.19           R-squared     =  0.3416
-------------+------------------------------           Adj R-squared =  0.2711
```

```
       Total |  56092579.5    31  1809438.05           Root MSE      =  1148.5

------------------------------------------------------------------------------
       HP21Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |    108.275    68.6335     1.58   0.126    -32.31436    248.8644
interventionAPRIL |    1336.64    811.286     1.65   0.111    -325.2039    2998.484
   postslopeAPRIL |  -297.9931   89.12516    -3.34   0.002    -480.5577   -115.4285
            _cons |   5625.933   624.0241     9.02   0.000     4347.678    6904.189
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1233.6708
Mean Absolute Errors     |   996.36376
Pseudo-R2                |      0.16269
-----------------------------------------

. reg HP21Median time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    3.97
       Model |  16749983.3     3  5583327.77           Prob > F      =  0.0177
    Residual |  39342596.2    28  1405092.72           R-squared     =  0.2986
-------------+------------------------------           Adj R-squared =  0.2235
       Total |  56092579.5    31  1809438.05           Root MSE      =  1185.4

------------------------------------------------------------------------------
       HP21Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   125.9802   78.58901     1.60   0.120    -35.00206    286.9625
interventionMARCH |   1064.811   837.3191     1.27   0.214    -650.3597    2779.981
   postslopeMARCH |  -282.8265   95.26972    -2.97   0.006    -477.9776   -87.67527
            _cons |   5531.505   669.1613     8.27   0.000     4160.791     6902.22
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1264.4955
Mean Absolute Errors     |   1001.816
Pseudo-R2                |      0.12727
-----------------------------------------
```

46 Article Group Cross Validation (48 minus Hamas and Fundamentalism articles)

RAW TOTAL MONTHLY VIEWS

```
reg T46NoFundaorHamasRaw time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   24.30
       Model |  3.2537e+12     3  1.0846e+12           Prob > F      =  0.0000
    Residual |  1.2496e+12    28  4.4628e+10           R-squared     =  0.7225
-------------+------------------------------           Adj R-squared =  0.6928
       Total |  4.5033e+12    31  1.4527e+11           Root MSE      =  2.1e+05

------------------------------------------------------------------------------
T46NoFunda~w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   37645.21   10458.63     3.60   0.001     16221.69    59068.73
intervention |  -683829.1   151002.1    -4.53   0.000    -993142.9   -374515.3
   postslope |  -60274.25   16394.18    -3.68   0.001    -93856.21   -26692.28
       _cons |    2261895     107169    21.11   0.000      2042370     2481421
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   226571.35
Mean Absolute Errors     |   189268.95
Pseudo-R2                |      0.63861
-----------------------------------------



. reg T46NoFundaorHamasRaw time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   18.61
       Model |  2.9994e+12     3  9.9980e+11           Prob > F      =  0.0000
    Residual |  1.5039e+12    28  5.3711e+10           R-squared     =  0.6660
-------------+------------------------------           Adj R-squared =  0.6303
       Total |  4.5033e+12    31  1.4527e+11           Root MSE      =  2.3e+05

----------------------------------------------------------------------------------
T46NoFundaorH~w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-----------------------------------------------------------------
           time |   40060.48   12568.78     3.19   0.004      14314.5    65806.46
interventionMAY |    -500903   164358.1    -3.05   0.005    -837575.4   -164230.7
    postslopeMAY |  -75642.86   17774.94    -4.26   0.000    -112053.2   -39232.55
          _cons |    2247404   121534.3    18.49   0.000      1998452     2496355
----------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   253115.47
Mean Absolute Errors     |   200434.46
Pseudo-R2                |      0.55147
-----------------------------------------
```

```
. reg T46NoFundaorHamasRaw time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   13.50
       Model |  2.6623e+12     3  8.8744e+11           Prob > F      =  0.0000
    Residual |  1.8410e+12    28  6.5749e+10           R-squared     =  0.5912
-------------+------------------------------           Adj R-squared =  0.5474
       Total |  4.5033e+12    31  1.4527e+11           Root MSE      =  2.6e+05

------------------------------------------------------------------------------------
T46NoFundaorHam~w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |      34170   15323.79     2.23   0.034     2780.649    65559.36
interventionAPRIL |    -203913   181135.6    -1.13   0.270    -574952.6    167126.5
   postslopeAPRIL |   -83616.5   19898.95    -4.20   0.000    -124377.7   -42855.34
            _cons |    2280783   139325.7    16.37   0.000      1995387     2566179
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   280248.05
Mean Absolute Errors     |   226329.8
Pseudo-R2                |      0.45312
-----------------------------------------

. reg T46NoFundaorHamasRaw time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.67
       Model |  2.5927e+12     3  8.6423e+11           Prob > F      =  0.0000
    Residual |  1.9106e+12    28  6.8236e+10           R-squared     =  0.5757
-------------+------------------------------           Adj R-squared =  0.5303
       Total |  4.5033e+12    31  1.4527e+11           Root MSE      =  2.6e+05

------------------------------------------------------------------------------------
T46NoFundaorHam~w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   29286.11   17318.75     1.69   0.102    -6189.741    64761.95
interventionMARCH |   265.6691   184520.9     0.00   0.999    -377708.4    378239.7
   postslopeMARCH |  -84909.06    20994.7    -4.04   0.000    -127914.7   -41903.37
            _cons |    2306830   147463.8    15.64   0.000      2004765     2608896
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   281781.73
Mean Absolute Errors     |   236816.42
Pseudo-R2                |      0.44585
-----------------------------------------
```

AVERAGE TOTAL MONTHLY VIEWS

```
reg T46NoFundHamasAVG time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   24.30
       Model |  1.5377e+09     3   512561551           Prob > F      =  0.0000
    Residual |   590556612    28  21091307.6           R-squared     =  0.7225
-------------+------------------------------           Adj R-squared =  0.6928
```

```
       Total |  2.1282e+09    31  68652944.1           Root MSE      =  4592.5

------------------------------------------------------------------------------
T46NoFundH~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   818.3775    227.364     3.60   0.001     352.6434    1284.111
intervention |  -14866.27   3282.691    -4.53   0.000    -21590.56   -8141.982
   postslope |  -1310.277   356.3993    -3.68   0.001    -2040.328   -580.2265
       _cons |   49171.72   2329.788    21.11   0.000     44399.37    53944.07
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   4925.5268
Mean Absolute Errors     |   4114.587
Pseudo-R2                |      0.63861
-----------------------------------------

. reg T46NoFundHamasAVG time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   18.61
       Model |  1.4175e+09     3   472494167           Prob > F      =  0.0000
    Residual |   710758766    28  25384241.6           R-squared     =  0.6660
-------------+------------------------------           Adj R-squared =  0.6303
       Total |  2.1282e+09    31  68652944.1           Root MSE      =  5038.3

---------------------------------------------------------------------------------
T46NoFundHama~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
           time |   870.8809    273.239     3.19   0.004     311.1762    1430.586
interventionMAY |  -10889.49   3573.063    -3.05   0.005    -18208.58   -3570.406
    postslopeMAY |  -1644.387   386.4183    -4.26   0.000    -2435.929   -852.8448
          _cons |    48856.7   2642.095    18.49   0.000     43444.61    54268.79
---------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   5502.6127
Mean Absolute Errors     |   4357.318
Pseudo-R2                |      0.55146
-----------------------------------------



. reg T46NoFundHamasAVG time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   13.50
       Model |  1.2582e+09     3   419393074           Prob > F      =  0.0000
    Residual |   870062045    28  31073644.5           R-squared     =  0.5912
-------------+------------------------------           Adj R-squared =  0.5474
       Total |  2.1282e+09    31  68652944.1           Root MSE      =  5574.4

-----------------------------------------------------------------------------------
T46NoFundHamasAVG |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+----------------------------------------------------------------
             time |   742.8179   333.1326     2.23   0.034     60.42674    1425.209
interventionAPRIL |  -4432.982   3937.811    -1.13   0.270    -12499.22    3633.259
   postslopeAPRIL |   -1817.73   432.5947    -4.20   0.000     -2703.86   -931.5995
```

```
           _cons |   49582.39   3028.882    16.37   0.000     43378.01    55786.77
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   6092.4748
Mean Absolute Errors     |   4920.2718
Pseudo-R2                |      0.45311
-----------------------------------------


. reg T46NoFundHamasAVG time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.66
       Model |  1.2253e+09     3   408423359           Prob > F      =  0.0000
    Residual |   902971189    28    32248971           R-squared     =  0.5757
-------------+------------------------------           Adj R-squared =  0.5303
       Total |  2.1282e+09    31  68652944.1           Root MSE      =  5678.8

------------------------------------------------------------------------------------
T46NoFundHamasAVG |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   636.6527    376.502     1.69   0.102    -134.5765    1407.882
interventionMARCH |    5.58394   4011.404     0.00   0.999    -8211.405    8222.573
   postslopeMARCH |  -1845.833   456.4154    -4.04   0.000    -2780.758   -910.9087
            _cons |    50148.6   3205.799    15.64   0.000     43581.82    56715.39
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   6125.807
Mean Absolute Errors     |   5148.2146
Pseudo-R2                |      0.44584
-----------------------------------------
```

MEDIAN TOTAL MONTHLY VIEWS

```
reg T46NoFundHamasMED time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.75
       Model |  76524624.7     3  25508208.2           Prob > F      =  0.0000
    Residual |  56024918.5    28  2000889.95           R-squared     =  0.5773
-------------+------------------------------           Adj R-squared =  0.5320
       Total |   132549543    31  4275791.72           Root MSE      =  1414.5

------------------------------------------------------------------------------
T46NoFundH~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   345.4044   70.02958     4.93   0.000     201.9553    488.8535
intervention |  -3798.599    1011.09    -3.76   0.001    -5869.723   -1727.474
   postslope |   -439.108   109.7733    -4.00   0.000    -663.9684   -214.2476
       _cons |   9221.419   717.5897    12.85   0.000     7751.503    10691.33
------------------------------------------------------------------------------

. cv_regress
```

```
Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1513.5372
Mean Absolute Errors     |   1253.5942
Pseudo-R2                |      0.45732
-----------------------------------------

reg T46NoFundHamasMED time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   14.61
       Model |  80886765.8     3  26962255.3           Prob > F      =  0.0000
    Residual |  51662777.4    28  1845099.19           R-squared     =  0.6102
-------------+------------------------------           Adj R-squared =  0.5685
       Total |   132549543    31  4275791.72           Root MSE      =  1358.3

------------------------------------------------------------------------------
T46NoFundHama~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   400.9118   73.66656     5.44   0.000     250.0127    551.8109
interventionMAY |  -3519.988   963.3153    -3.65   0.001     -5493.25   -1546.726
   postslopeMAY |  -529.4309   104.1803    -5.08   0.000    -742.8345   -316.0273
          _cons |   8888.375   712.3218    12.48   0.000      7429.25     10347.5
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1436.5901
Mean Absolute Errors     |   1220.6195
Pseudo-R2                |      0.50736
-----------------------------------------



. reg T46NoFundHamasMED time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    9.74
       Model |  67688935.8     3  22562978.6           Prob > F      =  0.0001
    Residual |  64860607.4    28  2316450.27           R-squared     =  0.5107
-------------+------------------------------           Adj R-squared =  0.4582
       Total |   132549543    31  4275791.72           Root MSE      =    1522

------------------------------------------------------------------------------
T46NoFundHamasMED |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   381.7107   90.95623     4.20   0.000     195.3953    568.0261
interventionAPRIL |  -1824.989   1075.153    -1.70   0.101     -4027.34    377.3625
   postslopeAPRIL |  -592.4264   118.1127    -5.02   0.000    -834.3693   -350.4835
            _cons |   8997.181   826.9851    10.88   0.000     7303.179    10691.18
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1645.5625
Mean Absolute Errors     |   1376.5626
Pseudo-R2                |      0.36082
```

```
-----------------------------------------



. reg T46NoFundHamasMED time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.37
       Model |    62686297     3  20895432.3           Prob > F      =  0.0004
    Residual |  69863246.2    28  2495115.94           R-squared     =  0.4729
-------------+------------------------------           Adj R-squared =  0.4165
       Total |   132549543    31  4275791.72           Root MSE      =  1579.6

------------------------------------------------------------------------------------
T46NoFundHamasMED |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |      349.2    104.726     3.33   0.002     134.6784    563.7216
interventionMARCH |  -326.1035   1115.794    -0.29   0.772    -2611.703    1959.496
   postslopeMARCH |  -609.2341   126.9544    -4.80   0.000    -869.2884   -349.1797
            _cons |   9170.571   891.7101    10.28   0.000     7343.986     10997.16
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1706.3859
Mean Absolute Errors     |   1444.0275
Pseudo-R2                |      0.31438
-----------------------------------------
```

44 Article Group Cross Validation (48 minus Hamas Fundamentalism Jihad Ammonium Nitrate articles)

RAW TOTAL MONTHLY VIEWS

```
. reg T44NoHamAmmJihFundRAW time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   23.63
       Model |  2.7271e+12     3  9.0903e+11           Prob > F      =  0.0000
    Residual |  1.0771e+12    28  3.8467e+10           R-squared     =  0.7169
-------------+------------------------------           Adj R-squared =  0.6865
       Total |  3.8042e+12    31  1.2272e+11           Root MSE      =  2.0e+05

------------------------------------------------------------------------------
T44NoHamAm~W |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |    30711.4   9709.946     3.16   0.004     10821.48    50601.32
intervention |  -593689.9   140192.6    -4.23   0.000    -880861.5   -306518.3
   postslope |  -52947.52   15220.61    -3.48   0.002    -84125.52   -21769.52
       _cons |    2196493   99497.33    22.08   0.000      1992682     2400304
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   210080.45
Mean Absolute Errors     |   174432.3
Pseudo-R2                |      0.63222
-----------------------------------------

. reg T44NoHamAmmJihFundRAW time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   18.46
       Model |  2.5266e+12     3  8.4221e+11           Prob > F      =  0.0000
    Residual |  1.2776e+12    28  4.5627e+10           R-squared     =  0.6642
-------------+------------------------------           Adj R-squared =  0.6282
       Total |  3.8042e+12    31  1.2272e+11           Root MSE      =  2.1e+05

---------------------------------------------------------------------------------
T44NoHamAmmJi~W |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
           time |    32588.3   11584.35     2.81   0.009     8858.842    56317.76
interventionMAY |    -429858     151485    -2.84   0.008    -740160.9     -119555
   postslopeMAY |  -66264.44   16382.74    -4.04   0.000    -99822.96   -32705.91
          _cons |    2185232   112015.3    19.51   0.000      1955779     2414685
---------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   232324.24
Mean Absolute Errors     |   185440.75
Pseudo-R2                |      0.55250
-----------------------------------------

. reg T44NoHamAmmJihFundRAW time interventionAPRIL postslopeAPRIL
```

```
      Source |       SS       df       MS              Number of obs =        32
-------------+------------------------------           F(  3,    28) =     15.12
       Model |  2.3522e+12     3  7.8406e+11           Prob > F      =    0.0000
    Residual |  1.4520e+12    28  5.1857e+10           R-squared     =    0.6183
-------------+------------------------------           Adj R-squared =    0.5774
       Total |  3.8042e+12    31  1.2272e+11           Root MSE      =   2.3e+05

------------------------------------------------------------------------------------
T44NoHamAmmJihF~W |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   31293.29   13608.94     2.30   0.029     3416.651    59169.93
interventionAPRIL |    -239699   160865.2    -1.49   0.147    -569216.3    89818.35
   postslopeAPRIL |  -74286.75   17672.11    -4.20   0.000    -110486.4   -38087.08
            _cons |    2192570   123734.1    17.72   0.000      1939113     2446028
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   247561.74
Mean Absolute Errors     |   199474.92
Pseudo-R2                |      0.49301
-----------------------------------------

. reg T44NoHamAmmJihFundRAW time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =        32
-------------+------------------------------           F(  3,    28) =     13.67
       Model |  2.2607e+12     3  7.5358e+11           Prob > F      =    0.0000
    Residual |  1.5434e+12    28  5.5123e+10           R-squared     =    0.5943
-------------+------------------------------           Adj R-squared =    0.5508
       Total |  3.8042e+12    31  1.2272e+11           Root MSE      =   2.3e+05

------------------------------------------------------------------------------------
T44NoHamAmmJihF~W |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   26300.13   15565.88     1.69   0.102    -5585.138     58185.4
interventionMARCH |   -35214.1   165845.2    -0.21   0.833    -374932.6    304504.4
   postslopeMARCH |  -76156.01   18869.78    -4.04   0.000      -114809   -37503.01
            _cons |    2219201   132538.7    16.74   0.000      1947707     2490694
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   253436.9
Mean Absolute Errors     |   211313.23
Pseudo-R2                |      0.46859
-----------------------------------------
```

AVERAGE MONTHLY VIEWS

```
. reg T44NoHamAmmJihFundAVG time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   23.63
       Model |  1.4086e+09     3   469536182           Prob > F      =  0.0000
    Residual |   556363558    28  19870127.1           R-squared     =  0.7169
-------------+------------------------------           Adj R-squared =  0.6865
       Total |  1.9650e+09    31  63386196.9           Root MSE      =  4457.6

------------------------------------------------------------------------------
T44NoHamAm~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   697.9975   220.6837     3.16   0.004     245.9475    1150.048
intervention |  -13493.04   3186.241    -4.23   0.000    -20019.76   -6966.322
   postslope |  -1203.355   345.9278    -3.48   0.002    -1911.956   -494.7537
       _cons |   49920.14   2261.335    22.08   0.000     45288.01    54552.27
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   4774.6186
Mean Absolute Errors     |   3964.3921
Pseudo-R2                |     0.63221
-----------------------------------------

. reg T44NoHamAmmJihFundAVG time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   18.46
       Model |  1.3051e+09     3   435018695           Prob > F      =  0.0000
    Residual |   659916019    28  23568429.3           R-squared     =  0.6642
-------------+------------------------------           Adj R-squared =  0.6282
       Total |  1.9650e+09    31  63386196.9           Root MSE      =  4854.7

---------------------------------------------------------------------------------
T44NoHamAmmJi~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
           time |     740.65   263.2848     2.81   0.009     201.3354    1279.965
interventionMAY |  -9769.475   3442.896    -2.84   0.008    -16821.93   -2717.022
    postslopeMAY |  -1506.013    372.341    -4.04   0.000    -2268.719   -743.3073
          _cons |   49664.22   2545.843    19.51   0.000      44449.3    54879.15
---------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   5280.1831
Mean Absolute Errors     |   4214.5736
Pseudo-R2                |     0.55248
-----------------------------------------

. reg T44NoHamAmmJihFundAVG time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.12
       Model |  1.2150e+09     3   404983797           Prob > F      =  0.0000
    Residual |   750020712    28    26786454           R-squared     =  0.6183
-------------+------------------------------           Adj R-squared =  0.5774
```

```
       Total |  1.9650e+09    31  63386196.9           Root MSE      =  5175.6

------------------------------------------------------------------------------------
T44NoHamAmmJihF~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   711.2107   309.2991     2.30   0.029     77.64032    1344.781
interventionAPRIL |  -5447.554   3656.086    -1.49   0.147    -12936.71    2041.599
   postslopeAPRIL |  -1688.336   401.6453    -4.20   0.000    -2511.069   -865.6025
            _cons |   49831.05   2812.184    17.72   0.000     44070.55    55591.55
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   5626.4981
Mean Absolute Errors     |   4533.5744
Pseudo-R2                |     0.49299
-----------------------------------------

. reg T44NoHamAmmJihFundAVG time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   13.67
       Model |  1.1677e+09     3   389240217           Prob > F      =  0.0000
    Residual |   797251454    28  28473266.2           R-squared     =  0.5943
-------------+------------------------------           Adj R-squared =  0.5508
       Total |  1.9650e+09    31  63386196.9           Root MSE      =    5336

------------------------------------------------------------------------------------
T44NoHamAmmJihF~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   597.7121   353.7757     1.69   0.102    -126.9645    1322.389
interventionMARCH |  -799.9311   3769.269    -0.21   0.833    -8520.929    6921.066
   postslopeMARCH |  -1730.808   428.8654    -4.04   0.000    -2609.299    -852.317
            _cons |   50436.37   3012.291    16.74   0.000     44265.97    56606.77
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   5760.0175
Mean Absolute Errors     |   4802.6243
Pseudo-R2                |     0.46858
-----------------------------------------
```

MEDIAN MONTHLY VIEWS

```
. reg T44NoHamAmmJihFundMED time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.36
       Model |  57879940.6     3  19293313.5           Prob > F      =  0.0000
    Residual |  35166601.3    28  1255950.05           R-squared     =  0.6221
-------------+------------------------------           Adj R-squared =  0.5816
       Total |  93046541.9    31  3001501.35           Root MSE      =  1120.7

------------------------------------------------------------------------------
T44NoHamAm~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   277.2255   55.48251     5.00   0.000     163.5747    390.8763
intervention |  -3722.905    801.059    -4.65   0.000      -5363.8    -2082.01
   postslope |  -315.3791   86.97037    -3.63   0.001    -493.5298   -137.2283
       _cons |     8611.5   568.5266    15.15   0.000     7446.926    9776.074
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1204.2531
Mean Absolute Errors     |   998.7613
Pseudo-R2                |      0.50971
-----------------------------------------

. reg T44NoHamAmmJihFundMED time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   20.34
       Model |  63777614.9     3    21259205           Prob > F      =  0.0000
    Residual |    29268927    28  1045318.82           R-squared     =  0.6854
-------------+------------------------------           Adj R-squared =  0.6517
       Total |  93046541.9    31  3001501.35           Root MSE      =  1022.4

---------------------------------------------------------------------------------
T44NoHamAmmJi~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
           time |   336.9338   55.44788     6.08   0.000      223.354    450.5137
interventionMAY |  -3670.966   725.0751    -5.06   0.000    -5156.215   -2185.717
    postslopeMAY |  -404.5265   78.41514    -5.16   0.000    -565.1526   -243.9003
          _cons |    8253.25   536.1555    15.39   0.000     7154.985    9351.515
---------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1079.2187
Mean Absolute Errors     |   935.57829
Pseudo-R2                |      0.60239
-----------------------------------------

. reg T44NoHamAmmJihFundMED time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   10.04
       Model |  48209351.7     3  16069783.9           Prob > F      =  0.0001
    Residual |  44837190.2    28  1601328.22           R-squared     =  0.5181
-------------+------------------------------           Adj R-squared =  0.4665
       Total |  93046541.9    31  3001501.35           Root MSE      =  1265.4
```

```
------------------------------------------------------------------------------------
T44NoHamAmmJihF~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   309.8536   75.62426     4.10   0.000     154.9443    464.7629
interventionAPRIL |   -1926.92   893.9207    -2.16   0.040    -3758.034   -95.80659
   postslopeAPRIL |  -468.0104   98.20312    -4.77   0.000    -669.1704   -266.8505
            _cons |   8406.705    687.585    12.23   0.000     6998.251    9815.159
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1387.7669
Mean Absolute Errors     |   1091.2975
Pseudo-R2                |      0.35549
-----------------------------------------

. reg T44NoHamAmmJihFundMED time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.08
       Model |  43173895.2     3  14391298.4           Prob > F      =  0.0005
    Residual |  49872646.7    28  1781165.95           R-squared     =  0.4640
-------------+------------------------------           Adj R-squared =  0.4066
       Total |  93046541.9    31  3001501.35           Root MSE      =  1334.6

------------------------------------------------------------------------------------
T44NoHamAmmJihF~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   283.9912   88.48334     3.21   0.003     102.7413    465.2411
interventionMARCH |  -632.2332   942.7373    -0.67   0.508    -2563.343    1298.877
   postslopeMARCH |  -489.7167   107.2642    -4.57   0.000    -709.4374    -269.996
            _cons |   8544.637   753.4085    11.34   0.000      7001.35    10087.92
------------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1445.5777
Mean Absolute Errors     |   1183.4531
Pseudo-R2                |      0.30089
-----------------------------------------
```

List of 48 Terrorism Articles with Privacy Sensitivity Scores (from Survey)

Topic Keyword	Wikipedia Articles	Government Trouble	Browser Delete	Privacy Sensitive	Avoidance	Average
Al Qaeda	http://en.wikipedia.org/wiki/Al-Qaeda	2.20	2.11	2.21	2.84	2.34
terrorism	http://en.wikipedia.org/wiki/terrorism	2.19	2.05	2.16	2.79	2.30
terror	http://en.wikipedia.org/wiki/terror	1.98	1.96	2.01	2.64	2.15
attack	http://en.wikipedia.org/wiki/attack	1.92	1.91	1.92	2.56	2.08
Iraq	http://en.wikipedia.org/wiki/iraq	1.60	1.74	1.76	2.25	1.84
Afghanistan	http://en.wikipedia.org/wiki/afghanistan	1.61	1.71	1.75	2.23	1.83
Iran	http://en.wikipedia.org/wiki/iran	1.62	1.73	1.78	2.25	1.85
Pakistan	http://en.wikipedia.org/wiki/pakistan	1.59	1.71	1.75	2.22	1.82
agro	http://en.wikipedia.org/wiki/agro	1.51	1.80	1.76	2.29	1.84
Environmental terrorism	http://en.wikipedia.org/wiki/Environmental_terrorism	2.20	2.20	2.24	2.92	2.39
Eco terrorism	http://en.wikipedia.org/wiki/Eco-terrorism	2.22	2.20	2.22	2.92	2.39
Conventional weapon	http://en.wikipedia.org/wiki/Conventional_weapon	2.03	2.16	2.07	2.81	2.27
Weapons grade	http://en.wikipedia.org/wiki/Weapons-grade	2.18	2.22	2.17	2.99	2.39
dirty bomb	http://en.wikipedia.org/wiki/Dirty_bomb	2.72	2.55	2.50	3.45	2.81
Nuclear Enrichment	http://en.wikipedia.org/wiki/Nuclear_enrichment	2.22	2.21	2.21	2.92	2.39
Nuclear	http://en.wikipedia.org/wiki/nuclear	1.84	1.97	1.91	2.55	2.07
Chemical weapon	http://en.wikipedia.org/wiki/Chemical_weapon	2.43	2.36	2.39	3.16	2.59
Biological weapon	http://en.wikipedia.org/wiki/Biological_weapon	2.44	2.39	2.39	3.18	2.60
Ammonium nitrate	http://en.wikipedia.org/wiki/Ammonium_nitrate	2.49	2.44	2.26	3.24	2.61
Improvised explosive device	http://en.wikipedia.org/wiki/Improvised_explosive_device	2.82	2.64	2.53	3.46	2.86
Abu Sayyaf	http://en.wikipedia.org/wiki/Abu_Sayyaf	2.02	1.96	1.99	2.57	2.14
Hamas	http://en.wikipedia.org/wiki/hamas	1.90	1.93	1.97	2.49	2.07
FARC	http://en.wikipedia.org/wiki/FARC	1.83	1.88	1.90	2.46	2.02
Irish Republican Army	http://en.wikipedia.org/wiki/Irish_Republican_Army	1.62	1.77	1.83	2.24	1.87
Euskadi ta Askatasuna	http://en.wikipedia.org/w/Euskadi_ta_Askatasuna	1.86	1.88	1.88	2.43	2.01
Hezbollah	http://en.wikipedia.org/wiki/hezbollah	1.86	1.90	1.96	2.46	2.05
Tamil Tigers	http://en.wikipedia.org/wiki/Tamil_Tigers	1.76	1.86	1.87	2.39	1.97
PLO	http://en.wikipedia.org/wiki/Palestine_Liberation_Organization	1.77	1.87	1.91	2.42	1.99
Palestine Liberation Front	http://en.wikipedia.org/wiki/Palestine_Liberation_Front	1.81	1.89	1.95	2.47	2.03
Car bomb	http://en.wikipedia.org/wiki/Car_bomb	2.72	2.61	2.50	3.40	2.81
jihad	http://en.wikipedia.org/wiki/jihad	2.15	2.19	2.17	2.89	2.35
Taliban	http://en.wikipedia.org/wiki/taliban	2.06	2.03	2.10	2.70	2.22
Suicide bomber	http://en.wikipedia.org/wiki/Suicide_bomber	2.25	2.31	2.24	2.97	2.44
Suicide attack	http://en.wikipedia.org/wiki/Suicide_attack	2.30	2.36	2.29	3.04	2.50
AL Qaeda in the Arabian Peninsula	http://en.wikipedia.org/wiki/Al-Qaeda_in_the_Arabian_Peninsula	2.01	1.98	2.06	2.63	2.17
Al Qaeda in the Islamic Maghreb	http://en.wikipedia.org/wiki/Al-Qaeda_in_the_Islamic_Maghreb	2.05	1.98	2.06	2.60	2.17
Tehrik-i-Taliban Pakistan	http://en.wikipedia.org/wiki/Tehrik-i-Taliban_Pakistan	1.96	1.96	1.97	2.59	2.12
Yemen	http://en.wikipedia.org/wiki/yemen	1.60	1.72	1.74	2.18	1.81
Pirates	http://en.wikipedia.org/wiki/pirates	1.44	1.67	1.67	2.10	1.72
Extremism	http://en.wikipedia.org/wiki/extremism	1.64	1.90	1.86	2.40	1.95
Somalia	http://en.wikipedia.org/wiki/somalia	1.50	1.68	1.67	2.12	1.74
Nigeria	http://en.wikipedia.org/wiki/nigeria	1.48	1.66	1.64	2.07	1.71
Political radicalism	http://en.wikipedia.org/wiki/Political_radicalism	1.75	1.91	1.97	2.48	2.03
Al-Shabaab	http://en.wikipedia.org/wiki/Al-Shabaab	1.84	1.89	1.89	2.48	2.03
nationalism	http://en.wikipedia.org/wiki/nationalism	1.48	1.71	1.73	2.20	1.78
Recruitment	http://en.wikipedia.org/wiki/recruitment	1.74	1.90	1.87	2.54	2.01
Fundamentalism	http://en.wikipedia.org/wiki/fundamentalism	1.60	1.79	1.80	2.32	1.88
Islamist	http://en.wikipedia.org/wiki/islamist	1.79	1.89	1.93	2.45	2.45
					MEDIAN	2.08

- 2.08 = Median of the Average Privacy Ratings for the 48 Articles
- 23 Most Privacy Sensitive Article Set includes all articles in the 48 Terrorism Group with combined privacy rating average above the 2.08 median.

Al Qaeda	http://en.wikipedia.org/wiki/Al-Qaeda	2.20	2.11	2.21	2.84	2.34
terrorism	http://en.wikipedia.org/wiki/terrorism	2.19	2.05	2.16	2.79	2.30
terror	http://en.wikipedia.org/wiki/terror	1.98	1.96	2.01	2.64	2.15
attack	http://en.wikipedia.org/wiki/attack	1.92	1.91	1.92	2.56	2.08
Environmental terrorism	http://en.wikipedia.org/wiki/Environmental_terrorism	2.20	2.20	2.24	2.92	2.39
Eco terrorism	http://en.wikipedia.org/wiki/Eco-terrorism	2.22	2.20	2.22	2.92	2.39
Conventional weapon	http://en.wikipedia.org/wiki/Conventional_weapon	2.03	2.16	2.07	2.81	2.27
Weapons grade	http://en.wikipedia.org/wiki/Weapons-grade	2.18	2.22	2.17	2.99	2.39
dirty bomb	http://en.wikipedia.org/wiki/Dirty_bomb	2.72	2.55	2.50	3.45	2.81
Nuclear Enrichment	http://en.wikipedia.org/wiki/Nuclear_enrichment	2.22	2.21	2.21	2.92	2.39
Chemical weapon	http://en.wikipedia.org/wiki/Chemical_weapon	2.43	2.36	2.39	3.16	2.59
Biological weapon	http://en.wikipedia.org/wiki/Biological_weapon	2.44	2.39	2.39	3.18	2.60
Ammonium nitrate	http://en.wikipedia.org/wiki/Ammonium_nitrate	2.49	2.44	2.26	3.24	2.61
Improvised explosive device	http://en.wikipedia.org/wiki/Improvised_explosive_device	2.82	2.64	2.53	3.46	2.86
Abu Sayyaf	http://en.wikipedia.org/wiki/Abu_Sayyaf	2.02	1.96	1.99	2.57	2.14
Car bomb	http://en.wikipedia.org/wiki/Car_bomb	2.72	2.61	2.50	3.40	2.81
jihad	http://en.wikipedia.org/wiki/jihad	2.15	2.19	2.17	2.89	2.35
Taliban	http://en.wikipedia.org/wiki/taliban	2.06	2.03	2.10	2.70	2.22
Suicide bomber	http://en.wikipedia.org/wiki/Suicide_bomber	2.25	2.31	2.24	2.97	2.44
Suicide attack	http://en.wikipedia.org/wiki/Suicide_attack	2.30	2.36	2.29	3.04	2.50
AL Qaeda in the Arabian Peninsula	http://en.wikipedia.org/wiki/Al-Qaeda_in_the_Arabian_Penins	2.01	1.98	2.06	2.63	2.17
Al Qaeda in the Islamic Maghreb	http://en.wikipedia.org/wiki/Al-Qaeda_in_the_Islamic_Maghreb	2.05	1.98	2.06	2.60	2.17
Tehrik-i-Taliban Pakistan	http://en.wikipedia.org/wiki/Tehrik-i-Taliban_Pakistan	1.96	1.96	1.97	2.59	2.12