COMMENTS

Are a Few Huge Outcomes Distorting Financial Misconduct Research?
Emre Kuvvet
1–34

A Response to “Are a Few Huge Outcomes Distorting Financial Misconduct Research?”
Andrew C. Call, Nathan Y. Sharp, and Jaron H. Wilde
35–36

Is the United States an Outlier in Public Mass Shootings? A Comment on Adam Lankford
John R. Lott, Jr., and Carlisle E. Moody
37–68

Confirmation That the United States Has Six Times Its Global Share of Public Mass Shooters, Courtesy of Lott and Moody’s Data
Adam Lankford
69–83

Do Right to Carry Laws Increase Violent Crime? A Comment on Donohue, Aneja, and Weber
Carlisle E. Moody and Thomas B. Marvell
84–96

RTC Laws Increase Violent Crime: Moody and Marvell Have Missed the Target
John J. Donohue, Abhay Aneja, and Kyle D. Weber
97–113

Unforced Errors: Tennis Serve Data Tells Us Little About Loss Aversion
Michał Krawczyk
114–123

Tennis Serve Data May Elude Some as Serves Get Too Fast
Nejat Anbarci, K. Peren Arin, and Christina Zenker
124–129
INTELLECTUAL TYRANNY OF THE STATUS QUO

Why Did Milton Friedman Win the Nobel Prize? A Consideration of His Early Work on Stabilization Policy

James Forder and Hugo Monnery

130–145

WATCHPAD

Edmund Burke as an Economist

Donal Barrington

146–154

Foreword to Burke’s “Thoughts and Details on Scarcity”

Daniel B. Klein

155–157

Thoughts and Details on Scarcity

Edmund Burke

158–179
Are a Few Huge Outcomes Distorting Financial Misconduct Research?

Emre Kuvvet

In generating meaningful empirical research, one problem is the existence of a few observations with extreme values, often referred to as outliers. Including extreme values in data is apt to produce misleading results. In his textbook *Econometric Analysis*, William Greene writes:

Even in the absence of multicollinearity or other data problems, it is worthwhile to examine one’s data closely for two reasons. First, the identification of outliers in the data is useful, particularly in relatively small cross sections in which the identity and perhaps even the ultimate source of the data point may be known. Second, it may be possible to ascertain which, if any, particular observations are especially influential in the results obtained. As such, the identification of these data points may call for further study. (Greene 2002, 60)

In financial misconduct research, a leading dataset of more than 1,100 observations is often used, and in that dataset, Enron, WorldCom, Cendant, Colonial Bancgroup, and a dozen or so more are especially influential observations. Researchers deal with the extreme-value problem in different ways. Some drop outliers from the data, while others cap extreme values at a certain level. If researchers decide simply to keep the observations as they are, they need to alert readers to them, clearly and early in their analysis. Also, they should report how the extreme values drive any of their results. In literature on corporate governance,

1. Nova Southeastern University, Fort Lauderdale, FL 33314.
accruals, and risk premia, for example, we find papers that show that extreme observations produce misleading empirical findings (Guthrie et al. 2012; Kraft et al. 2006; Knez and Ready 2012). Furthermore, extreme-value observations call for qualitative investigation to understand whether the variable in question really seemed to play a crucial role.

In the 2018 *Journal of Accounting Research* article “Whistleblowers and Outcomes of Financial Misrepresentation Enforcement Actions,” Andrew Call, Gerald Martin, Nathan Sharp, and Jaron Wilde (2018a) investigate whether the participation of a whistleblower affects the severity of enforcement outcomes. Enforcement outcomes are categorized into several forms, most notably (1) firm penalties, (2) other-penalties, (3) employee penalties, and (4) employee prison sentences. Call, Martin, Sharp, and Wilde (2018a)—which I abbreviate CMSW and treat grammatically as plural—find that enforcement outcomes are more severe in enforcement actions that are associated with whistleblowers. But the top one percent of those enforcement outcomes (11 observations) in their sample of 1,133 enforcement actions influence their results. Table 1 shows the top observations (numbering either 11 or 12) for the top percentile of each category of penalty as a percentage of total penalties among the 1,133 enforcement actions. The top 11 observations in firm penalties constitute 66 percent of the total (all 1,133 observations) firm penalties; the top 11 observations in other-penalties make up 92 percent of the total other-penalties; the top 12 observations in employee penalties make up 84 percent of the total employee penalties; and the top 11 observations in employee prison sentences make up 26 percent of the total employee prison sentences. In addition, all of the nine most extreme observations for the largest firm penalties have the variable in question, the *Whistleblower* dummy, coded as one.

<table>
<thead>
<tr>
<th>Type of Penalties</th>
<th>Total Penalties</th>
<th>Top 11 Observations (Top 1 percent)</th>
<th>Top 11 Observations as a Percentage of Total Penalties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firm Penalties ($ Million)</td>
<td>$14,496</td>
<td>$9,500</td>
<td>65.5%</td>
</tr>
<tr>
<td>Other-Penalties ($ Million)</td>
<td>$13,552</td>
<td>$12,419</td>
<td>91.6%</td>
</tr>
<tr>
<td>Employee Penalties ($ Million)</td>
<td>$25,710</td>
<td>$21,521</td>
<td>83.7%</td>
</tr>
<tr>
<td>Employee Prison Sentences (Months)</td>
<td>24,247</td>
<td>6,394</td>
<td>26.4%</td>
</tr>
</tbody>
</table>

I also provide in Table 2 the same analysis but restricted to the post-Sarbanes-Oxley Act (SOX) sample of 658 cases focused on by CMSW, and the extreme-values problem there is only slightly less pronounced.

---

2. When I use the hyphenated “other-penalties” I am referring specifically to the set of “other penalties” that Call, Martin, Sharp, and Wilde analyze, as defined in Section 5.3 of their paper (2018a, 151).
TABLE 2. Top 6 observations as a percentage of total penalties in the 658 enforcement actions

<table>
<thead>
<tr>
<th>Type of Penalties</th>
<th>Total Penalties</th>
<th>Top 6 Observations (Top 1 percent)</th>
<th>Top 6 Observations as a Percentage of Total Penalties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firm Penalties ($ Million)</td>
<td>$13,578</td>
<td>$7,396</td>
<td>54.5%</td>
</tr>
<tr>
<td>Other-Penalties ($ Million)</td>
<td>$11,107</td>
<td>$9,684</td>
<td>87.2%</td>
</tr>
<tr>
<td>Employee Penalties ($ Million)</td>
<td>$21,176</td>
<td>$16,383</td>
<td>77.4%</td>
</tr>
<tr>
<td>Employee Prison Sentences (Months)</td>
<td>17,181</td>
<td>3,658</td>
<td>21.3%</td>
</tr>
</tbody>
</table>

Yet CMSW do not mention this extreme-values issue in their abstract or introduction. They do not bring up the issue until the twelfth page of the article. In a lengthy post at a Columbia University blog they make no mention whatever of the issue (Call et al. 2018b). I suggest that in both pieces they should have been upfront about the problem of the extreme enforcement outcomes, and that CMSW should have shown how outliers affect their results.

On the twelfth page of their article, CMSW state:

Two challenges that arise when estimating outcomes of regulatory enforcement actions are the large number of zero-valued observations (i.e., enforcement actions without any resultant penalties or criminal prison sentences) and the severe positive skewness in the dependent variable (i.e., some extremely large penalties). Whereas other regression techniques using a log-transformed dependent variable plus a constant (e.g., Tobit or log-linear regression) suffer from potentially severe bias when estimating regressions using data with these attributes, prior research shows that the Poisson pseudo-maximum likelihood (PPML) estimator is a particularly effective modeling technique for data distributions characterized by a disproportionate number of zeros and severe skewness. (CMSW, 134, citations omitted)

Eight pages later they write:

a potential problem that arises when estimating outcomes of regulatory enforcement actions is the combination of a large number of zero-valued observations with a severe positive skew in the dependent variable (e.g., many observations with no penalties and a nontrivial number of very large penalties). In our sample, 474 (72.0%) of the enforcement actions have no penalties assessed against the firm, while each of the largest 20 actions has $100 million or more in firm penalties (with three actions exceeding $1 billion). Further, 208 (31.6%) actions have no penalties assessed against employees, while the largest 22 each exceed $100 million in employee penalties and the largest four each exceed $1 billion. Finally, 506 (76.9%) have no prison sentences assessed against employees, while 25 exceed 20 years. Notably, we find that 105 (16.0%)
actions result in no penalties (firm, employee, or agent firm/employee) and no prison sentences. These distributions (i.e., severe skewness and many observations with zeros) suggest that PPML is the best estimator for our regression analyses. (CMSW, 142ff.)

Later, in a footnote, they write:

In terms of economic significance, we find that whistleblowers are associated with an increase in predicted firm penalties from $8.7 million (without a whistleblower) to $30.5 million (with a whistleblower), an increase in predicted employee penalties increase from $22.8 million to $69.4 million, and an increase in predicted prison sentences increase from 22.5 months to 41.9 months. However, these estimates should be interpreted with caution because of severe skewness in distribution of both the outcome variables (firm penalties, employee penalties, prison sentences) and several of the control variables associated with outcomes of enforcement actions (e.g., Bribery, Organized crime). (CMSW, 146 n.17)

And finally, later they state that “the distribution of regulator penalties (monetary fines and prison sentences) exhibits severe skewness, which limits our ability to reliably quantify the economic impact of our findings” (CMSW, 164).

Thus CMSW do mention the extreme-values issue. But those mentions begin too late and are perhaps rather peripheral and brief, and merely cautionary. The authors do not identify the extreme observations, such as WorldCom and Enron, in their manuscript. The term “outlier” is not found in the paper. As for the morpheme “extreme,” it is found only once. They do repeatedly use the term “severe,” in connection to the word “skew.” But to speak of “severe skewness” does not clearly convey the simple fact of a few extreme observations. Also, although they mention “severe skewness” by the by, they nowhere investigate and show how extreme observations affect their results. We all understand that authors do not wish to accentuate possible weakness of their papers. Still, the authors could have spoken about the extreme-values issue from the start and shown how outliers affect the results.

**Several matters of context**

Before looking into the results of CMSW, I provide some background discussion regarding: (1) policy and research on financial misconduct, (2) examples of other financial-misconduct studies vulnerable to the extreme-values problem, (3) government programs that incentivize whistleblowing, (4) how I came to write
this paper, (5) the leading dataset on enforcement outcomes, and (6) the Whistleblower coding in the dataset used by CMSW.

Remarks on policy and research on financial misconduct

Since 1933, there have been several major laws and regulations made in hopes of reducing financial misconduct at publicly traded firms, including these:

- Securities Act of 1933
- Securities and Exchange Act of 1934
- Foreign Corrupt Practices Act of 1977
- Insider Trading and Securities Fraud Enforcement Act of 1988
- Securities Enforcement Remedies and Penny Stock Reform Act of 1990
- Private Securities Litigation Reform Act of 1995
- Sarbanes-Oxley Act of 2002
- Dodd-Frank Act of 2010

Financial misconduct research has become an attractive research area for many finance and accounting researchers, as the findings can hold significance for public policy. A study with even a remote chance of influencing policy can gather significant interest. The recent availability of financial misconduct databases from such sources as the Government Accountability Office, Audit Analytics, Stanford’s Securities Class Action Clearinghouse, and the Securities and Exchange Commission’s Accounting and Auditing Enforcement Releases have made it easier for researchers to pursue this line of research. Jonathan Karpoff, Allison Koester, D. Scott Lee, and Gerald Martin (2017) report that “more than 150 papers have been published in top accounting and finance journals that use one or more of these data sources.” Searching with Google Scholar for the term “financial misconduct” (with the quotation marks) delivers about 4,250 hits (as of January 17, 2019). One team of authors write: “The growth in datasets and strong industry and public policy implications for research on financial market misconduct and corporate fraud suggests that the demand for high quality research on corporate fraud and financial market misconduct will continue to grow significantly in the future” (Cumming, Dannhauser, and Johan 2015, 165). Recent findings have attracted great attention not only from media but from lawmakers and regulators such as the Securities and Exchange Commission (SEC). Findings are often used to initiate new policies or justify existing policies. I’ve toiled extensively in the
literature, and I really cannot name a single financial-misconduct paper in a top accounting or finance journal that reflects badly on a current law or regulation.

**Other financial misconduct research with extreme-values problems**

In this subsection I briefly discuss a few other papers that might well suffer from the extreme-values problem, to indicate that the issue raised in this paper extends beyond CMSW.

A paper by Frank Yu and Xiaoyun Yu (2011), which has 295 Google Scholar citations (as of January 18, 2019), looks at the relationship between corporate lobbying and fraud detection. Yu and Yu find that firms’ lobbying activities make a significant difference in the number of days between the commencement at a firm of a later-perceived fraud and the commencement of the detection of the fraud:

> Our study sheds light on the recent debate about whether to improve the transparency in corporate political spending. Many firms have argued against detailed disclosure of political spending, citing objections such as the possibility of revelation of corporate strategy to competitors, distractions to management, and negligible impact on shareholder values. Our results suggest that political spending does affect the welfare of investors and that there is a need for more transparency in corporate political spending. (Yu and Yu 2011, 1867)

In other words, the study suggests reform to require firms to disclose their political spending in greater detail. The sample of their study, however, includes extreme observations such as Enron and WorldCom. Including the extreme observations might generate misleading results, as for example Enron and WorldCom spent large amounts of money in lobbying activities and also avoided fraud detection for a very long time.

In 2010, the SEC’s Enforcement Division announced a new policy named the “Cooperation Program.” The program includes various measures designed to encourage greater cooperation by individuals and companies in SEC investigation and enforcement action. It provides incentives to individuals and companies who come forward and offer valuable information to SEC investigators. On its website, the SEC asserts: “There is a spectrum of tools available to the Commission and its staff for facilitating and rewarding cooperation by individuals and entities. These benefits to cooperators can range from reduced charges and sanctions in enforcement actions to taking no enforcement action at all” (link).

3. On Enron, see Tran 2002.
Given the existence of that program, the paper by Rebecca Files, Gerald Martin, and Stephanie Rasmussen (2018) set out to examine the benefits of cooperation. Files et al. (2018) look at the association between the severity of enforcement outcomes and firm cooperation in the enforcement action. They find that a firm’s credit for having cooperated is negatively associated with firm monetary penalties assessed by the SEC and the Department of Justice (DOJ). Their estimates suggest that firms with cooperation credit realize an average penalty reduction of $23.8 million. The authors state “our results provide important insight into what constitutes meaningful cooperation with regulators and suggest that the benefits can be substantial for firms deemed to be cooperative” (2018, 1).

Although Files et al. (2018) try to give us insights about the usefulness of cooperation for the cooperators, they do a less than satisfactory job of addressing the extreme-value problem in their results. Further investigation of the effect of extreme firm penalties on the relation between the cooperation and firm penalties is warranted, as the study can be interpreted as justifying the existence of the current SEC’s Cooperation Program. The study also uses the same enforcement action data as CMSW.

Maria Correia (2014) finds that firms with long-term political connections, as measured by contributions and lobbying, face lower monetary penalties when they are prosecuted by the SEC. Correia suggests that an increase of $100,000 in political action committee money in the five pre-violation years is linked to an 11 percent decrease in monetary penalties by the SEC. However, this paper also has not considered the effect of the extreme observations on its results. The Los Angeles Times published an article based on this study titled “Politically Connected Companies Get a Break from the SEC, Study Says” (Hiltzik 2014). The Times article suggests “some (SEC) chairs have tried to get permission from Congress to self-fund fees, but Congress isn’t that dumb,” and proposes that this is because self-funding will eliminate Congress’s ability to pressure the SEC as suggested in the study. Correia (2014) also uses the same enforcement action data as CMSW.

Ed deHaan, Simi Kedia, Kevin Koh, and Shivaram Rajgopal (2015) look at the association between enforcement outcomes and career opportunities for SEC trial lawyers in civil cases involving accounting misrepresentation. They find that revolving-door incentives do not appear to undermine the prosecution of civil cases against accounting misrepresentations. They write:

[These results provide preliminary input to the discussion among the press, policy makers, and Congress about whether revolving doors are detrimental to the SEC’s regulatory efforts. In our particular setting, future job prospects, on average, appear to make SEC lawyers increase their enforcement efforts in trying civil cases. These results can potentially inform the SEC’s policy on]
However, the SEC regards the further harm to injured shareholders as an important consideration in the determination of whether or not to impose monetary penalties. That means zero or low monetary damages do not imply lax enforcement by the SEC lawyers. That makes the deHaan et al. (2015) dependent variable of monetary penalty a questionable measure of enforcement effort. Their measure of the monetary penalty variable also comes from the same enforcement action data as CMSW.

**Government programs that incentivize whistleblowing**

Section 806 of the Sarbanes-Oxley Act of 2002 was enacted to protect any employee of a publicly traded company or subsidiary who provides evidence of fraud. It authorizes the U.S. Department of Labor to protect whistleblowers against employers who retaliate. Section 1107 of the Act permits the Department of Justice to criminally charge those responsible for the retaliation. On the surface, this make sense. For instance, Alexander Dyck, Adair Morse, and Luigi Zingales (2010) find that fraud detection does not rely on standard corporate governance actors such as the SEC and auditors, but rather it takes a village, including several nontraditional actors such as employees and media. Given the perception that whistleblowing is an effective way to expose fraud, the Dodd-Frank Act of 2010 created the SEC Whistleblower Program. The program rewards individuals who submit tips related to violations of the federal securities laws, with awards in the range of 10 to 30 percent of the monetary sanctions collected. It also provides whistleblowers with employment protection and allows them to report the wrongdoings anonymously. The program is managed by the newly established SEC Office of the Whistleblower (link).

Now comes the focus of my paper: CMSW—that is, Call, Martin, Sharp, and Wilde (2018a). The study has been considered as providing indirect evidence of the effectiveness of whistleblower programs. CMSW investigate whether a whistleblower’s participation affects the severity of enforcement outcomes such as firm penalties, employee penalties, other-penalties, and employee prison sentences. They find enforcement outcomes are more severe in enforcement actions associated with whistleblowers. The authors state:

Examine the role of whistleblowers in securities enforcement is important because policy makers continue to enact legislation attempting to encourage whistleblower involvement and because regulators dedicate significant resources to promoting and rewarding whistleblowing activity. For example, the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010
(Dodd-Frank Act) requires the SEC and the Commodity Futures Trading Commission (CFTC) to establish whistleblower offices that provide a formal venue through which whistleblowers can voice complaints and share evidence with regulators. Rewards for whistleblowers who come forward with original information about corporate misconduct can be large, ranging from 10% to 30% of monetary sanctions over $1 million stemming from investigations facilitated by whistleblowers’ information, documentation, or cooperation. (CMSW, 124, citations omitted)

CMSW write: “Our findings are important to legislators considering the efficacy of current whistleblower policies and the determination of budgets for whistleblower programs, to regulators who design enforcement programs, to SEC and DOJ prosecutors evaluating the merits of using information from whistleblowers in their investigations, and to firms in assessing the consequences of potential enforcement actions” (CMSW, 128; see also 164).

The CMSW study attracted a lot of attention. The Wall Street Journal ran an article based on it titled “Firms Hit With Bigger Penalties When Whistleblowers Involved,” which never mentions the problem of extreme observations. One of the co-authors of the paper, Nate Sharp, is quoted as saying: “Even after holding all those things constant, we see that whistleblowers have a very big effect” (Ensign 2014).

Recently, the authors wrote a blog post on Columbia Law School’s Blog on Corporations and the Capital Markets titled “Financial Enforcement Actions and the Role of Whistleblowers” and based on the CMSW study. There they write:

Our findings are relevant and timely in light of the U.S. federal government’s extensive investments in whistleblowing programs. Section 922 of the Dodd-Frank Wall Street Reform and Consumer Protection Act offers significant monetary incentives (10 percent to 30 percent of monetary sanctions collected via criminal or civil proceedings) to prospective whistleblowers, and also established the SEC Investor Protection Fund to provide funding for this program. As of the end of 2017, the balance in this fund was $321 million, and the government had paid out a total of $160 million to 46 different whistleblowers since the passage of Dodd-Frank. In addition, while the U.S. generally offers the most aggressive whistleblowing rewards, other countries are following suit. For example, in 2016, the Ontario Securities Commission adopted a whistleblowing program and began offering financial incentives to prospective whistleblowers in Canada. As such, large-scale evidence on the usefulness of whistleblowers in the enforcement process (at least in terms of

---

4. Much of the attention paid to the study occurred in 2014, shortly after it was posted as a working paper (link).
enforcement outcomes) is relevant to regulators and legislators who continue
to promote whistleblowing programs and reward those who assist in
enforcement actions. (Call et al. 2018b)

The intention of CMSW is noble, as policymakers have made a significant
attempt to establish or increase rewards for whistleblowers in the area of financial
misconduct. Few studies have examined the benefits of those whistleblower
programs. However, CMSW should not be interpreted as providing solid support
for the whistleblower programs, as I find that the top one percent (11 observations)
of those enforcement outcomes in CMSW’s large sample of 1,133 enforcement
actions influence their results. By illustrating the effect of extreme observations on
CMSW’s findings, I show that further investigation is warranted. This is important
because policies such as the Whistleblower Program also have negative
consequences, and policy judgments should consider both benefits and costs.

The SEC asks for comment letters when it proposes a new rule. For the
implementation of the Whistleblower Provisions of Section 21F, the SEC received
a comment letter from the U.S. Chamber of Commerce. The letter makes it
obvious that companies are concerned about the unintended consequences and the
cost of whistleblower programs. The letter states:

If implemented as proposed…the rule would have a number of harmful
consequences, including eviscerating corporate compliance and reporting
programs; giving rise to unjustified negative publicity about, and unnecessary
SEC investigations of, a large number of innocent companies; and
overwhelming the Commission with an avalanche of poor-quality
information. These results are directly contrary to the well-documented fact
that companies and employees benefit, and scarce government enforcement
dollars are preserved, when companies have the first chance to address
financial wrongdoing. These outcomes would also fly in the face of the
legislative purpose reflected in Section 301 of SOX, which requires public
companies to develop sophisticated internal reporting programs. (Hirschman
and Rickard 2010, 2–3, emphasis in original)

The list of possible harmful effects could be extended much further; here I relegate
mention of some to a footnote.5

5. In addition to possible consequences mentioned in the Chamber of Commerce quotation, here are some
others: a culture of suspicion amongst firm employees and a chilling effect on communication within the
firm; disgruntled employees abusing this power by whistleblowing anonymously to disrupt the company;
employees abusing this power by threatening to whistleblow and shakedown the firm (this threat could
be effective even if the firm were to know itself innocent); overregulation opens the door to political
abuse or shakedown of disfavored firms; competitors accusing a company, or putting someone up to it,
wasting the company’s time and resources; and suffocation/crowding out of other private-sector means of
How I came to this project

My research is in the political economy of finance, especially financial misconduct and related policy issues, including a piece (Kuvvet 2015) that used the aforementioned leading dataset. I became interested in the literature on the severity of enforcement outcomes. But I notice that the extreme-values problem is often buried. I have searched for papers that address it rigorously and have found only one (Files 2012). In 2018, I came across CMSW in the Journal of Accounting Research. The journal requires researchers to publish their data online upon acceptance of their paper, and so the data used in that paper are now publicly available there (link). Using that data I explored whether the extreme-value problem has any influence on CMSW’s results. I found that it does, wrote a comment paper on CMSW, and submitted it to the Journal of Accounting Research. I received a rejection letter but also two fruitful referee reports that helped me to make improvements reflected in the present paper. The earlier version of my paper focused only on firm penalties. After receiving the comments from the referees I expanded my analysis to address employee penalties, other penalties, and employee prison sentences. The journal did not, however, indicate an interest in receiving a revised version, so I tried Econ Journal Watch.

The Karpoff, Koester, Lee, and Martin dataset

The leading financial misconduct dataset, used by CMSW in their study, comes from Jonathan Karpoff, Allison Koester, D. Scott Lee, and Gerald Martin (hereafter KKLM). The data have been hand-collected by Martin, also one of the coauthors of CMSW. The data include financial misrepresentation enforcement actions under Section 13(b) of the Securities Exchange Act, created by the 1977 Foreign Corrupt Practices Act between 1978 and 2012. The first publications to use the data were authored by Karpoff, Lee, and Martin (2008a; 2008b).

To access the full KKLM dataset or any part that has not been publicly released one must contact its creators. If they agree to share their data with the researcher, the researcher has to sign a one-time use-of-data agreement, meaning that the researcher can only use the data only once for a study as stated in the correcting the problem. It is important to avoid double standards by recognizing that government officials, just like private-sector actors, have limitations and imperfections: they wield great power but face little accountability, and their conduct can be capricious, self-interested, unreasonable, or politically biased. I acknowledge, however, that the $1 million penalties threshold for payouts to whistleblowers helps to mitigate some of these possible pathological consequences. Additional commentary on these matters has been published by think tanks and other sources skeptical of government activism (see, e.g., Calomiris, Scott, and Spatt 2011; Katz 2011).
agreement. The agreement includes the following terms of use: (1) “The data will be used only for the research project described below (please provide a short description and title);” (2) “List all coauthors and their affiliations on the research project below;” and (3) “The data will not be shared with any other person without prior written consent” from Karpoff, Lee, and Martin.

The data have been widely used in the corporate misconduct literature. I found more than 20 published papers that use the KKLM database, including papers in top-tier journals such as the *Journal of Accounting Research, Journal of Accounting and Economics, Journal of Finance, Journal of Financial Economics,* and *Journal of Financial and Quantitative Analysis* (e.g., CMSW; Call et al. 2016; Correia 2014; Files 2012; Kedia and Raigopal 2011; Karpoff et al. 2008a; 2008b). Many published and working papers attempting to find the determinants of the severity of the enforcement outcomes use KKLM’s data.

The KKLM data has become rather the gold standard for corporate misconduct data after Karpoff, Koester, Lee, and Martin’s 2017 article in *Accounting Review*. The article compares their KKLM database to those from four popular sources (Government Accountability Office, Audit Analytics, Securities Class Action Clearinghouse, and Accounting and Auditing Enforcement Releases) used in the corporate misconduct literature. They suggest that one can get different results from empirical tests depending on which of those four databases is used in a study. Although Karpoff et al. (2017) do not explicitly say that their KKLM data is superior to the four other databases, one can clearly infer that conclusion from the article. They claim that the KKLM data, unlike the others, does not suffer from the following issues: misidentified event dates, missing relevant information, errors of omission, duplicate events for the same instance of misconduct, and inclusion of events unrelated to misconduct. Unlike Audit Analytics data, KKLM data are not available for researchers to purchase, and unlike GAO data, KKLM data are not freely available to the public. But today if a researcher uses one of the other four popular databases, referees are prone to bring up Karpoff et al.’s (2017) *Accounting Review* paper and point to the weaknesses of the database as grounds for rejection.

### The *Whistleblower and Tipster* variables

The KKLM data goes back to 1978 and contains 1,133 enforcement actions. We now turn to CMSW’s primary analysis which focuses on the 658 enforcement actions in the post-SOX period (2002–2012); they confined their primary analysis to that period because most of the cases coded as whistleblowing took place during

---

6. As I noted above (p. 11), in 2018 some of the KKLM data became publicly available, to meet the requirement of the *Journal of Accounting Research*, upon the publication of CMSW.
that period.

CMSW’s basis for coding the existence (and timing) of whistleblowing is complicated and less than clear. Figure 1 reproduces a portion of their Table 2 (CMSW, 139). One sees the 658 post-SOX “Total enforcement actions.” Of those, 148 have been coded Whistleblower. Those codings are arrived at by three types of sources: 110 are attributed to “OSHA FOIA”, 13 to “Qui tam,” and 25 to “As noted in enforcement proceedings.” I begin by explaining those attributed to the OSHA source.

**Figure 1.** Partial reproduction of Table 2 from CMSW (p. 139)

<table>
<thead>
<tr>
<th>Panel A: Source of whistleblower action</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td></td>
</tr>
<tr>
<td>OSHA FOIA whistleblower complaints received</td>
<td>984</td>
</tr>
<tr>
<td>Total enforcement actions</td>
<td>658</td>
</tr>
<tr>
<td>No whistleblower</td>
<td>510</td>
</tr>
<tr>
<td>Whistleblower</td>
<td>148</td>
</tr>
<tr>
<td>Whistleblower Cases by Source:</td>
<td></td>
</tr>
<tr>
<td>OSHA FOIA</td>
<td>110</td>
</tr>
<tr>
<td>Qui tam</td>
<td>13</td>
</tr>
<tr>
<td>As noted in enforcement proceedings</td>
<td>25</td>
</tr>
<tr>
<td>Whistleblower Cases by Type:</td>
<td></td>
</tr>
<tr>
<td>Tipster</td>
<td>74</td>
</tr>
<tr>
<td>Nontipster</td>
<td>74</td>
</tr>
</tbody>
</table>

The OSHA source does not refer to any direct report from SEC about whether whistleblowing was involved in the investigation, since SEC keeps that secret. Rather, the coding is based on an inference from complaints by employees of having been discriminated against for having blown a whistle. After the passage of Sarbanes-Oxley Act, the Occupational Safety and Health Administration (OSHA) became responsible for handling employee complaints of having been discriminated against for having blown the whistle on alleged financial misconduct. OSHA is required to communicate those discrimination complaints to the SEC. Given the fact that OSHA stores all those employee complaints of discrimination for whistleblowing, Andrew Call, one of the coauthors of CMSW, used Freedom of Information Act (FOIA) requests to obtain those complaints from each of the OSHA offices across the nation. Each of the ten OSHA offices responded to Call’s request with the information for that particular region.

The information requested by Call only concerned complaints for discrimination-due-to-whistleblowing that fall under Section 806 of the Sarbanes-Oxley Act of 2002. Call requested all such complaints contained in OSHA’s database. The information provided by OSHA includes the date the employee filed the complaint with OSHA and the name of the firm complained about. The
The main variable of interest in CMSW is *Whistleblower*. It is an indicator variable equal to one if the researchers deem a whistleblower to have been associated with the enforcement action, and zero otherwise. CMSW also examine the association between whistleblowing and the severity of enforcement outcomes conditional on the timing of the whistleblower’s discrimination complaint. They create two additional whistleblowing variables, namely *Whistleblower (Tipster)* and *Whistleblower (Nontipster)*. CMSW use the filing date of complaints of discrimination due-to-whistleblowing as the relevant date for determining whether the whistleblower is a tipster or a nontipster. They consider whistleblowers to be nontipsters if the complaint date is after the earliest known regulatory investigation or enforcement inquiry date.

CMSW (p. 169) treat an OSHA complaint as a tipster if:

1. the complaint date is unknown,
2. the complaint date precedes the end of the violation period (that is, the period which, according to SEC’s determinations, the firm engaged in the misconduct), or
3. the complaint date precedes the earliest known investigation, by the SEC or DOJ, into the misconduct.  

To generate the codings, CMSW merge the OSHA data with KKLM enforcement action data. The filing date of complaint comes from Call’s OSHA data. The end of the violation period, the earliest known regulatory investigation, and the enforcement inquiry date come from the KKLM enforcement action data. Merging those datasets with those mentioned dates creates major issues, some of which are acknowledged by CMSW.

Let me first point out a matter of possibly confusing terminology. CMSW
refer to allegations of discrimination-due-to-whistleblowing as “whistleblowing allegations.” What they really mean here is discrimination complaints, but specifically complaints of having been discriminated against for having blown a whistle on firm financial misconduct. The one making the complaint/“allegation” is saying that what motivated the discrimination was her having blown a whistle. The data do not establish that she in fact blew a whistle, nor what conduct on the part of the firm any such whistleblowing may have concerned, nor that actual discrimination occurred. Such is the nature of what CMSW call a “whistleblowing allegation.” Empirical research is not always sensitive to fuzziness in data, but the research in question here, exhibiting the extreme-values problem, can be highly sensitive to such fuzziness: The miscoding of just a few of the extreme values can change the results. Researchers should be clear and upfront about such sensitivity.

On page 126, CMSW state that “because we cannot directly observe whether regulators actually used the information from each OSHA whistleblower, these whistleblower allegations [read: complaints of discrimination due to whistleblowing] reflect only potential whistleblower involvement in an enforcement action.” On page 164, CMSW also state that “most of the whistleblower allegations in our sample are obtained from OSHA, and we cannot directly observe whether the SEC or DOJ actually used the information from each OSHA whistleblower. As a result, these cases represent potential whistleblower involvement in an enforcement action.” In other words, even though OSHA is required to communicate the employee discrimination complaint for blowing the whistle on alleged financial misconduct to the SEC, the SEC is not required to act on any such allegation. The SEC is especially likely not to act if it regards the allegation as frivolous.

Here is something that confuses me: Call published a paper in *Journal of Accounting and Economics* with Simi Kedia and Shivaram Rajgopal in 2016. They find that firms grant more rank-and-file stock options when involved in financial reporting violations, consistent with management’s incentives to discourage employee whistleblowing (blowing a whistle would reduce the value of the stock). That paper does not use the OSHA data. Instead:

We use a LexisNexis search to construct our sample of whistle-blowing firms. We follow Bowen et al. (2010) and search every combination of the following sets of terms: (1) ‘whistle,’ ‘whistle-blowing,’ ‘whistleblower,’ and ‘whistleblower’ and (2) ‘financial,’ ‘accounting,’ and ‘fraud.’ We perform this search over the calendar years 1992 through 2010. We augment the sample with the employee-based whistleblowing events identified by Dyck et al. (2010), yielding a total of 153 whistle-blowing events. (Call, Kedia, and Rajgopal 2016, 286)

Thus, Call and his 2016 coauthors do not use his already-available OSHA data
for that paper. Instead, they rely on LexisNexis to collect whistleblowing involvements. In a footnote, the authors say why, citing Bowen, Call, and Rajgopal (2010): “Bowen et al. (2010) evaluate the efficacy of whistleblowing complaints filed with OSHA and conclude that these complaints are generally frivolous. Hence, we do not employ OSHA-related whistle-blowing events in our data analysis” (Call, Kedia, and Rajgopal 2016, 287 n.20). Yet, CMSW uses the OSHA data.

Another problem arises when CMSW merges the OSHA data with the KKLM enforcement actions data. CMSW merge the two by using the date the employee filed the complaint with OSHA. If the date falls between the beginning of the violation period and the last regulatory proceedings of an enforcement outcome for that company, CMSW consider the enforcement action as associated with that particular complaint. That is, even if that date is just one day before the final regulatory proceedings of an enforcement outcome, CMSW treats that enforcement outcome as if a whistleblower played a role, albeit, in that case, as a nontipster. Second, as Call, Kedia, and Rajgopal (2016, 287, n20) point out, such OSHA complaints are often frivolous. Thus, they are not likely to be used by the SEC.

Other issues arise when CMSW try to classify whistleblowers as tipsters versus nontipsters based on the filing date of the complaint with OSHA. First of all: Of the 148 whistleblowing cases in CMSW, 13 do not have the date for the whistleblowing (CMSW, 147 n.18). One might think that such cases should simply be excluded. But CMSW count them as tipsters, without explaining why they do so. Second, the filing date of the OSHA complaint is not necessarily the same as the date the complainer began to assist the SEC investigation even if we assume that the complaint is used by SEC for the investigation.

Also, it is also not clear from the manuscript what happens if there is more than one complainer/inferred-whistleblower. For instance, suppose there is a tipster and then over the course of the investigation a nontipster also provides information. CMSW do not explain how to code such a case—which certainly seems a plausible scenario for major misconduct that catches fire.

But my greatest concern with the OSHA-derived data of CMSW is this: Their whistleblower identities for many observations conflict with Dyck et al. (2010)’s whistleblower data. For instance, the Dyck et al. (2010) data suggest that the ‘whistleblower’ for Enron is a newspaper, but CMSW’s coding shows an employee as a tipster whistleblower. In other words, the CMSW data says that the investigation of Enron was started because of an employee allegation. Again, the dubious coding of one case can make a huge difference when the data suffers from an extreme-values problem.

In addition to the 110 complaints to OSHA for alleged discrimination due to
whistleblowing, CMSW also add 13 *qui tam* whistleblowing cases to their sample. Under the False Claims Act of 1863, also known as the Lincoln Law, people who are not affiliated with the government, known as “relators” under the law, can file actions on behalf of government against persons and companies that defraud government programs and be paid a percentage of the settlement. However, a relator does not have to be an employee of the fraudulent company; any person with the knowledge of a company defrauding government program can file a *qui tam* lawsuit. CMSW do not state, and it is not otherwise clear, whether their 13 *qui tam* cases are all related to employee whistleblowing.

CMSW also add 25 additional whistleblowing cases directly referred to in administrative and legal proceedings in the enforcement actions. However, CMSW do not specify whether those whistleblowing cases are all employee-related or what those administrative and legal proceedings are.

Of the 148 cases coded *Whistleblower* in CMSW, 13 do not have the date for the whistleblowing. CMSW count them as tipsters. It is not clear from the manuscript whether those 13 cases come from OSHA cases, *qui tam* cases, or those additional 25 cases. We also do not know whether any of those 13 cases without the whistleblowing dates is an extreme observation. Again, these uncertainties are important, because a few codings can matter a lot when a few extreme observations drive results.

For any given case, we simply do not know the lineage of the coding. As we have seen, uncertainties abound in CMSW’s inferring of whistleblowing and of tipster status. The extreme-values problem calls for such disclosure, because, again, a few dubious codings could make a big difference in the results. It is puzzling that CMSW do tell the lineage for each of their 148 whistleblowing codings. They do not say why they do not tell the lineage.

### Results when extreme observations are removed

A total of 1,133 enforcement actions are included in the KKLM database, 658 of which occurred after the passage of the Sarbanes-Oxley Act in 2002. One of the types of enforcement action is firm penalties, which are fines levied against the fraudulent firm. Table 3 shows the largest 11 firm penalties (top one percent) in CMSW’s sample of SEC and DOJ enforcement actions between 1978 and 2012 that are associated with alleged financial misrepresentation. The sum of the three largest firm penalties is $5,459 million. The sum of the firm penalties for those 11 enforcement actions is $9,500 million. The sum of all firm penalties in the
entire sample of 1,133 enforcement actions is $14,496 million. Thus, those 11 enforcement actions constitute 66 percent of the total firm penalties in the sample. (The number of enforcement actions which resulted in $0 in firm penalties was 911, or 80 percent.) Thus, the few extreme observations are likely to have a significant effect on the severity of firm penalties if whistleblowers are also associated with them. Indeed, among those 11 observations, nine (81 percent) of the largest firm penalties have a value of one for the \(Whistleblower\) dummy. Again, we do not know how those nine codings were arrived at.

### TABLE 3. Largest firm penalties (Top 1 percent) in CMSW’s sample of SEC and DOJ enforcement actions between 1978 and 2012 associated with alleged financial misrepresentation

<table>
<thead>
<tr>
<th>Record ID</th>
<th>Company Name</th>
<th>Firm Penalties ($ Million)</th>
<th>Whistleblower</th>
<th>Whistleblower (Tipster)</th>
<th>Whistleblower (Nontipster)</th>
</tr>
</thead>
<tbody>
<tr>
<td>582</td>
<td>WorldCom Inc</td>
<td>2,278</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>991</td>
<td>Siemens AG</td>
<td>1,659</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>566</td>
<td>Enron Corp (2)</td>
<td>1,522</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>662</td>
<td>American International Group Inc. (1)</td>
<td>825</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>970</td>
<td>Halliburton Co</td>
<td>600</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>586</td>
<td>Reliant Resources Inc</td>
<td>512</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>735</td>
<td>Time Warner Inc. (2) [America Online Inc. (1)]</td>
<td>510</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>710</td>
<td>Bristol Myers Squibb Co</td>
<td>450</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>811</td>
<td>Federal National Mortgage Association (Fannie Mae)</td>
<td>400</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>228</td>
<td>National Medical Enterprises Inc</td>
<td>379</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1116</td>
<td>ENI SpA</td>
<td>365</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Note: Record ID is an SEC identifier that relates to the specific action in the SEC enforcement action dataset.*

In Table 4, the dependent variable is *Firm Penalties*, in increments of millions of dollars. Firm penalties are the total civil and criminal monetary penalties assessed against the firm, its parent, and its subsidiaries, and they consist of disgorgement, prejudgment interest, civil and criminal fines, and criminal restitution. Following CMSW, Table 4 uses the Poisson pseudo-maximum likelihood regressions. Model 1 in Table 4 successfully replicates CMSW’s principal results for firm penalties using the full sample of 1,133 enforcement actions. The coefficients and \(\xi\)-statistics for Model 1 are the same as those of Model (1) in Table A2 in CMSW’s online appendix.
TABLE 4. Replicating the main results of CMSW without the 11 largest firm penalties (Top 1 percent) by using the 1,133 SEC and DOJ enforcement actions between 1978 and 2012 associated with alleged financial misrepresentation (Full Sample)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient ($z$)</td>
<td>Coefficient ($z$)</td>
<td>Coefficient ($z$)</td>
<td>Coefficient ($z$)</td>
</tr>
<tr>
<td>Intercept</td>
<td>−8.751*** (−4.45)</td>
<td>−6.941*** (−5.67)</td>
<td>−7.268*** (−5.81)</td>
<td>−8.402*** (−6.03)</td>
</tr>
<tr>
<td>Whistleblower (Tipster)</td>
<td>0.995* (1.67)</td>
<td>−0.329 (−1.08)</td>
<td>0.237 (0.54)</td>
<td></td>
</tr>
<tr>
<td>Whistleblower (Nontipster)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-dealing</td>
<td>−1.546* (−1.82)</td>
<td>−0.208 (−0.66)</td>
<td>−0.334 (−1.02)</td>
<td>−0.512 (−0.96)</td>
</tr>
<tr>
<td>% Blockholder ownership</td>
<td>−1.855* (−1.19)</td>
<td>−1.246*** (−2.79)</td>
<td>−1.063*** (−2.29)</td>
<td>−0.444 (−0.48)</td>
</tr>
<tr>
<td>% Initial abnormal return</td>
<td>0.996 (0.93)</td>
<td>0.176 (0.15)</td>
<td>0.107 (0.09)</td>
<td>0.597 (0.55)</td>
</tr>
<tr>
<td>Violation period</td>
<td>0.738*** (3.66)</td>
<td>0.616*** (3.08)</td>
<td>0.724*** (3.41)</td>
<td>0.726*** (4.17)</td>
</tr>
<tr>
<td>Bribery</td>
<td>0.411 (1.01)</td>
<td>0.86 (1.58)</td>
<td>0.872* (1.74)</td>
<td>0.474 (1.04)</td>
</tr>
<tr>
<td>Organized crime</td>
<td>−3.998*** (−2.89)</td>
<td>−2.773*** (−2.39)</td>
<td>−2.615*** (−2.26)</td>
<td>−3.11*** (−2.65)</td>
</tr>
<tr>
<td>Deterrence</td>
<td>−0.078 (−0.18)</td>
<td>0.334 (1.08)</td>
<td>0.243 (0.83)</td>
<td>0.083 (0.2)</td>
</tr>
<tr>
<td># C-level respondents</td>
<td>0.515 (1.34)</td>
<td>0.756* (2.47)</td>
<td>0.783* (2.57)</td>
<td>0.663* (1.96)</td>
</tr>
<tr>
<td># Code violations</td>
<td>1.422*** (2.97)</td>
<td>0.501* (1.83)</td>
<td>0.398 (1.52)</td>
<td>0.905* (2.39)</td>
</tr>
<tr>
<td>Fraud</td>
<td>−0.658* (−1.75)</td>
<td>−0.229 (−0.52)</td>
<td>−0.187 (−0.47)</td>
<td>−0.201 (−0.5)</td>
</tr>
<tr>
<td>Misled auditor</td>
<td>0.556 (1.45)</td>
<td>0.076 (0.24)</td>
<td>0.065 (0.21)</td>
<td>−0.013 (−0.04)</td>
</tr>
<tr>
<td>Big N auditor</td>
<td>2.939*** (1.99)</td>
<td>0.645 (1.36)</td>
<td>0.654 (1.41)</td>
<td>1.521*** (2.3)</td>
</tr>
<tr>
<td>Exec respondent terminated</td>
<td>−0.201 (−0.39)</td>
<td>−0.799*** (−2.23)</td>
<td>−0.814*** (−2.3)</td>
<td>−0.727 (−1.7)</td>
</tr>
<tr>
<td>Cooperation</td>
<td>0.542 (1.59)</td>
<td>0.651* (1.71)</td>
<td>0.633* (1.7)</td>
<td>0.432 (1.44)</td>
</tr>
<tr>
<td>Impeded investigation</td>
<td>0.022 (0.05)</td>
<td>1.966*** (4.22)</td>
<td>2.056*** (4.65)</td>
<td>0.756 (1.6)</td>
</tr>
<tr>
<td>% Independent directors</td>
<td>−0.584 (−0.83)</td>
<td>0.917 (1.5)</td>
<td>1.179* (1.96)</td>
<td>1.077 (1.54)</td>
</tr>
<tr>
<td>Recidivist</td>
<td>−0.112 (−0.33)</td>
<td>−0.387 (−1.45)</td>
<td>−0.224 (−0.83)</td>
<td>−0.054 (−0.16)</td>
</tr>
<tr>
<td>Market capitalization</td>
<td>0.261*** (2.74)</td>
<td>0.461*** (6.32)</td>
<td>0.442*** (6.09)</td>
<td>0.400*** (5.2)</td>
</tr>
<tr>
<td>Market-to-book assets</td>
<td>−0.445*** (−2.2)</td>
<td>−0.300*** (−2.24)</td>
<td>−0.314*** (−2.43)</td>
<td>−0.197 (−1.54)</td>
</tr>
<tr>
<td>Leverage ratio</td>
<td>0.579*** (3.06)</td>
<td>0.200 (1.38)</td>
<td>0.207 (1.26)</td>
<td>0.420*** (3.06)</td>
</tr>
<tr>
<td>Distance from regulator</td>
<td>−0.079 (−1.36)</td>
<td>−0.003 (−0.06)</td>
<td>0.000 (0.01)</td>
<td>−0.058 (−1.02)</td>
</tr>
<tr>
<td>Post-Sarbanes Oxley</td>
<td>0.016 (0.03)</td>
<td>−0.312 (−0.48)</td>
<td>−0.253 (−0.4)</td>
<td>−0.417 (−0.82)</td>
</tr>
<tr>
<td>FF 12 Industry Fixed Effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>1,133</td>
<td>1,122</td>
<td>1,122</td>
<td>1,133</td>
</tr>
</tbody>
</table>

Notes: The dependent variable is Firm Penalties. My model specification is the same as that in Table 4 of CMSW (p. 145). Model [1] uses the full sample of 1,133 enforcement actions. Model [2] and Model [3] omit the top one percent of Firm Penalties from the sample (11 observations). Model [4] replaces each of those 11 extreme observations with the 99th percentile value ($338 million) of Firm Penalties for the full sample. Variables are as defined in CMSW’s Appendix. These are exponential regression results, with $z$-statistics shown, using robust standard errors. *, **, and *** denote statistical significance at the 10-percent, 5-percent, and 1-percent levels, respectively.
However, after omitting the top one percent of firm penalties from the sample (11 observations) and rerunning the regression, in Model 2, the estimated coefficient for the Whistleblower dummy is statistically insignificant (\(z = -1.08\)). In contrast to CMSW’s findings, the sign of the Whistleblower dummy is also negative. Model 3 examines the association between Firm Penalties without the 11 largest penalties and both Whistleblower (Tipster) and Whistleblower (Nontipster). In Model 3, the estimated coefficient for Whistleblower (Tipster) is negative and statistically significant (\(z = -2.23\)) and the coefficient for Whistleblower (Nontipster) is statistically insignificant (\(z = 0.06\)), in contrast to CMSW’s findings of a statistically significant positive association between Firm Penalties and both Whistleblower (Tipster) and Whistleblower (Nontipster). Model 4 replaces each of those 11 extreme observations with the 99th percentile value ($338 million) of Firm Penalties for the full sample, and the Whistleblower results remain statistically insignificant (\(z = 0.54\)).

CMSW’s primary analysis focuses on the 658 enforcement actions in the post-SOX period, because most of the whistleblowing allegations took place during that period. Therefore, Model 1 in Table 5 reproduces the main results for Firm Penalties of CMSW’s Model (1) in their Table 5 (p. 145) without the extreme observations for Firm Penalties by using the 658 enforcement actions that occurred between 2002 and 2012, after the passage of SOX. Models 1 and 2 exclude the 10 post-SOX extreme observations from the 11 pre- and post-SOX cases used in Table 3. In Model 1, the estimated coefficient for the Whistleblower dummy is statistically insignificant (\(z = -0.63\)), which suggests that whistleblowers have no effect on the severity of firm penalties. In Model 2, the estimated coefficient for Whistleblower (Tipster) is negative and statistically significant (\(z = -1.86\)). Again, this contrasts with CMSW, who find a statistically significant positive association between Whistleblower (Tipster) and the severity of Firm Penalties. The coefficient for Whistleblower (Nontipster) in Model 2 is again statistically insignificant (\(z = 0.14\)). Model 3 replaces each of the ten extreme observations with the 99th percentile value ($338 million) of Firm Penalties for the full sample. The Whistleblower dummy remains statistically insignificant.

One might argue that I should use the extreme observations (top one percent) in the sample of the 658 enforcement actions in Table 5, based on the post-SOX period, rather than the full sample with 1,133 enforcement actions, as CMSW’s main analysis (their Table 4) uses the post-SOX sample of 658 enforcement actions. To address this argument in Models 4 and 5, I show the outcome when I omit only the six largest firm-penalty cases, that is, the 99th percentile value for Firm Penalties ($510 million) based on the 658-observation sample because its enforcement action is pre-SOX.

---

8. One of the 11 extreme observations for firm penalties, National Medical Enterprises, is not in the 658-observation sample because its enforcement action is pre-SOX.
sample. In addition, I show the results using alternative models, rather than using the Poisson pseudo-maximum likelihood. Model 6 is an OLS model, and Model 7 is a Tobit model. The Whistleblower dummy remains statistically insignificant after excluding those 10 extreme observations in Models 6 and 7.

CMSW also examine the relationship between whistleblowers and the severity of other enforcement outcomes such as other-penalties, employee penalties, and employee prison sentences, and find statistically significant but slightly weaker results. Part A of Table 6 shows the largest 11 values (top one percent) for the Other-Penalties variable in CMSW’s sample of SEC and DOJ enforcement actions between 1978 and 2012 associated with alleged financial misrepresentation. Other-penalties are the total civil and criminal monetary penalties assessed against the agent firms and/or respondents (e.g. audit firm, bankers, suppliers) in connection with the financial misrepresentation of the target firm, in increments of millions of dollars. The largest amount of other-penalties (Colonial BancGroup Inc.) is $7,532 million. The sum of the other-penalties for those 11 enforcement actions is $12,419 million. The sum of all other-penalties in the entire sample of 1,133 enforcement actions is $13,552 million. Those 11 enforcement actions constitute 92 percent of the total other-penalties in the sample.

Part B of Table 6 shows the largest 12 values (top one percent) for Employee Penalties in CMSW’s sample of SEC and DOJ enforcement actions between 1978 and 2012 associated with alleged financial misrepresentation. Employee penalties are the total civil and criminal penalties assessed against all employees—consisting of disgorgement, prejudgment interest, civil fines, criminal restitution, and criminal fines—in increments of millions of dollars. The largest employee penalty (Cendant Corp) is $6,557 million. The sum of the employee penalties for those 12 enforcement actions is $21,521 million. The sum of all employee penalties in the entire sample of 1,133 enforcement actions is $25,710 million. Thus, those 12 enforcement actions constitute 84 percent of the total employee penalties in the sample.
TABLE 5. Replicating CMSW’s main results without the largest firm penalties (top 1 percent) using the 658 SEC and DOJ enforcement actions between 2002 and 2012 associated with alleged financial misrepresentation (Post-SOX)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (z)</td>
<td>Coefficient (z)</td>
<td>Coefficient (z)</td>
<td>Coefficient (z)</td>
<td>Coefficient (z)</td>
<td>Coefficient (t)</td>
<td>Coefficient (t)</td>
</tr>
<tr>
<td>Whistleblower</td>
<td>−0.185 (−0.63)</td>
<td>0.615 (1.55)</td>
<td>0.064 (0.22)</td>
<td>2.920 (0.86)</td>
<td>7.430 (1.12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whistleblower</td>
<td>−0.636* (−1.86)</td>
<td>−0.570* (−1.88)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Tipster)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whistleblower</td>
<td>0.042 (0.14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Nontipster)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Variables</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>FF 12 Industry</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fixed Effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>648</td>
<td>648</td>
<td>658</td>
<td>652</td>
<td>652</td>
<td>648</td>
<td>648</td>
</tr>
</tbody>
</table>

Notes: The dependent variable is Firm Penalties. Model 1 and Model 2 exclude the 10 post-SOX extreme observations that are among the 11 pre- and post-SOX cases used in Table 3. Model [3] replaces each of those ten extreme observations with the 99th percentile value ($338 million) of Firm Penalties for the full sample. Model 4 and Model 5 omit only the six largest firm-penalties cases, that is, the top one percent of Firm Penalties from the post-SOX sample. Exponential regression results are shown for Models 1, 2, 3, 4, and 5. Model 6 and Model 7 exclude the same 10 extreme observations excluded as in Models 1 and 2; Model 6 is an OLS model, and Model 7 is a Tobit model. Z- or t-statistics are shown, using robust standard errors. *, **, and *** denote statistical significance at the 10-percent, 5-percent, and 1-percent levels, respectively.
TABLE 6 Part A. Largest other-penalties (top 1 percent) in CMSW’s sample of SEC and DOJ enforcement actions between 1978 and 2012 associated with alleged financial misrepresentation

<table>
<thead>
<tr>
<th>Record ID</th>
<th>Company Name</th>
<th>Other-Penalties ($ Million)</th>
<th>Whistle-blower (Tipster)</th>
<th>Whistle-blower (Nontipster)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1217</td>
<td>Colonial BancGroup Inc.</td>
<td>7,532</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>148</td>
<td>American Continental Corp. (Lincoln Savings &amp; Loan)</td>
<td>2,215</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>909</td>
<td>Refco Inc/Refco Group Ltd.</td>
<td>689</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>566</td>
<td>Enron Corp. (2)</td>
<td>518</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>531</td>
<td>Franklin American Corp.</td>
<td>466</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>706</td>
<td>Allou Health &amp; Beauty Care Inc.</td>
<td>326</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>682</td>
<td>Suprema Specialties Inc.</td>
<td>153</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>970</td>
<td>Halliburton Co.</td>
<td>149</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>971</td>
<td>KBR Inc.</td>
<td>149</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>575</td>
<td>PNC Financial Services Group Inc.</td>
<td>122</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>123</td>
<td>Crazy Eddie Inc.</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

TABLE 6 Part B. Largest employee penalties (top 1 percent) in CMSW’s sample of SEC and DOJ enforcement actions between 1978 and 2012 associated with alleged financial misrepresentation

<table>
<thead>
<tr>
<th>Record ID</th>
<th>Company Name</th>
<th>Employee Penalties ($ Million)</th>
<th>Whistle-blower (Tipster)</th>
<th>Whistle-blower (Nontipster)</th>
</tr>
</thead>
<tbody>
<tr>
<td>442</td>
<td>Cendant Corp.</td>
<td>6,557</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>909</td>
<td>Refco Inc/Refco Group Ltd.</td>
<td>4,887</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>148</td>
<td>American Continental Corp. (Lincoln Savings &amp; Loan)</td>
<td>3,476</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>584</td>
<td>Adelphia Communications Corp.</td>
<td>1,477</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>589</td>
<td>Mortgage Corporation of America (MCA Financial Corp.)</td>
<td>1,217</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>684</td>
<td>Computer Associates International Inc.</td>
<td>889</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>531</td>
<td>Franklin American Corp.</td>
<td>850</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>170</td>
<td>Sahlen &amp; Associates Inc.</td>
<td>532</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1217</td>
<td>Colonial BancGroup Inc.</td>
<td>506</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>706</td>
<td>Allou Health &amp; Beauty Care Inc.</td>
<td>500</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>553</td>
<td>Lason Inc.</td>
<td>325</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>400</td>
<td>Centennial Technologies Inc.</td>
<td>305</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Part C of Table 6 shows the largest 11 values (top one percent) for Employee Prison Sentences in CMSW’s sample of SEC and DOJ enforcement actions between 1978 and 2012 associated with alleged financial misrepresentation. Employee prison sentences are the total incarcerations, consisting of jail, prison, home detention, and halfway house sentences, in increments of months imposed upon employee respondents named in the enforcement action. The sum of the employee prison sentences for those 11 enforcement actions is 6,394 months. The sum of all employee prison sentences in the entire sample of 1,133 enforcement actions is 24,247 months. Those 11 enforcement actions constitute one-fourth of the total employee prison sentences in the sample.

In Part D of Table 6, I examine the relationship between Whistleblower and the severity of Other-Penalties, Employee Penalties, and Employee Prison Sentences, without using those extreme observations stated above. I use Poisson pseudo-maximum likelihood regressions in the Table, and I show that Whistleblower is not associated with the severity of Other-Penalties, Employee Penalties, and Employee Prison Sentences after omitting those extreme observations.
### TABLE 6 Part D. Replicating the results of CMSW without extreme (top 1 percent) other-penalties, employee penalties, and employee prison sentences using the 1,133 SEC and DOJ enforcement actions between 1978 and 2012 associated with alleged financial misrepresentation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Other-Penalties ($ Million)</td>
<td>Coefficient (z)</td>
<td>Coefficient (z)</td>
<td>Coefficient (z)</td>
<td>Coefficient (z)</td>
<td>Coefficient (z)</td>
<td>Coefficient (z)</td>
</tr>
<tr>
<td>Other-Penalties ($ Million)</td>
<td>−0.162 (−0.33)</td>
<td>0.208 (0.67)</td>
<td>0.254 (0.89)</td>
<td>0.280 (0.63)</td>
<td>0.331 (0.90)</td>
<td>0.208 (0.67)</td>
</tr>
<tr>
<td>Employee Penalties ($ Million)</td>
<td>0.146 (0.31)</td>
<td>0.131 (0.47)</td>
<td>0.173 (0.49)</td>
<td>0.280 (0.63)</td>
<td>0.331 (0.90)</td>
<td>0.173 (0.49)</td>
</tr>
<tr>
<td>Employee Prison Sentences (Months)</td>
<td>1,122</td>
<td>1,122</td>
<td>1,121</td>
<td>1,121</td>
<td>1,122</td>
<td>1,122</td>
</tr>
</tbody>
</table>

Notes: These are exponential regression results, with z-statistics shown, using robust standard errors. *, **, and *** denote statistical significance at the 10-percent, 5-percent, and 1-percent levels, respectively.

### TABLE 7. Alternative models

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Other-Penalties ($ Million)</td>
<td>Coefficient (t)</td>
<td>Coefficient (t)</td>
<td>Coefficient (t)</td>
<td>Coefficient (t)</td>
<td>Coefficient (t)</td>
<td>Coefficient (t)</td>
</tr>
<tr>
<td>Other-Penalties ($ Million)</td>
<td>−0.075 (−0.14)</td>
<td>0.83 (0.58)</td>
<td>2.575 (1.12)</td>
<td>3.553 (1.37)</td>
<td>4.657 (0.75)</td>
<td>10.659 (0.57)</td>
</tr>
<tr>
<td>Employee Penalties ($ Million)</td>
<td>1,122</td>
<td>1,122</td>
<td>1,121</td>
<td>1,121</td>
<td>1,122</td>
<td>1,122</td>
</tr>
</tbody>
</table>

Notes: Models 1, 3, and 5 are OLS models, and Models 2, 4, and 6 are Tobit models. T-statistics are shown, using robust standard errors. *, **, and *** denote statistical significance at the 10-percent, 5-percent, and 1-percent levels, respectively.
I also demonstrate the results by using alternative models. Models 1, 3, and 5 in Table 7 are OLS models, while Models 2, 4, and 6 are Tobit models. The key results remain statistically insignificant.

Although it would be considered uncommon, it is also possible to remove extreme observations of one enforcement-outcome variable to investigate the robustness of the results on another enforcement-outcome variable. For instance, one can remove the extreme observations of Firm Penalties from the sample and show the robustness of CMSW’s results for Employee Penalties and Employee Prison Sentences. Models 1 and 2 of Table 8 show those results. If we drop the extreme observations this way, the results of CMSW remain robust. There is certainly not one right way to do robustness checks. But we need to observe that none of the 11 extreme observations of Firm Penalties matches with any extreme observations of Employee Penalties and only one of the extreme observations of Firm Penalties and Employee Prison Sentences is the same.

As for results on mere incidence, that is, yes/no on the involvement of a whistleblower on enforcement outcomes, CMSW write:

We find that whistleblower involvement is positively associated with the incidence of firm penalties ($p < 0.05$) and prison sentences ($p < 0.10$), but we do not find a significant result for the incidence of employee penalties. These results suggest that whistleblower involvement is associated with an 8.58% increased likelihood that the SEC imposes monetary sanctions on the firm and a 6.64% increased likelihood of criminal sanctions against the targeted employees. (CMSW, 157)

Given the natural correlations we would expect among severity of misconduct, likelihood of penalties, and whistleblowing, somewhat like the correlations among the severity of health emergencies, the likelihood of medical interventions, and calls to 9-1-1, it is surprising that CMSW did not find a statistically significant correlation for one of the enforcement-outcome categories. The other correlations are perhaps less than one would expect. I examine whether whistleblowing is associated with the incidence of firm penalties after dropping the extreme values of Firm Penalties. I use a logit model with a binary dependent variable indicating whether a firm penalty was assessed. Model 3 of Table 8 shows, unsurprisingly, since we just are dropping a few cases in the large sample of yes-or-no data, that the findings of CMSW for the incidence of Firm Penalties remain.
TABLE 8. Replicating the results of employee penalties, employee prison sentences, and the incidence of firm penalties without extreme (top 1 percent) firm penalties using the 1,133 SEC and DOJ enforcement actions 1978–2012 associated with alleged financial misrepresentation

<table>
<thead>
<tr>
<th></th>
<th>Employee Penalties ($ Million)</th>
<th>Employee Prison Sentences (Months)</th>
<th>Pr (Firm Penalties)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient ($z$)</td>
<td>Coefficient ($z$)</td>
<td>Coefficient (Chi-squared)</td>
</tr>
<tr>
<td>Whistleblower</td>
<td>1.295*** (2.87)</td>
<td>1.056*** (3.29)</td>
<td>0.867*** (9.55)</td>
</tr>
<tr>
<td>Control Variables</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>FF 12 Industry Fixed Effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>1,122</td>
<td>1,122</td>
<td>1,122</td>
</tr>
</tbody>
</table>

Note: Exponential regression results are reported for Models 1 and 2. Model 3 is a logit model. Z-statistics and chi-squared are shown, using robust standard errors. *, **, and *** denote statistical significance at the 10-percent, 5-percent, and 1-percent levels, respectively.

I also acknowledge the difficult issue involved in dropping extreme observations from any data. Some might liken that to throwing the baby out with the bathwater. Extreme observations in some settings can be highly economically important. In terms of frequency among firms, both whistleblowing and enforcement action cases would be considered uncommon in the population of public firms. The combination of these events can have big economic implications. Extreme observations can also help us understand a unique phenomenon. Although many extreme-setting studies in accounting and finance such as those by Merle Erickson, Michelle Hanlon, and Edward L. Maydew (2004), John R. Graham and Alan L. Tucker (2006), and Karthik Ramanna and Sugata Roychowdhury (2010) are non-generalizable to all firms, those studies still provide us some insights. We can learn something valuable from extreme settings. Most extreme observations may be especially important in helping us to understand a phenomenon. Nine of the 11 extreme observations for firm penalties in CMSW are purported to involve whistleblowers. Thus, one might argue that whistleblowers are particularly important in helping uncover the most severe violations, thus supporting the conclusion of CMSW’s study.

It is possible, however, that characteristics such as whistleblowing just happen to be present in extreme observations. One would want to look, by qualitative investigation, at the extreme observations and see whether whistleblowing was crucial. Katherine Guthrie, Jan Sokolowsky, and Kam-Ming Wan (2012) did such an investigation in their 2012 *Journal of Finance* paper. Guthrie et al. (2012) reexamine the results of Vidhi Chhaochharia and Yaniv Grinstein (2009). Chhaochharia and Grinstein (2009) find that CEO pay decreases by 17 percent more in firms whose boards were not compliant with the recent NYSE/NASDAQ independence requirements than in firms that were compliant. How-
ever, Guthrie et al. (2012) investigate two extreme observations in Chhaochharia and Grinstein’s (2009) sample and prove that in neither of the two cases could the board structure have been a reason that CEO compensation changed. After dropping those observations, Guthrie et al. (2012) find that board independence has no effect on the level of CEO pay. A similar type of investigation for those extreme observations, that is whether whistleblowing played an important role in those enforcement actions, would be ideal for CMSW. On page 164, however, CMSW state “most of the whistleblower allegations in our sample are obtained from OSHA, and we cannot directly observe whether the SEC or DOJ actually used the information from each OSHA whistleblower. As a result, these cases represent potential whistleblower involvement in an enforcement action.” In other words, we may not be able to investigate whether whistleblowers played a crucial role in those nine extreme cases given the nature of CMSW’s whistleblower data, though ideally we would like to see a qualitative investigation. We do not know whether whistleblowing in the few extreme observations was like 9-1-1 calls following a heart attack.

However, we can rely on the extensive analysis of whistleblowing cases for the alleged corporate misconduct by Dyck et al. (2010) to do our qualitative investigation for some of those extreme observations of CMSW. In this excellent and well-known study (957 Google Scholar citations as of January 25, 2019) Dyck et al. (2010) look at what actors bring corporate fraud to light by gathering data on a comprehensive alleged corporate fraud that took place in the U.S. firms with 750 million dollars in assets between 1996 and 2004. Dyck et al. (2010)’s data includes cases such as Enron and WorldCom, which are also extreme observations in CMSW. Thus, we can compare the codings for some of CMSW’s extreme observations with Dyck et al. (2010)’s classification of actors bringing fraud to light. Dyck et al. (2010)’s classifications include analyst, auditor, client or competitor, employee, equity holder, firm, industry regulator/government agency or self-regulatory organization, law firm, newspaper, SEC, and short-seller. Dyck et al. (2010) is an excellent example of how qualitative investigation should be conducted. I shall provide a lengthy quotation from the paper to show their way of clarifying and explaining how they classify each fraud case:

Our key variable is the identity of the actor who brings each fraud to light. To uncover the fraud detectors for each of our 216 cases, we search Factiva for news wires and articles over the period from 3 months prior to the class action period (defined as the period over which the suit claims misbehavior) to the settlement date or the current date, if the case is still pending. Our searches return approximately 800 articles per case. The point to reading so many articles for each case is to understand, as much as possible, the circumstances of the alleged fraud and the detector who reveals the
information.

In a number of cases, we find that the whistleblower is not the person labeled by the media as such. A chain of events initiated by another party may already be forcing the scandal to light when an individual expedites the process by disclosing internal information. For instance, Enron’s whistleblower by our classification is the Texas edition of the *Wall Street Journal*, not Sherron Watkins, who is labeled the Enron whistleblower. Of course, we do not wish to undercredit the importance of individuals who contribute details as the fraud emerges. However, our aim is to identify the initial force that causes a scandal to come to light.

To mitigate potential concerns about subjectivity in identifying the first actor to bring each fraud to light, we implement a meticulous procedure. The initial coding of each case was done by a research assistant (a law student) and, independently, by at least one of the authors. Where judgment was required, all three authors analyzed the case until a consensus was reached. A year after the initial coding we divided the cases into thirds and each of the authors recoded cases without referencing the prior coding. Again, when the coding was at all unclear, all three authors read the case to ensure consistency in interpretation. In the process of verifying our coding, we went back to our sources and created a list of the news article(s) that were most informative in pointing to which player was the whistleblower. We sent this document to academic colleagues who work in corporate governance and to the NBER corporate governance list soliciting comments regarding the details of particular cases. (Dyck et al. 2010, 2218–2222)

The online appendix of Dyck et al. (2010) shows the summaries of 216 fraud cases in their sample (link). In each of the summaries, they identify the responsible actor along with a representative quote from the available evidence. Table 3 of my paper shows the nine largest firm penalties of CMSW with employee whistleblowers. Six out of those nine cases are also in Dyck et al. (2010)’s sample; the three that are not are Siemens, AIG, and Fannie Mae.

Let’s take a closer look at those six cases: Enron, WorldCom, Halliburton, Reliant Resources, America Online, and Bristol Myers Squibb. For Enron, CMSW data show an employee as a tipster whistleblower. In other words, CMSW data claims that because of the employee allegation, the investigation for Enron was started. However, if we look at the sample of Dyck et al. (2010), we see that a newspaper brought the fraud to light and not the employee as suggested by CMSW’s whistleblowing data. In their online appendix, Dyck et al. (2010) provides qualitative evidence for a newspaper having been the ‘tipster’ in the Enron case:

---

9. I have omitted from the quotation here a parenthetical pointer to the appendix containing the list. Now the appendix is available on Adair Morse’s website (link).
The Texas version of the Wall Street Journal publishes a story in the fall of 2000 asking whether the profits from companies like Enron are just artifacts of the firms’ manipulation of marking assets to market. A few months later, Fortune and the New York Times publish articles questioning the ultimate origin of value in the stock run-up of Enron and covering Enron’s incredulous behavior in the California energy crisis, respectively. Meanwhile, following the Texas WSJ article, short sellers begin increased scrutiny of the firm and, in particular, into Enron’s financing entities. Not long after the CEO resigns in August 2001, information comes to light that the firm had misrepresented the value of its assets by billions of dollars, and related party transactions were siphoning value from the firm to the benefit of executives. A number of other improprieties emerge. Shareholders have claimed $30 billion in damage from now-defunct Enron; litigation continues. Officers settle with SEC for $64.4 million and Auditors and investment banks settle with SEC for $7.3 billion. (Dyck et al. 2010, online appendix p. 21)

Similarly, Dyck et al. (2010) indicate that their classifications for WorldCom, Halliburton, Reliant Resources, America Online, and Bristol Myers Squibb were, respectively: SEC, Newspaper, Industry Regulator, Auditor, and Firm. Thus, Dyck et al. show no employee tipsters for these five cases. CMSW code all five of these cases as having an employee whistleblower.

Only 26 out of 216 cases classified in Dyck et al. (2010) data are employee whistleblowers. Andrew Call, one of the coauthors of CMSW, used those 26 employee-based whistleblowing events as identified by Dyck et al. (2010) in Call, Kedia, and Rajgopal (2016), so he knew of Dyck et al. (2010). In CMSW he should have at least mentioned the inconsistencies. Dyck et al. (2010)’s data only goes through 2004, but some of CMSW’s cases, which go through 2012, such as Enron, are coded as employee tipster by CMSW but not by Dyck et al., which instead cites a newspaper as the tipster.

Conclusion

With the abundance of data in our time, the number of empirical research studies in accounting and finance that show some statistically significant relationship has increased substantially. Many policy makers use those findings to initiate new policies or justify existing policies. Government laws and regulations that aim to prevent corporate misconduct can burden the economy and do little good for their intended purpose. Some researchers even question the efficacy of the government’s enforcement of securities laws relative to private enforcement. For instance, Rafael La Porta, Florencio Lopez-De-Silanes, and Andrei Shleifer
(2006) find that private enforcement is more important than public enforcement for financial market development.

In this paper, I suggest that my investigations and considerations should move all of us further toward doubt and skepticism about the findings of CMSW. We should not reach a definitive conclusion about the value of whistleblowing programs based on CMSW.

Unfortunately, the extreme-values problem is not unique to CMSW. This issue is very common in the financial misconduct literature (e.g., Files, Martin, and Rasmussen 2018; Correia 2014; Yu and Yu 2011). Therefore, I recommend six practices that financial misconduct researchers should implement in their future studies to address the problem and related issues:

1. **Be upfront about the extreme-values problem**: Financial misconduct scholars should give adequate attention to the extreme-values issue in their analysis. Extreme values in the data should be identified and readers should be alerted clearly and early in the analysis about the identities and directions of those extreme observations. Mentions of extreme values in the study that are brief, peripheral, and too late can be considered as deceptive. Scholars can show the extreme observations such as WorldCom and Enron, together with the main variable of interest, in a table as I have done in Table 3.

2. **Show the results without extreme observations in an analysis**: Extreme observations such as Enron, WorldCom, Cendant, Colonial Bancgroup are here to stay. Scholars should investigate and show how extreme observations influence their main findings. I could find only one paper (Files 2012) in the literature that shows in an analysis that the main findings of a paper remain robust after excluding the top one percent of enforcement outcomes from the data.

3. **Be clear about the lineage of each and every coding**: Especially when plagued by the extreme-values problem, it is crucial to reduce and resolve any fuzziness about how the coding of each observation was arrived at.

4. **Do a qualitative investigation of extreme-value observations**: Financial misconduct scholars should do a qualitative investigation of extreme-value observations to understand whether the variable in question really play a crucial role (e.g., Guthrie et al. 2012). For instance, the variable in question in CMSW is Whistleblower and CMSW’s data show that the top nine observations for firm penalties have employee whistleblowers. If CMSW had done a qualitative investigation, they might have seen that it was other actors (e.g., the SEC, newspapers, industry regulators, auditors, and firms) and not employees that played a crucial role for
some of those extreme cases (e.g., Reliant and Enron).

5. **Be cautious in identifying implications of findings.** Financial misconduct scholars should realize that because of the extreme-value problem in enforcement outcomes, the findings of the literature cannot be generalized. Scholars should not present (either implicitly or explicitly) their findings as providing solid support for government policies. They need to be cautious not to overstretch their findings. For instance, in the third paragraph of their conclusion, CMSW accurately state that “our study should not be interpreted as an examination of the efficacy of...any whistleblowing program,” but in the final paragraph they state that “this study makes important contributions...to policy discussions on the efficacy of...formal whistleblowing programs.” Which one of those statements should we take seriously? Scholars should avoid this type of ambiguous statement of the implications of an analysis. When scholars give an interview about their studies to the media, they should also mention the extreme-value problem and other limitations of their studies.

6. **Avoid one-sided analysis:** Policies such as the SEC Whistleblower Program have negative as well as positive consequences. Financial misconduct scholars should consider both benefits and costs of policy judgments in their analysis. We should never suppose that private enterprise and free markets work perfectly, but even less should we assume that government does.

**References**


Calomiris, Charles, Kenneth Scott, and Chester Spatt. 2010. Beyond Dodd-Frank:


Hiltzik, Michael. 2014. Politically Connected Companies Get a Break from the SEC, Study Says. Los Angeles Times, August 12. Link


About the Author

Emre Kuvvet is an Associate Professor of Finance at Nova Southeastern University. He has published articles in International Business Review, Financial Review, Journal of Investing, Review of Pacific Basin Financial Markets and Policies, Journal of Trading, and Journal of Financial Planning. He received the Outstanding Paper Award from Southern Finance Association in 2010. His works have been noticed in articles in the Chicago Tribune, the New York Times, and the CFA Digest. Before joining NSU, he was a visiting assistant professor of finance at Texas A&M University. He received his B.S. from Marmara University in his native country of Turkey, his M.S. from Rochester Institute of Technology, and Ph.D. from the University of Memphis. His email address is ekuvvet@nova.edu.

Discuss this article at Journaltalk: https://journaltalk.net/articles/5978/
A Response to “Are a Few Huge Outcomes Distorting Financial Misconduct Research?”

Andrew C. Call¹, Nathan Y. Sharp², and Jaron H. Wilde³

Kuvvet’s paper discusses extreme observations in research on financial misconduct and also examines the robustness of the findings in Call, Martin, Sharp, and Wilde (2018, Journal of Accounting Research) (hereafter CMSW) to the removal of these observations.

The published version of CMSW empirically addresses the role of extreme observations in enforcement actions with an estimator designed specifically to handle skewed data (Poisson pseudo-maximum likelihood) and with additional robustness tests, including one focused on the incidence rather than the magnitude of penalties. Other claims offered by Kuvvet reflect a misunderstanding of both the enforcement action setting and the whistleblower designations in CMSW. For example, unlike many other settings in accounting, finance, and economics where the focus is often on the average firm, the enforcement action setting is inherently extreme. As another example and as explained in the published paper, CMSW’s “tipster” and “non-tipster” whistleblower designations are defined specifically in relation to the enforcement process and are not an attempt to identify the individual who first uncovered the misconduct. Lastly, Kuvvet argues that CMSW’s findings speak to correlation rather than causation. The published version of CMSW makes this point clearly throughout the paper.

¹. Arizona State University, Tempe, AZ 85281.
². Texas A&M University, College Station, TX 77843.
³. University of Iowa, Iowa City, IA 52242.
References


About the Authors

Andrew C. Call is a professor at Arizona State University. His email address is andycall@asu.edu.

Nathan Y. Sharp is an associate professor at Texas A&M University. His email address is nsharp@mays.tamu.edu.

Jaron H. Wilde is an associate professor at the University of Iowa. His email address is jaron-wilde@uiowa.edu.

Go to archive of Comments section
Go to March 2019 issue

Discuss this article at Journaltalk: https://journaltalk.net/articles/5979/
Is the United States an Outlier in Public Mass Shootings?  
A Comment on Adam Lankford

John R. Lott, Jr. 1 and Carlisle E. Moody 2

In 2016, Adam Lankford published an article in *Violence and Victims* titled “Public Mass Shooters and Firearms: A Cross-National Study of 171 Countries.” In the article he concludes: “Despite having less than 5% of the global population (World Factbook, 2014), it [the United States] had 31% of global public mass shooters” (Lankford 2016, 195). Lankford claims to show that over the 47 years from 1966 to 2012, both in the United States and around the world there were 292 cases of “public mass shooters” of which 90, or 31 percent, were American. Lankford attributes America’s outsized percentage of international public mass shooters to widespread gun ownership. Besides doing so in the article, he has done so in public discourse (e.g., Lankford 2017).

Lankford’s findings struck a chord with President Obama:

“I say this every time we’ve got one of these mass shootings: This just doesn’t happen in other countries.” —President Obama, news conference at COP21 climate conference in Paris, Dec. 1, 2015 (link)

2. College of William and Mary, Williamsburg, VA 23187; Crime Prevention Research Center, Alexandria, VA 22302. We would like to thank Lloyd Cohen, James Alan Fox, Tim Groseclose, Robert Hansen, Gary Kleck, Tom Kovandzic, Joyce Lee Malcolm, Craig Newmark, Scott Masten, Paul Rubin, and Mike Weisser for providing helpful comments. We appreciate the able help compiling these cases that we received from Rujun Wang, Roger Lott, Andrzej Zdzitowiecki, and Daniel Beets. Rujun and Roger also helped out in many other ways with this research.
"The one thing we do know is that we have a pattern now of mass shootings in this country that has no parallel anywhere else in the world." —President Obama, interview that aired on CBS Evening News, Dec. 2, 2015 (link)

"You don't see murder on this kind of scale, with this kind of frequency, in any other advanced nation on Earth." —President Obama, speech at U.S. Conference of Mayors, June 19, 2015 (link)

The Obama administration justified these and similar claims by citing the then-unpublished version of Lankford’s paper (Lee 2015).

Lankford’s study was also a big hit with the media. Beginning in the summer of 2015, it received uncritical coverage in hundreds of news stories, in at least 35 different countries.3 Headlines accepted his findings as fact. Here are a few prominent examples:

- *Wall Street Journal* (subheading): “U.S. produces more mass shootings than other countries” (Palazzolo and O’Connell 2015)
- *Los Angeles Times*: “Why the U.S. is No. 1—in mass shootings” (Healy 2015a)
- *Sydney Morning Herald*: “Why the U.S. is No. 1 in Mass Shootings” (Healy 2015b)
- *Time* magazine: “Why the US has 31% of the World’s Mass Shootings” (Basu 2015)

When Lankford’s research began to receive news coverage in the summer of 2015, one of us (Lott) asked to see the paper, but without success. Lankford’s paper was published at the end of January 2016. Lankford has refused many requests to share his data. His dataset of 292 cases (90 U.S., 202 non-U.S.) remains unavailable.

Worldwide, from 1966 through 2012, the number of non-U.S. shooters killing at least four people (not including the perpetrators) that today the English-

3. For example: Australia, Austria, Argentina, Armenia, Brazil, Canada, China, Colombia, Costa Rica, Denmark, Egypt, Finland, France, Germany, Hungary, India, Indonesia, Iran, Ireland, Japan, Malaysia, Mexico, Peru, Portugal, Russia, Slovenia, South Africa, Spain, Sweden, Turkey, UK, Venezuela, Vietnam, and Cuba. The information on the worldwide coverage for Lankford’s work is available on his website (link).
speaking world could aptly identify as “public mass shooters” is vastly more than 202, as we show below. The number is very hard to determine with any accuracy. However, we are comfortable saying that the number is upwards of 2,000. Yet Lankford reports a mere 202 non-U.S. public mass shooters. How did he arrive at that number?

Only very recently, in February 2019, did we begin to get some clarification of Lankford’s definitions, of what his 292 cases are cases of. The chief editor of Econ Journal Watch, Daniel Klein, wrote to Lankford, requesting the data and replies to 13 questions of clarification. Although Lankford declined to release his data, he provided replies to the 13 questions, with permission to post them online (link). We shall refer to that document containing Lankford’s replies as “the Q&A” (and cite it as Lankford 2019). Even with the Q&A, it is still unclear how Lankford arrived at his numbers.4

Our assessment of Lankford (2016) comes to the conclusion that Lankford implicitly defined a concept of a lone-wolf public mass shooter (see Lankford’s answer A1 in the Q&A). It is true that the United States shows an outsized number of lone-wolf shooters. But once a concept of lone-wolf shooter is made explicit, one would naturally ask whether there might be explanations other than gun prevalence for why the United States has an outsized number of lone-wolf shooters.

Another explanation presents itself: Magnets for dangerous individuals are much more commonly found in other countries, magnets which then make packs of wolves, as it were—magnets such as terrorist networks, ethnic and clan groups, insurgency groups, and so on. Around the world, mass shootings occur pervasively, but fewer of the lone-wolf sort. Understanding the dynamics of social conflict around the world exposes the irresponsibility of saying that the United States has more lone wolves because it has more guns. Rather, maybe the United States has more lone shooters because it has more loners in general. The United States is less clannish and less rooted; it is more ethnically diverse and less kin-based; its culture and social structure is more oriented, even exceptionally so, toward autonomy and individuality.

In this article, we suggest that Lankford has misled readers by defining and using terms in unconventional ways. While Lankford’s paper includes terrorist shooters in the United States such as the Islamic extremist Nidal Hasan of the 2009 Fort Hood massacre, he strips out almost all—we do not know how much—of the entire spectrum of terrorism-related shootings outside the United States. Even though Hasan had had, for example, extensive communications with the radical Islamist

---

4. In the time up to the completion and publication of the present paper we did not actually see Lankford’s answers in the Q&A, but rather have heard some of them read to us by the editor Daniel Klein. Where Lankford’s answers are quoted in the present paper, Klein mediated by inserting the exact text.
imam Anwar al-Awlaki, who, according to U.S. government officials, has planned terrorist operations of al-Qaeda, Hasan was included by Lankford apparently because Hasan’s attack was not “sponsored.”

However, established, official definitions of ‘public mass shooting’ and similar terms do not exclude any incidents of terrorism, irrespective of whether they are ‘sponsored.’ Despite claiming that he followed standard definitions, Lankford’s semantic move to exclude “sponsored acts of…terrorism” was made cryptically—only in those few words, found at the top of page 191 of his 2016 article. The exclusion is especially hard to understand given that Lankford consistently claims that cases such as the Columbine and Sandy Hook shootings are “functionally similar to terrorism” (p. 188). Virtually all of the media coverage simply missed or neglected that exclusion of terrorism, and its significance. In none of the seven media articles bulleted above does the word terror or its variants appear—and that is typical of the news coverage of his study.\(^5\)

Non-problematic aspects of Lankford’s “public mass shooter”

Lankford calls his unit of investigation, that is, the thing to be counted, “public mass shooter.” Lankford takes some aspects of the definition directly from the New York City Police Department’s (NYPD) 2012 Active Shooter report. Lankford (2016, 190) aptly notes that “active shooters” are also known as “rampage shooters” or “public mass shooters.” In defining that creature which goes by several names, Lankford (ibid.) first of all follows the NYPD report, saying: “According to the formal definition, their attacks must have (a) involved a firearm, (b) appeared to have struck random strangers or bystanders and not only specific targets, and (c) not occurred solely in domestic settings or have been primarily gang-related, drive-by shootings, hostage-taking incidents, or robberies (Kelly, [NYPD] 2012).” On that, we have no quarrel. Furthermore, Lankford (2016, 191) says that he will count only shooters who shot and killed at least four (other) persons. We have no quarrel with that, either. But before turning to the problematic matter of Lankford’s treatment of terrorism, let’s look at the established and conventional definitions of ‘public mass shooter.’

---

5. Here is video of Lankford (2015) being asked how he did his study and saying absolutely nothing about terrorism.
Established definitions do not exclude terrorism

Conventional and official sources, including all those that Lankford cites, do not exclude any kind of terrorism from their definitions of ‘active shooter’ or ‘public mass shooter.’ Lankford (2016, 190) says: “Data for this study were drawn first from the New York City Police Department’s (NYPD) 2012 Active Shooter report.” That report states quite clearly its unit of investigation:

The NYPD included only those incidents carried out by attackers that met the DHS [Department of Homeland Security] definition of an active shooter: an individual actively engaged in killing or attempting to kill people in a confined and populated area. The NYPD further restricted this definition to exclude: gang-related shootings, shootings that solely occurred in domestic settings, robberies, drive-by shootings, attacks that did not involve a firearm, and attacks categorized primarily as hostage-taking incidents. (NYPD 2012, 10)

The NYPD thus does not exclude terrorism.

Lankford (2016, 191) goes on to say that, for his study, “the NYPD report was therefore supplemented with additional data from the FBI’s 2014 active shooter report (Blair & Schweit 2014) and with data gathered on incidents from other countries. All efforts were made to ensure that the same data collection methodology employed by the NYPD was used to obtain this information.” The FBI report contains a lengthy clarification of its unit of investigation, which we need to quote extensively:

The agreed upon definition of an active shooter by U.S. government agencies—including the White House, U.S. Department of Justice/FBI, U.S. Department of Education, and U.S. Department of Homeland Security/Federal Emergency Management Agency—is ‘an individual actively engaged in killing or attempting to kill people in a confined and populated area.’ The FBI extends this definition to include individuals, because more than one shooter could be involved in some incidents. Implicit in the definition is that the subject’s criminal actions undertaken include the use of a firearm. Though the federal definition includes the word confined, the FBI excluded this word when considering active shooter incidents. This is because the term confined

6. Lott and Landes (2001; 2003) and Lott (2010; 2018) are other studies on public mass shootings that also include terrorism cases and are in agreement with Lankford (2016, 188) that attacks such as Columbine are “functionally similar to terrorism.”
could be interpreted to omit incidents that occurred outside a building, when in actuality, many incidents originated outside or progressed from indoors to outdoors, or vice-versa, or occurred entirely along a route of travel or at various locations.

The FBI developed discriminating factors to further differentiate potential active shooter incidents, considering for inclusion:

- Shootings in public places;
- Shootings occurring at more than one location;
- Shootings where the shooters’ actions did not appear to be another criminal act;
- Shootings resulting in a mass killing;
- Shootings indicating an apparent spontaneity by the shooter;
- Shootings where the shooters appeared to methodically search for potential victims; or
- Shootings that appeared focused on injury to people, not buildings or objects.

Because the risk to civilians in active shooter incidents appears to do with the apparent randomness of so many victims, for purposes of this study, an event was excluded if research established it involved primarily the following factors:

- Conflicts arising from self-defense;
- Gang violence;
- Contained residential or domestic disputes;
- Controlled barricade/hostage situations;
- Crossfire as a byproduct of another ongoing criminal act; or
- Drug violence. (Blair and Schweit 2014, 44)

Thus, the FBI does not exclude terrorism.

Besides the NYPD and FBI sources that Lankford relies on, we can further bolster the claim that ‘public mass shooter’ does not exclude terrorism. In a short article at Vox titled “The Debate Over How to Define Mass Shootings Is Ridiculous,” German Lopez (2015) refers to criminologists and considers fine points over whether to “exclude domestic, gang, and drug violence,” how many killed, etc. Nowhere in Lopez’s article does the word terror or any of its variants appear—excluding terrorism is not even considered. Likewise, articles in Mother Jones (Follman et al. 2019), PolitiFact California (Nichols 2017), the Washington Post (Ingraham 2015), and elsewhere do not exclude terrorism. We know of only one source (Bjelopera et al. 2013) that excludes terrorism from its definition of ‘active shooter’
or ‘public mass shooter.’

What Lankford says about his handling of terrorism

In explaining his unit of investigation, Lankford provides the NYPD definition. But he then adds the following sentence:

For this study, attackers who struck outdoors were included; attackers who committed sponsored acts of genocide or terrorism were not. (Lankford 2016, 190–191, our emphasis).

The “outdoors” clause is insignificant; what matters is the “sponsored acts” clause. Lankford provides no clarification or elaboration. Casual readers in the United States might not take note of the importance of this additional feature, because terrorist mass shootings have been very uncommon in the United States. But Lankford’s study is an international comparison. In other countries terrorism-related mass shootings have been much more common—and Lankford has quietly excluded most of them from his non-U.S. count, under the cover of the cryptic sentence just quoted.

That sentence appears on the fifth page of his 14-page article. After that sentence, “terror” never again appears. As in his title, abstract and elsewhere, Lankford goes on to speak of public mass shooters/shootings as though his unit of investigation is the same as the NYPD/FBI (with the clarification of four or more people shot and killed). Thus, not only does Lankford fail to clarify the “sponsored acts” condition, he fails to call attention to its critical significance of excluding all but lone-wolf incidents.

What does Lankford mean when he says that he excluded “sponsored acts”? What does it mean for an act to have been “sponsored”? In his article Lankford

---

7. Bjelopera et al. (2013, 1) says, “this report uses its own definition for public mass shootings. These are incidents occurring in relatively public places, involving four or more deaths—not including the shooter(s)—and gunmen who select victims somewhat indiscriminately. The violence in these cases is not a means to an end—the gunmen do not pursue criminal profit or kill in the name of terrorist ideologies, for example” (see also ibid., 6–7). We consider this definition to be unworkable, since any such act can be seen as a means to an end. Moreover, the example given at the end of quotation would clearly exclude the 2009 Fort Hood shooting, which Lankford includes.

8. From the Q&A (see Q9 and Q10), it is clear that Lankford included the “outdoors” clause merely to clarify that the word “confined” in the NYPD’s expression “a confined and populated area,” quoted in Lankford (2016, 190), was not to be understood as meaning that the research counted only those acts that occurred indoors (Lankford 2019, 4).
says absolutely nothing to address that critical question. In the Q&A, Klein asked him to clarify (Q8); Lankford’s answer (2019, 3–4) confirms that some such line is being drawn but it does not address the issue of what constitutes the line between sponsored and not-sponsored acts of terrorism.

For any given act of mass shooting, one might guess at a broad range of interaction that might be deemed sponsorship, from encouraging to informing to training to equipping to planning and funding. Furthermore, researchers have imperfect information about the preparations of an act, which requires further clarification of how the researchers classify acts. One might argue, for example, that if a terrorist organization claims credit for the act, as al-Qaeda did for Hasan’s Fort Hood massacre, that is evidence of sponsorship. One has to draw a line between sponsored and not sponsored, explain it, and then apply it consistently.

Lankford’s response to Q8 amounts to saying that the 2012 Sandy Hook elementary school shooting falls on one side of his line, while the 2008 Mumbai attacks fall on the other (Lankford 2019, 3–4). It is as though a researcher said that he did an analysis that separated the United States into two regions, and, when asked to clarify the line between the two regions he replies only that New York is in one region and Los Angeles is in the other. Obviously, the way to clarify Lankford’s line would be to cite, not cases far from the line, but cases close to it. Lankford declined to do so.

But it gets worse. Now we quote the sentence just quoted, but also the one right after it:

For this study, attackers who struck outdoors were included; attackers who committed sponsored acts of genocide or terrorism were not. This is consistent with the criteria employed by the Federal Bureau of Investigation (FBI) in its 2014 active shooter report (Blair & Schweit, 2014). (Lankford 2016, 190–191, our emphasis)

The “This is consistent with…” sentence gives the impression that Lankford’s criterion—which excludes “sponsored acts”—is the same as the FBI’s—which does not involve any such exclusion. Klein queried Lankford about the matter in the Q&A. Here is Q4 and Lankford’s full response:

Q4: In the quotation above, I have put the pronoun “This” in boldface. Please clarify what it is that we should understand to be the antecedent of “This”. That is, what is it that is consistent with the criteria employed by the FBI in its 2014 active shooter report?
A4: The FBI list also included attackers who struck outdoors and did not include cases where individuals, rather than killing of their own volition, were engaging in sponsored (or commissioned, if you prefer that word) acts of
genocide or terrorism. (Lankford 2019, 3)

The domain of the FBI report is limited to the United States. Notice how Lankford chose to say “[t]he FBI list…did not include cases where,” and not: the criterion was such-and-such. One may say that the application of Lankford’s criteria on the domain of the FBI report yields a list “consistent with” the application of the criteria actually used by the FBI. But what Lankford says is untrue: The two sets of criteria (Lankford’s, FBI’s) are not consistent; they are importantly different. What are consistent are their respective applications on the particular domain. Notice how he speaks of the FBI and not the NYPD report, which includes non-U.S. cases. If he had included the NYPD report in the sentence there would not be a smoke-screen interpretation for what he had written.

After the “sponsored acts” and “This is consistent” sentences, in the next paragraphs Lankford does shift back to focusing on the NYPD report, saying that he supplemented that information with his own searches for mass shootings abroad. He writes: “All efforts were made to ensure that the same data collection methodology employed by the NYPD was used to obtain this information” (2016, 191). But without qualifying that statement by adding the clause excluding “sponsored acts” of terrorism/genocide, that sentence is terribly misleading. The NYPD method does not exclude terrorism, and their list includes many obvious cases of terrorism. For example, in addition to the 2008 Mumbai attack, they include the following:

May 15, 1974: Terrorists from the Popular Front for the Liberation of Palestine opened fire at an elementary school in a series of attacks that killed 26 people and wounded 70 others. (NYPD 2012, 147)

December 27, 1985: Three gunmen belonging to the Abu Nidal Organization opened fire at the El-Al ticket counter at Vienna’s Schwechat Airport, killing three people and wounding 30 others. (ibid., 203)

March 6, 2008: Alaa Abu Dhein opened fire in a crowded library at the Mercaz Harav Yeshiva in Jerusalem, killing eight teenage students and wounding 11 others. (ibid., 102)

December 13, 2011: Nordine Amrani opened fire and threw four stun grenades into a crowd at Saint-Lambert square in Liege, Belgium, killing six people and wounding 125 others. (ibid., 34)

Thus, directly after slipping the belated “sponsored acts” condition into his description of method, Lankford immediately talks in a way that, at best, obscures it and, at worst, aims at erasing any awareness the reader may have snatched from
the terse “sponsored acts” clause.

Lankford claims that his paper’s findings are “based on its quantitative analysis of all known public mass shooters who attacked anywhere on the globe from 1966 to 2012 and killed a minimum of four victims (N=292)” (2016, 188, our emphasis). That sounds authoritative, but apparently “all known” means simply ‘all the cases known to me that are not sponsored terrorism according to my idiosyncratic definition.’ Likewise, Lankford says, “Complete data were available for 171 countries, and they averaged 1.7 public mass shooters per country from 1966 to 2012” (ibid., 192). This average indicates that Lankford is focusing primarily on ‘lone wolf’ attacks.9 It also sounds very precise, but it is unlikely that he could have found data on mass shootings in, say, Mozambique in 1977. For less developed parts of the world such as Africa or Latin America, it can be very difficult to obtain news stories from even a decade ago. It is virtually impossible to obtain news stories on all of the cases of four or more people being killed in the 1960s or 1970s.10 Finally, Lankford makes no use of the obvious best source for terrorist mass shootings, the University of Maryland’s Global Terrorism Database (GTD).

Is it sensible to exclude terrorism?
The importance of magnets for dangerous individuals

Would it be sensible to exclude all terrorist incidents except when deemed by Lankford not to have been “sponsored”? Would there then be sense in an international comparison of public mass shooters so qualified? We will use the Greek letter lambda to designate Lankford’s concept, thus λ-shooters, conveniently indicating both Lankford as creator of the concept and ‘lone wolf’ as the type of shooter that Lankford implicitly means when he says “public mass shooter.”

Lankford finds that λ-shooters are much more common in the United States and that is because of its relatively liberal gun policy and widespread gun ownership. That is why the media hyped his study.

---

9. Lankford mentions Columbine as an example of a public mass shooting (Lankford 2016, 187). It was perpetrated by two shooters, and hence not a lone wolf. So Lankford’s shooters are not lone wolves, strictly speaking, but his answer A1 in the Q&A says that in his data set the ratio of shooters to shootings is “approximately 1:1” (Lankford 2019, 1).

10. The U.S. has computerized databases of news stories, but even these are greatly limited prior to 1991. For 1991, there are at least 389 newspapers included in the Nexis database. Just prior to 1991, there are only 31 newspapers. The number quickly gets smaller and smaller as one goes further back in time. And, of course, the English-language news media of decades ago couldn’t be counted on to cover public mass shootings in Europe, let alone Africa or other parts of the world.
Having created the concept of a λ-shooter, Lankford makes international comparisons—most notably, a U.S./non-U.S. comparison—to draw a lesson about U.S. gun policy. We see big problems in using international comparisons of λ-shooters to address gun policy, even assuming that the idea is well-defined and accurately measured for both the U.S. and the rest of the world.

It is important to emphasize that terrorist mass shootings are much less common in the United States than in the rest of the world. Counting only λ-shooters removes most of the terrorist mass shooters, a maneuver that alters the non-U.S. picture much more than it alters the U.S. picture.

We have two points against drawing a lesson as Lankford does from the U.S./non-U.S. comparison of λ-shooters. First, it is plausible that there is a causal mechanism from gun policy, gun ownership, and gun carrying to the prevalence of terrorist mass shootings. If terrorists open fire in the United States, it is more likely, relative to, say, in Europe, that those fired on, or someone else, will fire back. That hazard has a deterrent effect before the fact; it also means that when a terrorist starts shooting, he is less likely to kill four victims. A bumper sticker for the point would be: More guns, less terrorism.

But there is a more important causal mechanism that vitiates the lesson that Lankford and others draw from λ-shooter comparisons. Suppose that every national population, on every continent, has its share of angry, violence-prone, even suicidal individuals, most of whom are young men. In many parts of Europe, Asia, Africa, the Middle East, and South America, such a dangerous individual is likely to find a welcome, and an outlet, in a terrorist network. Terrorist groups are magnets for dangerous individuals.

If a dangerous individual enters a terrorist network abroad and then commits a mass shooting, that shooter is removed by Lankford because he is deemed “sponsored.” That angry young shooter is not a λ-shooter. A parallel young man in the United States, who does not readily find a terrorist welcome and outlet, proceeds, let’s say, to commit a mass shooting—as a ‘lone wolf’ and a λ-shooter. He is counted by Lankford, but his counterpart abroad is not. Here the bumper sticker is: More magnets for dangerous individuals, fewer λ-shooters.

Around the world, mass shootings occur pervasively, but many fewer as lone-wolf mass shootings. Understanding the dynamics of social conflict around the world exposes the irresponsibility of saying that the United States has more lone wolves because it has more guns. Perhaps because the United States has more lone shooters because it has more loners in general; it is less clannish and less rooted; more ethnically diverse and less kin-based; its the culture and social structure is more oriented toward autonomy and individuality.

Support for that idea—more magnets for dangerous individuals, fewer λ-shooters—comes from Lankford himself. He has done ample research with that
implication, including his book *The Myth of Martyrdom: What Really Drives Suicide Bombers, Rampage Shooters, and Other Self-Destructive Killers* (Lankford 2013). Here is the book description:

> For decades, experts have told us that suicide bombers are the psychological equivalent of America’s Navy SEALs—men and women so fully committed to their cause or faith that they cease to fear death. In *The Myth of Martyrdom*, Adam Lankford corrects this misconception, arguing that terrorists are driven to suicide for the same reasons any civilian might be: depression, anxiety, marital strife, or professional failure. He takes readers on a journey through the minds of suicide bombers, airplane hijackers, ‘lone wolf’ terrorists, and rampage shooters, via their suicide notes, love letters, diary entries, and martyrdom videos. The result is an astonishing account of rage and shame that will transform the way we think of terrorism forever. Lankford convincingly demonstrates that only by understanding the psychological crises that precipitate these acts can we ever hope to stop them. (Lankford 2013, dust jacket)

The point is made again in the 2016 article, when Lankford says that public mass shootings “are typically premeditated attacks that strike random, innocent victims. *This makes them functionally similar to terrorism*” (p. 188, our emphasis). We agree with Lankford that premeditated attacks that strike innocent victims with the desire to get media attention are a lot like terrorism. In the U.S. we might be worried about a lone wolf who wants to kill as many innocent people as possible in a public place, whereas in France we might be worried about a terrorist who wants to kill as many innocent people as possible in a public place. Being shot in a public place by a terrorist in London or Paris is, to use Lankford’s expression, “functionally similar” to being shot by a λ-shooter in Los Angeles or New York City.

Lankford has made similar comments to the press, arguing that all the mass shooters in his data set share a common set of traits such as “a sense of victimization, a pattern of seeking negative attention, and being suicidal or not caring whether they live” (Barrett and Berman 2018, paraphrasing Lankford). He also argues that these shared psychological traits may be more important than their agendas (ibid.).

Lankford’s insights about psychological preconditions have profound implication for any U.S./non-U.S. comparison. Most importantly, it casts doubts on the usefulness or reasonableness of excluding terrorist acts. However, if one is going to exclude some terrorist acts one should explain why. If the researcher

---

11. This exact text can be found at, e.g., the Amazon.com page for the book ([link](http://example.com)), while a longer variant is found at the site of the publisher, Macmillan ([link](http://example.com)).
decides to exclude some, he should explain his reasoning on which to exclude. Once the researcher has decided on a line for excluding some terrorist acts, he should try to make clear the line used and how it was implemented in relation to source information. Lankford (2016) does none of this: He does not explain why any terrorist acts should be excluded nor, given that some are excluded, the issues involved in making that decision. All he tells the reader is that “sponsored” acts were excluded, and even that statement is made fleetingly and given none of the prominence it deserves.

We have suggested that magnets for dangerous individuals reduce the number of \( \lambda \)-shooters. Because such magnets are generally much more prevalent outside the United States, the ‘magnet factor’ represents a major possible explanation for why the United States has an outsized number of lone-wolf shooters. One type of magnet is terrorist groups and networks, which we focus on because terrorist shooters are not excluded from standard definitions of ‘public mass shooter.’ But it is important to realize that the magnet factor goes beyond terrorism. There are other sorts of magnets that are excluded from the standard definitions, such as insurgent groups, genocidal actions, state-sponsored violence, kidnapping rings, and gang violence. With the possible exception of gang violence, these types of shootings, too, are more prevalent outside the United States. As a rival explanation for differential rates of lone wolf shooters, then, the magnet factor extends well beyond terrorism.

Our empirical investigation

Moving from shooters to shootings

Whereas Lankford’s unit of investigation is shooter (with four or more killed, excluding perpetrators), our unit of investigation is cases/incidents of shootings (with four or more killed, excluding perpetrators).

We choose to work primarily with shootings, rather than shooters, for a number of reasons. First, the official definition of public mass shooting does not exclude terrorist acts, which often involve multiple shooters. Investigators often do not know exactly who was killed by whom, and reports often do not specify all that investigators might know. The information on attacks worldwide, reaching back decades, are often unclear on the number of shooters, and typically will not include the detailed knowledge required to determine the exact number of shooters and how many people were killed by each. Second, news and other sources usually report by incident, and describe the overall event. Third, criminologists and other
researchers typically quantify cases or incidents, not shooters.

As we turn to our own empirical investigation, then, the unit is “shootings.” One issue that was relatively common among cases in Africa and some other less developed countries is that many news stories only reveal the total killed and the number of places attacked. Without more information, we cannot determine whether each target meets the criterion of four or more people being killed. Twenty people may have been killed on different days in three different towns that are many miles apart; while it is possible that all three attacks satisfy our definition, we took the more conservative route and counted this as only one attack. This causes a slight underestimate of the total number of shootings.

How we collected our data

Our primary source is the GTD, which has data on over 170,000 attacks from 1970 to 2016 (Global Terrorism Database 2018; LaFree et al. 2015). The GTD defines terrorist attacks as “the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation.” The database lists attacks that were carried out using firearms, incendiary, knives, bombs, vehicles, chemical, biological, or radiological weapons. We included only those cases that indicated firearms as the principal weapon used in the incident.

GTD divides its list of attacks into six categories: 12

1. Terrorism
2. Insurgency/Guerilla Action
3. Other Crime Type
4. Intra/Inter-group Conflict
5. Lack of Intentionality
6. State Actor

Excluded entirely from all our metrics are: Insurgency/Guerilla Action, Lack of Intentionality, and State Actor (categories 2, 5, and 6). As for Intra/Inter-group Conflict (category 4), we excluded all but ten cases because our investigation of the reporting of those ten cases found little indication of intra- or inter-group conflict in the sense of two groups engaged in mutual hostilities. This leaves us with two categories: Terrorism (category 1) and Other Crime Type (category 3), plus the ten cases from Intra/Inter-group Conflict (category 4). But we also exclude individual

12. The five categories other than Terrorism are the designations for the coding of GTD’s “Doubt Terrorism Proper” field (see GTD 2018, 11).
cases within those categories if the criteria of the NYPD and FBI would exclude them; thus for example we exclude robberies and gang violence.

Lankford’s study period was 1966 through 2012. Since we know it is almost impossible to find information on all mass shooting incidents before the advent of the World Wide Web, we examined the last 15 years of Lankford’s period of study: 1998 to 2012.\textsuperscript{13} We started with the GTD list of cases for that period. We then reviewed each case using Nexis (\url{link}) and general web searches to determine whether they met our definition (which is the same as the NYPD/FBI definition, except that we excluded insurgency-related shootings). More than 50 percent of the shooting cases identified by the GTD fell in categories 2, 4, 5, or 6, and thus less than 50 percent met our definition of public mass shootings.\textsuperscript{14}

We included 66 cases that involved kidnapping that satisfied our criteria for a public mass shooting.\textsuperscript{15} At one extreme, attackers start killing people and then take hostages when the police or military arrive—a type of case clearly within the purview of this data. At the other extreme, attackers kidnap people and then kill them—a type that is less obvious, though the NYPD includes two cases where a kidnapping preceded a shooting and in one of those cases the kidnapping clearly precipitated the shooting.\textsuperscript{16}

Focused as it is on terrorism, the GTD does not have a complete list of shootings. For the 1998 to 2012 period, we found 43 attacks in the U.S. whereas the GTD lists just three: the 1999 Columbine High School shooting, the 2009 Fort Hood massacre, and the 2012 Sikh Temple attack in Oak Creek, Wisconsin. The Columbine attack is classified as Other Crime Type, and the other two are classified as Terrorism. But the GTD readily admits that they do not have a comprehensive list of ‘other crime types,’ causing them to miss cases such as the 2012 Sandy Hook Elementary School attack that would fall into that category.

Over the fifteen years studied here, the GTD also misses 29 cases in Europe, presumably because the GTD does not identify them as terrorist attacks. In Germany, there were two school massacres.\textsuperscript{17} Finland, a country with less than $1/\ldots$
50th of the U.S.’s population, suffered ten people shot to death at a college in 2008 and five people fatally shot at a mall in 2009. Also, for some countries outside of Europe, such as the Solomon Islands (three cases 1998–2012), the GTD misses all of the cases.

To obtain cases missed by the GTD, we used our own Nexis and web searches for mass shootings in Europe and the United States and for large-scale public mass shootings where at least 15 people were killed. For some parts of the world we found Wikipedia entries on rampage killers (link) and mass shootings (link). We also employed researchers to conduct searches in Chinese, French, Polish, Russian, and Spanish in an attempt to reduce an English-only reporting bias.

Neither the NYPD report nor Lankford discuss what search terms they used. We employed Nexis to search for cases by year and our search terms were “mass W/10 shooting,” “mass W/10 firearm*,” “mass W/10 gun,” “multiple W/10 shooting*,” “multiple W/10 firearm*,” and “multiple W/10 gun.” All told, we found 114 cases not included in the GTD.

We likely missed many public mass shootings around the world 1998 to 2012. Consider the numbers found for Central America. The GTD has listed only six Central American and Caribbean public mass shootings (2 for Haiti, 1 for Honduras, and 3 for Mexico), and we only picked up two more cases for Mexico with Nexis. Many Central American countries have very high homicide rates. While it is possible that countries with high homicide rates could have low rates of public mass shootings, it is also possible that the news media in these countries don’t provide much coverage of a shooting with four fatalities.

We are confident that we have all the public mass shootings for the U.S. and perhaps for Europe, but we do not have all of the cases for the rest of the world. No incidents are identified in 91 countries, but that might simply be because we missed them, due to language challenges and poor information sources. While we will show that the rate of public mass shootings in the rest of the world is much higher than in the U.S., we do so even while almost certainly significantly underestimating the prevalence of gun violence in the rest of the world.

**Main results, 1998–2012: U.S. vs. non-U.S.**

The list of all of our 1,491 cases from 1998 to 2012 is provided in Appendices 1 and 2. The main results are presented in Table 1.

---

18. These two attacks in Finland were at a vocational college in Kauhajoki, Finland, Sept. 23, 2008 and the Sello shopping center in Espoo, Finland, Dec. 31, 2009.
19. “W/10” is a Nexis operator that finds items if the search terms are within 10 words of each other (link).
TABLE 1. Mass shootings, U.S. vs. non-U.S.

<table>
<thead>
<tr>
<th>Measure</th>
<th>U.S.</th>
<th>Non-U.S.</th>
<th>Total</th>
<th>Percent U.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lankford’s numbers:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>λ-shooters (1966–2012)</td>
<td>90</td>
<td>202</td>
<td>292</td>
<td>31%</td>
</tr>
<tr>
<td>Our numbers (1998–2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incidents</td>
<td>43</td>
<td>1,448</td>
<td>1,491</td>
<td>2.88%</td>
</tr>
<tr>
<td>Killed</td>
<td>331</td>
<td>15,095</td>
<td>15,426</td>
<td>2.15%</td>
</tr>
<tr>
<td>Shooters</td>
<td>45</td>
<td>10,699</td>
<td>10,744</td>
<td>0.42%</td>
</tr>
<tr>
<td>Population (thousands)</td>
<td>295,156</td>
<td>6,253,801</td>
<td>6,548,957</td>
<td>4.5%</td>
</tr>
</tbody>
</table>

Notes: The shooters for our data are the number of shooters involved in incidents in which four or more people are killed, excluding perpetrators. If the number of shooters is unknown, we assume that there were two shooters. Population is the average, in thousands, for 1998–2012, from the United Nations (link).

While the U.S. had about 4.5 percent of the world’s population during this period, it had just 2.9 percent of the public mass shootings—or even less, since our non-U.S. data is surely missing many cases. The United States was host to a still smaller share of people killed in these attacks (2.1 percent). Finally, with respect to the number of shooters, the U.S. has a minuscule 0.4 percent. Again, these percentage-U.S. numbers are upper bounds.

Calculating shooters/killers

Although we excluded cases for which the reporting indicated means other than firearms, such as bombs, trucks, knives, etc., we acknowledge that in the included cases some of the killing may have been by means other than firearms. So our shooters are killers in attacks that were primarily gun attacks.

The NYPD list that Lankford used contains 32 attacks outside of the U.S., perpetrated by at least 56 killers from 1966 to 2012, the same period that Lankford studied, an average of 1.8 killers per attack. The 2008 Mumbai attack tops the list with 10 killers. It isn’t possible to determine the exact number of attackers in the NYPD list, because in one case—Israel in 1974—we only know that there was more than one killer. If there were two killers in that Israeli attack, and the NYPD average held for Lankford’s entire sample, 202 shooters would amount to at most 112 attacks. We have over 12 times the number of cases for 15 years than Lankford had for 47 years.

Out of our 1,491 cases, news reports provide the number of killers involved in the attack in only 380 instances. In 98 cases, a lone killer was identified, and that

---

20. The usual procedure is to replace the missing values with the mean of the observed values so as not to affect the mean. However, the number of shooters is highly skewed (the mean is 22). An alternative is to use the median (4 shooters), in which case the number of shooters is 12,803.
is 26 percent of the cases that list a number of attackers. Another 42 attacks had two killers and 27 had three, indicating that 44 percent of the cases where the number of killers was identified had between one and three shooters. Meanwhile, 107 cases were identified as having more than 10 killers, which is 28 percent of the cases with the number of killers stated. In larger-scale attacks, numbers of perpetrators are virtually always reported as multiples of ten, making their accuracy doubtful. Witnesses and reporters are most likely just making rough guesses. News reports for 1,068 of the cases simply indicate that there were multiple attackers, with no specific number provided.

In the U.S., just 45 shooters perpetrated the 43 public mass shootings between 1998 and 2012. If we take the conservative estimate that there were only two shooters in each of the attacks outside the U.S. with an indeterminate number, our list shows that there would have been 10,699 shooters worldwide from 1998 to 2012.\textsuperscript{21} So the most conservative estimate is that the number of shooters is 37 times greater than Lankford’s over less than a third of his time period, and the U.S. would account for less than one percent of shooters.

Results per capita

Per capita, public mass shootings occur with 35 percent less frequency and result in 41 percent fewer casualties in the U.S. compared with the rest of the world. Appendix 3 lists the per capita attack and murder rates in the 89 countries where we identified public mass shootings. The U.S. ranks 58th in attack rate and 62nd in murder rate. Norway, Finland, Switzerland, and Russia are major European countries with rates of murder from public mass shootings that are at least 45 percent higher than the United States. The rates in Pakistan and India are respectively 555 percent and 76 percent higher than the U.S. rate. Appendix 4 shows the absolute number by country.

Breakdown by geographic region

Breaking down the cases by geographic regions, we find that the United States ranks roughly in the middle in terms of the number of public mass shootings (see Figures 1A to 1D). We use the sixteen geographic regions provided by the Population Reference Bureau (link). Not surprisingly, Western Asia ranks first since it is largely comprised of Middle Eastern countries such as Iraq, which has per

\textsuperscript{21} The NYPD list does not include any public mass shootings with more than 10. If we ignore such cases then the number of shooters would be 3,121 + 45 = 3,166, of which the United States would account for 1.4 percent.
capita rates of attacks and deaths that are respectively 702 percent and 858 percent higher than those of the United States. Both Northern Africa and sub-Saharan Africa also have dramatically higher rates than the United States. While attacks occur more frequently in Northern Africa, they are more deadly in sub-Saharan Africa (the average number of people killed per attack is 16.1 in sub-Saharan Africa and 9.3 in Northern Africa).

Of particular interest are comparisons between Europe and the United States. There are huge differences in public mass shooting rates across Northern, Western, Eastern, and Southern Europe. While the attack rate in Northern Europe is only 36 percent of the rate in the U.S., 20.4 people were killed per attack in Northern Europe versus 7.6 in the United States. Consequently, the fatality rate from public mass shootings is the same in both Northern Europe and the United States. The fatality rates in the other parts of Europe were lower than the United States.

Attacks in the United States are less deadly than in most of the rest of the world (see Figure 2). There are a number of possible explanations for this. Consistent with our reasoning about magnets for dangerous individuals, attacks by multiple gunmen are more common in the rest of the world. On the other hand, better medical care in the U.S. could cause the death rate for people wounded in mass attacks to be lower here. And again, maybe victims and others start shooting back sooner in the United States.

**Figure 1A.** Public mass shooting murders by geographic region (per 100,000 people)
Figure 1B. Public mass shooting woundings by geographic region (per 100,000 people)

Figure 1C. Public mass shooting casualties by geographic region (per 100,000 people)
Figure 1D. Public mass shooting attacks by geographic region (per 1 million people)

Figure 2. How deadly are public mass shootings in different parts of the world?: Number of people killed per attack
Reason to believe that non-U.S. shootings are significantly under-reported

In South America, people are more than twice as likely to die from public mass shootings, and attacks occur 87 percent more frequently, despite what appears to be a serious lack of news coverage of crime. For example, in Venezuela, not only was the official homicide rate 9.6 times higher than the U.S. rate, but the government has gone to great lengths to prevent the media from reporting on murders. The newspaper *El Universal* reported in 2010 that the Venezuelan police were supposed to tell “relatives of victims who are in the morgue of Caracas not to make statements to the press in exchange for expediting the procedures to recover the bodies” (*Mundo* 2010, our translation). 22

There is evidence of this also happening in China. We have found three large-scale public mass shootings in China in years outside of the 1998 to 2012 period: 1994, 28 killed; 1981, 21 killed; and 1979, 16 killed. 23 We know of no other country that exhibited such large public mass shootings, yet reported no incidents in the 1998–2012 period. Victor Mair, a University of Pennsylvania professor who specializes in China, told us:

I’m almost certain that they had mass public shootings of all sizes up to the three big ones, but such things just don’t get recorded in the media. … The Chinese government is very good about hiding the news. Of course, it’s easier to hide the news for smaller incidents, but much harder for larger incidents, because more people would have noticed them. 24

As an example, Mair claims that friends of his in China have been “forbidden to talk about” a recent knife attack on school children. 25

Does gun prevalence explain mass shootings?

Lankford reports four negative binomial regressions based on a cross section of the countries in his data set. 26 The dependent variable is the number of shooters,

---

22. This is a quotation of an article in the Madrid newspaper *El Mundo*, which cited the Venezuelan daily *El Universal* for the reporting.
23. Beijing and Jiajuomen, China, September 9, 1994; Fudong, China, February 17, 1981 (link); and Qingyang, China, September 24 and 25, 1979 (link).
24. Email correspondence, Victor Mair to John Lott, May 1, 2018. Mair contacted other academics who made similar statements.
26. See also Lott (2018) for further regressions that explain variations in public mass shooters across
the independent variable of interest is gun prevalence— the number of firearms per capita—as measured by the Small Arms Survey (SAS). 27 Lankford’s control variables are the country’s homicide rate, suicide rate, sex ratio, and percent urban, none of which are significant (at the .05 level two-tailed). The only significant variables are the SAS gun prevalence measure and population.

We estimate negative binomial regressions on the number of shooters, the number of incidents, the number of people killed, and the number of people wounded in mass shooting incidents summed over the years 1998–2012 for the 175 countries for which SAS firearm prevalence data are available. 28 Since Lankford finds that none of his control variables are significant, we estimate simple negative binomial models with no controls, except that we include population as an exposure variable with coefficient equal to unity. The data are summarized in Table 2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of shooters</td>
<td>175</td>
<td>52.25</td>
<td>44764</td>
</tr>
<tr>
<td>Number of incidents</td>
<td>175</td>
<td>6.91</td>
<td>684</td>
</tr>
<tr>
<td>Number killed</td>
<td>175</td>
<td>68.70</td>
<td>64077</td>
</tr>
<tr>
<td>Number wounded</td>
<td>175</td>
<td>34.03</td>
<td>18996</td>
</tr>
<tr>
<td>Guns per capita</td>
<td>175</td>
<td>10.10</td>
<td>145</td>
</tr>
</tbody>
</table>

Note: Entries refer to the totals for each country over the period 1998–2012.

The variance for all of the dependent variables is much larger than the mean, indicating overdispersion and the need for the negative binomial regression model. The data are graphed in Figure 3. There does not appear to be any obvious pattern. The regression results are presented in Table 3. None of the coefficients in Table 3 are significant at the .05 level, although the coefficient for guns per capita in the model for the number wounded has a p-value of .09.

27 We are very concerned about the accuracy of the SAS data. They refuse to provide their sources for the data from the vast majority of countries. The firearms data for the countries that they do have are for years quite different from what they report them as being for. Finally, they don’t explain what adjustments are made and how they made those adjustments to get the values for other years.
28 We do this fully aware of all the problems that could plague such regressions, i.e., endogeneity, unobserved heterogeneity, measurement errors, etc.
TABLE 3. Negative binomial regression results (incident rate ratios)

<table>
<thead>
<tr>
<th>SAS firearms per capita</th>
<th>Shooters</th>
<th>Incidents</th>
<th>Killed</th>
<th>Wounded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>0.987</td>
<td>1.023</td>
<td>1.025</td>
<td>1.035</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.018</td>
<td>0.021</td>
<td>0.021</td>
<td>0.021</td>
</tr>
<tr>
<td>T-ratio</td>
<td>−0.72</td>
<td>1.13</td>
<td>1.23</td>
<td>1.71</td>
</tr>
</tbody>
</table>

We did a number of additional robustness tests. We estimated Poisson models, despite the overdispersion. We estimated negative binomial models using the default standard errors and using bootstrap standard errors. We limited the sample to those countries that experienced mass shootings. We also estimated ordinary least squares models using per capita rates of shooters, incidents, killed, and wounded. Finally, we dropped the United States and re-estimated the original negative binomial model. The results were unchanged. None of the coefficients on firearms per capita were significantly different from zero in any of our regressions. Data and programs used in this article are available here.

There is apparently no significant relationship internationally between firearms per capita and the number of shooters, number of incidents, number killed, or number wounded in public mass shootings.
Conclusion

Our paper has shown the importance of semantics and definitions. Lankford cryptically infused his expression “public mass shooter” with an idiosyncratic meaning, that of lone-wolf mass shooter. It is true that the United States is an outlier in lone-wolf mass shooters. But it is not true that the United States is an outlier in public mass shooters.

Following the conventional definitions of public mass shootings, we find that, while in 1998–2012 the United States had about 4.5 percent of the world's population, it had less than one percent of the public mass shooters, 2.1 percent of their murders, and 2.9 percent of their attacks. The United States has fewer public mass shooters, fewer public mass shootings, and fewer murders from these attacks than the average for the rest of the world.

These data not only have implications for how the United States compares to other countries but also to previous claims about what might be responsible for these attacks. Lankford’s claim that higher rates of gun ownership are associated with more public mass shooters falls apart when more complete data on worldwide public mass shootings are used.

Then there is the question of why the United States has more lone-wolf mass shooters. We have suggested that the major reason is not gun prevalence, but rather that in other countries there are more magnets for dangerous individuals, making packs of wolves rather than lone wolves.

Social scientists have a responsibility to make their data easily available so that other researchers can understand and check their findings. The obligation is particularly important after the research has been published or received media or other public attention.

Data and code

Data and code for this research can be downloaded here.

Appendices 1 and 2

Appendix 1: List of public mass shootings and references for countries other than the United States (link).
Appendix 2: List of public mass shootings and references for the United States (link).
Appendix 3

Countries with public mass shootings from 1998 through 2012:
Ranking by per capita rate of attacks and people murdered

<table>
<thead>
<tr>
<th>Rank</th>
<th>Country</th>
<th>Number of attacks per 100,000 people</th>
<th>Rank</th>
<th>Country</th>
<th>Number of people murdered per 100,000 people</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Northern Mariana Islands</td>
<td>1.569</td>
<td>1</td>
<td>Northern Mariana Islands</td>
<td>6.275</td>
</tr>
<tr>
<td>2</td>
<td>Iraq</td>
<td>0.625</td>
<td>2</td>
<td>Iraq</td>
<td>6.007</td>
</tr>
<tr>
<td>3</td>
<td>Solomon Islands</td>
<td>0.600</td>
<td>3</td>
<td>Angola</td>
<td>5.221</td>
</tr>
<tr>
<td>4</td>
<td>Guyana</td>
<td>0.500</td>
<td>4</td>
<td>Guyana</td>
<td>4.000</td>
</tr>
<tr>
<td>5</td>
<td>Afghanistan</td>
<td>0.405</td>
<td>5</td>
<td>Solomon Islands</td>
<td>4.000</td>
</tr>
<tr>
<td>6</td>
<td>Algeria</td>
<td>0.299</td>
<td>6</td>
<td>Sierra Leone</td>
<td>3.309</td>
</tr>
<tr>
<td>7</td>
<td>Somalia</td>
<td>0.291</td>
<td>7</td>
<td>Burundi</td>
<td>2.936</td>
</tr>
<tr>
<td>8</td>
<td>West Bank and Gaza Strip</td>
<td>0.271</td>
<td>8</td>
<td>Algeria</td>
<td>2.808</td>
</tr>
<tr>
<td>9</td>
<td>Burundi</td>
<td>0.256</td>
<td>9</td>
<td>Afghanistan</td>
<td>2.783</td>
</tr>
<tr>
<td>10</td>
<td>Colombia</td>
<td>0.180</td>
<td>10</td>
<td>Somalia</td>
<td>2.581</td>
</tr>
<tr>
<td>11</td>
<td>Angola</td>
<td>0.175</td>
<td>11</td>
<td>Sudan</td>
<td>2.184</td>
</tr>
<tr>
<td>12</td>
<td>Yemen</td>
<td>0.140</td>
<td>12</td>
<td>West Bank and Gaza Strip</td>
<td>1.988</td>
</tr>
<tr>
<td>13</td>
<td>Sri Lanka</td>
<td>0.132</td>
<td>13</td>
<td>Colombia</td>
<td>1.752</td>
</tr>
<tr>
<td>14</td>
<td>Uganda</td>
<td>0.119</td>
<td>14</td>
<td>Norway</td>
<td>1.457</td>
</tr>
<tr>
<td>15</td>
<td>Israel</td>
<td>0.113</td>
<td>15</td>
<td>Uganda</td>
<td>1.420</td>
</tr>
<tr>
<td>16</td>
<td>Sierra Leone</td>
<td>0.109</td>
<td>16</td>
<td>Sri Lanka</td>
<td>1.335</td>
</tr>
<tr>
<td>17</td>
<td>Lebanon</td>
<td>0.105</td>
<td>17</td>
<td>Guinea</td>
<td>1.126</td>
</tr>
<tr>
<td>18</td>
<td>Armenia</td>
<td>0.100</td>
<td>18</td>
<td>Yemen</td>
<td>0.971</td>
</tr>
<tr>
<td></td>
<td>Sudan</td>
<td>0.100</td>
<td>19</td>
<td>Rwanda</td>
<td>0.874</td>
</tr>
<tr>
<td>20</td>
<td>Pakistan</td>
<td>0.086</td>
<td>20</td>
<td>Dem. Rep. of the Congo</td>
<td>0.863</td>
</tr>
<tr>
<td>21</td>
<td>Philippines</td>
<td>0.061</td>
<td>21</td>
<td>Chad</td>
<td>0.825</td>
</tr>
<tr>
<td>22</td>
<td>Kosovo</td>
<td>0.059</td>
<td>22</td>
<td>Pakistan</td>
<td>0.718</td>
</tr>
<tr>
<td>23</td>
<td>Finland</td>
<td>0.058</td>
<td>23</td>
<td>Nigeria</td>
<td>0.701</td>
</tr>
<tr>
<td>24</td>
<td>Nigeria</td>
<td>0.057</td>
<td>24</td>
<td>Armenia</td>
<td>0.700</td>
</tr>
<tr>
<td>25</td>
<td>Nepal</td>
<td>0.051</td>
<td>25</td>
<td>Lebanon</td>
<td>0.684</td>
</tr>
<tr>
<td>26</td>
<td>Macedonia</td>
<td>0.050</td>
<td>26</td>
<td>South Sudan</td>
<td>0.641</td>
</tr>
<tr>
<td></td>
<td>Namibia</td>
<td>0.050</td>
<td>27</td>
<td>Nepal</td>
<td>0.630</td>
</tr>
<tr>
<td>28</td>
<td>Dem. Rep. of the Congo</td>
<td>0.049</td>
<td>28</td>
<td>Israel</td>
<td>0.606</td>
</tr>
<tr>
<td>29</td>
<td>Azerbaijan</td>
<td>0.048</td>
<td>29</td>
<td>Mauritania</td>
<td>0.581</td>
</tr>
<tr>
<td></td>
<td>Central African Republic</td>
<td>0.048</td>
<td>30</td>
<td>Philippines</td>
<td>0.524</td>
</tr>
<tr>
<td>31</td>
<td>Georgia</td>
<td>0.044</td>
<td>31</td>
<td>Finland</td>
<td>0.442</td>
</tr>
<tr>
<td>32</td>
<td>Syria</td>
<td>0.043</td>
<td>32</td>
<td>Syria</td>
<td>0.397</td>
</tr>
<tr>
<td>33</td>
<td>Rwanda</td>
<td>0.034</td>
<td>33</td>
<td>Honduras</td>
<td>0.389</td>
</tr>
<tr>
<td>Rank</td>
<td>Country</td>
<td>Number of attacks per 100,000 people</td>
<td>Rank</td>
<td>Country</td>
<td>Number of people murdered per 100,000 people</td>
</tr>
<tr>
<td>------</td>
<td>--------------------</td>
<td>-------------------------------------</td>
<td>------</td>
<td>-------------------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>34</td>
<td>Mauritania</td>
<td>0.032</td>
<td>34</td>
<td>Liberia</td>
<td>0.364</td>
</tr>
<tr>
<td>35</td>
<td>Chad</td>
<td>0.031</td>
<td>35</td>
<td>Azerbaijan</td>
<td>0.321</td>
</tr>
<tr>
<td>36</td>
<td>Liberia</td>
<td>0.030</td>
<td>36</td>
<td>Kenya</td>
<td>0.317</td>
</tr>
<tr>
<td>37</td>
<td>Tajikistan</td>
<td>0.029</td>
<td>37</td>
<td>Niger</td>
<td>0.314</td>
</tr>
<tr>
<td></td>
<td>Peru</td>
<td>0.029</td>
<td>38</td>
<td>Kosovo</td>
<td>0.293</td>
</tr>
<tr>
<td>39</td>
<td>Cote d'Ivoire</td>
<td>0.027</td>
<td>39</td>
<td>Central African Republic</td>
<td>0.262</td>
</tr>
<tr>
<td>40</td>
<td>Bosnia</td>
<td>0.026</td>
<td>40</td>
<td>Macedonia</td>
<td>0.250</td>
</tr>
<tr>
<td>41</td>
<td>South Sudan</td>
<td>0.025</td>
<td>41</td>
<td>Cote d'Ivoire</td>
<td>0.225</td>
</tr>
<tr>
<td>42</td>
<td>Haiti</td>
<td>0.024</td>
<td>42</td>
<td>Georgia</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>Russia</td>
<td>0.024</td>
<td>44</td>
<td>India</td>
<td>0.193</td>
</tr>
<tr>
<td>45</td>
<td>South Africa</td>
<td>0.023</td>
<td>45</td>
<td>Switzerland</td>
<td>0.189</td>
</tr>
<tr>
<td></td>
<td>Croatia</td>
<td>0.023</td>
<td>46</td>
<td>Laos</td>
<td>0.169</td>
</tr>
<tr>
<td>47</td>
<td>Norway</td>
<td>0.022</td>
<td>48</td>
<td>Ethiopia</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>Thailand</td>
<td>0.022</td>
<td>49</td>
<td>Niger</td>
<td>0.162</td>
</tr>
<tr>
<td></td>
<td>Guinea</td>
<td>0.021</td>
<td>50</td>
<td>Croatia</td>
<td>0.159</td>
</tr>
<tr>
<td>51</td>
<td>Kyrgyzstan</td>
<td>0.019</td>
<td>52</td>
<td>Bosnia</td>
<td>0.158</td>
</tr>
<tr>
<td></td>
<td>India</td>
<td>0.019</td>
<td>53</td>
<td>Peru</td>
<td>0.154</td>
</tr>
<tr>
<td></td>
<td>Yugoslavia</td>
<td>0.019</td>
<td>54</td>
<td>South Africa</td>
<td>0.149</td>
</tr>
<tr>
<td></td>
<td>Serbia</td>
<td>0.019</td>
<td>55</td>
<td>Slovakia</td>
<td>0.130</td>
</tr>
<tr>
<td>56</td>
<td>Senegal</td>
<td>0.017</td>
<td>56</td>
<td>Senegal</td>
<td>0.128</td>
</tr>
<tr>
<td></td>
<td>Laos</td>
<td>0.017</td>
<td>57</td>
<td>Turkey</td>
<td>0.122</td>
</tr>
<tr>
<td>58</td>
<td>United States</td>
<td>0.015</td>
<td>58</td>
<td>Serbia</td>
<td>0.121</td>
</tr>
<tr>
<td>59</td>
<td>Honduras</td>
<td>0.014</td>
<td>59</td>
<td>Haiti</td>
<td>0.120</td>
</tr>
<tr>
<td></td>
<td>Switzerland</td>
<td>0.014</td>
<td>60</td>
<td>Saudi Arabia</td>
<td>0.118</td>
</tr>
<tr>
<td>61</td>
<td>Turkey</td>
<td>0.012</td>
<td>61</td>
<td>Thailand</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>Iran</td>
<td>0.012</td>
<td>62</td>
<td>United States</td>
<td>0.110</td>
</tr>
<tr>
<td>63</td>
<td>Tunisia</td>
<td>0.010</td>
<td>63</td>
<td>Iran</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>Belgium</td>
<td>0.010</td>
<td>64</td>
<td>Mali</td>
<td>0.104</td>
</tr>
<tr>
<td>65</td>
<td>Saudi Arabia</td>
<td>0.008</td>
<td>65</td>
<td>Kyrgyzstan</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>Zimbabwe</td>
<td>0.008</td>
<td>66</td>
<td>Egypt</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>Uzbekistan</td>
<td>0.008</td>
<td>67</td>
<td>Venezuela</td>
<td>0.067</td>
</tr>
<tr>
<td>68</td>
<td>Venezuela</td>
<td>0.007</td>
<td>68</td>
<td>Uzbekistan</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>Mali</td>
<td>0.007</td>
<td>69</td>
<td>Belgium</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>Kazakhstan</td>
<td>0.007</td>
<td>70</td>
<td>Zimbabwe</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>France</td>
<td>0.007</td>
<td>71</td>
<td>Germany</td>
<td>0.040</td>
</tr>
</tbody>
</table>
## Appendix 4

Countries with public mass shootings from 1998 through 2012:
### Ranking by number of attacks and people killed

<table>
<thead>
<tr>
<th>Rank</th>
<th>Country</th>
<th>Number of attacks</th>
<th>Rank</th>
<th>Country</th>
<th>People killed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>India</td>
<td>208</td>
<td>1</td>
<td>India</td>
<td>2130</td>
</tr>
<tr>
<td>2</td>
<td>Iraq</td>
<td>180</td>
<td>2</td>
<td>Iraq</td>
<td>1730</td>
</tr>
<tr>
<td>3</td>
<td>Pakistan</td>
<td>139</td>
<td>3</td>
<td>Pakistan</td>
<td>1166</td>
</tr>
<tr>
<td>4</td>
<td>Afghanistan</td>
<td>121</td>
<td>4</td>
<td>Nigeria</td>
<td>922</td>
</tr>
<tr>
<td>5</td>
<td>Algeria</td>
<td>98</td>
<td>5</td>
<td>Algeria</td>
<td>921</td>
</tr>
<tr>
<td>6</td>
<td>Colombia</td>
<td>83</td>
<td>6</td>
<td>Sudan</td>
<td>878</td>
</tr>
<tr>
<td>7</td>
<td>Nigeria</td>
<td>75</td>
<td>7</td>
<td>Afghanistan</td>
<td>832</td>
</tr>
<tr>
<td>8</td>
<td>Philippines</td>
<td>52</td>
<td>8</td>
<td>Colombia</td>
<td>806</td>
</tr>
<tr>
<td>9</td>
<td>United States</td>
<td>43</td>
<td>9</td>
<td>Angola</td>
<td>804</td>
</tr>
<tr>
<td>10</td>
<td>Sudan</td>
<td>40</td>
<td>10</td>
<td>Dem. Rep. of the Congo</td>
<td>525</td>
</tr>
<tr>
<td>11</td>
<td>Russia</td>
<td>34</td>
<td>11</td>
<td>Philippines</td>
<td>444</td>
</tr>
<tr>
<td>Rank</td>
<td>Country</td>
<td>Number of attacks</td>
<td>Rank</td>
<td>Country</td>
<td>People killed</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------</td>
<td>-------------------</td>
<td>------</td>
<td>--------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>12</td>
<td>Uganda</td>
<td>32</td>
<td>12</td>
<td>Uganda</td>
<td>382</td>
</tr>
<tr>
<td>14</td>
<td>Yemen</td>
<td>29</td>
<td>14</td>
<td>Sri Lanka</td>
<td>263</td>
</tr>
<tr>
<td>15</td>
<td>Angola</td>
<td>27</td>
<td>15</td>
<td>Burundi</td>
<td>229</td>
</tr>
<tr>
<td>16</td>
<td>Sri Lanka</td>
<td>26</td>
<td>16</td>
<td>Russia</td>
<td>227</td>
</tr>
<tr>
<td>17</td>
<td>Somalia</td>
<td>25</td>
<td>17</td>
<td>Somalia</td>
<td>222</td>
</tr>
<tr>
<td>18</td>
<td>Burundi</td>
<td>20</td>
<td>18</td>
<td>Yemen</td>
<td>201</td>
</tr>
<tr>
<td>19</td>
<td>Thailand</td>
<td>14</td>
<td>19</td>
<td>Sierra Leone</td>
<td>182</td>
</tr>
<tr>
<td>20</td>
<td>Nepal</td>
<td>13</td>
<td>20</td>
<td>Nepal</td>
<td>160</td>
</tr>
<tr>
<td>21</td>
<td>South Africa</td>
<td>11</td>
<td>21</td>
<td>Ethiopia</td>
<td>127</td>
</tr>
<tr>
<td>22</td>
<td>Indonesia</td>
<td>9</td>
<td>22</td>
<td>Guinea</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>Turkey</td>
<td>9</td>
<td>23</td>
<td>Kenya</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>West Bank and Gaza Strip</td>
<td>9</td>
<td>24</td>
<td>Turkey</td>
<td>89</td>
</tr>
<tr>
<td>25</td>
<td>Iran</td>
<td>8</td>
<td>25</td>
<td>Chad</td>
<td>80</td>
</tr>
<tr>
<td>26</td>
<td>Israel</td>
<td>8</td>
<td>26</td>
<td>Rwanda</td>
<td>76</td>
</tr>
<tr>
<td>27</td>
<td>Kenya</td>
<td>8</td>
<td>27</td>
<td>Indonesia</td>
<td>74</td>
</tr>
<tr>
<td>28</td>
<td>Peru</td>
<td>8</td>
<td>28</td>
<td>Thailand</td>
<td>74</td>
</tr>
<tr>
<td>29</td>
<td>Syria</td>
<td>8</td>
<td>29</td>
<td>Iran</td>
<td>73</td>
</tr>
<tr>
<td>30</td>
<td>Sierra Leone</td>
<td>6</td>
<td>30</td>
<td>Syria</td>
<td>73</td>
</tr>
<tr>
<td>31</td>
<td>Ethiopia</td>
<td>5</td>
<td>31</td>
<td>South Africa</td>
<td>70</td>
</tr>
<tr>
<td>32</td>
<td>Cote d’Ivoire</td>
<td>5</td>
<td>32</td>
<td>Norway</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Mexico</td>
<td>5</td>
<td>33</td>
<td>West Bank and Gaza Strip</td>
<td>66</td>
</tr>
<tr>
<td>34</td>
<td>Azerbaijan</td>
<td>4</td>
<td>34</td>
<td>Egypt</td>
<td>56</td>
</tr>
<tr>
<td>35</td>
<td>Egypt</td>
<td>4</td>
<td>35</td>
<td>South Sudan</td>
<td>52</td>
</tr>
<tr>
<td>36</td>
<td>France</td>
<td>4</td>
<td>36</td>
<td>Brazil</td>
<td>46</td>
</tr>
<tr>
<td>37</td>
<td>Guyana</td>
<td>4</td>
<td>37</td>
<td>Niger</td>
<td>44</td>
</tr>
<tr>
<td>38</td>
<td>Lebanon</td>
<td>4</td>
<td>38</td>
<td>Israel</td>
<td>43</td>
</tr>
<tr>
<td>39</td>
<td>Armenia</td>
<td>3</td>
<td>39</td>
<td>Peru</td>
<td>43</td>
</tr>
<tr>
<td>40</td>
<td>Bangladesh</td>
<td>3</td>
<td>40</td>
<td>Mexico</td>
<td>42</td>
</tr>
<tr>
<td>41</td>
<td>Brazil</td>
<td>3</td>
<td>41</td>
<td>Cote d’Ivoire</td>
<td>41</td>
</tr>
<tr>
<td>42</td>
<td>Chad</td>
<td>3</td>
<td>42</td>
<td>Germany</td>
<td>33</td>
</tr>
<tr>
<td>43</td>
<td>Finland</td>
<td>3</td>
<td>43</td>
<td>Guyana</td>
<td>32</td>
</tr>
<tr>
<td>44</td>
<td>Niger</td>
<td>3</td>
<td>44</td>
<td>Saudi Arabia</td>
<td>29</td>
</tr>
<tr>
<td>45</td>
<td>Rwanda</td>
<td>3</td>
<td>45</td>
<td>Honduras</td>
<td>28</td>
</tr>
<tr>
<td>46</td>
<td>Solomon Islands</td>
<td>3</td>
<td>46</td>
<td>Azerbaijan</td>
<td>27</td>
</tr>
<tr>
<td>47</td>
<td>South Korea</td>
<td>3</td>
<td>47</td>
<td>Lebanon</td>
<td>26</td>
</tr>
<tr>
<td>48</td>
<td>Central African Republic</td>
<td>2</td>
<td>48</td>
<td>Finland</td>
<td>23</td>
</tr>
<tr>
<td>49</td>
<td>Georgia</td>
<td>2</td>
<td>49</td>
<td>Armenia</td>
<td>21</td>
</tr>
<tr>
<td>50</td>
<td>Germany</td>
<td>2</td>
<td>50</td>
<td>Bangladesh</td>
<td>21</td>
</tr>
<tr>
<td>51</td>
<td>Guinea</td>
<td>2</td>
<td>51</td>
<td>France</td>
<td>20</td>
</tr>
<tr>
<td>Rank</td>
<td>Country</td>
<td>Number of attacks</td>
<td>Rank</td>
<td>Country</td>
<td>People killed</td>
</tr>
<tr>
<td>------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>------</td>
<td>--------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>52</td>
<td>Haiti</td>
<td>2</td>
<td>52</td>
<td>Solomon Islands</td>
<td>20</td>
</tr>
<tr>
<td>53</td>
<td>Myanmar</td>
<td>2</td>
<td>53</td>
<td>Mauritania</td>
<td>18</td>
</tr>
<tr>
<td>54</td>
<td>Saudi Arabia</td>
<td>2</td>
<td>54</td>
<td>Myanmar</td>
<td>18</td>
</tr>
<tr>
<td>55</td>
<td>Senegal</td>
<td>2</td>
<td>55</td>
<td>Venezuela</td>
<td>18</td>
</tr>
<tr>
<td>56</td>
<td>Serbia</td>
<td>2</td>
<td>56</td>
<td>Yugoslavia</td>
<td>18</td>
</tr>
<tr>
<td>57</td>
<td>South Sudan</td>
<td>2</td>
<td>57</td>
<td>South Korea</td>
<td>17</td>
</tr>
<tr>
<td>58</td>
<td>Tajikistan</td>
<td>2</td>
<td>58</td>
<td>Uzbekistan</td>
<td>17</td>
</tr>
<tr>
<td>59</td>
<td>Uzbekistan</td>
<td>2</td>
<td>59</td>
<td>Senegal</td>
<td>15</td>
</tr>
<tr>
<td>60</td>
<td>Venezuela</td>
<td>2</td>
<td>60</td>
<td>Mali</td>
<td>14</td>
</tr>
<tr>
<td>61</td>
<td>Yugoslavia</td>
<td>2</td>
<td>61</td>
<td>Switzerland</td>
<td>14</td>
</tr>
<tr>
<td>62</td>
<td>Argentina</td>
<td>1</td>
<td>62</td>
<td>Serbia</td>
<td>13</td>
</tr>
<tr>
<td>63</td>
<td>Belgium</td>
<td>1</td>
<td>63</td>
<td>Liberia</td>
<td>12</td>
</tr>
<tr>
<td>64</td>
<td>Bosnia</td>
<td>1</td>
<td>64</td>
<td>United Kingdom</td>
<td>12</td>
</tr>
<tr>
<td>65</td>
<td>Cameroon</td>
<td>1</td>
<td>65</td>
<td>Central African Republic</td>
<td>11</td>
</tr>
<tr>
<td>66</td>
<td>Canada</td>
<td>1</td>
<td>66</td>
<td>Tajikistan</td>
<td>11</td>
</tr>
<tr>
<td>67</td>
<td>Croatia</td>
<td>1</td>
<td>67</td>
<td>Haiti</td>
<td>10</td>
</tr>
<tr>
<td>68</td>
<td>Honduras</td>
<td>1</td>
<td>68</td>
<td>Laos</td>
<td>10</td>
</tr>
<tr>
<td>69</td>
<td>Italy</td>
<td>1</td>
<td>69</td>
<td>Georgia</td>
<td>9</td>
</tr>
<tr>
<td>70</td>
<td>Kazakhstan</td>
<td>1</td>
<td>70</td>
<td>Croatia</td>
<td>7</td>
</tr>
<tr>
<td>71</td>
<td>Kosovo</td>
<td>1</td>
<td>71</td>
<td>Slovakia</td>
<td>7</td>
</tr>
<tr>
<td>72</td>
<td>Kyrgyzstan</td>
<td>1</td>
<td>72</td>
<td>Zimbabwe</td>
<td>7</td>
</tr>
<tr>
<td>73</td>
<td>Laos</td>
<td>1</td>
<td>73</td>
<td>Belgium</td>
<td>6</td>
</tr>
<tr>
<td>74</td>
<td>Liberia</td>
<td>1</td>
<td>74</td>
<td>Bosnia</td>
<td>6</td>
</tr>
<tr>
<td>75</td>
<td>Macedonia</td>
<td>1</td>
<td>75</td>
<td>Kazakhstan</td>
<td>6</td>
</tr>
<tr>
<td>76</td>
<td>Malaysia</td>
<td>1</td>
<td>76</td>
<td>Netherlands</td>
<td>6</td>
</tr>
<tr>
<td>77</td>
<td>Mali</td>
<td>1</td>
<td>77</td>
<td>Cameroon</td>
<td>5</td>
</tr>
<tr>
<td>78</td>
<td>Mauritania</td>
<td>1</td>
<td>78</td>
<td>Italy</td>
<td>5</td>
</tr>
<tr>
<td>79</td>
<td>Namibia</td>
<td>1</td>
<td>79</td>
<td>Kosovo</td>
<td>5</td>
</tr>
<tr>
<td>80</td>
<td>Netherlands</td>
<td>1</td>
<td>80</td>
<td>Kyrgyzstan</td>
<td>5</td>
</tr>
<tr>
<td>81</td>
<td>Northern Mariana Islands</td>
<td>1</td>
<td>81</td>
<td>Macedonia</td>
<td>5</td>
</tr>
<tr>
<td>82</td>
<td>Norway</td>
<td>1</td>
<td>82</td>
<td>Malaysia</td>
<td>5</td>
</tr>
<tr>
<td>83</td>
<td>Slovakia</td>
<td>1</td>
<td>83</td>
<td>Ukraine</td>
<td>5</td>
</tr>
<tr>
<td>84</td>
<td>Switzerland</td>
<td>1</td>
<td>84</td>
<td>Argentina</td>
<td>4</td>
</tr>
<tr>
<td>85</td>
<td>Tunisia</td>
<td>1</td>
<td>85</td>
<td>Canada</td>
<td>4</td>
</tr>
<tr>
<td>86</td>
<td>Ukraine</td>
<td>1</td>
<td>86</td>
<td>Namibia</td>
<td>4</td>
</tr>
<tr>
<td>87</td>
<td>United Kingdom</td>
<td>1</td>
<td>87</td>
<td>Northern Mariana Islands</td>
<td>4</td>
</tr>
<tr>
<td>88</td>
<td>Vietnam</td>
<td>1</td>
<td>88</td>
<td>Tunisia</td>
<td>4</td>
</tr>
<tr>
<td>89</td>
<td>Zimbabwe</td>
<td>1</td>
<td>89</td>
<td>Vietnam</td>
<td>4</td>
</tr>
</tbody>
</table>
References


Lankford, Adam. 2015. Adam Lankford on Study of Mass Shooting in US: We Have Too Many Guns [interview by Sonali Kolhatkar]. *Uprising with Sonali*, Free Speech TV (Denver), August 27. [Link](#)


Lankford, Adam. 2019 (Q&A). Correspondence with Daniel Klein, February 10. [Link](#)

Lee, Michelle Ye Hee. 2015. Obama’s inconsistent claim on the ‘frequency’ of mass shootings in the U.S. compared to other countries. *Washington Post*, December 3. [Link](#)


Ownership Rates. Working paper.


**Mundo.** 2010. Venezuela favorece a los familiares de fallecidos que no informan a la prensa. *El Mundo* (Madrid), August 22. Link


**Palazzolo, Joe, and Vanessa O’Connell.** 2015. 5 Things About Mass Shootings in the U.S. *Wall Street Journal*, October 2. Link

### About the Authors

**John R. Lott, Jr.,** is the founder and president of the Crime Prevention Research Center, a research and education organization based in Alexandria, Virginia. Lott is an economist who has held research and/or teaching positions at the University of Chicago, Wharton, Yale, Stanford, UCLA, and the University of Maryland. He has published over 100 articles in refereed journals on a range of topics from law and economics, crime, finance, education, and industrial organization. His email address is johnrlott@crimeresearch.org.

**Carlisle E. Moody** is Professor of Economics at the College of William & Mary, where he teaches mathematical economics, econometrics, and time series analysis. His research is primarily in the economics of crime, especially the relationship between guns and crime. His email address is cemood@wm.edu.
Confirmation That the United States Has Six Times Its Global Share of Public Mass Shooters, Courtesy of Lott and Moody’s Data

Adam Lankford

LINK TO ABSTRACT

John Lott and Carlisle Moody (2019) have unwittingly replicated a major finding from my study and confirmed its accuracy: the United States has far more than its global share of public mass shooters (Lankford 2016). To understand how they did this without realizing it, you have to know only one thing about this specific type of criminal—which as Lott and Moody acknowledge, is similar to “active shooters” or “rampage shooters” and traditionally defined by having killed four or more victims, along with several other criteria.

They almost always attack alone. This is such common knowledge that I am surprised it requires any comment. Most laypeople already know this without my needing to say so, and certainly all researchers with experience in this area recognize this simple fact. It is one of the things that makes public mass shootings so terrifying: they are one of the most vivid demonstrations of just how much death and destruction a single person can cause on his own.

How frequently do public mass shooters attack alone?

Of course, there are rare exceptions, such as the two Columbine shooters

1. The University of Alabama, Tuscaloosa, AL 35487.
who attacked in 1999. But independent reports published by the Federal Bureau of Investigation (Blair and Schweit 2014; FBI 2018), Congressional Research Service (Bjelopera et al. 2013), Rockefeller Institute of Government (Schildkraut, Formica, and Malatras 2018), and New York City Police Department (NYPD 2012) all show that 95–98% of these crimes are committed by solo perpetrators acting alone.

The same thing is demonstrated by open-source data hosted by the *Washington Post* (Berkowitz, Lu, and Alcantara 2019) and *Mother Jones* (Follman, Aronsen, and Pan 2019), which anyone can analyze for themselves, as well as previous research on these types of shootings in the United States (Capellan et al. 2018; Duwe 2016) and beyond (Böckler et al. 2013; Lemieux 2014).

In Table 1, I have listed the frequency with which public mass shootings, active shootings, and rampage school shootings have been committed by a single perpetrator, according to a dozen separate studies and data sources, including my own. I have also listed the corresponding frequencies from the list of cases that Lott (2018a) compiled in an attempt to justify his claims prior to co-authoring with Moody.

Readers are encouraged to peruse Table 1 and play a simple game. Ask yourself: What is wrong with this picture? Which one doesn’t fit? Who seems to be counting some other type of crime?

Of course, if someone stretches the definition of a ‘public mass shooter’ beyond its established notion, then the nature of these incidents and perpetrators would be dramatically altered. At the extreme, someone could theoretically label many soldiers ‘public mass shooters’ based on their participation in armed conflict or war, because they do engage in public violence that results in more than four people being killed. Someone could also add other perpetrators of group violence, including paramilitary fighters, armed rebels, militia group members, and terrorist strike teams.

But that would distort the notion of this crime and defy common sense. In 2004, 300 rebels from the Lord’s Resistance Army in Uganda attacked a camp for internally displaced people and killed at least 54 civilians and two enemy soldiers (GTD 2018). In 2007, one student at Virginia Tech university shot and killed 32 people. There are major differences in the psychology, behavior, weapons acquisition, underlying causes, and prevention strategies that apply to these distinct types of violence (see, for example: Duquet 2018; Eichstaedt 2009; Hoffman 1998; Lankford 2015; Moghadam 2005; Silver, Simons, and Craun 2018; Stein 2017). Studying attacks by the Lord’s Resistance Army will not help us understand and prevent the next Virginia Tech shooting, or vice versa.

If all participants in group violence were counted, that would also result in the inclusion of many people who were far less lethal than public mass shooters who personally killed four or more victims themselves. Should all 28 guardsmen
who were reportedly involved in four deaths at Kent State in 1970 be labeled public mass shooters, even though they averaged killing 0.14 victims each? Should they all be put in the same category as mass shooters from Parkland, Sandy Hook, and Las Vegas who personally killed 17, 27, and 58 victims, respectively? To analyze these distinct forms of violence together would be a textbook example of comparing apples and oranges.

**TABLE 1. Frequency of public mass shootings, active shootings, and rampage school shootings committed by someone attacking alone**

<table>
<thead>
<tr>
<th>Source</th>
<th>Frequency of shootings committed by someone attacking alone</th>
<th>Number of shootings by someone attacking alone/Number of total incidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Federal Bureau of Investigation (Blair and Schweit 2014)</td>
<td>98.8%</td>
<td>158/160</td>
</tr>
<tr>
<td>Federal Bureau of Investigation (2018)</td>
<td>98.4%</td>
<td>246/250</td>
</tr>
<tr>
<td>New York City Police Department (2012)</td>
<td>95.4%</td>
<td>271/284</td>
</tr>
<tr>
<td>Congressional Research Service (Bjelopera et al. 2013)</td>
<td>96.2%</td>
<td>75/78</td>
</tr>
<tr>
<td>Rockefeller Institute of Government (Schildkraut, Formica, and Malatras 2018)</td>
<td>97.6%</td>
<td>332/340</td>
</tr>
<tr>
<td>Washington Post (Berkowitz, Lu, and Alcantara 2019)</td>
<td>97.6%</td>
<td>162/166</td>
</tr>
<tr>
<td>Mother Jones (Follman, Aronsen, and Pan 2019)</td>
<td>97.3%</td>
<td>107/110</td>
</tr>
<tr>
<td>Duwe (2016)</td>
<td>95.6%</td>
<td>153/160</td>
</tr>
<tr>
<td>Capellan et al. (2018)</td>
<td>98.1%</td>
<td>310/316</td>
</tr>
<tr>
<td>Lemieux (2014)</td>
<td>98.3%</td>
<td>117/119</td>
</tr>
<tr>
<td>Böckler et al. (2013)</td>
<td>97.5%</td>
<td>117/120</td>
</tr>
<tr>
<td>Lankford’s (2016) U.S. cases</td>
<td>98.9%</td>
<td>88/89</td>
</tr>
<tr>
<td>Lankford’s (2016) foreign cases</td>
<td>99.0%</td>
<td>198/200</td>
</tr>
<tr>
<td>Lankford’s (2016) total cases</td>
<td>99.0%</td>
<td>286/289</td>
</tr>
<tr>
<td>Lott’s (2018a) U.S. cases</td>
<td>95.3%</td>
<td>41/43</td>
</tr>
<tr>
<td>Lott’s (2018a) foreign cases</td>
<td>6.7%</td>
<td>97/1447</td>
</tr>
<tr>
<td>Lott’s (2018a) total cases</td>
<td>9.3%</td>
<td>138/1490</td>
</tr>
</tbody>
</table>

*Notes:* These studies and data sources all focused on public mass shootings or active shootings, except Böckler et al.’s (2013) study, which examined rampage school shootings. Data indexed by the Washington Post and Mother Jones are updated regularly; these results are current through March 1, 2019. Lankford (2016) included 292 public mass shooters from 289 incidents. One of Lott’s (2018a) foreign cases was removed prior to these calculations because it was a duplicate of the same incident (#960, #961).

I believe this is one of the major reasons why the FBI (Blair and Schweit 2014; FBI 2018) and most other researchers have not included gang violence or
other group violence in their studies: group behavior is so profoundly different from that of individuals. I tried to follow their lead by similarly applying consistent criteria to all cases worldwide, and therefore excluded gang violence, along with sponsored acts of terrorism or genocide that did not appear self-initiated by the perpetrator, because group behavior plays such an important causal role in those other types of crimes.

Back in 2015, Lott claimed he also cared about the integrity of cross-national analyses. “To make a fair comparison with American shootings, I have excluded terrorist attacks that might be better classified as struggles over sovereignty, such as the 22 people killed in the Macedonian town of Kumanovo last month,” he wrote at the time (Lott 2015). But now he has abandoned that pretense, without providing any justification. According to Lott’s (2018a) own coding, and now that of Lott and Moody (2019), nearly 500 foreign attacks that stemmed from battles over sovereignty have been included in their dataset.

Lott and Moody (2019) lump seemingly everything into their list of incidents from other countries: attacks by militia groups, paramilitary fighters, terrorist cells, and more. They include the aforementioned 2004 Lord’s Resistance Army attacks in Uganda, as well as hundreds of other acts of group violence. They even include attacks by “soldiers” in Nigeria (case #333) and by “a squad of uniformed troops” in Colombia (case #324). Here are a few more examples from their dataset:

- “300 heavily armed Pokot raiders attacked a village in the Suam sub-county, killing people, burning as many as 200 houses and stealing at least 300 head of cattle” (case #465).
- “Approximately 250 militants from Fuerzas Armadas Revolucionarias de Colombia (FARC) attacked the Alto de San Juan village, Colombia. Fifteen people died as a result” (case #264).
- “Over 200 gunmen of the Ogaden National Liberation Front (ONLF) attacked a Chinese-run oil field in Abole, Ethiopia. Seventy-four people were killed” (case #825).
- “Armed Arab militia members riding horses and camels attacked the Aro Sharow refugee camp in Sudan’s West Darfur Province, killing at least 29 people and injuring 10 others. The perpetrators, numbering 300, burned 80 shelters and sent thousands of refugees fleeing into the countryside” (case #634).
- “The Democratic Karen Buddhist Army (DKBA) attacked the Huay Kaloke Myanmar Refugee Camp in the Mae Sot District of Thailand with guns and grenades, causing severe damage. Four people were killed and 39 were injured” (case #12).
- “Around 150 Taliban militants stormed a government building killing
at least four policemen, wounding five, and abducting a tribal elder” (case #957).

Overall, more than 95 percent of the incidents Lott and Moody count from the United States were committed by a single perpetrator (41 out of 43). They do not count any U.S. incidents with more than two killers. But when it comes to foreign attacks, less than 7 percent of the incidents they count from other countries were committed by a solo attacker. In fact, they admit that they do not even know the number of shooters for all incidents they are counting. Those they do know about had an average of 22 perpetrators and a median of four perpetrators per incident (Lott and Moody 2019, 53 n.20).

What do Lott and Moody’s own data show about public mass shooters who attack alone?

Fortunately, Lott and Moody’s data can speak the truth they deny. Just focus on their list of public mass shooters who attacked alone. Of course, a few legitimate dual-perpetrator cases would not be included in that analysis, but previous research indicates that focusing on shooters who attack alone would account for 95–99% of the entire phenomenon (Bjelopera et al. 2013; Blair and Schweit 2014; Böckler et al. 2013; Capellan et al. 2018; Duwe 2016; FBI 2018; NYPD 2012; Lemieux 2014; Schildkraut et al. 2018). This is certainly sufficient for estimating how the United States compares to the rest of the world.

Of the 1,448 cases Lott and Moody (2019) compiled from foreign countries, 98 involved a single perpetrator. Readers can confirm this for themselves by sorting the “no. of perpetrators” column in Lott’s original dataset (which I provide here, as Appendix A) and counting the results. Of these single perpetrator cases, two were duplicate entries of the same incident (#960 and #961), which leaves them with 97 foreign cases.

It is then simple arithmetic to calculate the American proportion. Lott and Moody’s own data show that from 1998–2012, 41 of all 138 public mass shootings by single perpetrators worldwide were committed in the United States. That represents 29.7 percent. Because America had in those years approximately 4.5 percent of the world’s population (according to Lott and Moody’s calculations), this indicates that based on their own data, the United States had more than six times its global share of public mass shooters who attacked alone (29.7/4.5 = 6.6). Another way of understanding this is that, if the United States had its proportionate share of these mass shooters, it should have had 4.5 percent of the 138 total cases,
which would be 6. It actually had 41.

This finding clearly demonstrates the magnitude of America’s mass shooting problem, but Lott’s history suggests he will attempt to spin it anyway. For instance, Snopes, a fact-checking service, has had to warn people about Lott’s mass shooting claims (MacGuill 2018). Yes—the same fact-checkers who warned consumers not to believe that the poltergeist curse is real or that food companies use aborted babies in their flavor additives also had to warn the public not to believe John Lott.

In that instance, it was because Lott calculated mass-shooting death rates in small countries that experienced only a single incident, and then used them “to create the false impression that mass shootings are less frequent and less deadly in the United States than in European countries” (MacGuill 2018). Lott “uses inappropriate statistical methods to obscure the reality that mass shootings are very rare in most countries, so that when they do happen they have an outsized statistical effect,” Snopes concluded (ibid.).

Lott and Moody play the same game when they claim the Northern Mariana Islands has a mass shooting rate more than 100 times greater than that of the United States, even though the Northern Mariana Islands had only one qualifying incident from 1998–2012, according to their findings (2019, 66). By Lott and Moody’s view, the smaller the population of the place where a mass shooting occurs, the larger the rate, and presumably the risk. The same logic would suggest that Sutherland Springs, Texas—which is the home of approximately 600 people but saw 26 killed in a terrible 2017 church shooting—must be one of the most dangerous places in the world, rather than the spot of a tragic aberration.

To get around these high variance challenges when calculating rates of rare events, it is most reliable to compare larger sample sizes (i.e., large population areas) to each other. Lott half-admitted this in 2015 when he stated, “If you are going to compare the U.S. to someplace else, if you are going to compare it to small countries, you have to adjust for population. Alternatively, compare the U.S. to Europe as a whole” (quoted in Lee 2015).

I have taken up this latter challenge, using Lott and Moody’s own data, but they will not like the results. As shown in Table 2, their data indicate that from 1998–2012, the United States was the site of more public mass shooters who attacked alone than all of Europe, even though Europe has more than twice the U.S. population. In fact, the United States had more public mass shootings by perpetrators attacking alone than all of Europe, Africa, South America, or Oceania. Asia was the only continent with more of these crimes than the United States, and its population is over ten times as large.
TABLE 2. How the United States compares with five continents, according to Lott and Moody’s data on public mass shooters who attacked alone

<table>
<thead>
<tr>
<th>Location</th>
<th>Number of public mass shooters who attacked alone</th>
<th>Population (2010 est.)</th>
<th>Public mass shooters who attacked alone, per 10 million people</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>41</td>
<td>310 million</td>
<td>1.323</td>
</tr>
<tr>
<td>Europe</td>
<td>25</td>
<td>739 million</td>
<td>0.338</td>
</tr>
<tr>
<td>Oceania</td>
<td>1</td>
<td>37 million</td>
<td>0.270</td>
</tr>
<tr>
<td>Africa</td>
<td>15</td>
<td>1,030 million</td>
<td>0.146</td>
</tr>
<tr>
<td>Asia</td>
<td>50</td>
<td>4,157 million</td>
<td>0.120</td>
</tr>
<tr>
<td>South America</td>
<td>4</td>
<td>391 million</td>
<td>0.102</td>
</tr>
</tbody>
</table>


Lott and Moody’s unwitting replication and confirmation of Lankford’s (2016) findings

I will not spend much time on Lott and Moody’s (2019) attempts to rationalize their approach or discredit mine. Could a few phrases in my original study be rewritten for clarity? Sure: “is consistent” (Lankford 2016, 191) could be reworded as “appears consistent,” “sponsored acts of genocide or terrorism” (ibid., 191) could be accompanied by a brief explanation of the differences between violence by self-directed individuals and group/organizational violence, and “complete data” (192) could be rewritten as “complete data on independent variables.” As Lott and Moody (2019) acknowledge, I did include shooters with terrorist motives (like the 2009 Fort Hood shooter) as long as their behavior appeared self-initiated, even though some researchers do not count any terrorist shootings (e.g., Bjelopera et al. 2013, in their report for the Congressional Research Service).

Fortunately, the global distribution of public mass shooters has an objective reality that is not dependent on such issues. This is confirmed by the fact that, when we focus on the same thing, we have the same results. As noted above, Lott and Moody’s own data show that 29.7 percent of the entire world’s public mass shootings by single perpetrators were committed in the United States, and that America had more than six times its share of the world’s public mass shooters who attacked alone. This is remarkably similar to my original study’s published result: I found that 30.8 percent of public mass shooters attacked
in the United States (Lankford 2016), which would also be more than six times our share of the world’s public mass shooters (30.8/4.5 = 6.8).

Our results are not identical, and the accuracy of our conclusions diverges widely. But these findings are far closer than any independent replication would be expected to produce.

In Table 3, I demonstrate how strongly these findings persist, whether people use (a) my original dataset (Lankford 2016), (b) my original dataset’s information on public mass shooters who attacked alone, (c) Lott and Moody’s (2019) list of public mass shooters who attacked alone, (d) Lott and Moody’s (2019) list of public mass shooters who attacked alone, not counting cases they coded as “battles over sovereignty” (which Lott claimed in 2015 should not be compared with American mass shootings), or (e) the combination of Lankford’s (2016) and Lott and Moody’s (2019) lists of public mass shooters who attacked alone. No matter which approach is selected, the United States had more than six times its global share of public mass shootings by single perpetrators.

TABLE 3. The United States’ global share of public mass shooters worldwide

<table>
<thead>
<tr>
<th>Data source</th>
<th>Time period</th>
<th>The United States’ global share of offenders worldwide</th>
<th>Number of U.S. offenders/Number of total offenders worldwide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lankford’s (2016) data on public mass shooters</td>
<td>1966–2012</td>
<td>30.8%</td>
<td>90/292</td>
</tr>
<tr>
<td>Lankford’s (2016) data on public mass shooters who attacked alone</td>
<td>1966–2012</td>
<td>30.8%</td>
<td>88/286</td>
</tr>
<tr>
<td>Lott and Moody’s (2019) list of public mass shooters who attacked alone, not counting cases they coded as “battles over sovereignty” (which Lott claimed in 2015 should not be compared with U.S. mass shootings)</td>
<td>1998–2012</td>
<td>31.3%</td>
<td>41/131</td>
</tr>
<tr>
<td>Combination of Lankford’s (2016) and Lott and Moody’s (2019) data on public mass shooters who attacked alone</td>
<td>1998–2012</td>
<td>28.3%</td>
<td>52/184</td>
</tr>
</tbody>
</table>

Note: The United States has approximately 4.5% of the world’s population (according to Lott and Moody), so even by the most conservative finding listed above, the United States had more than six times its global share of public mass shootings by single perpetrators (28.3/4.5 = 6.3).

So that anyone can see this evidence for themselves, I have attached my original study’s data on public mass shooters and Lott and Moody’s data on public mass shooters who attacked alone as Appendices B (link) and C (link), along with our combined lists on single perpetrator shootings from 1998–2012 as Appendix
D [link]. My original dataset included 46 cases from 1998–2012 not on Lott and Moody’s list, and their list includes 37 cases (of varying suitability) not on mine. I will not vouch for all of the cases on their list, because seven are “battles over sovereignty,” at least eleven appear to be sponsored or group-influenced attacks where the shooter was not acting of his own volition, and several additional cases are questionable for other important reasons.

It is not only my empirical research and Lott and Moody’s unwitting replication which show that the United States has a disproportionate number of these crimes. For instance, Nils Böckler and coauthors (2013) found that “more [rampage] school shootings have occurred to date in the United States than in all other countries combined. By the end of 2011, the U.S. total had reached 76 (63% of all recorded cases), while there had been 44 cases in the rest of the world (37%).” Similarly, Frederic Lemieux (2014) compared mass shootings in the United States with those in 24 other industrialized countries, and found that the U.S. had more than double the number of attacks in “all other 24 countries combined in the same 30-year period.”

Lott finally admits some of this in *Econ Journal Watch*—seemingly for the first time ever. “It is true that the United States shows an outsized number of lone-wolf shooters,” he and Moody confess (Lott and Moody 2019, 39), before scrambling to minimize the importance of this monumental fact. Perhaps Lott admits this now because he knows he has been undone by his own data, which he published in August 2018 before fully realizing what they showed.

**Why does the United States have such a disproportionate number of public mass shooters who attack alone?**

Like the findings from my original study (Lankford 2016), the findings I have presented from Lott and Moody’s data have powerful implications. The United States has a disproportionate number of public mass shooters who attack alone, and this demands an explanation.

Of course, anyone can speculate about why some countries have more perpetrators than others. Lott and Moody quickly drum up a hypothesis: the United States “has more loners” than anywhere else (2019, 39, 47). They do not support this hypothesis with any citations on cross-cultural differences in loneliness; nor do they bother to test it. Does the United States have six times its global share of loners, and more loners than any entire continent except Asia? They do not say—probably because the answer is no.
Lott and Moody (2019, 46–49) then discuss what they call “magnets,” suggesting that in the United States, dangerous individuals attack alone, but outside the United States, dangerous individuals are drawn to join groups before attacking. This is certainly true in some cases, and Lott and Moody cite my own research (Lankford 2013) as a source of the idea. The shape of violence varies across cultures, and individuals may be more prone to seek assistance or support in some contexts than in others. It is also clear that some countries have far more active and influential paramilitary organizations, rebel groups, and terrorist organizations than the United States has, and they do attract some people with violent intentions.

At the same time, Lott and Moody exaggerate the extent of these differences between the United States and the rest of the world. America does not have a domestic corollary for the Lord’s Resistance Army, Ogaden National Liberation Front, Democratic Karen Buddhist Army, or United Self-Defense Units of Colombia, but nor do many other developed countries. Yet for some other reason, we suffer far more public mass shootings by single perpetrators than our peers. It is also not the case that mass killings are some sort of inevitability, and that the only question is what form they will take. There are many countries that almost never experience mass killings perpetrated by individuals or groups.

However, even if the magnet hypothesis is partially correct, this does not mean we should avoid studying public mass shooters who attack alone. It just means that using the number of public mass shooters who attack alone to measure the total number of dangerous individuals would provide some significant underestimates, because group actors would not be counted. But that seems obvious. I have never implied that my findings explained the variation in all dangerous individuals worldwide.

As to the question of why some countries have more attacks by soldiers, uniformed troops, paramilitary fighters, armed rebels, and terrorist organizations than others, I would not pretend to know. That was clearly beyond the scope of my study, and I have never suggested otherwise. But Lott and Moody have not answered that question, either. They claim to have tested the relationship between this particular type of violence and firearm ownership rates, and to have found no significant relationship. Perhaps they are correct—although I would not trust that without further verification. However, this merely clarifies what does not explain the type of mass shootings the United States does not have, anyway.

I am more interested in understanding and preventing the type of public mass shootings that have plagued America for more than 50 years, which is why I studied the types of attackers I did. So far, only one explanation for the cross-national variation in these mass shooters has been empirically demonstrated, and that is firearm ownership rate. As Lemieux (2014, 82) summarized based on his comparative analysis of the United States and 24 other industrialized nations,
“mass shootings and gun ownership rates are highly correlated ($r = 0.75; p < 0.01$) …the higher the gun ownership rate, the more a given country is susceptible to experience mass shooting incidents.” My study independently found the same thing: firearm ownership rate appeared to be the most significant factor, even though I also tested homicide rates, suicide rates, urbanization, and sex ratios (Lankford 2016). Lemieux (2014) and I also both independently found that this association between mass shooters and national firearm ownership rates was so strong that it explained the variation across other countries, even when the United States was not considered.

This firearms explanation is also just common sense. By definition, firearms are needed for people to commit mass shootings, so in countries where it is easier for dangerous or disturbed individuals to legally purchase firearms—like the United States—there is an increased likelihood of an attack. That isn’t rocket science.

In fact, it is the public mass shooters who attack alone—the students, office employees, factory workers, and so on—whose behavior is most likely to be explained by national firearm ownership rates. These perpetrators are civilians who usually get their guns legally (Silver, Simons, and Craun 2018), so they are directly affected by national gun restrictions, or the lack thereof. By contrast, the participants in group violence who more commonly attack in other countries—the paramilitary fighters, armed rebels, militia group members, and terrorist strike teams—seem less likely to be affected by firearms legislation. Their groups often operate in open defiance of the local government and its laws, and they appear much more likely to obtain weapons through illegal methods, such as smuggling (Duquet 2018; Eichstaedt 2009).

The Lord’s Resistance Army does not get its firearms by walking into a Sunrise Tactical Supply store and slapping down some cash or a credit card on the counter. But that is exactly what the Parkland school shooter did, and many public mass shooters who attack alone in the United States are unfortunately similarly enabled.

**Conclusion**

It might be easy to assume my disagreement with Lott and Moody is mostly about definitions. They have given that impression in their *EJW* comment, which contains none of the personal attacks or slanderous accusations that Lott has levied against me in other forums.

However, I believe this is actually the case of one researcher who conducted an honest study and let the empirical results guide his conclusions being opposed.
by others who are primarily driven by ideological motives. My track record is clear: I published hundreds of thousands of words about crime and violence in two books and many articles before examining public mass shooters worldwide, but never had any interest in debating firearms or gun control. To this day, I am far more interested in studying behavior than weapons, and the latter occupies only a tiny portion of my research. There are certainly impassioned crusaders on all sides of this issue, but I am not one of them. I just followed the evidence where it led, and then was willing to speak publicly about what I found.

By contrast, Lott has a long track record of denying any consequences of the United States’ world-leading firearm ownership rate, and this appears directly related to his repeated refrain that America’s mass shooting problem is exaggerated (Lott 2014; 2015; 2018b). Before my study was even released, Lott published an op-ed entitled “Myths of American Gun Violence” in which he insisted that “many European countries actually have higher rates of death in mass public shootings” than the United States. This is the same line of research that the fact-checking service Snopes eventually lambasted for using “inappropriate statistical methods” and creating a “false impression” (MacGuill 2018). That was not solely a matter of definitions or semantics.

I suspect that ever since my findings were publicly reported, Lott has been looking for a way to discredit them. As a reminder, he even changed how he counted these attacks prior to posting his criticism of my work (Lott 2018a)—including nearly 500 battles over sovereignty, after claiming in his prior analysis that they should be excluded (Lott 2015).

Regardless of the crime or context, Lott always seems to come to the same conclusion: firearms are not part of the problem. In fact, he has long insisted they are the solution: “more guns, less crime” he has proclaimed for years, and “more guns, less terrorism” he and Moody assert now (2019, 47). However, their rush to turn this newest motto into a marketable “bumper sticker” (ibid.) reveals their willingness to prioritize ideology over accuracy. According to their own findings on mass shootings writ large (primarily group terrorist attacks and other group violence): “There is apparently no significant relationship internationally between firearms per capita and the number of shooters, number of incidents, number killed, or number wounded” (ibid., 60). So they simultaneously claim “more guns, less terrorism,” and that guns are not statistically related to terrorism.

Is this simply a mistake, or something more telling? Am I wrong, or are they? Whose judgment, analysis, and findings can you trust? I hope other researchers will weigh in, but ultimately, people will have to decide for themselves.
Appendices

Appendix A: Lott (2018a) complete unedited dataset (link).

All four appendices also can be downloaded in one compressed file (link).

References

2019: Data from Mother Jones’ Investigation. February 15. Link

Global Terrorism Database (GTD). 2018. Incident Summary. National Consortium for the Study of Terrorism and Responses to Terrorism (College Park, Md.). Link


Lott, John R. Jr. 2018b. The Problem With the FBI’s ‘Active Shooter’ Data. RealClearPolitics, October 23. Link


About the Author

Adam Lankford is an associate professor of criminology and criminal justice at The University of Alabama. He is the author of two books and many peer-reviewed journal articles on various types of criminal behavior, including mass murder, mass shootings, and terrorism. His research has examined perpetrators’ psychological tendencies, mental health problems, suicidal motives, fame-seeking tactics, copycat behavior, and weapons acquisition—along with the strategies that might be used to prevent their attacks. His email address is adam.lankford@ua.edu.
Do Right to Carry Laws Increase Violent Crime? A Comment on Donohue, Aneja, and Weber

Carlisle E. Moody and Thomas B. Marvell

Along with coauthors, John J. Donohue of the Stanford Law School regularly releases empirical research finding that certain laws concerning the carrying of firearms, those known as right-to-carry or RTC laws, increase violent crime. We have engaged with the previous releases to show that the latest findings are not robust, and here we do so again. This time we address something maintained in the previous releases, namely the weighting of fixed-effects regressions by population, and something that is new to the last two versions, namely a synthetic control procedure. Perhaps after another round or two it will be time to recap all the robustness criticisms to date, but here we confine ourselves to the two new criticisms.

Donohue, Abhay Aneja, and Kyle Weber (hereafter DAW) have produced two revisions (2018a; b) to their 2017 working paper, which itself is part of a series of papers dating back 15 years (Donohue 2003; Aneja, Donohue, and Zhang 2010; 2011; 2012; 2014; see also Aneja, Donohue, Pepper, et al. 2012). That paper adds a synthetic control analysis and finds that RTC laws increase violent crime. They conclude, as they state in the abstract of the latest revision:

Our preferred panel data regression specification…generates statistically significant estimates showing RTC laws increase overall violent crime. Our synthetic control approach also strongly confirms that RTC laws are associated with 13–15 percent higher aggregate violent crime rates…

1. College of William and Mary, Williamsburg, VA 23187.
average RTC state would need to roughly double its prison population to
gen, the increase in violent crime caused by RTC adoption. (DAW 2018b, 1,
italics in original)

We argue that DAW’s results are fragile and most likely incorrect. We inves-
tigate two features of the research: the weighting of their fixed-effects regressions
by population, and the use of synthetic controls. Both topics apply to numerous
other difference-in-differences (DD) policy studies, so the discussion here has
relevance beyond the DAW papers.

Weighting by population

DAW (2018a; b) base their conclusions on DD regressions that are weighted
by population, a common practice in this literature and one which we have often
used unquestioningly. They do not give reasons for weighting, nor to our knowl-
edge have other researchers in this area, even though the results may well depend
on the choice of weight.

Donohue has noted that weighting causes large states to dominate DD
results (e.g., Ayres and Donohue 2003, 1276), but without stating that such
emphasis is the purpose for weighting. Such might be the purpose if the research
goal is to estimate the overall national impact of a policy change, because then
weighting can be justified by arguing that the impact of laws in large states should
be emphasized simply because they affect more people, or more populous states
should be weighted more heavily because their data have smaller variance. On the
other hand, if policy makers in states without an RTC law want information about
whether a new law will have an impact in their state, the results of a weighted
regression are not particularly useful unless their state happens to be large. If
Donohue and his colleagues desire to emphasize the influence of large states, they
should specify as much.

Weighting is also used to address heteroskedasticity, a potential problem in
any regression analysis, and perhaps that is what DAW aim to do. Traditionally
the standard econometric approach to heteroskedasticity was to test for it, and if
found, to weight by some function of the variables in the model. This procedure
is known as weighted least squares (WLS) or feasible generalized least squares
(FGLS). For the crime equation, one common approach was to assume that the
underlying unobserved individual crime equation had independent and identically
distributed (i.i.d.) errors, which when summed and divided by the population to
produce per capita crime rates within states, generated heteroskedasticity of the
form $\sigma^2/\text{pop}$; for states $i=1…N$, so that multiplying the state variances by state
population would produce homoskedastic errors. However, it is not clear that this approach is invariably correct. If the individual crime equation errors do not have i.i.d. errors then the population-weighted least squares approach would be incorrect and may exacerbate the problem.

The availability of heteroskedasticity-consistent standard errors (White 1980) has led many to abandon WLS. The Stata command for clustered standard errors includes the White correction, and Donohue and his colleagues cluster by state. In practice researchers never really know the true form of the error variance. Consequently, as Jeffrey Wooldridge notes, “more and more researchers simply use OLS and compute robust standard errors when estimating models using per capita data” (2016, 258–259). Also, according to James Stock and Mark Watson, “it is simple to use heteroskedasticity-robust standard errors, and the resulting inferences are reliable under very general conditions…without needing to specify a functional form for the conditional variance. For these reasons, it is our opinion that, despite the theoretical appeal of WLS, heteroskedasticity-robust standard errors provide a better way to handle potential heteroskedasticity in most applications” (2007, 696). Similarly, Joshua Angrist and Jörn-Steffen Pischke write, “Any efficiency gain from weighting is likely to be modest, and incorrect or poorly estimated weights can do more harm than good” (2009, 94).

On the other hand, Edward Leamer calls this practice “White-washing” and says “we should be doing the hard work of modeling the heteroskedasticity and the time dependence to determine if sensible reweighting of the observations materially changes the locations of the estimates of interest as well as the widths of the confidence intervals” (2010, 43). Joseph Romano and Michael Wolf (2017) propose what they call “adaptive least squares” which requires using a standard test (e.g., White 1980) for heteroskedasticity. If the test does not reject the null of homoskedasticity, use OLS with robust standard errors. If the test rejects, model the variance function and use the inverse of the predicted value as the weight, with inference based on robust standard errors. Wooldridge (2016, 262–264) also recommends combining FGLS with robust standard errors.

Does it make a difference whether one weights the per capita crime equation by population? We use the DAW “preferred” model. We made some slight modifications because we do not have exactly the same data as DAW. For example, DAW use the percent of the population living in MSAs to measure population density; we use the population-to-area ratio. They use lagged total police employment to measure police presence; we use the lagged number of sworn officers. Finally, they use six demographic variables: the proportion of black, white, and other males in two age groups (15–19 and 20–39). We use the percent of the

2. Despite two requests we were unable to get the DAW data and do files.
population 15–39 by five-year intervals and the proportion of the total population that is black. Also, we have data from 1970 as opposed to DAW whose data start in 1977. Like DAW, we include a trend and state and year fixed effects, and also like DAW we use robust clustered standard errors.

DAW estimate two versions of their crime equation. The first is a standard difference in differences analysis using a dummy variable to represent the presence or absence of an RTC law. The second uses a spline (post-law trend) to capture the effects of the law over time. We estimated each model with and without population weights and using FGLS. All models are estimated with errors clustered at the state level. We also tested for heteroskedasticity using the finite version of the White (1980) test. For all the models, the tests were invariably highly significant, indicating the presence of heteroskedasticity, despite the weighting.

DAW, like many other researchers, assume that the variance function is \( \text{Var}(u_{it}) = (1/\text{pop}_{it}) \sigma^2 \), where \( \sigma^2 \) is the homoskedastic error variance. If that assumption is incorrect then the resulting estimators are inefficient. We can investigate this assumption by estimating a more general form of the variance function including population:

\[
u_{it}^2 = X_{it} \beta + \lambda(1/\text{pop}_{it}) + \alpha_i + \gamma_t + \nu_{it}
\]  
(1)

where \( \nu_{it}^2 \) is the squared regression error, the matrix \( X_{it} \) includes all the control variables in the DAW “preferred” model including the time trend, and \( \alpha_i \) and \( \gamma_t \) are state and year fixed effects. If \( \lambda \) is significant and equal to one, and all of the other variables are jointly insignificant, then DAW’s assumption is correct. We estimated equation (1) using the squared residuals from the OLS regression as the dependent variable. Although not reported to conserve space, there were many significant variables in these regressions, but the inverse of population was significant only in the violent crime models and it was never the only significant coefficient.

Following Leamer (2010), we use FGLS to estimate the appropriate weights for the crime equation, replacing the squared errors in (1) with the corresponding squared residuals and taking logs to guarantee positive variance estimates.

\[
\log(\hat{\nu}_{it}^2) = X_{it} \beta + \lambda(1/\text{pop}_{it}) + \alpha_i + \gamma_t + \omega_{it}
\]  
(2)

The FGLS weight is the inverse of the antilog of the predicted value from (2). We also estimated an unweighted OLS model.

The results of this experiment for the murder rate and the violent crime rate

---

3. All our Stata do files, log files and data are available at the Econ Journal Watch website [link].
equations are shown in Table 1, where the coefficients refer to the RTC dummy and spline variables.

<table>
<thead>
<tr>
<th>Dummy</th>
<th>Murder Weighted (1)</th>
<th>Murder OLS (2)</th>
<th>Murder FGLS (3)</th>
<th>Violent Weighted (4)</th>
<th>Violent OLS (5)</th>
<th>Violent FGLS (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>-0.466</td>
<td>-1.754</td>
<td>0.808</td>
<td>11.440***</td>
<td>-3.658</td>
<td>-1.440</td>
</tr>
<tr>
<td>Std. error</td>
<td>(4.113)</td>
<td>(3.985)</td>
<td>(2.309)</td>
<td>(3.679)</td>
<td>(4.451)</td>
<td>(1.979)</td>
</tr>
<tr>
<td>P-value</td>
<td>0.910</td>
<td>0.662</td>
<td>0.728</td>
<td>0.003</td>
<td>0.415</td>
<td>0.470</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Spline</th>
<th>Coefficient</th>
<th>Violent Weighted (4)</th>
<th>Violent OLS (5)</th>
<th>Violent FGLS (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>0.352</td>
<td>-0.283</td>
<td>-0.115</td>
<td>1.148**</td>
</tr>
<tr>
<td>Std. error</td>
<td>(0.414)</td>
<td>(0.454)</td>
<td>(0.253)</td>
<td>(0.436)</td>
</tr>
<tr>
<td>P-value</td>
<td>0.399</td>
<td>0.536</td>
<td>0.651</td>
<td>0.012</td>
</tr>
</tbody>
</table>

Note: All models include control variables equivalent to the DAW “preferred” model as well as state and year fixed effects. Coefficients and standard errors are reported in percentage terms. Weighted indicates population-weighted least squares. FGLS uses the inverse of the antilog of the predicted values of the log variance equation (2) as the weight. Standard errors are robust to heteroskedasticity and clustered at the state level for all models. * p < .1, ** p < .05, *** p < .01.

The DAW population-weighted least squares estimates are reported in DAW (2018b, 21, Table 3). For the murder equation, they find no significance. For the violent crime model, they find that only the dummy variable is significantly different from zero.

How one weights the regression does in fact make a difference. Our attempted replication of DAW, using population weights, yields similar results in terms of significance, except that we find that both the dummy and spline are significant in the violent crime equation. The OLS model finds no significant coefficients. The FGLS model, which has the most precise estimates, also finds no significant coefficients.4

Thus, we corroborate the DAW result that the RTC law has no significant effect on murder, but we also find that their conclusion that violent crime is increased by the passage of a RTC law is fragile insofar as it is dependent on the use of population weights.

---

4. We also tried using population instead of the inverse of population in the variance functions for both the squared residuals and the log of the squared residuals, the results reported in Table 1 were unchanged.
Fixed effects vs. synthetic controls

In the second part of their paper, DAW use the synthetic control (SC) method which also finds that RTC laws increase violent crime rates. The SC analysis also confirms that RTC laws do not significantly increase murder.

The synthetic control method compares the outcome (e.g., violent crime rate) of a treated state (i.e., a state with an RTC law) to a weighted average of the outcomes of the control states (i.e., states with no RTC law). The resulting weighted average, formed by using a matching algorithm (Abadie, Diamond, and Hainmueller 2010), is called the “synthetic control state.” The result is the difference or “gap” between \( y_t^{\text{treated}} \) the outcome of the treated state and the weighted average of the outcomes of the control states \( y_t^{\text{synthetic}} \):

\[
gap_t = y_t^{\text{treated}} - y_t^{\text{synthetic}}
\]

Assume that the outcome is a function of the policy being investigated, measured by \( z_{it} (i=1,\ldots,N; t=1,\ldots,T) \) which could be a continuous or a dummy variable, a set of trends \( \{t_r\} \), a set of control variables \( \{X_{kit}\}, k=1,\ldots,K \), a set of state fixed effects \( \{\alpha_i\} \), a set of year fixed effects \( \{\delta_t\} \), and a random i.i.d. error term \( \{\epsilon_{it}\} \). The outcomes of the control states are also functions of \( z_{it} \) trends, control variables, state and year fixed effects and a random error term.

Therefore,

\[
y_{it} = f(z_{it}, X_{kit}, \alpha_i, \delta_t, t_r, \epsilon_{it})
\]

where \( y_{it} \) is the outcome for state \( i \) in year \( t \). Denoting the treated state as \( i=1 \) and the control states as \( i=2,\ldots,N^* < N \),

\[
gap_t = f(z_{1t}, X_{11t}, \alpha_1, \delta_1, t_{r1}, \epsilon_{1t}) - \sum_{i=2}^{N^*} c_i f(z_{it}, X_{kit}, \alpha_i, \delta_i, t_r, \epsilon_{it})
\]

where the weights, \( c_i \), are nonnegative and sum to one.

The state and year fixed effects as well as the control variables and trends are in both the treated and control outcomes. Thus the gap is a function of all those factors. For the gap to measure the effect of the policy requires one to assume that all of the variables except for the policy dummy are constant in the treatment period, a very strong assumption.

DAW implicitly recognize this limitation when they argue that their gap measurements are a function of at least three exogenous control variables that are not held constant: “If one adjusts the synthetic controls estimates to control
for the increased rates of police and incarceration that follow RTC adoption, the RTC-induced increases in murder are almost nine percent with a p-value of 0.089” (2018b, 42 n.64). Also, “Specifically, if...the death penalty is a powerful deterrent one might be concerned that Texas’s far greater use of the death penalty during the post-passage period than in the states comprising synthetic Texas might bias downward the prediction that RTC laws increased crime by 16.9 percent in Texas. Conversely, the greater increases in incarceration and police in synthetic Texas would lead to the opposite bias” (2018a, 39). Finally, “Figure 3 makes clear what Texas is being compared to, and we can reflect on whether this match is plausible and whether anything other than RTC laws changed in these three states during the post-passage decade that might compromise the validity of the synthetic control estimate of the impact of RTC laws” (DAW 2018b, 30–31). Thus, DAW acknowledge that omitted exogenous control variables in both the treatment and control states could bias the results. In any case, the SC model fails to control for police, prison, and executions, as well as all the other control variables, in the treatment period.

Also, states have certain permanent, or at least long-lived, attributes such as history, tradition, culture, climate, attitudes toward firearms, attitudes toward self-defense, etc., that could affect both crime and the policies adopted to combat crime. Louisiana is fundamentally different from Massachusetts, Utah is very different from Nevada, Texas is different from Wisconsin, Alaska and Hawaii are different from each other and from every other state, and so on. The SC gap is determined in part by the difference in the mean of the dependent variable (e.g., violent crime) for the treated state and the weighted average of the corresponding means of violent crime for the states making up the synthetic control state. Thus the gap is a cross-section difference which is a function of the state fixed effects, implying that the SC technique suffers from unobserved heterogeneity. According to Manuel Arellano,

[Unobserved heterogeneity] has been a pervasive problem in cross-sectional regression analysis. If characteristics that have a direct effect on both left- and right-hand side variables are omitted, explanatory variables will be correlated with errors and regression coefficients will be biased measures of the structural effects. Thus researchers have often been confronted with massive cross-sectional data sets from which precise correlations can be determined but that, nevertheless, had no information of policy interest. (Arellano 2003, 8)

The only way to avoid this type of omitted variable bias is to use panel data with fixed effects, which gives each cross-section unit its own intercept term, allowing these permanent effects to be partialed out of the parameter estimate corresponding to the policy variable.
According to Angrist and Pischke (2009, 244 n.9), the synthetic control model is a nonparametric version of the lagged dependent variable model:

\[
y_{it} = \alpha + \beta_0 z_{it} + \sum_{k=1}^{K} \beta_k X_{kit} + \sum_{i=1}^{N-1} \gamma_i \tau_i + \sum_{j=1}^{p} \varphi_j y_{i-j} + \varepsilon_{it}
\]

where the lagged dependent variables control for unobserved time-varying confounders (Abadie et al. 2010, 495).

Obviously, in the SC model there is the possibility that the control variables could vary significantly during the post-treatment period, altering the gap between the treated and control states; also states could have different trends with respect to the outcome; and the method fails to control for unobserved heterogeneity. The problem is that there is nothing held constant in the treatment period, so the gap is a function of the trends, the changing control variables, and the state fixed effects.

On the other hand, the standard fixed-effects model estimates a linear approximation of the function \( f \).

\[
y_{it} = \alpha_j + \beta_0 z_{it} + \sum_{k=1}^{K} \beta_k X_{kit} + \sum_{i=1}^{N-1} \gamma_i \tau_i + \varepsilon_{it}
\]

The estimate of the effect of the policy is the coefficient \( \hat{\beta}_0 \), where the effects of the control variables, trends, and state and year fixed effects have been partialled out. These two approaches are very different. The fixed-effects regression model, properly specified, controls for all relevant factors, pre- and post-treatment, including trends and state and year fixed effects.\textsuperscript{5} This cannot be said for the synthetic control model.

Nevertheless, we estimate the SC model of the violent crime rate and murder rate. Matching in the pre-treatment period is a function of the same control variables from DAW’s “preferred model,” that we used in the fixed effects analyses above. Following DAW we use the Stata synth program, limiting the post-treatment period to 10 years, and we include as potential control states all those states without RTC laws during the 10-year post-treatment period, including states that later adopted RTC laws. We use the default settings, except that we use both the nested matching option and the default regression method. We also use the regression option for those states for which the nested option fails. We do not force the gap to be zero for the year in which the policy was implemented, because that makes that one year more important than all the other years. Also, according to DAW, “As it turns out, the choice we made to subtract off the initial-year crime discrepancy

\textsuperscript{5} Including lagged dependent variables in the fixed effects model will control for unobserved time-varying confounders, but will cause bias (Nickell 1981). However, the bias is small for panels with N>20 and T>30 (Judson and Owen 1999).
is a conservative one, in that the estimated crime increases from RTC laws would be greater without subtraction” (2018b, 35 n.55; italics in original). Finally, as noted above, we have seven more years of data in the pre-treatment period.

We determine whether the gap is significant for each state by using a t-test of the null hypothesis that the mean gap over the 10 years of the post-treatment period is equal to zero. We use a t-test on the mean of the 10-year sums over the 33 treated states to determine the overall effect of the law.

As a first step, we do a SC analysis of violent crime for the state of Texas, an example featured in DAW (2018b, 28–30) so we can easily compare our results to theirs. We replicate DAW by limiting the analysis to 1977–2007. The results are shown in Figure 1.

**Figure 1.** Synthetic control applied to Texas

The top left graph shows the Texas violent crime rate (solid line) and the weighted average of the violent crime rates for the three states chosen by the SC matching algorithm, using the weights reported by DAW (2018b, 30): California (.577), Nebraska (.097), and Wisconsin (.326). 6 This appears to be exactly the figure

---

6. Even though both California and Wisconsin differ from Texas in terms of history, tradition, culture, climate, attitudes toward firearms, attitudes toward self-defense, etc.
presented by DAW (2018b, 30, 110).

The top right graph shows the result of our application of the SC method to the Texas data where we use all the data from 1970 to 2007 but limited to the three control states chosen by DAW. The two graphs are virtually identical despite the fact that the SC program chose a slightly different synthetic Texas (California .578, Wisconsin .422). In the lower left graph we expand the number of control states to include all the states without RTC laws until 2007. Again the graph is similar to the first two. All three show a large positive gap after 1996 apparently indicating that the Texas RTC law increased violent crime. A robust t-test reveals that the mean over the post-treatment period is significantly positive. This confirms that our analyses are equivalent to those of DAW.

The situation is much different for the murder rate, shown in the lower right-hand corner. The rate for the real Texas is below that of synthetic Texas for all post-treatment years. The robust t-test indicates that the gap is significantly negative. The murder rate for the synthetic Texas was below that for real Texas until 1993, before the passage of its RTC law, after which the real Texas murder rate fell below the synthetic Texas rate, indicating that the RTC law is unlikely to be entirely responsible for the decline. However, there was no increase in the murder rate after 1996 that could be attributed to the Texas RTC law.

The results of the full SC analyses are reported in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
<th>Positive Significant</th>
<th>Negative Significant</th>
<th>Mean Gap</th>
<th>T-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violent</td>
<td>15</td>
<td>18</td>
<td>14</td>
<td>12</td>
<td>122.6</td>
<td>0.55</td>
</tr>
<tr>
<td>Murder</td>
<td>13</td>
<td>20</td>
<td>8</td>
<td>11</td>
<td>0.85</td>
<td>0.38</td>
</tr>
<tr>
<td>Default</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violent</td>
<td>16</td>
<td>17</td>
<td>13</td>
<td>16</td>
<td>−157.7</td>
<td>−0.45</td>
</tr>
<tr>
<td>Murder</td>
<td>17</td>
<td>16</td>
<td>13</td>
<td>11</td>
<td>0.505</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Notes: “Significant” means p < .05 for the robust t-statistic across the post-treatment period. The mean gap is the mean of the sums of the 10 yearly post-treatment gaps for each state across all 33 states. The T-statistic refers to the robust t-test of the hypothesis that the net effect across the 33 treated states is zero.

There is no obvious significance here. For violent crime, the number of negative results (35 reductions in crime) is greater than the number of positive results (31 increases in crime) as a result of the RTC law. However, with respect to significant results it is a virtual tie (27 significantly positive and 28 significantly negative). The results are similar for murder, the number of negative results (36) is again greater than the number of positive effects (30), and the number of significant effects is almost the same (21 positive, 22 negative). The one consistent result across all four
models is that the net effect across states is not significantly different from zero. In all, the SC analysis suggests that the RTC law cannot be shown to have a significant effect on either violent crime or murder.

**Summary and conclusion**

Our examination of the evidence indicates that there is no significant relationship between RTC laws and either murder or violent crime. The fixed-effects models estimated using our version of the DAW “preferred model” is fragile. Without weighting by population, there is no significant effect of RTC laws on violent crime. Using feasible generalized least squares confirms the OLS results that RTC laws do not increase violent crime or murder.

We also have a problem with the SC model that DAW use to buttress their population-weighted FE model. In the critical treatment period, the SC model fails to control for any of the major factors that cause crime rates to vary. It also suffers from unobserved heterogeneity. Nevertheless, when we use the SC model we find that the claim that RTC laws increase either murder or violent crime is not supported. We find states where crime increased after the implementation of the RTC law and we find more states in which crime decreased after the law. Our tests reveal that there is no significant overall net effect of the RTC laws on murder or violent crime across all 33 states that have implemented such laws.

**References**


About the Authors

Carlisle E. Moody is Professor of Economics at the College of William & Mary, where he teaches mathematical economics, econometrics, and time series analysis. His research is primarily in the economics of crime, especially the relationship between guns and crime. His email address is cemood@wm.edu.

Thomas B. Marvell is a lawyer-sociologist. His email address is marvell@cox.net.

Donohue, Aneja, and Weber's reply to this article
Go to archive of Comments section
Go to March 2019 issue

Discuss this article at Journaltalk:
https://journaltalk.net/articles/5982/
RTC Laws Increase Violent Crime: Moody and Marvell Have Missed the Target

John J. Donohue¹, Abhay Aneja², and Kyle D. Weber³

LINK TO ABSTRACT

Donohue, Aneja, and Weber (2018), released as National Bureau of Economic Research working paper 23510, uses two distinct methodologies to provide the latest and most comprehensive evaluation of the impact on crime of state laws that confer on citizens a right to carry concealed weapons—so-called right-to-carry or RTC laws. Its most robust finding is that RTC laws increased violent crime: our preferred panel data estimate indicates a 9 percent increase, while our synthetic control analysis indicates that violent crime rose by about 14 percent in the first decade after RTC adoption.

In a comment on the Donohue, Aneja, and Weber (hereafter DAW) paper, Carlisle Moody and Thomas Marvell (hereafter MM) concede that the uniform approach of using population weights in panel data estimates of crime shows a strongly statistically significant increase of RTC laws on crime in the DAW model (MM 2019, 88). They make an unconvincing argument that the uniform practice should now be rejected and then proceed to show that simplistic panel data models not weighted by population (and using badly miscoded data) would diminish the strength of the finding that RTC laws increase violent crime (ibid., 85–88). We show that both of the proffered MM models violate the basic ‘parallel trends’ requirement of a valid panel data analysis, so their resulting estimates must be rejected. But even with these serious flaws, a more nuanced implementation and
evaluation of the MM models with attention to the requirements of panel data can illustrate and buttress the basic finding of the DAW panel data analysis that RTC laws increase violent crime.

MM (2019, 89–94) then present their own synthetic control analysis, which purports to establish that 14 states show statistically significant increases in violent crime while 12 states show statistically significant decreases. We have many criticisms of their implementation of the synthetic control analysis, from using inappropriate states as potential controls to failing to account for major pre-treatment differences. These problems cause MM to generate many severely inaccurate predictions, particularly for small states. Nonetheless, a simple aggregation of MM’s overall synthetic controls results—whether weighted by state population or the inverse of the pre-treatment error fit—reveals a strong pattern of increasing violent crime in the decade following RTC adoption.

We discuss these points in turn and then summarize in the final section.

**DAW’s population-weighted model is superior to MM’s models, and it provides clear evidence that RTC laws increase crime**

**Weighting by population is conceptually superior**

The uniform practice in the literature on estimating the impact of RTC laws on crime from the early work of John Lott through the DAW paper has been to present population-weighted panel data estimates. Every regression run by the authors of the National Research Council (2005) report examining RTC laws was weighted by population. In fact, this is the standard practice in virtually all panel data studies looking at state or county crime data, including in prior work by MM on RTC laws. In their current paper, however, they argue that the standard practice should now be rejected, and they would repose confidence in regressions that are not designed to reflect the relative population of each state.

MM acknowledge the reason that all researchers have used population-weighted regressions:

---

4. For just two very recent examples, see Chalfin and McCrary 2018; Anderson, Sabia, and Tekin 2018.
If the research goal is to estimate the overall national impact of a policy change, ... then weighting can be justified by arguing that the impact of laws in large states should be emphasized simply because they affect more people. (MM 2019, 85)

Put simply, we are trying to estimate the impact that RTC laws have had on Americans, and this can only be identified by a population-weighted regression. Following the unweighted approach that MM have suddenly decided to champion would imply that the impact of RTC laws on 600,000 individuals in Wyoming is considered to be equally important as the impact on 28 million Texans. To illustrate the importance of weighting by population, consider the MM synthetic control estimates of the impact on violent crime of the RTC laws in these two states. Using their non-normalized synthetic control approach, MM would predict that the Texas RTC law increased violent crime by 19.5 percent after ten years but that the Wyoming law had generated a 36 percent decrease in violent crime over the decade following adoption (although they never show these estimates in their paper). While we discuss below why we think MM's Wyoming estimate is so flawed, the decision to equally weight Texas and Wyoming, as MM would have us do, generates a prediction that the combined RTC laws reduced crime by 8.25 percent. A population-weighted average would show the total effect on the residents of these states to be an 18.3 percent increase in violent crime. In this example, the 18.3 percent increase would reflect the effect of RTC laws on the average American who experienced this legal adoption, and a population-weighted analysis alone would generate this estimate. MM's approach would badly mischaracterize the impact of RTC laws, heralding a significant decline in violent crime when in fact the two RTC laws led to a combined large increase in violent crime.

Having conceded the key reason for population weighting in the panel data regressions, MM (2019, 85–86) then mention a second possible advantage of population weighting: it may serve to address the problem of heteroskedasticity. This is not the primary rationale, but it is often—although not always—a secondary benefit of weighting by population. Since MM conclude that the White test indicates the presence of heteroskedasticity in the DAW population-weighted regressions, MM present estimates using a non-weighted regression approach (their OLS results) and a non-population-weighted approach that seeks to directly

---

6. MM's wildly inaccurate Wyoming estimate stems from their failure to normalize their synthetic control estimate, which leads them to attribute pre-treatment differences between the fit of the synthetic control and the treatment state to the effect of the treatment. Our DAW synthetic control estimates for the impact of RTC laws on violent crime showed a 16.9 percent increase for Texas and a 15.9 percent increase for Wyoming after ten years. The comparable normalized MM synthetic control estimates for these two states are a 13.4 percent increase for Texas and a 9.1 percent increase for Wyoming.
control for heteroskedasticity (Feasible Generalized Least Squares, or FGLS). Neither of these approaches can succeed in our primary mission, which is to estimate the experience of the average American exposed to RTC laws. But in addition to the conceptual flaw in failing to weight by population, both of the MM suggested alternatives have further problems, including the second problem that they both fail the very test for homoskedasticity that MM advocate using.\(^7\)

**The importance of investigating the parallel trends assumption**

While that second problem underscores that the MM regressions are still marred by heteroskedasticity (or some specification error), a third problem with the simplistic MM models results from MM’s failure to attend to the parallel trends assumption, which is critical to generating valid panel data estimates.

This third problem with MM’s two new panel data regressions can be highlighted by comparing them to the results of the DAW population-weighted violent crime regression. The DAW paper provides the year-by-year effect on violent crime following RTC adoption from that regression (2018, 25), which we reproduce here as Figure 1 below. This figure illustrates the critical feature of a valid panel data model that the estimated values on the states that end up adopting RTC laws is virtually zero in the years prior to adoption. Not only are the deviations from zero small, but crucially there is virtually no slope to these pre-adoption values in the years prior to RTC passage. This is important because a panel data estimate will only reveal the causal effect of the RTC law if we can assume that the trends in crime between our two sets of states (adopters and non-adopters) would evolve similarly in the absence of the law.

Three lessons emerge from the Figure 1 DAW violent crime regression. First, we see an almost perfect pre-treatment pattern confirming the critical parallel trends assumption for a panel data regression. Controlling for an array of factors (the DAW explanatory variables), violent crime is flat prior to RTC adoption. Second, Figure 1 also reveals that there is a change in the previously stable relationship of crime in the RTC and non-RTC states, and that this change begins exactly in the year of adoption of the RTC laws. If RTC laws had no impact on violent crime, one would expect that flat pattern seen in the years before adoption would continue thereafter. If some factor other than RTC laws (and the array of explanatory variables controlled for in the DAW model) led to worse violent crime performance in RTC states, you would see an elevation in the violent crime estimates, but there

---

\(^7\) This is true for both the MM unweighted OLS regression and for their FGLS regression, both of which badly fail the White test for homoskedasticity with \(p\)-values < 0.00000001.
is no reason to think it would occur in exactly the year that the RTC law goes into effect. Figure 1 makes clear that a sharp secular increase in violent crime commences at the time of RTC adoption, again buttressing a causal interpretation of these results. Third, this increase in violent crime is statistically significant beginning in the first year after RTC adoption and every year thereafter.

**Figure 1.** The impact of RTC laws on violent crime, DAW model, 1979–2014 (population-weighted)

---

**Evaluating MM’s simple panel models**

We can now compare the two alternative models—OLS and FGLS—that MM offer in place of the DAW violent crime estimates reflected in Figure 1. We must first discard all of the MM estimates because of serious coding errors they made in their panel data analysis. Specifically, the MM panel data analysis miscodes both North Dakota and South Dakota as having never adopted an RTC law during the 1977–2014 data period they analyze, even though North Dakota and South Dakota both adopted RTC laws in 1985. The error is perplexing because, in their subsequent synthetic control analysis, MM generate estimates for states adopting RTC laws, including both North and South Dakota, based on that actual year of adoption. MM also code the date of adoption for Virginia differently in their two analyses. They give Virginia a starting date of 1996 in their synthetic control

---

8. MM also have a less precise coding of their RTC law than we use in our DAW paper: they simply use a zero-one dummy that becomes one the first full year the RTC is in effect, while we use an RTC dummy that takes the value of the fraction of the year an RTC law was in effect during the year it was adopted. MM also exclude DC from their panel analysis, while we only exclude DC from our synthetic control analysis.
analysis, which is consistent with their protocol of turning on their RTC indicator in the year after adoption. In their panel data analysis, however, MM use a Virginia date of 1995, which is doubly wrong in being both a violation of their own protocol and inconsistent with their treatment of Virginia in their synthetic control analysis.

Figure 2 shows violent crime estimates using the preferable DAW data but following MM’s “OLS” approach, which does not weight by population. Three lessons emerge from this analysis. First, Figure 2 reveals a substantial violation of the critical parallel trends assumption: the red line illustrates the sharply sloping downward linear trend in crime for RTC states prior to RTC adoption.

**Figure 2.** The impact of RTC laws on violent crime, DAW model, 1979–2014 (not weighted by population)

Second, the dashed continuation of this line shows the predicted path of violent crime in RTC states had their pre-RTC-adoption trend continued, and by assumption of panel data analysis, the dashed line of Figure 2 suggests that crime would have fallen (relative to non-adopting states) by 7.2 percent after ten years without RTC adoption. Instead we see that the observed post-adoption crime path
is always above this predicted downward trend, suggesting RTC laws increased crime relative to trend.

Third, by the sixth year after adoption and beyond, the estimated increase in violent crime is always statistically significantly above this trend (at the .05 level). But instead of providing this more nuanced analysis, MM simply look at one number for the OLS violent crime estimate: they run a single dummy model for this unweighted regression, which generates the small positive estimate of 0.65 (as shown in the legend to Figure 2). But by failing to realize that such a simple model is marred by the violation of the parallel trends assumption, they merely present an inaccurate and misleading estimate of the impact of RTC laws on violent crime. In other words, MM’s violent crime unweighted OLS estimate (MM 2019, 88, Table 1, row 1, column 5) is inaccurate and misleading.  

MM also include an FGLS model designed to address the problem of heteroskedasticity (although we have already noted this model’s extreme failure of the White test). Figure 3 shows the DAW violent crime year-by-year estimates using this FGLS approach. What are the lessons from this MM-suggested model? First, unlike in the DAW model in Figure 1 where all the pre-treatment values are close to zero and flat in the years prior to RTC adoption, Figure 3 reveals both greater variability in those values and another departure from the ideal parallel trends as captured again in the downward-sloping red line in the period prior to RTC adoption. Indeed, this FGLS model fails the most basic test of parallel trends since its pre-trend dummy values are not jointly zero.

Second, the dashed continuation of this line shows the predicted path of violent crime in RTC states had their pre-RTC-adoption trend continued, and it suggests that crime would have fallen (relative to non-adopting states) by 1.2 percent after ten years without RTC adoption. As in Figure 2, we see that the observed post-adoption crime path is always above this predicted downward trend, again suggesting RTC laws increased violent crime relative to trend, and that this

---

9. MM also present an additional row of “spline” estimates in their Table 1, which is a practice that also dates back to the initial Lott and Mustard (1997) paper. Since the RAND Corporation (2018) study on gun violence research has now argued that “spline” results should not be relied upon, we ignore that component of the MM paper (and have also dropped this model in our own forthcoming work). The RAND analysis of gun research identifies “the use of spline and hybrid effect codings that do not reveal coherent causal effect estimates” as a limitation of earlier studies (2018, xxvii).

10. The most basic statistical test of the assumption of parallel trends uses an F-test of the null hypothesis that the pre-period dummies are jointly equal to zero. Applying this test in Figure 3 generates a p-value of .057, which is too low to support the parallel trends assumption. For this very permissive initial test, one would typically like this p-value to be greater than .50 and certainly no lower than .20, so the Figure 3 FGLS model fails this test badly in a way that obscures any increase in violent crime resulting from RTC adoption. For comparison, the p-value on the same F-test for our far superior Figure 1 DAW violent crime population-weighted regression is .87.
reversal in the path of violent crime occurred in the year of RTC adoption.

Figure 3. The impact of RTC laws on violent crime, DAW model, 1979–2014 (FGLS)

Third, Figure 3 shows that after the seventh year following RTC adoption, the estimated increase in violent crime is always statistically significantly (at the .05 level) above the dashed projected downward trend. Again, if one were to run the single dummy model for this FGLS regression and ignore the violation of parallel trends as MM do, one would not be presenting valid results. Accordingly, the small positive estimate of 2.12 (as shown in the legend to Figure 3) that emanates from this flawed model again yields an inaccurate and misleading picture of the true path of increased violent crime after RTC adoption. In other words, the MM violent crime regressions (2019, 88, Table 1, row 1, columns 5–6)—other than the population-weighted regression which shows a statistically significant increase in violent crime—are inaccurate and misleading. But note that both the Figure 2 and Figure 3 models that are merely more informative versions of the overly simplistic OLS and FGLS that MM present (using their badly miscoded data) still lead us to a
very clear conclusion: regardless of the flaws or limitations of the two models that MM present, their more accurate and revealing versions in Figures 2 and 3 can still detect that RTC states experience statistically significant increases in violent crime relative to pre-existing trends within a decade of adoption.

In other words, MM would reject the DAW panel data estimates that RTC laws increase violent crime by roughly 9 percent by instead offering regressions with key miscodings of RTC states that are conceptually inferior because they don’t address the primary question of interest (which is the impact of RTC laws on Americans), empirically unsophisticated by virtue of their failure to address the parallel trends assumption, and offer no benefit in addressing the problem of heteroskedasticity.

MM’s discussion of heteroskedasticity is largely a distraction from a more important issue: that the difference in results between the population-weighted and unweighted regressions is likely signaling a specification issue. This finding provides an additional reason to turn to the synthetic control analysis, which can give insight into this concern and also provide potentially superior estimates, at least for those states for which good pre-treatment matches can be found. But before turning to the synthetic control estimates, it is important to highlight once again that the DAW violent crime panel data model dominates the MM models both conceptually and econometrically for the reasons set out above.

Evaluating MM’s synthetic control analysis, which despite its flaws is shown to reveal that RTC laws increase violent crime

The DAW synthetic control analysis aggregated across all RTC-adopting states generates a year-by-year prediction of the impact of RTC laws on violent crime over the ten years following adoption (2018, 36), shown here in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1.076)</td>
<td>(1.310)</td>
<td>(1.848)</td>
<td>(2.068)</td>
<td>(2.499)</td>
<td>(3.135)</td>
<td>(2.823)</td>
<td>(3.831)</td>
<td>(3.605)</td>
<td>(2.921)</td>
</tr>
<tr>
<td>N</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Pseudo p-value</td>
<td>0.936</td>
<td>0.274</td>
<td>0.220</td>
<td>0.192</td>
<td>0.094</td>
<td>0.106</td>
<td>0.060</td>
<td>0.038</td>
<td>0.032</td>
<td>0.032</td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses. Column numbers indicate post-passage year under consideration; N = number of states in sample. Dependent variable is the difference between the percentage difference in the violent crime rate in treatment and synthetic control states at given post-treatment interval and at time of the treatment. See DAW (2018, 37–38) regarding how the pseudo p-value is estimated. * p < 0.10, ** p < 0.05, *** p < 0.01.
The synthetic control analysis of Table 1 shows that after RTC laws have been in effect for a year, violent crime starts steadily rising (relative to the synthetic control state). After ten years, the DAW synthetic controls analysis estimates that violent crime is about 14.3 percent higher than it would be in the absence of the RTC law. Note that even though Figure 1 (panel data) and Table 1 (synthetic control analysis) are derived from entirely different methodologies, they both estimate that RTC laws increasingly elevate violent crime in the ten years after adoption, which mutually reinforces this conclusion.

Moreover, DAW (2018) showed that the synthetic control result was extremely robust. Indeed, one would generate very similar estimates whether one used the control variables of DAW (those used to derive the estimates shown in Table 1) or those of other papers examining the impact of RTC laws, such as those by Lott and David Mustard (1997) and the Brennan Center (Roeder et al. 2015), or an earlier Moody and Marvell paper (2008). Similarly, one could drop any single control state from the analysis or even completely drop New York and California from the set of potential controls and the results remained strong: RTC laws consistently led to statistically significant increases in violent crime after a decade.

DAW (2018) also showed that the result that RTC laws increase violent crime was not sensitive to whether one normalized the synthetic control estimates to be zero at the time of adoption or simply allowed the estimates to emerge from the matching protocol without adjustment. Similarly, the result was robust to efforts to trim off treatment states for which the synthetic control did not well match the target state in the period prior to RTC adoption. DAW also showed the violent crime results remained strong whether one used any of four different approaches designed to improve the fit of the synthetic control by including pre-treatment values of violent crime in the matching protocol or whether one included none of these values.

Since our finding was so strong and robust, we were surprised that Moody and Marvell (2019) offered their own synthetic control analysis that appeared to question the DAW results. Unfortunately, MM’s analysis has gone astray, and the short answer is that they have not undermined the synthetic control finding that RTC laws increase violent crime in the first decade following adoption.

**MM’s flaws in implementing their synthetic control analysis**

The first step in a successful synthetic control analysis is to denote a set of possible states—called donor states—from which the synthetic control can be constructed. MM got off on the wrong foot by making a mess of that process. In total, we found 57 erroneous donor pool decisions by MM. Sometimes a state
that should not be in the donor pool was included; other times, states that should have been included were left out. For example, in their synthetic control analysis, MM erroneously treat Alabama as not becoming an RTC state until 2014 while the dominant coding that we employ treats Alabama as an RTC state as of 1975 (which MM also did in their panel data analysis). Accordingly, as an RTC state, Alabama cannot serve as a control, yet MM treat it as a potential donor state for 26 out of the 33 RTC states they analyze (and a component of the synthetic control in 14 of those 26 RTC adopters). Seventeen states have some other difference between the donor pool used by Moody and Marvell (2019) and the appropriate states used by DAW (2018). Out of 33 states in the analysis, MM used only five donor pools identical to the correct pools used by DAW.

While the various problems in the MM synthetic control analysis are not worth extended discussion, we just want to highlight how their abbreviated presentation omits any discussion of some of the major pitfalls in their approach. One obvious problem can be seen by examining their own synthetic control estimate of the impact of RTC laws on violent crime in Idaho. MM indicate that Idaho had a violent crime rate of 290 per 100,000 during the first full year of having a RTC law in 1991. Unfortunately, their poorly fit synthetic control had an estimated value of 500 per 100,000 that year. For the next two years, that rather wide disparity between the actual and MM synthetic control estimates of violent crime remained roughly stable, suggesting there had been little impact on crime in those two years, yet under MM’s assumptions these were years of more than 40 percent crime drops engineered by the adoption of RTC laws! In other words, MM attributed the massive discrepancy between violent crime in synthetic Idaho and actual Idaho before Idaho’s RTC law was adopted—resulting from their poor fit—as a crime-reducing benefit of the RTC law.

Over the ten-year period following RTC adoption, the violent crime drop

---

11. While there is some ambiguity in the appropriate date that Alabama should be coded as having an RTC law, we believe that MM were correct in their treatment of Alabama in their panel data analysis but wrong in using a 2014 RTC date for the state in their synthetic control analysis. The Rand Corporation’s Gun Policy in America initiative “developed a longitudinal data set of state firearm laws” that codes the start of Alabama’s RTC law as occurring in 1975, as we do (see https://www.rand.org/pubs/tools/TL283.html for the downloadable database). This is also consistent with the codings used by the National Rifle Association (NRA), John Lott, and the NRC Firearms and Violence report. Indeed, if one looks at Lott’s estimated percentage of citizens with concealed carry permits, Alabama ranked first among all the states for which he had data. Lott lists the Alabama percentage as greater than 8 percent for 2007—seven years before the date that MM use for Alabama in their synthetic control analysis (Lott 2010, 238). Moreover, the 2014 date that MM use would imply that Alabama was one of the last states in the union to adopt a RTC law, which would not be consistent with the gun politics of the region nor the estimated percentage of permit holders in the state seven years prior to 2014. The NRA clearly would have successfully pushed for an RTC law in Alabama decades ago if Alabama was thought not to have one.

12. DAW (2018, 60, Table A1) provides the complete list of dates for RTC adoption.
in MM’s synthetic Idaho was estimated to be over 35 percent (from 501 to 324), which was substantially better than the far smaller 16 percent drop in actual Idaho (from 290 to 243). Yet MM treat this as evidence of statistically significant and substantial crime drops caused by Idaho’s RTC law. Note that the DAW synthetic controls analysis was superior because it produced a much better fit (the DAW initial year synthetic Idaho estimate was 344, versus the MM estimate of 501!), but also because DAW did not treat that pre-existing difference as evidence that the RTC law immediately caused a major drop in crime. By doing so, MM were able to mask the fact that their own analysis frequently showed that the synthetic control performed much better (with either larger crime drops or smaller crime increases) than the comparable RTC-adopting state over the ten years following adoption.

Aggregating MM’s synthetic control estimates reveals that RTC laws increase violent crime

This unpromising beginning ends in an array of synthetic control estimates that on the whole are considerably less promising than those contained in the DAW synthetic control analysis. Essentially, MM got some very bad fits on small states and then used those poor fits to argue that there is no support for the DAW position because 14 states adopting RTC laws experienced statistically significant increases in crime and 12 experienced decreases. (Note that our more accurate synthetic control analysis would show a 15-to-8 advantage for RTC laws causing statistically significant increases in crime, which grows to 16-to-4 if one limits the

---

13. One can see this same problem illustrated in MM’s synthetic control graph of the murder rate in Texas (MM 2019, 92, Figure 1). MM’s poorly fitting synthetic Texas has a substantially higher murder rate than actual Texas at the time of adoption of the Texas RTC law. Their graph highlights that this occurred because Texas enjoyed a substantial drop in murder relative to the synthetic control—prior to the adoption of the RTC law! The MM calculus treats that ill-fitting differential as a benefit of the law, even though if one examined how crime changed in both Texas and synthetic Texas in the aftermath of RTC adoption, no such murder-reduction benefit would be observed.

14. Since we were trying to show whether the panel data finding that RTC laws increased crime was supported by a synthetic control analysis, it was important to use the same 1979–2014 time period for both approaches, which we did. Extending the data set further backwards creates data problems for variables such as poverty and unemployment, which were either not available or not consistently gathered prior to 1979. Disregarding these concerns, MM started their panel data analysis in 1977, and, without explanation, used a different time period (extending back to 1970) for their synthetic control analysis.

15. MM show that, for their statistically significant results, the majority of states experienced an increase in violent crime using the preferred “nested approach” but then go on to present inferior “default” results perhaps because the inferior estimates weakened their finding of a 14-to-12 state dominance for RTC laws increasing violent crime. This is not good practice, and the “default” estimates, which are only appropriate when “nested” results cannot be computed, should be ignored in the MM paper. See the documentation for the Stata synth program, which states that the nested option offers "better performance" than the default option (Abadie, Diamond, and Hainmueller 2014).
analysis to six to ten years after adoption, reflecting the consistent pattern that the harm of RTC laws rises over the decade following adoption.) Even though the errors in implementation invalidate the MM synthetic control analysis, if MM had simply computed how much violent crime was estimated to have changed in aggregate for the 33 RTC-adopting states for each of the ten years using their own estimates, they would have generated the estimated impacts of RTC laws on violent crime shown in Table 2.

Table 2 presents the aggregate, population-weighted impact of RTC laws on violent crime using MM’s own data and synthetic control methodology (which does not normalize the estimates to equate the actual and synthetic control crime rates at the time of RTC adoption). In other words, Table 2 just takes MM’s actual individual state estimates—which they fail to show—and aggregates them. The finding is clear: RTC laws consistently generated a statistically significant increase in violent crime, rising from a 7.2 percent increase in the first year to 16.3 percent in the tenth year. Note that this is even a larger violent crime increase than that predicted in the DAW synthetic control table reproduced in Table 1 above. Remarkably, MM have completely disguised the key finding of their own synthetic control analysis, which is that, in aggregate, RTC laws are estimated to have substantially increased violent crime.

Table 2. The impact of RTC laws on the violent crime rate, MM synthetic control methodology and data, 1970–2016

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average non-normalized TEP</td>
<td>7.21*</td>
<td>7.61*</td>
<td>6.64</td>
<td>8.06</td>
<td>9.81*</td>
<td>10.97**</td>
<td>11.01**</td>
<td>12.55**</td>
<td>14.86**</td>
</tr>
<tr>
<td>(3.82)</td>
<td>(4.05)</td>
<td>(4.21)</td>
<td>(4.72)</td>
<td>(4.78)</td>
<td>(4.76)</td>
<td>(4.79)</td>
<td>(5.41)</td>
<td>(5.05)</td>
<td>(4.80)</td>
</tr>
<tr>
<td>N</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>P-value</td>
<td>0.07</td>
<td>0.07</td>
<td>0.13</td>
<td>0.10</td>
<td>0.05</td>
<td>0.03</td>
<td>0.03</td>
<td>0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses. Column numbers indicate post-passage year under consideration; N = number of states in sample. Dependent variable is the percentage difference in the violent crime rate in treatment and synthetic control states at given post-treatment interval. * p < 0.10, ** p < 0.05, *** p < 0.01.

16. The MM (2019) synthetic control analysis goes astray so badly because their non-normalized violent crime estimates tend to be large and positive for big states (for example, four of the five highest population states have positive estimates and three of those four are bigger than 15 percent by the fifth year after RTC adoption) and large and negative for small states (four of the five lowest population states have negative estimates by the fifth year, ranging from −29 percent for Wyoming to −78 percent for North Dakota). Not surprisingly, the unrealistically large negative results tend to be found in the states with the worst pre-treatment fits between synthetic control and treatment states. The DAW (2018) paper documents the ratio of the root mean-squared prediction error (RMSPE) to the mean violent crime rate as a measure of goodness of pre-treatment fit and indicated particular concern when this value rose above 19 percent. To highlight how the MM synthetic control model was doing a particularly bad job for generating plausible controls for small states, note that the error ratio averaged a whopping 48.3 percent for MM’s estimates for
We are quite confident that the DAW (2018) paper has the best available synthetic control estimates of the impact of RTC laws on crime because our synthetic control analysis is done with greater care, with more accurate coding of RTC law adoption dates, and with a far more probing array of robustness checks than the MM analysis.

**Conclusion: The best evidence shows that RTC laws increase violent crime**

We have shown that the DAW population-weighted panel data estimates shown in Figure 1 satisfy the parallel-trends assumption of a valid panel data analysis, while neither of the alternative models advanced by MM do. This is on top of the serious miscoding problems of the MM panel data analysis. Nonetheless, a proper interpretation of the two MM models (shown in Figures 2 and 3) can reveal that RTC laws alter the path of violent crime starting at the date of adoption and generate statistically significant deviations from prior trends within a decade of passage.

Of course, the fact that our Figure 1 is the best panel data model does not mean it is perfect, and we take the MM critique as providing another reason to be interested in the results of the synthetic control approach to gain insight into the difficult problem of specification that exists in every panel data analysis. While we find the MM synthetic control approach to be too flawed and primitive to rival the more accurate, thorough, and sound analysis in the DAW paper, it is encouraging to see that their analysis conducted over a longer time frame (1970–2016, while ours extended from 1977–2014) and using a non-normalized set of estimates (in contrast to our normalized estimates) still found that a majority of states experienced statistically significant increases in violent crime from RTC adoption. It is likewise encouraging that the aggregated impact across all states mimicked our own analysis in finding strongly increasing violent crime over the decade following RTC adoption (compare our estimates, shown in Table 1, with those aggregated from the MM results, shown in Table 2).

In summary, there is consistent evidence that RTC laws elevate violent crime in the decade after adoption whether one looks at DAW’s panel data estimates (Figure 1) or synthetic controls estimates (Table 1) or the properly interpreted
panel data results using MM’s suggested non-population weighted or FGLS approaches (Figures 2 and 3) or the MM synthetic controls estimates (Table 2). Policymakers and citizens should recognize that the best available empirical data to date supports the view that RTC laws have resulted in statistically significant increases in violent crime in the ten-year period after adoption.

References


### About the Authors

**John J. Donohue III** is an economist as well as a lawyer and is well known for using empirical analysis to determine the impact of law and public policy in a wide range of areas, including civil rights and antidiscrimination law, employment discrimination, crime and criminal justice. Before rejoining the Stanford Law School faculty in 2010, Professor Donohue was the Leighton Homer Surbeck Professor of Law at Yale Law School. He is a member of the American Academy of Arts and Sciences, and the former editor of the *American Law and Economics Review* and president of the American Law and Economics Association and the Society of Empirical Legal Studies. He is also a Research Associate of the National Bureau of Economic Research. His email address is donohue@law.stanford.edu.

**Abhay Aneja** is a J.D./Ph.D. candidate studying at Stanford Law School and the University of California, Berkeley. His email address is aneja@berkeley.edu.
Kyle D. Weber is currently a doctoral student in Economics at Columbia University, having previously worked as a research fellow at Stanford Law School. His primary research interests are industrial organization and media economics. His email address is kdw2126@columbia.edu.
Unforced Errors: 
Tennis Serve Data Tells Us Little About Loss Aversion

Michał Krawczyk

In two recent papers, Nejat Anbarci and collaborators (Anbarci, Arin, Okten, and Zenker 2017; Anbarci, Arin, Kuhlenkasper, and Zenker 2018)—I refer to the papers henceforth as A17 and A18—argue that male tennis players’ behavior is consistent with aversion to losses rather than maximization of expected utility. To prove this point they sketch simple theoretical models. They propose that when a player is behind in score more effort will be exerted than when ahead if and only if losses loom larger than gains. They then find that serve speeds of male participants in one tournament are indeed higher (which arguably requires more effort) when the server is behind, for example is losing 0–1 in sets.

However interesting the case may be, I believe the analysis is flawed for three main reasons: First, they implicitly assume that under expected utility, starting from a tied score, losing a game or a set or a match should be as bad as winning it is good. That would mean that if we find that such losses in fact loom larger than such gains, this may be seen as loss aversion. However, as I show in the next (second) section, this assumption is clearly incorrect, when applied to losing or winning games at any rate. Also, they interpret greater serve speed in terms of greater effort. In the

1. University of Warsaw, 00-927 Warsaw, Poland. I gratefully acknowledge the advice I received from Adam Romer, Editor-in-Chief of Poland’s leading tennis magazine Teniski klub, which was particularly instrumental in the development of the argument provided in the “Stronger serve...” section. I have also greatly benefitted from comments and suggestions made by M. Daniele Paserman, Joanna Tyrowicz, and Łukasz Woźny. Dominika Zychowicz provided research assistance. Still, all the errors are mine, and the above-mentioned individuals do not necessarily endorse the opinions expressed here.
third section I argue that, counterintuitively, a stronger serve ultimately means less total effort. Lastly, in their theoretical analysis, they disregard behavior of the returning player (a problem discussed here in the fourth section) and make other important mistakes (the fifth section). Overall, I do not believe that Anbarci and coauthors have shown that the notion of loss aversion is useful in understanding tennis servers’ behavior.

Server’s lost games should loom larger than won games

Consider a single game. The same player keeps serving throughout the game, until one player wins four points while the other has at most two or until one player wins \( k > 4 \) points while the other has only \( k - 2 \) points. The confusing convention apparently stemming from medieval France is that, upon winning the first point, the score is given as “15:0” rather than “1:0,” followed by “30:0” and “40:0” (of course assuming the server keeps winning). The score of 40:40, and any subsequent tie, is called deuce; one point above that is called advantage. Suppose now a game is currently tied at 30:30 or deuce, which are strategically equivalent (A17 and A18 call this state \( t \), for tied). Eventually the game may be won (state \( w \)) or lost (state \( l \)) by the server. A17 normalize their value function such that \( V(w) = 1 \) and \( V(t) = 0 \). They then claim that if \( V(l) = -\lambda \) is lower than \(-1\), the player exhibits loss aversion (because losing is more painful than winning is enjoyable). As I show below, however, such a valuation is fully consistent with the standard expected utility model. Because eventually the game will be won (value 1) or lost (value \(-\lambda\)), then the value of \( t \) depends on the server’s probability of ultimately winning the game, conditional on the game being tied now, say, \( p \):

\[
V(t) = p + (1-p)(-\lambda)
\]

Now, \( p \) is not 50 percent in tennis. It is considerably higher because the server has the upper hand—it is easier to win a point when serving. Therefore setting \( V(t) = 0 \) immediately implies that \( \lambda > 1 \). For example, for \( p = 2/3 \), we have \( \lambda = 2 \).

I do not have statistics for the fraction of games that are ultimately won for each current score. One may, however, easily do the following exercise.\(^2\) First, using data from TennisAbstract.com or elsewhere, calculate the overall fraction of games won by the server. In turns out that on average, it is about 75 percent

---

\(^2\) A more sophisticated way of calculating each point’s importance, i.e., how it changes the probability of ultimately winning the match, is used by Paserman (2010).
for males and 63 percent for females (whose serves tend to have lower velocity). Second, basing on these statistics and assuming, as a handy approximation, that the server’s probability of winning each point is identical and independent, one may calculate the probability of ultimate success for each current score. For example, starting from 30:30, the server will win the game with probability 71 percent in male tennis and probability 60 percent in female tennis (note that these will always be subtly different from those for other tied scores, 0:0 for example, but A17 seem to treat all tied scores jointly, which would be a mistake). It is only logical then, that, starting from a 71 percent chance of success, a loss looms much larger than a gain. This consideration not only explains the effect that A17 wrongly call loss aversion but also predicts that it be stronger for males—as they indeed find.

Analogous analysis can account for, as A17 put it, the “‘diminishing sensitivity’ in score difference such that incremental gains in \((T_s - T_r)\) above the reference point, that is, the tied score, result in progressively smaller utility improvements” (A17, 3551). For example, in a typical male tennis match the increase in probability of eventually winning the game when one moves, say, from 30:0 to 40:0 is very low—less than five percentage points—because the chance of winning is very high already at 30:0. Again, contrary to what the authors seem to imply, this has nothing to do with risk preference and all to do with the fact that (by far) not all the points in a tennis match are equally important in terms of the ultimate success. Maximization of expected value is a reasonable benchmark for decisions with uncertain outcomes expressed in money, because (for a price taker) every dollar can buy the same basket of goods. In tennis, the goal is to win the match, not to maximize the number or the fraction of points won, so clearly not all points are equally important: there is no reason to purport, for example, that the utility difference between 15:0 and 30:0 is the same as the utility difference between 30:0 and 40:0. It seems therefore much more fruitful to analyze players’ choices not in terms of utilities of intermediate goals such as winning or losing a point, but in terms of probabilities of the ultimate success (winning the match). Such a strategy was employed by M. Daniele Paserman (2010).

Similar reasoning explaining apparent “loss aversion” and “diminishing sensitivity” applies to the score in games and sets. To illustrate the former effect, note that being the first to serve in a set has been reported to raise the probability of winning this set to 54 percent in males and 53 percent in females (Jensen 2014). Thus if each player has won the same number of games within a set so far (e.g.,

---

3. The numbers depend on the surface but these overall averages should be about right for the hard courts of the Dubai tournament that Anbarci and coauthors analyze.
4. For example, between 1990 and 2014, Roger Federer lost 24 matches in which he scored more points than the opponent (Rodenberg 2014).
zero), losing the set should rationally loom slightly larger to the currently serving player (i.e., the one who served on the very first game of the set) than winning it, because $54\% > 100\% - 54\%$. To illustrate “diminishing sensitivity,” consider a player currently losing 0–4 in games. The chances to win this set are already extremely slim, so a perfectly rational expected utility maximizer will care very little about losing another game within the same set. By contrast, losing a game at 2–2 is painful.

**Stronger serve means less, not more, total effort**

It would seem an obvious manifestation of the laws of physics that sending the ball with a greater speed generally requires more energy. However, the rally does not usually stop there and it seems prudent to assume that experienced, well-paid professionals have a planning horizon which is longer than one or two seconds. Now, there are two main consequences of a stronger serve. First, it is less precise, so there is a greater risk that it fails to go in. Second, it leaves the opponent less time to react, so that she is less likely to successfully return (O’Donoghue and Ballantyne 2004), which is of course why tennis players hit the ball so hard. Because of both of these effects, the rally is expected to be considerably shorter after a stronger serve—in all probability reducing the server’s total effort. I was not able to find direct verification of this claim in existing literature, but there is a lot of corroborative evidence. For example, playing on grass—the fastest surface—results in shorter rallies, both in terms of the number of strokes and duration (Hughes and Clarke 1995). Likewise, males, who serve more strongly on average, generally have shorter rallies than females (O’Donoghue and Liddle 1998).

I did some back-of-the-envelope calculations based on data available at TennisAbstract.com for two matches, one female and one male, played in the 2013 Dubai tournament that A17 and A18 focused on, namely Samantha Stosur vs. Ekaterina Makarova and Roger Federer vs. Malek Jaziri. The website provides the fraction of rallies won by the server conditional on the rally being at least one, two, etc. strokes long, separately for the first and second serve. From this data, one can calculate the average number of strokes after the first, stronger, serve vs. after the second, weaker, serve. While player-specific numbers differ a lot, for all four players the second serve results in longer rallies. On average, the difference in the

---

5. Although Paserman in a personal communication said he remembered that indeed there was clearly such a pattern in the data used in his 2010 paper.
6. This data is provided in a way that makes processing it quite cumbersome but in principle the same exercise could be repeated for many games.
number of times the server had to hit the ball after the second serve and after the first serve equaled 0.85.

Now, W. Ben Kibler (2009), basing on the work of Richard Schönborn (2000), states that the effort involved in a tennis serve is comparable to that of one forceful groundstroke. If by hitting the ball somewhat harder on the serve one can considerably reduce subsequent number of groundstrokes, this appears clearly beneficial from the viewpoint of overall effort, precisely contrary to the authors’ operationalization of effort. In particular, if players were to hit the ball on the second serve as strongly as they do on the first serve, they would save on effort.

As a side note, modern-day professionals are superbly fit athletes who can generally endure two or three sets of top performance. Thus in most matches they do not try to economize on effort but simply give their best in every point. The observed non-trivial within-match within-player variance of the (first) serve speed may be best explained in terms of random variations of performance and the use of mixed strategies (the receiver is in a more difficult situation when she cannot be sure if the next serve will be as strong as possible vs. somewhat weaker but heavily rotated, etc.).

The receiver is also there

Of course, a stronger serve may also affect the expected effort on part of the receiver. As for me, for example, I would probably not manage to exert any effort whatsoever if the opponent served at 150 mph (except for the anxious heartbeat that is). Reducing the opponent’s effort may well be a bad thing for the server, but the authors do not even try to consider the opponent’s effort.

Most importantly, the authors implicitly assume that the receiver’s effort does not depend on current score. This is a very questionable assumption. A priori, there is not a clear reason why she should respond to it any less than the server does. Moreover, as it is easy to see, the cross elasticity of server’s probability of winning (A17, 3550, eq. 1; A18, 4, eq. 2) with respect to the efforts of the two players does not vanish, so the server should adjust her effort in part in response to the anticipated change in the receiver’s effort. In other words, disregarding the changes in receiver’s effort most probably leads to incorrect conclusions concerning server’s optimal effort. Unfortunately, the authors do not provide any relevant game-

7. It should be noted that although Anbarci and coauthors generally interpret greater serve speed in terms of greater effort, they at times seem to cavalierly switch to interpreting it as greater risk acceptance instead, for example in their Result 3 of A17 (p. 3552).
8. Exceptions will include very long tournaments, very long matches, low-stakes matches, matches with a clear dominator, and matches involving older or injured players.
theoretic analysis and do not have any data on receiver’s effort. Again, the analysis of Paserman (2010) seems superior here. As a side note, Paserman also uses a larger sample and obtains different results, a fact upon which A17 and A18, surprisingly, do not comment in any way.

Note that all of the considerations of the second, third, and fourth sections here are strongly discipline-specific. A17 and A18 seem to try to partly build credibility for their approach on its similarity to that of Devin G. Pope and Maurice E. Schweitzer (2011), a well-known study published in a prestigious journal, by citing it 15 times in total and emulating its catchy title. Now, Pope and Schweitzer (2011) analyze golf; the only similarity with tennis seems to be that it involves millionaires trying to hit a ball of some sort. There are at least three important features of golf that make it quite different from tennis, in a manner relevant for our analysis: there is no interaction with other players; only the total score (number of strokes) matters, so natural reference points at individual holes (i.e., “pars”) are purely symbolic; and it is often the case that a player needs to exert a lot of (mental) effort to reach the hole in one putt and no effort at all to make it in two. All of these render the approach of Pope and Schweitzer (2011) one that makes perfect sense for golf but useless for tennis.

Remaining issues and conclusion

It should also be noted that the reading of both A17 and A18 is obstructed by various mistakes and omissions in their theoretical analysis. For example, the first form of the right-hand side of equations 3a and 3b in A17 (p. 3550) is obviously incorrect as the probabilities of winning and losing do not add up to one. It is also not clear whether $W'$ stands for (1) any advantageous score, e.g., 30:15, or (2) only the case in which winning the game is just one point away, e.g., 40:15, or (3) only the case in which both winning and tie are just one point away, precisely 40:30 or the strategically equivalent advantage. I sent the authors an email asking for clarification and was informed that interpretation (2) was correct. When I pointed out that this is inconsistent with their equations 3a and 3b, I was told that interpretation (3) was correct. Yet, their Result 1 seems to use interpretation (1)! Similar interpretational obscurity may trouble the readers of A18.

Both in A17 (p. 3551) and A18 (p. 5) the authors also define $\Delta_s = T_s - T_r$ and $-\Delta_r = T_r - T_s$. It must thus be the case that $\Delta_r = \Delta_s$, so it is not clear why they need to define both of them. The authors also claim (in A17) that

$$\Delta_r = -\Delta_s = \Delta \text{ if and only if } T_r - T_s = -(T_r - T_s).$$


The right-hand side of the equivalence is trivially always satisfied (so it is not clear why this condition is mentioned at all). The left-hand side is meaningless because $\Delta$ is not defined. Possibly this is exactly where the authors want to define $\Delta$ (that’s what they do in A18, where they also remove the minus sign in $\Delta = -\Delta$) but then it would not be clear how it is defined if $\Delta \neq -\Delta$. Moreover, $\Delta = -\Delta$ will in general not be true, while the second part of the equivalence is always true, as mentioned before, so the equivalence is not only useless but also incorrect.

The new theoretical elements in A18 leave the reader even more confused. In equation 6 (A18, p. 5) expected utility “[w]hen it is a game (or set or match) point favoring the server,” is defined as

$$p_s V(W) + (1 - p_s) V(t) - c,$$

whereby $p_s$ is the probability of winning a point, undefined $V(W)$ probably denotes utility after the game (or set or match) has been won and $V(t)$ means the utility when the game (or set or match) is currently tied, and $c$ stands for the cost of effort. This seems misconceived, because the game (or set or match) is usually neither over nor tied after just one additional point in tennis. In equation 9 (A18, p. 6) the authors seem to plug in $(-\lambda)$ for $V(\ell)$, although $V(\ell)$ is defined to be equal to something else in equation 4 (A18, p. 5). The authors also might have inadvertently skipped $\Delta^\ell$; this guess is corroborated by the fact that it reappears when first derivatives are taken in equation 10 (A18, p. 6). Analogous mistakes seem to have been made in equations 6, 7, and 8. The second epsilon in equation 6 as well as second and third in equation 8 should be preceded with a minus sign. It is not clear if it is on purpose that the $\ell$ is not capitalized whereas $W$ in an analogous position is capitalized (A18, 5–6).

Whereas in preparation of the Propositions, the authors speak in terms of effort when serving, Propositions 1 and 3 concern server’s “risk averse serves” (A18, 6–8). It is not clear if “more risk averse serves” are supposed to be the same thing as serves involving less effort. No proof is provided to any of the Propositions anyway. The authors ask the reader to “[o]bserve that when $\Delta_s = \Delta_r \equiv \Delta$, the first-order conditions still imply…” (A18, 6), which is confusing because there is no $\Delta$ term in the first-order conditions anyway. And on and on.

Inappropriate treatment of the distinction between the first and the second serve is also a major problem. This distinction, empirically, makes by far the largest impact on within-player variation in serve speed; note, e.g., the huge estimate for the second-serve dummy in the regressions. Yet, astonishingly, it plays no role at all in their theoretical analysis (unlike in other economic analysis of tennis, e.g., Klaassen and Magnus 2009)! If conservation of effort was the main driver behind players’ preferred serve speed, then one would expect the speed difference between
the first and the second serve to be drastically reduced when the possibility of winning or losing the match draws near; there is little reason to economize on effort at this point. Yet I was not able to find any hint that this is indeed observed. More generally, the first and second serve are so different that clearly interactions with other variables should be considered.

The empirical analysis is also troubled by omission of crucial variables. For example, the rules of tennis stipulate that after an even number of rallies within a game, in particular when the score is tied, the server must be placed to the right of the center mark and must hit the “deuce court.” The opposite is true after an odd number of rallies, in particular when the score is 40:30, 30:40, or advantage: the “ad court” must be targeted. This has important consequences for some players at least, Roger Federer included (Hodgkinson 2016). The side of the court would be important even for players with completely symmetric serving skills, as, other things being equal, it is generally beneficial to target opponent’s backhand; players do so particularly often in the second serve (Rioult et al. 2015). Obviously, targeting the backhand requires playing down the T (near the center of the court) at deuce and serving wide (near the side line) at advantage, a distinction that can also mean different optimal serve speed. Of course, the situation is reversed if the receiver is left-handed, yet another crucial factor neglected by A17 and A18.

Likewise, they could have easily controlled for the number of rallies since the beginning of the match, as a proxy for the time lapsed and for fatigue. Clearly, this variable can be expected to affect the speed of serve and to be (strongly) correlated with the set score dummies (D2, D3, and D4). The estimates on these variables in the current models are thus biased. Without controls for time and fatigue, by the way, the most telling is the comparison between the scores 0–1 and 1–0, not between 0–1 (or 1–0) and 0–0. It would thus seem more important for the interpretation of the data that the coefficients on D2 and D3 are not significantly different from each other, not that D2 or D3 is significantly different from zero (i.e., from the baseline of 0–0 in sets).

Finally, they do not explicitly include a dummy for being the first player to serve in the current set. This is surprising, because in the first set of the match this would represent a useful exogenous randomization device—the winner of a coin toss serves in the first game, so that the other player is more likely to be behind in games during this set (“in the domain of losses”). *Nota bene*: If, in line with the authors’ interpretation of loss aversion, this leads to more effort being taken by the loser of the coin toss, we should expect that she is more likely to win the first set. In fact, the opposite is true, as mentioned before (Jensen 2014). By contrast, who is the first server in the second and subsequent sets is not exogenously randomized—it is the player who was not serving in the last game of the previous set, so more often than not it is the loser of the previous set. Omitting this variable is thus likely to lead...
to biased estimates on the set score dummies (D2, D3, and D4), at least in models 1 and 4 of A17 where the difference in games is not controlled for. For example, if D2, the dummy indicating the 0–1 set score, has a positive coefficient, it could be because the player who is the first one to serve in the (second) set tends to serve faster balls.

To summarize, the theoretical model suffers from several deficiencies; the empirical analysis disregards variables which are extremely important for tennis tactics; interpretation of the key variable is highly doubtful; and the results obtained are inconsistent with previous, more thorough analysis.

Overall, Anbarci et al. have not provided evidence that Roger Federer (or anybody else) is loss averse (although he may well be). Would such papers have been accepted if their conclusion was that there was no loss aversion after all? Or is it a case for the “bias bias in behavioral economics” (Gigerenzer 2018)? Is it enough to place Serena Williams and loss aversion in the title and some gender differences in the abstract and conclusion to compensate for deficiencies of the analysis? That would certainly not serve the scientific community well.

References


O’Donoghue, P., and A. Ballantyne. 2004. The Impact of Speed of Service in Grand Slam Singles Tennis. In Science and Racket Sports III, ed. Adrian Lees, Jean-Francois Kahn,
Michał Krawczyk received his Ph.D. in economics from the University of Amsterdam. He is currently associate professor at the University of Warsaw. His research interests encompass several subfields of behavioral economics. Most of his studies involve laboratory and field experiments. He has published i.a. in American Economic Review, Behavioral and Brain Sciences, Experimental Economics, Journal of Economic Behavior and Organization, Journal of Economic Psychology, Journal of the European Economic Association, Journal of Public Economics, MIS Quarterly, and PLoS ONE. His email address is mkrawczyk@wne.uw.edu.pl.

About the Author

Discuss this article at Journaltalk:
https://journaltalk.net/articles/5984/
LINK TO ABSTRACT

We are grateful for the opportunity to reply to Michał Krawczyk’s (2019) comments on our two papers, Anbarci, Arin, Okten, and Zenker (2017) and Anbarci, Arin, Kuhlenkasper, and Zenker (2018), hereafter AAOZ17 and AAKZ18, respectively. Using different methodologies, our papers each show that loss aversion plays a major role in tennis serves and that gender is an important demographic characteristic in understanding loss aversion. Krawczyk (2019) criticizes our papers as follows:

1. Our theoretical models are “simple,”
2. Loss aversion is nonexistent in tennis,
3. A stronger serve means less effort, and
4. There is an omitted variable bias due to not incorporating the receiver.

In our reply below, we will address each of his criticisms.

Alleged shortcomings of our theoretical models

While Krawczyk (2019, abs., 114) claims that we sketch “simple” theoretical

1. Durham University, Durham DH1 3LE, United Kingdom.
2. Zayed University, Abu Dhabi, United Arab Emirates.
3. University of St. Gallen, CH-9000 St. Gallen, Switzerland.
models, in reality we fully incorporate Devin Pope and Maurice Schweitzer’s (2011) model and apply it to the tennis setting by also embedding a Tullock contest function (the most popular analytical tool in modelling any strategic contest), which takes the effort levels of the server and the receiver into consideration. Our theoretical model in AAKZ18 also incorporates the location of the serve, where placing a serve closer to sidelines of the service box would give an advantage to the server but would also entail additional cognitive effort as well as additional risk taking for him.

Moreover, Krawczyk (2019, 116) mistakenly argues that we treat all tied points jointly. In reality, we focused on crucial points and to that end we stated that: “In particular, a serving player derives the following expected utility when he/she has an advantageous score (e.g. 40–30) while serving for the game (or set or match); that is, when it is a game (or set or match) point favoring the server, where $W$ denotes this state” (AAOZ17, 3550). Likewise, we do the same when the player faces a disadvantageous score. Then, we also indicated that similar value functions could be constructed for other (i.e., uncrucial) scores as well (AAOZ17, 3550 n.3).

### Alleged lack of loss aversion in tennis

Interestingly, Krawczyk (2019) acts as if there is no loss aversion in tennis and it is a scandal to state that there is one, giving the impression that we are the only ones to do so. First of all, we are not the only ones; a recent paper by Graham Mallard (2016) studies loss aversion and decision fatigue at the Wimbledon tennis championship. More importantly, we would like to emphasize that Krawczyk (2019) ignores our empirical results that clearly show a strong behavioral pattern. Our two papers, using different empirical methodologies (one linear, one non-parametric), repeatedly show that players, both male and female, serve faster when they are behind in score compared to when they are ahead, although the timing of this aforementioned behavior differs between male and female players.

### Stronger serve purportedly meaning less effort

Here, we believe Krawczyk (2009) is confusing ‘effort’ with ‘risk.’ We would like to emphasize again that our empirical results have repeatedly shown that both male and female players serve faster when behind in score. Krawczyk (2019, 118) claims: “In particular, if players were to hit the ball on the second serve as strongly as they do on the first serve, they would save on effort.” We find this claim particularly irrelevant, given the fact that it is common knowledge that tennis
players almost always hit second serves considerably slower, and this is also evident in our empirical results, with the regression coefficient of second serve being negative and statistically significant at 1 percent, in both papers.

Other unsubstantiated blanket claims by Krawczyk include: “Thus in most matches they do not try to economize on effort but simply give their best in every point. The observed non-trivial within-match within-player variance of the (first) serve speed may be best explained in terms of random variations of performance and the use of mixed strategies” (2019, 118). This claim, like many of his other claims, is not supported by any data or previous research. Further, our empirical results are completely contrary to this claim. There is a systematic increase in the serve speed and change in the placement of serve for players who are behind in score.

Receiver also being there and other potential omitted variable biases in the empirical analysis

Krawczyk (2019, 118) contends: “Most importantly, the authors implicitly assume that the receiver’s effort does not depend on current score. This is a very questionable assumption.” In reality, we use a Tullock contest function in our theoretical model, which incorporates the effort levels of both the server and the receiver at the same point. Moreover, the panel nature of our dataset allows us to control for unobserved heterogeneity at the player and at the match level by adding fixed effects at both levels. Therefore, Krawczyk’s claim is, once again, unsubstantiated.

Krawczyk (2019, 121) furthermore argues that there are other important omitted variables in the empirical analysis, such as a player being left-handed, which may give such players an advantage when serving at the ad side while the righty players would have an advantage when serving at the deuce side. This seems rather trivial as a server serves an approximately equal number of times from both sides. Moreover, while in AAOZ17 we do not control for the location of the serve, in AAKZ18 we do.

We should once again emphasize that our results are robust to the inclusion of both match and player dummies which control for unobserved heterogeneity on both levels. While we agree that there is some merit in adding “number of rallies” as an additional control variable (Krawczyk 2019, 12) and exploring the interaction of the aforementioned variable with other variables, we believe that this is an avenue for future research and beyond the scope of our paper.
Finally, Krawczyk (2019) claims a dummy should be included for the first player to serve in the current set. Jan Magnus and Franc Klaasen (1999) did some research on the topic, and they established that: “Overall only 48.2% of the sets played in the men’s singles are won by the player who begins to serve in the set. In the ladies’ singles the percentage is 50.1%. The standard errors of the two estimates are 1.6% and 2.2%, respectively.” We believe these results show that not including a dummy for the player serving first does not create a significant omitted variable bias.

Other criticisms and our responses

We are genuinely surprised by Krawczyk’s (2019) claim that we have excessive citations to Pope and Schweitzer (2011). It should not be too surprising that we mention that paper as often as we do (nine times in AAOZ17, and six times in AAKZ18) since our papers build on the Pope and Schweitzer article, not to mention the fact that it was published in the *American Economic Review*—the top journal in the field—and has already gathered more than 350 citations within eight years of publication. On the other hand, Krawczyk (2019) cites M. Daniele Paserman (2010)—an unpublished paper—six times in his comment, when Paserman’s document is not even half the length of either of our papers.

Krawczyk (2019) repeatedly highlights the differences between golf and tennis. We do not deny that indeed tennis and golf are very different in their competitive nature. As we mention in AAOZ17, “in golf one competes against the whole field (‘open play’), while in tennis one competes against only one opponent/team at a time (‘match play’)” (p. 3547). In addition, in AAKZ18 we also add that “contrary to golf where a player has full control of every shot, in tennis the only action where a player has full control is the serve” (p. 2). That’s actually why we focused on serves in our studies. Nevertheless, both golf and tennis are both high-stakes, competitive sports, and players may be loss averse, and choose riskier options. Further, in both papers, we consider “the natural and well-defined ‘reference point’ in tennis—as the counterpart of ‘par’ in golf. It is the ‘tied score’” (AAOZ17, 3547; see also AAKZ18, 3).

Krawczyk (2019, 119, 120) claims that in our theoretical models there are mistakes in the calculation of some probabilities (in AAOZ17) and some other inconsistencies in some notation (in AAKZ18). We admit that they were overall caused by our own typos; nevertheless, they are not consequential in that all of our theoretical results are still intact.

Lastly, Krawczyk (2019) claims that there are inconsistencies between our two papers concerning differences in loss aversion between male and female
players. In both papers, we find that loss aversion influences the behavior of both genders over a longer time horizon (i.e., within a match), while for male players it also holds in the shorter run (i.e., within a set and within a game).

References


About the Authors

Nejat Anbarci is currently a professor of economics at Durham University in the UK, which he joined in August 2018. He received his Ph.D. in economics from the University of Iowa in 1988. He taught at the State University of New York at Buffalo and at Florida International University in the U.S. prior to joining Deakin University in 2008 in Australia, where he taught until August 2018. Anbarci has worked extensively in game theory (especially in bargaining theory), behavioral economics (especially in experimental economics), and political economy of natural disasters. Anbarci has published more than 60 papers at journals such as the *Quar-
K. Peren Arin received his Ph.D. from Louisiana State University in 2003. He worked at Massey University, Auckland, New Zealand, and currently he is employed by Zayed University, Abu Dhabi, UAE, and he is a research associate at the Centre of Applied Macroeconomic Analysis, Australian National University, Canberra, Australia. His area of specialization is macroeconomics, and he has wide research interests in the areas of monetary and fiscal policy, open economy macroeconomics and public economics. He has published over 20 articles in journals such as the *Journal of Public Economics*, *Journal of Industrial Economics*, *Journal of Applied Econometrics, Economics Letters, Journal of Management, Public Choice, Journal of Economic Behavior and Organization, Journal of Macroeconomics, Open Economics Review, Applied Economics*, and *World Development*. His email address is Kerim.Arin@zu.ac.ae.

Christina Zenker is currently the Executive Director of the Bachelor of Business Administration at the University of St. Gallen in Switzerland. She received her Ph.D. at the University of Basel in Switzerland in 2008. Dr. Zenker taught at Salem College in North Carolina for three years before joining the College of Business at Zayed University in the United Arab Emirates, where she worked as the Chair of the Department of Finance and Economics and as Assistant Professor of Economics until December 2018. Dr. Zenker also has extensive experience in project management in the financial industry. Her research interests lie in institutional and behavioral economics. She has published in reputed journals such as the *Journal of Economic Behavior and Organization, Applied Economics*, and the *Journal of Comparative Economics*. Her email address is czenszenker@gmail.com.
Why Did Milton Friedman Win the Nobel Prize? A Consideration of His Early Work on Stabilization Policy

James Forder¹ and Hugo Monnery²

LINK TO ABSTRACT

It is hard to imagine that many economists would dissent from the view that the combination of *A Theory of the Consumption Function* (Friedman 1957) and *A Monetary History of the United States 1867–1960* (Friedman and Schwartz 1963)—two books of fundamental importance in the postwar development of economics—were quite sufficient to earn Milton Friedman his Nobel Prize in 1976. The Nobel citation ([link](https://journaltalk.net/articles/5986/)), though, whilst clearly pointing to his work in those areas did not stop there. It reads in full:

for his achievements in the fields of consumption analysis, monetary history and theory, and for his demonstration of the complexity of stabilization policy.

It might seem that “demonstration of the complexity of stabilization policy” signifies Friedman’s work on the Phillips curve and changing expectations. There, his *American Economic Review* article “The Role of Monetary Policy” (1968) is often said to be crucial in virtue of introducing the idea that ongoing inflation would be incorporated in the wage bargain so that any expansionary effects of inflation would dissipate—the ‘accelerationist hypothesis’ as it is sometimes called. There

---

¹ Balliol College, University of Oxford, Oxford OX1 3BJ, UK. We are grateful for comments on an earlier draft from Roger E. Backhouse, Edward Nelson, and three referees.

² Graduate student, Balliol College, University of Oxford, Oxford OX1 3BJ, UK.
are clear reasons, however, to reject the view that this is the work to which the citation referred.

Although the opposite has often been claimed, as of the middle or late 1960s, there was nothing new in the expectations argument. In any case, it was the argument for rules rather than discretion, not discussion of the Phillips curve, that was the focus of Friedman’s paper. Perhaps more importantly than those points, however, the idea that a policymaker might seek to hold unemployment below equilibrium with inflationary policy is not quite stabilization policy at all—it is an attempt to maintain a disequilibrium. And if that is treated as being a variety of stabilization policy, Friedman’s argument was not that it was complex, but that it was impossible. Furthermore, the Nobel press release accompanying the announcement linked discussed the three contributions listed in the citation, and then the question of expectations and the Phillips curve, followed by that of flexible exchange rates. These last two were therefore distinct from the ideas in the citation.

And finally, there is a much better candidate to be Friedman’s “demonstration of the complexity of stabilization policy,” since there is a line of thinking—or collection of lines of thinking—in Friedman’s work that points precisely at such complexity. It is one of the bases of his advocacy of rules rather than discretion in policymaking, and in so far as it is summed up in the expression “long and variable lags” (see, e.g., Friedman 1948, 254), it might be said to be a well-recognized line of thinking. But that expression, in the context of Friedman’s work, most commonly just labels an empirical observation about lags in the effect of monetary changes.

In his work considered more broadly, there is rather more to the difficulty of stabilization policy than that, and also to the ways in which Friedman’s thinking on the matter developed. The single most notable contribution was “The Effects of a Full-Employment Policy on Economic Stability: A Formal Analysis,” one of his Essays in Positive Economics (Friedman 1953/1951), but an earlier contribution of some interest is in a book review in the Review of Economics and Statistics (Friedman 1944), and there are later indications that he continued to think along the same lines, including most notably in A Program for Monetary Stability (Friedman 1960) but also in some of the less-noted passages of the 1968 paper.

3. The contrary claim, that Friedman (1968), or perhaps Friedman (1966) or else one or other of the works of Phelps, such as Phelps (1967), was revolutionary in making this point became well established in textbooks and other sources during the 1970s. In fact, there are numerous, prominent statements of the same point before these authors. A number were presented in Forder (2010), and more in Forder (2014, ch 4, part 1). Phelps (1968, 682) actually said that argument was not original to him. The award of the Nobel Prize to Phelps in 2006 is often said to have been for the same thing. If it was, that was a mistake. The citation in fact says “for his analysis of intertemporal tradeoffs in macroeconomic policy.” The specifics of that analysis was original.

4. This point is established in Forder (2018).
The review of Oscar Altman’s

*Saving, Investment, and National Income*

Friedman (1944) first addressed the complexity of stabilization policy in a little-noticed review of Altman (1941). Altman’s book was a monograph on aggregate demand. It presented a simple account of the Keynesian theory, focused on the point that saving and investment must be equal; and a large collection of statistics tending to suggest that the American economy would generate excess saving. The implication was therefore that full employment would not be maintained without public action to achieve sufficient investment.

Friedman’s review said Altman’s book was “less significant in its own right than as an expression of the Keynesian saving-investment theory” (1944, 101). Indeed, Friedman had rather little to say about the specifics of the book, but seems to have regarded the review as an opportunity to record some objections to the Keynesian theory. Noting that Altman’s presentation of the model was simplified, he briefly described a fuller version in which the equilibrium level of income equalized saving and investment, and said that changes in investment were treated as the major determinant of income, with saving simply being a function of income. He drew attention to theoretical limitations of this view, and questioned whether available data could be used to confirm it, but also clearly noted that the focus on the equilibrium of the system presupposed that it was ‘quickly and smoothly attained’ (ibid., 102). He had it in mind, presumably, that a dynamic, rather than equilibrium, analysis of the pattern and rhythm of the business cycle, like that of Wesley Clair Mitchell (1913) and the later National Bureau studies including, eventually, Friedman and Anna Schwartz’s *Monetary History*, could be a preferable alternative.

On the question of the quick and smooth attainment of equilibrium Friedman wrote a long footnote presenting a numerical example of adjustment in a model characterized by specific lags in behavioural responses. Friedman presented the model verbally, but it can be summarised by the four equations:

\[
\begin{align*}
\epsilon_t &= 0.85y_{t-1} \\
q_t &= q_{t-1} + (s_T - s_t) \\
s_t &= s_{t-1} + q_t - \epsilon_t \\
y_t &= q_t + I_t
\end{align*}
\]

where $\epsilon$ is consumption, $y$ is income, $q$ is production, $s$ is inventory stock, $I$ is investment, $s_T$ is the target level of inventory, and subscripts indicate time periods.
In Friedman’s example the initial position was one of equilibrium at \( y_t = 100, c_t = 85, \) and \( I_t = 15, \) and he considered the effect of an exogenous fall in investment to 10. The equilibrium level of output would then be \( 66\frac{2}{3} \) and consumption \( 56\frac{2}{3}. \) He reported the output level generated in the first 34 periods. They varied between 10 and 221, in a seemingly random way, with a mean of 79 (and a standard deviation of nearly 70, though he did not report that). Friedman noted the similarity to the ‘cobweb’ model, and he could have pointed to a model by Paul Samuelson (1939) based on a second-order difference equation as compared to Friedman’s third-order one. He also said, though, that simple as it was, the model was entirely in the spirit of models of the Keynesian system of the time. Indeed, Arthur Smithies (1942) offered one of that kind, and another was the widely noted econometric study of the American economy of Jan Tinbergen (1939) which, as it happens, Friedman (1940) had also reviewed.

There is perhaps some curiosity value in aspects of Friedman’s presentation of the example. One little point is that there is what is presumably an error transcribing his results since the sequence he gave, and hence his calculation of the mean, was not quite correct. Another is that had he produced a longer series of values he would have found income settling into a periodic oscillation between values of 10 and 123. So although intended to illustrate a point about a slow and volatile approach to equilibrium, Friedman’s model in fact never reaches equilibrium. It is hard to say whether Friedman would have thought that damaged or strengthened his criticism.

In any case, Friedman had clearly called the value of such models into question. If the model needed a lag structure of the kind Friedman said, then with the volatility of outcomes it generated, it could hardly be of much use to policymakers. Actual lags, including lags in the operation of policy itself were surely more complex than those Friedman suggested, and that meant the outcome of policy would be, in practical terms, wholly unpredictable.

The “Formal Analysis” of the effects of a full employment policy

In “The Effects of a Full-Employment Policy on Economic Stability: A
Formal Analysis,” Friedman (1953/1951) again presented basic Keynesian theory. This time he was responding particularly to John Maurice Clark, Smithies, and coauthors (1949)—a report by a group of experts for the United Nations. Like Altman’s, the book was partly expository and partly hortatory, it described Keynesian theory in simple terms, and it carried the same emphasis on the possibility or danger of investment being insufficient to absorb full employment savings.

Here, though, Friedman (1953/1951) expanded on remarks he had made earlier (Friedman 1947; 1949), and he was more specific in saying the theory suggested that fiscal policy could straightforwardly compensate for variations in investment. He said that as the model was usually presented all that was required for full employment was a contemporaneous variation of fiscal policy to offset any exogenous variation of private investment. Noting that the assumed absence of any relevant lags was an important aspect of the model, he went on to say that the appearance of the theory was that so long as policy usually operated in the opposite direction to shocks and was not more powerful than the shocks, it would be stabilizing. However, that view, he observed, was in error for the purely statistical reason that (in the additive model he considered) the variance of outcomes after policy intervention is the sum of the variance of output without policy, the variance of the effect of policy, and twice the covariance of the two.7

Taking the policymaker’s ability to determine the correct direction of policy as imperfect and given, Friedman then showed that attempts at stabilization policy could easily be too powerful, and as a result be destabilizing, in the sense of increasing the variance of output. This demonstrated the fallacy in the common view that so long as policy had a weaker effect than the shock itself and was usually in the right direction, it was bound to be stabilizing. Actually, it needed to be much more accurately aimed than that.

So the overall success of stabilization policy depends on both the magnitude of its effects and its correlation with shocks. The strength of policy action, Friedman said, could be reasonably controlled since, for example, a larger fiscal response would be a more powerful policy. The magnitude of effect raised a more difficult matter since that would depend on the consequential changes brought by the policy action itself. Yet more difficult, though, was the question of whether policy would push in the right direction. Friedman treated that question in terms of the “timing” of policy, saying that “if the need for action could be recognized immediately, the recognition translated immediately into action, and the action

---

7. His model was ‘additive’ in the sense that if output under a full employment policy is $Z$, output without policy is $X$ and the effect of policy is $Y$, then $Z = X + Y$. In that case $\text{Var}(Z) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$. The essential point he was making would carry over to non-additive cases.
immediately effective” (1953/1951, 129), then the covariance of shock and policy could be close to its ideal value of −1, and that this was implicitly the assumption made by those who use that model. He said that in reality, however, achieving a covariance of that kind of value would require effective forecasting of outcomes both with and without policy action, and including all the lagged effects of past policy actions.

On the premise that such effective forecasting would not happen—as Friedman clearly and with much justice believed—he suggested that making the lags as short as possible was the only way to address the problem. Here, he noted that the lags in the effects of policy would be spread over some time and might be variable. Then, rephrasing what he had said in “A Monetary and Fiscal Framework for Economic Stability” (Friedman 1948), he distinguished three lags: the lag in recognizing need for action, the lag in taking it; and the lag in its having its effects. Friedman (1948) had argued in favour of the abandonment of discretionary policy and its replacement by a system of what would later be called ‘automatic fiscal stabilizers,’ with variations in the government budget met entirely by variations in the quantity of money; in making this case, he had argued that lags would on the whole be shorter with automatic rather than discretionary policy. In the “Formal Analysis” he contemplated the possibility that his proposal would nevertheless be destabilizing as a result of being too powerful relative to the proportion of occasions when it pushed in the right direction (1953/1951, 130)—a point he had not made in the 1948 piece. This point he contrasted with the claim of Clark, Smithies, et al. (1949) that automatic stabilizers would not be powerful enough to balance shocks and that only additional action could stabilize demand, saying “In the light of our analysis this statement is, at best, misleading; at worst, downright wrong” (Friedman 1953/1951, 131). Misleading, it certainly was.8

Friedman (1953/1951) did not draw attention to the kind of calculation he had made in his review of Altman, and whereas in the 1944 piece the key lags were in the behaviour of the economy, in the later piece they were in the response of the policymaker and in the effects of policy action. In that sense, the orientation of the two is distinct. On the other hand, both turn on an appreciation of the importance of lags in what others were thinking of as an analysis in comparative statics. And in both cases, drawing attention to these lags suggests that policymaking would be much more difficult than it might seem.

In neither case did Friedman offer any pretense of being able to describe the details of the lags, or indeed of thinking it might be a practical possibility to

8. Clark, Smithies, et al. (1949) certainly had not appreciated the kind of points Friedman was making and therefore did not specifically say that they were assuming policy always moved in the right direction and without lags, but allowing them that assumption, what they said was correct.
do so. In this respect he was less ambitious, though perhaps more realistic, than A. W. H. Phillips (1954; 1957). Phillips clearly hoped to make progress on the actual estimation of lag structures with a view to refining policy, although in the end he was unable to answer the questions he had raised. For Friedman though it was not the precise specification of lag structures that was at issue, but rather the drawing of attention to the sensitivity to them of optimal policy. In 1944 the volatility in the behaviour of the model suggested policymakers could hardly know what they were attempting to do; in the later version, the dependence of the point on a specific version of the Keynesian model had been replaced by a much more general point about the dangers of attempting stabilization based on any model. Lags would arise in any system, and Friedman’s way of looking at that was to think of them as impairing the timing of policy and the policymaker’s ability to control the magnitude of its effect. The lags thereby both move the covariance of policy and shock away from its ideal value and made it hard to be confident that policy would not then be too powerful. This argument of statistical theory, combined with Friedman’s discussion of the various lags in policy, made the case that stabilization policy is much more complex than it seems, and goes to the heart of policymaking presumptions at least as powerfully as do *A Theory of the Consumption Function* and *A Monetary History of the United States*.

**The later neglect of Friedman (1953/1951)**

It seems very notable that in recent times, the formal core of the Friedman (1953/1951) argument has not been nearly so noted as might be expected. A vaguer understanding of the importance he attached to long and variable lags is certainly widespread; but in the period after the award of the Nobel Prize, this key paper itself has not been nearly so noted as it might have been. Allan Meltzer (2016) raised the question of the reason for the award of the Prize itself but his reaction to the point about stabilization policy was to defer to Harry Johnson (1976). Meltzer said he had summarized Friedman’s relevant work, including “his demonstration of the complexity of stabilization policy,” especially his December 1967 Presidential Address to the American Economic Association,” but made no mention of Friedman (1953/1951). Actually, Johnson had said that the theme ran through Friedman’s work, but indeed, “The Role of Monetary Policy” (Friedman 1968) was the only piece he specifically identified in that connection. James Tobin (1976) wrote a piece that accompanied Johnson’s, but made no mention at all of the complexity of stabilization policy. Perhaps most notably, Niels Thygesen (1977), specifically describing Friedman’s work in the context of his award of the Prize, made nothing of it. Thygesen discussed the importance of lags in Friedman’s think-
ing about monetary policy, but principally only as a point leading to various controversies (1977, 77). Friedman (1953/1951) was referred to only towards the end of the paper in a round-up of undiscussed topics emphasizing the diversity of Friedman’s work (Thygesen 1977, 82); even then the paper was not actually identified but only presented in the disguise of being one of the “two essays on a stable framework” in Essays on Positive Economics. Edward Nelson (2018, 494) stands out as having recognized the point when he described the paper as being “alluded to” in the Nobel citation, though he said no more about the relation of the paper to Friedman’s award.

But it is not merely that the reason for the Prize has been overlooked; Friedman’s broader thought on this issue has been much neglected. Some might be persuaded by the point that Friedman (1953/1951) has been cited only something like one thirtieth as often as Friedman (1968). Citation counts are always treacherous, and in this case particularly so, but here they tell a story consistent with other evidence.

The point is much more clearly apparent from of the failure of authors to refer to Friedman (1953/1951) in contexts where, had its point been appreciated, they would surely mention it. There are those who do discuss it, of course, but many notable cases of those who do not. Eamonn Butler (1985), writing at the time the only book-length biography of Friedman focusing on presenting a distillation of his economics, does not mention it. Robert Cord and J. Daniel Hammond (2016) edited a volume of 860 pages in 40 chapters on Friedman, and none of the chapters mentions the paper. One would certainly expect the paper to appear in accounts of the rules and discretion debate, yet it hardly does. David Laidler (2017) noted Friedman’s (1960) advocacy of rules. Laidler (2017, 20–21 n.22) also commented in a footnote that an anonymous referee had pointed out that Friedman’s case for rules arose from uncertainty about the operation of the monetary system, but despite noting the importance of lags in Friedman’s thinking, he did not otherwise take the discussion in the direction of Friedman (1953/1951). John Taylor’s NBER working paper “Rules Versus Discretion: Assessing the Debate Over the Conduct of Monetary Policy” (2017) mentions Friedman (1948) and Friedman (1960), but not the 1953 paper. Surveys of the rules and discretion literature include Victor Argy (1988), Stanley Fischer (1990), and William Van Lear

---

9. The data is from Google Scholar (Feb. 23, 2019), which shows the earlier piece cited 287 times, as against 8886 for the later one.
10. Friedman (1953/1951) was originally published in French (Friedman 1951), and its later English publication came in a book (Friedman 1953), which in itself makes it less likely than an AEA Presidential Address to attract attention. Secondly, the 1968 paper has come to be cited for its presumed influence on the discussion of the Phillips curve. As Forder and Somme (2019) show, it ended up being cited for very much the wrong reasons, so that its true influence is called into question.
(2000). Of them, only Argy cited Friedman (1953/1951) at all, and then only for the existence of a variety of lags, not for Friedman’s point of statistical theory. And an article by David Glasner (2017) that is perhaps not exactly a survey but is titled “Rules Versus Discretion in Monetary Policy Historically Contemplated” also made no mention of Friedman (1953/1951).

Perhaps a particularly interesting case is William Brainard (1967), who adopted a theoretical disposition similar to Friedman’s. Brainard’s presentation had aspects of greater formalism than Friedman’s, and there are differences in detail, but the essence of his argument is very similar. The idea of “Brainard conservatism”—suggesting that in the presence of uncertainty, policymakers should act less firmly than they otherwise would—became fairly well understood. That is not the way Friedman put his argument, but it is equally an implication of it, and the theoretical explanation in the two papers is very similar. Brainard himself, though, did not cite Friedman (1953/1951). Brainard’s paper was well-cited, perhaps thereby diverting attention from Friedman’s earlier insight.

And then evidence of a slightly different kind comes from Thomas Sargent and Neil Wallace (1976). They were considering the consequences of the assumption of rational expectations and noted that with adaptive expectations, but not with rational expectations, there would always be a feedback rule for policy that would be superior to a simple money growth rule. On the assumptions they were making, which left no room for the considerations motivating Friedman’s position, that is correct. But the interesting point is that this led them to say: “Within the context of macroeconometric models as they are usually manipulated, Friedman’s advocacy of a rule without feedback seems indefensible” (Sargent and Wallace 1976, 170). They would hardly have been making a point like that in relation to Friedman’s position had they had the rationale of his view in mind.

### Explaining the neglect of Friedman (1953/1951)

There is therefore something of a puzzle as to why Friedman (1953/1951) has in this later period been so little noted. Part of the answer may be that its importance was obscured by the creation of an impression that Friedman’s crucial contribution related to the Phillips curve. As Forder (2018) shows, even in Friedman’s writings that is something that occurs only in 1975—just about the

---

11. The relationship between Friedman’s and Brainard’s papers, and Brainard’s non-citing of Friedman, are also discussed by Nelson (2018, 428).
same time as the Nobel award. But it is also interesting that even when discussing the complexity of stabilization policy, Friedman himself seems to have downplayed the argument of statistical theory when data accumulated to suggest that actual policy outcomes were poor.

In Friedman (1944), Friedman (1948), and Friedman (1953/1951), his discussions of the importance of lags were all essentially theoretical. But in Congressional testimony in 1958, Friedman noted that since 1907 the peak rate of growth of the quantity of money had preceded the peak of the business cycle by between thirteen and twenty-four months, and that in the case of troughs the range was between five and twenty-one months (Friedman 1969/1958, 180). His first conclusion was that the lags were too variable to expect active stabilization policy to be effective.

At that time, the further conclusion Friedman drew as regards policy was that the lag was apt to create an incorrect impression that monetary policy was generally powerless. He cited the inflation of 1956 and 1957, which had occurred while money was tight, and the lack of growth of the money supply in 1928, which was simultaneous with a period of expansion but, he said, contributed to the severity of the downturn in 1929. Furthermore, he said that this kind of outcome had the effect of misleading policymakers so that they were induced to take stronger action than appropriate. Referring to the same two incidents, and saying that 1920 offered an even clearer instance, he said that the appearance that policy was not having the desired effect led policymakers to believe that stronger measures were required, and so to take action which turned out to be excessive. On the other hand, he also said that in 1932 the failure of a mild monetary easing to bring results created the impression that monetary policy was ineffective, and that impression contributed to the Federal Reserve allowing the money supply to decline. These things, he clearly felt, led to a failure to recognize the long-term importance of monetary policy. The two conclusions together naturally led to the view that achieving a steady rate of growth of the quantity of money was the best practical option.

Friedman’s views on lags brought a critical comment from John Culbertson (1960) to which Friedman (1961) responded, again addressing the question of the relationship between long lags and the quality of policy. Citing Friedman (1953/1951), Friedman said that his view was not that policy interventions were generally perverse but that “they are largely random relative to the actions that in retrospect would have been appropriate. The result is to convert actions taken for countercyclical purposes into additional and unnecessary random disturbances” (1961, 464).

Friedman (1960) made an argument along similar lines to that of the 1958 testimony, whilst proposing that policy be set to achieve a constant rate of growth of the money supply. He said, “I doubt that many, if any, informed students of
monetary affairs would disagree with the judgment that the actual behaviour of the money stock has clearly been decidedly worse than the behaviour that would have been produced by the simple rule” (1960, 93), and over the next few pages listed a selection of cases where the rule would have done better than actual policy. This led to the conclusion, amongst others, that in a rough and ready way, one might think that in the postwar period, policy had pushed in the appropriate direction 47 percent of the time. Citing Friedman (1953/1951), he pointed out that anything like a 50-50 result meant that the rule would have been decidedly better, saying that the variability of money supply growth resulting from the absence of a rule “is simply a disturbance that introduces instability” (1960, 96).

Friedman’s (1960) advocacy of the money supply rule involved a retraction of his proposals from Friedman (1948). There, he had proposed a range of reforms, including the implementation of 100-percent reserve banking and the cessation of the issuance of government debt other than currency. The centerpiece, though, was a plan for automatic variation of the money supply. Friedman (1948) had proposed that spending and taxation rules would be fixed so that the business cycle would automatically generate fiscal imbalance—an arrangement for ‘automatic stabilizers’ as such fiscal responses would later be called. But rather than varying the quantity of interest bearing debt, Friedman proposed that fiscal imbalance be reflected in variations in the money supply. There would therefore be an automatic response of the money supply to business fluctuations.

Friedman (1948) had described his principal objective as the removal of discretionary policy from policymakers—something which he, following Henry Simons (1948/1936), thought essential to the maintenance of democracy. It was a happy coincidence that, as he argued, the plan would be likely to bring effective stabilization policy as well. The argument that the proposal would bring good economic results was, however, conjectural. Friedman argued that such automatic policy would be subject to shorter lags than discretionary policy, but there was no reason to suppose that the lags would be short enough, or that responses to those policy actions would be such as to bring stability. The proposal was, in principle, subject to the central criticism of Friedman (1953/1951): that the covariance of shock and policy might be such as to make the whole scheme destabilizing. But that was not true of the argument for a money supply rule made in Friedman (1960), because there the argument was empirical. It was a fact—so he argued—that actual policy had performed worse than the rule would have done. As Friedman (1960, 98) admitted, he had no very persuasive argument that targeting the money supply would work in theory, but the case for it was that it worked in practice.

The same idea—that evidence showed that policy was pushing in the right direction insufficiently often—appeared in Friedman (1968). Indeed, perhaps there Friedman came close to saying policy was perverse rather than merely ran-
dom. He said that policy should have been expansionary before 1965 when it was not, but when it became so in 1966 it went too far: “And this episode is no exception. Time and again this has been the course followed—as in 1919 and 1920, in 1937 and 1938, in 1953 and 1954, in 1959 and 1960” (1968, 16). The lesson was, then, that policymakers acting with discretion were not coming close to improving outcomes. This time there was not even a mention of Friedman (1953/1951). The whole case had become one of a lesson of experience: Stabilization policy was much more complex than it seemed, and that was shown not abstractly but by the record of failure.

The idea that “long and variable lags” create a significant difficulty for stabilization policy is a recurring theme of Friedman’s work, and surely widely recognized as such. It would not be surprising if some find it natural to see the Nobel citation as referring to that. That phrase, however, does not in itself invoke the analysis of Friedman (1953/1951). It might seem to suggest only the practical experience of the difficulties of policy rather that the argument of statistical theory that Friedman also made.

**Conclusion**

The lines of Friedman’s thinking arising from Friedman (1953/1951) and some of his later work on the variability of lags fits the description of the Nobel citation much better than do stories about the Phillips curve. It might at first sight seem questionable whether they are weighty enough to warrant an appearance in the Nobel citation, but it is to be recalled that, as of 1976, debate raged over monetarist proposals, and the case for rules rather than discretion was one of the issues at stake. In retrospect those lines of thinking may seem to have left less of a mark than *A Theory of the Consumption Function* or *A Monetary History of the United States*, but perhaps such marks are not reliable indicators of economic insight.

And in fact, the explanation of the muted appreciation of those lines of thinking may lie in the development of Friedman’s own thought and his presentations of the matter. He had, of course, a strong predisposition towards rule-governed policy from an early stage, for reasons of political accountability, rather than economic effectiveness. The little model of Friedman (1944) offers no more than an illustration, albeit one that captured a problem with the then-fashionable models. In Friedman (1953/1951), reacting to very much the same model, he had an argument of much more general significance, although still with the havoc caused by lags at its heart. It did mean, though, that his own proposals from Friedman (1948) were called into question as well, as there was no very clear reason to believe automatic responses would be better correlated with required policy than
discretionary ones. Then, though he continued to cite Friedman (1953/1951), his position changed in two ways. One is that he adopted the simple money supply rule, and the other is that the case he presented for preferring the rule to discretion became very largely empirical.

That case was, firstly, that retrospective observation revealed that the simple money supply rule would in fact have delivered better policy than what was actually done—a comparative result that hardly could have been argued in Friedman (1948) because it then was much less clear what outcomes would have arisen from the whole combination of reforms he had proposed and hence impossible to compare them with the actual results of policy. Secondly, there was the point that not merely would the rule have sometimes performed better than discretion, but errors in discretionary policy were frequent. The case was, precisely as Friedman (1960) put it, that rules appeared to work in practice, even though a convincing theoretical case was hard to devise. Once Friedman saw sufficient evidence of the failure of stabilization policy, the point of statistical theory from Friedman (1953/1951) may have seemed redundant to the making of his case. The point that policy which was right half the time was nowhere near good enough was certainly still important, but it was like a bonus from Friedman’s point of view, rather than being the basis for conjecture about how difficult stabilization might be. In his own treatments, he then did not greatly press that argument.

References


Laidler, David E. W. 2017. Economic Ideas, the Monetary Order and the Uneasy Case for


About the Authors

James Forder is Andrew Graham Fellow and Tutor in Political Economy, Balliol College Oxford. He has written on central bank independence, European integration, the history of the Phillips curve, and the work and reception of the work of Milton Friedman. In *Macroeconomics and the Phillips Curve Myth* (OUP, 2014) he has argued that the commonly told story of the discovery of a negative relationship between wages and unemployment in 1958, a Keynesian adoption of the curve as a policy menu, and Friedman’s refutation of that idea is historical nonsense. His email address is james.forder@balliol.ox.ac.uk.

Hugo Monnery is an MPhil student in Economics at Balliol College, Oxford. He previously studied Philosophy, Politics, and Economics. Alongside the history of economic thought, he is interested in productivity, growth, and business cycles. His email address is hugo.monnery@balliol.ox.ac.uk.
Edmund Burke as an Economist

Donal Barrington

“‘The age of chivalry is gone: that of sophisters, economists, and calculators, has succeeded; and the glory of Europe is extinguished forever.’”¹

It seems almost sacrilegious to accuse the writer of this famous passage of being an economist himself, and not many people have attempted to do so. The romantic language in which Burke has clothed much of his philosophy might lead one to imagine that he was not interested in the ordinary affairs of mankind, and as economics is concerned with very little else, that he was not interested in economics. Indeed, Melchior Palyi, discussing the relation between Burke and Adam Smith, wrote, “their ideas, their methods, even their problems, were decidedly different, as different as the men themselves and their personal careers.”²

Such an impression is false. One can over-emphasise the romantic element in Burke. In fact he was so concerned with the happiness of ordinary men and women that some writers have called him a utilitarian. That, I think, is putting the point too strongly, but utilitarianism was an important factor in his political philosophy.

Moreover, an interest in the nobler and more spiritual aspects of man’s nature does not preclude one from being acquainted with the manner in which he earns his bread. Indeed the career of Lord Keynes shows that one can become a great economist through sheer annoyance at man’s having to spend so much time earning his living.

Burke had a sound grasp of the central principles of political economy. He

¹. This article was originally published in Economica, n.s. vol. 21 no. 83, pp. 252–258, © 1954 London School of Economics. It is reprinted here with kind permission from John Wiley & Sons, Inc.
was the first great English statesman to preach Free Trade and he is entitled to be considered as one of the pioneers of economic science.

How Burke acquired his great knowledge of economics is not easy to explain. But it is not much more difficult to account for his grasp of this subject, than it is to trace the origins of his vast knowledge of almost every other branch of human learning. “Single speech” Hamilton, who employed Burke for six years, from 1759 to 1765, as “companion” in his studies, said of him that he understood everything in the world except music and gaming. On another occasion Hamilton remarked to a friend that, though he himself was a Lord of Trade, though he had access to all the official documents, and though he had studied them conscientiously, nevertheless he felt at a loss when talking to Burke, so great was Burke’s knowledge of this subject.

Burke probably spent part of his six years with Hamilton studying political economy. But long before this, while still a student in Trinity College, Dublin, he seems to have envisaged that the scattered principles of economics could be gathered together into a definite science. In *The Reformer* he lamented the sad condition of Irish trade and suggested that the nobility should patronise the study of trade just as they patronised the other branches of learning. Burke approved of the activities of bodies like the Dublin Society but thought that it was also necessary for private individuals to take up the study of trade as a definite branch of learning—

“Nothing comes to its height at first, and the Spirit of encouraging Trade, may at length rise to Science. What has been done hitherto has been by Bodies of Men, few have had the courage singly to venture any Thing, tho’ private Men have always been the Support of Works of Politeness.”

This was written in 1748 when Burke was eighteen years old. Eleven years later, Dr. Markham, the headmaster of Westminster, writing to the Duchess of Queensbury to recommend Burke for the post of consul in Madrid, remarked that his chief application had been to the knowledge of public business and to the commercial interests of Great Britain; that he seemed to have a most extensive knowledge, with extraordinary talents for business, and to want nothing but ground to stand upon to do his country very important services. Some years later the “all-knowing” Jackson said to Dr. Johnson that his “Journey to the Western Islands of

Scotland” contained more good sense about trade than would be heard for a whole
year in Parliament “except from Burke.”

On his entry to Parliament in 1766 Burke had quickly established a reputation
for himself as an expert on problems relating to trade and commerce. Shortly after
his first speech, General Lee wrote to the Prince Royal of Poland that an Irishman,
one Mr. Burke, had sprung up in the House of Commons, who had astonished
everybody with the power of his eloquence, his comprehensive knowledge of all
the exterior and interior politics of Great Britain and of her commercial interests.

In October 1766 the Duke of Grafton wrote to Chatham suggesting that they
try to attach Burke to the Ministry by making him Lord Commissioner of the Board
of Trade. Chatham rejected this suggestion. He admitted that Burke was a man
of great ability, but considered that his views on trade were too unorthodox. “As
to his notions and maxims of trade,” he wrote, “they can never be mine. Nothing
can be more unsound and repugnant to every true principle of manufacture and
commerce, than the rendering so noble a branch as the Cottons, dependant for the
first material upon the produce of French and Danish Islands, instead of British.”
The great debate on Free Trade had begun.

In his speech on his arrival in Bristol in 1774 Burke himself bore testimony
to his economic studies. When first he devoted himself to the public service, he had
considered how he should render himself fit for it: and this he did by endeavouring
to discover what it was that gave England the rank she held in the world. He found
that her prosperity and dignity arose principally, if not solely, from two sources, her
constitution and her commerce. Both these he had spared no study to understand,
and no endeavour to support.

Many years later, when, in his “Letter to a Noble Lord,” he was answering the
attacks which the Duke of Bedford and the Earl of Lauderdale (the noted econo-
mist) had made on his pension, he wrote that he had not come into Parliament
to con his lesson. He had earned his pension before he set foot in St. Stephen’s
Chapel. During his first session in Parliament he had found it necessary to analyse
the whole commercial, financial, constitutional and foreign interests of Great
Britain and its empire.

He continued with the following most interesting passage:—

“Does his grace think that they who advised the crown to make my retreat

10. Chatham to Grafton, October 19th, 1766, in Autobiography and Political Correspondence of Augustus Henry,
easy, considered me only as an economist? That, well understood, however, is a good deal. If I had not deemed it of some value, I should not have made political economy an object of my humble studies, from my very early youth to near the end of my service in parliament, even before (at least to any knowledge of mine) it had employed the thoughts of speculative men in other parts of Europe. At that time it was still in its infancy in England, where, in the last century, it had its origin. Great and learned men thought my studies were not wholly thrown away, and deigned to communicate with me now and then on some particulars of their immortal works. Something of these studies may appear incidentally in some of the earliest things I published. The house has been witness to their effect, and has profited of them more or less for above eight and twenty years.”

The remark about the “great and learned men” who communicated with Burke on some particulars of their “immortal works” almost certainly refers to Adam Smith and the Wealth of Nations. The two men greatly admired each other. When Smith read Burke’s essay on the “Sublime and Beautiful”—the first work which Burke published under his own name—he remarked that the author would be a great acquisition to Glasgow University “if he would accept of a chair.” Some time later, in the year 1759, Hume, who was attempting to popularise Adam Smith’s Theory of Moral Sentiments in London, sent Burke a copy, pretending that the present had come originally from Smith. Burke was very much taken by the book and asked Hume for Smith’s address. In the same year he reviewed the book very favourably in the Annual Register.

The Wealth of Nations was also reviewed in the Annual Register but we cannot be sure that this review was written by Burke. Burke was still in charge of the Annual Register in 1776 and was, in fact, still writing most of it. But he had an assistant, so that we cannot be sure that he wrote this particular review. On the other hand it is quite probable that he did, especially as Smith was a friend of his, and as he was so interested in economics. The style in which the review is written would also appear to support the theory that Burke wrote it.

The reviewer praises the work highly. He also shows himself to be acquainted with the writings of the Physiocrats. “The French economical writers,” he says, “undoubtedly have their merit. Within this century they have opened the way to a rational theory, on the subjects of agriculture, manufactures, and commerce. But no one work has appeared amongst them, nor perhaps could there be collected

16. Ibid., p. 312.
from the whole together, anything to compare with the present performance, for sagacity and penetration of mind, extent of views, accurate distinction, just and natural connection and dependence of parts.”

Burke and Smith may have met as early as 1759. Certainly they knew each other by 1775 when Smith was elected a member of Dr. Johnson’s Club. They became close friends and they saw a lot of each other during the two years Smith spent in London after the publication of the Wealth of Nations. Burke went to Glasgow in 1784, upon his election as Lord Rector of Glasgow University. He spent about ten days in Scotland. Where he stayed we do not know, but he certainly spent most of his time in the company of Smith, who was then Professor of Logic at the University.  

Bisset gives an account of an interview which a “very eminent literary gentleman” had with Burke shortly before the latter’s death. Burke, it seems, spoke in very flattering terms of Smith, described him as a man of profound and extensive learning, and said that his work would be of great value. He added that Smith’s heart had been “equally good with his head” and his manners “peculiarly pleasing.”

Dr. Johnson did not consider an occasional sally into the economic field beneath him, and with Burke and Smith such keen students of economics, it seems strange that that subject was so seldom mentioned in the Club. Perhaps the reason was that Johnson strongly disliked Smith, while Smith had ample reason to resent Johnson’s rudeness. Or perhaps it was that Johnson, who liked to argue for victory, preferred to avoid a subject about which Smith and Burke knew so much more than he did.

Burke and Smith being friends, it might seem reasonable to conclude that Burke borrowed his ideas on economics from Smith. No doubt Smith did influence him, but, nevertheless, Burke appears to have reached his main conclusions independently. We have seen that while still in college he had been interested in the study of trade, and that he himself claimed to have studied political economy long before, to any knowledge of his, it had engaged the thoughts of speculative men elsewhere in Europe. These claims are corroborated by a remark attributed to Smith. He is reported as having said that Burke was the only man he ever met who thought exactly as he himself did on economic problems without any prior communication having passed between them.

Indeed, it would seem that Burke not only arrived at his conclusions indepen-

---

dently of Smith but that he actually assisted Smith on some points. Thomas Moore reports in his *Memoirs* an interesting remark of Wordsworth’s. Wordsworth complained of the ignorance of politicians. Burke, however, he said, had been an exception. He was by far the greatest man of his age. Not only did he abound in knowledge himself, but he helped his most able contemporaries in almost every field of study, “assisting Adam Smith in his Political Economy and Reynolds in his Lectures on Painting.” Wordsworth’s statement that Burke assisted Reynolds is correct, and his statement about Smith may also be accurate.

In fact it fits in very well with Burke’s remark about the “great and learned men” who had discussed economic problems with him. But we have another piece of evidence. This is contained in the preface to an edition of Burke’s *Thoughts and Details on Scarcity* which was published in 1800. This little pamphlet is the most important of Burke’s economic writings, but it was not published during Burke’s lifetime. In its original form it was a memorial which Burke presented to Pitt in 1795. Burke had intended expanding it into a series of Letters on Rural Economics addressed to Mr. Arthur Young. Unfortunately, he had hardly begun this task when he heard of the peace negotiations between France and England and he abandoned it to write his Letters on a Regicide Peace.

Upon Burke’s death in 1797 his two most loyal disciples, Dr. French Laurence and Dr. Walker King, were entrusted with the care of his papers. Three years later the *Thoughts and Details on Scarcity* appeared. It consists of the memorial addressed to Pitt in 1795 expanded to include some fragments, which the editor had discovered, of Burke’s proposed Letters on Rural Economics.

The editor does not give his name but the Preface is dated “Beaconsfield, Nov. 1, 1800.” “Beaconsfield,” of course, was the home of the Burke family, and it is more than probable that the editor was either Dr. French Laurence or Dr. Walker King. Of the two it is more probable that the editor was Dr. French Laurence as he is known to have been interested in economics. That Laurence was in fact the editor is virtually proved by a remark made in a letter quoted by Arthur Young in his *Autobiography*. The letter is dated November 30, 1800, and refers to Burke’s pamphlet “very lately published by Dr. Lawrence.” Laurence is a very valuable witness because, as I pointed out already, he was one of Burke’s most loyal and intimate disciples and was also interested in economic problems himself.

The preface speaks of Burke’s vast knowledge of the British commercial system. It refers to the great reputation which he quickly won in Parliament as an

---

expert on economic affairs. Yet, it continues, despite his early fame as an econ-
mist, he studied political economy all his life, and made use of every opportunity
which turned up to examine the economic systems of other countries. As a result
of these studies he daily became more convinced that the “unrestrained freedom
of buying and selling is the great animating principle of production and supply.”
But, most important of all, the preface states that Burke was “consulted, and the
greatest deference was paid to his opinions, by Dr. Adam Smith, in the progress of
the celebrated work on the Wealth of Nations.”

Burke’s economic ideas might have been expected to attract the attention of
merchants and, in actual fact, they did. During the course of his career addresses
of thanks poured in on him from nearly all the mercantile cities of Ireland and
England, and he received the freedom of many of the most important centres of
trade. His economic opinions also got him into difficulties, as when he lost his seat
at Bristol largely through advocating Free Trade for Ireland.

It would serve no useful purpose to examine in detail the numerous ad-
dresses of thanks which Burke received. But it is perhaps worth while to have
a glance at one specimen. The following is a portion of an address which the
merchants of Lancaster sent to Burke in June 1766, six months after he entered
Parliament:—

“Sir,

With hearts full of gratitude and respect, we, the merchants of Lan-
caster, beg leave to return you our most sincere thanks for the great attention
you have given to the commercial interest of Great Britain and her colonies,
during the last long and laborious session of parliament, both by removing
obstructions that lay in the way of commerce, and opening new sources of
trade, unknown in former times; from which we have the most sanguine
hopes, that, in future, the manufactures and navigation of this kingdom will be
greatly increased and extended. …”

I think we may safely conclude, therefore, that Burke had a considerable
grasp of the principles of political economy; that he was a Free Trader; that he
arrived at his conclusions independently of the work of Adam Smith; that Adam
Smith respected his opinions on this subject and even consulted him on points;
and that he is therefore entitled to be considered as a pioneer of economic science.
Moreover, by his speeches in Parliament, he introduced the new principles of
economics to the politicians of the day, and the amazingly rapid success of the

26. Ibid., p. VI.
Wealth of Nations may have been, in part, due to the fact that Burke had prepared men’s minds for it.

References

Donal Barrington (1928–2018) was a barrister and judge. He was a member of the Supreme Court of Ireland from 1996 to 2000, and after retirement from the Court he served as the first president of the Irish Human Rights Commission. Barrington studied law at University College Dublin, where he wrote a master’s thesis on Edmund Burke. He was an active entrepreneur in civic life, helping to found the think tank Tuairim in 1954 and the watchdog Irish Council on Civil Liberties in 1976. An obituary in the Irish Times says that during his legal career he “achieved a cast-iron reputation as an advocate for whom constitutional rights, particularly when under threat, were of primary concern.” On the occasion of his being awarded an honorary degree by the National University of Ireland, the introductory speech stated that “His social concern for great principles is matched by his empathy for individuals and their idiosyncrasies.”
Foreword to Burke’s “Thoughts and Details on Scarcity”

Daniel B. Klein

Edmund Burke’s literary executors French Laurence and Walker King issued “Thoughts and Details on Scarcity” in 1800, three years after Burke’s death in 1797. The document—reproduced in entirety here though without their informative Preface—comes principally from a memorial Burke wrote to Prime Minister William Pitt in 1795, but almost half comes from other draft material, intended for the public but to be framed as letters addressed to Burke’s friend Arthur Young.

The material began as a timely warning against interventionist measures in the face of dearth, including a locally administered minimum-wage scheme (referred to as a “tax” by Burke, because employers pay more for labor). But the interpolations from the letters are more of the nature of general political economy. The final document, “Thoughts and Details on Scarcity,” then, is an admixture—“Details,” the more specific facts from the memorial, including testimony of Burke the farmer, which work as illustration of the “Thoughts,” formulated especially in the material that the executors had drawn from the subsequent draft letters.

We have made a few very minor corrections to Burke’s text, and we include most of the footnotes added by Francis Canavan for the Liberty Fund edition (Burke 1999), which relate the text to affairs of the moment. We thank Liberty Fund for their kind permission to reproduce Canavan’s notes.

Why do we draw attention to “Thoughts and Details on Scarcity”?

Adam Smith’s “liberal plan” or “liberal system” (WN, 664, 538–539) is cen-
tered on the idea of liberty, which is a flipside of commutative justice. Liberty is others not messing with one’s stuff, or as Smith puts it “allowing every man to pursue his own interest his own way” (WN, 664; cf. 687). A presumption of liberty is central to the original political meaning of “liberal.”

Presupposed by that liberalism, however, is a stable, integrated, functional polity. Smith’s “science of a legislator” (468) is a philosophy of policy reform, within a settled and integrated system of political authority.

Burke is famous for decrying the French Revolution and worrying about the destabilization of political systems. He often speaks to matters outside of what Smithian liberalism presupposes. Many of Burke’s writings concern not policy reform so much as polity reform. On the basis of his writings in this domain he may aptly be considered conservative—although one should not imagine that Burke favors the conservation of constitutional or fundamental political institutions per se, that is, even terrible ones. And when polity reform seems ripe, he may favor it, as he came to favor letting the American colonies go their own way.

But conservatism in polity reform may coexist with liberalism in policy reform. Burke was indeed a liberal. Burke’s policy sensibilities were liberal, seen throughout his life and career, notably in “Thoughts and Details on Scarcity.”

In philosophy and political ideology, David Hume, Smith, and Burke are three peas in a pod (see, e.g., Miller 1981, 196–203). But Hume and Smith operated in philosophy, speculation, scholarship, science, offering among other things “the science of a legislator.” Burke operated in practical legislating and advising. Burke was “that insidious and crafty animal, vulgarly called a statesman or politician, whose councils are directed by the momentary fluctuations of affairs” (WN, 468). Burke did, of course, also have a foot in philosophy, as well as one in journalism, publishing, and propaganda.

Alexis de Tocqueville (1856, 145) suggested that Britain, especially in contrast to France, was exceptional in the coordination between these two groups of players, the philosophers and the politicos: “In England writers on the theory of government and those who actually governed co-operated with each other, the former setting forth their new theories, the latter amending or circumscribing these in light of practical experience.” The present document by Edmund Burke helps us appreciate the outlook common to liberals at work in both realms, and it might inspire such cooperation today in all countries.

When reading Burke, we should mind whether he is treating polity reform or policy reform, and we should keep track of his multiple roles: Sometimes, somewhat the philosopher who treats what is relatively timeless, but also the politico or publicist, tending the timely and fluctuating.

Even in the latter, though, there’s a timelessness in the manner of his words and deeds.
References

Thoughts and Details on Scarcity

Edmund Burke

Of all things, an indiscreet tampering with the trade of provisions is the most dangerous, and it is always worst in the time when men are most disposed to it—that is, in the time of scarcity. Because there is nothing on which the passions of men are so violent, and their judgment so weak, and on which there exists such a multitude of ill-founded popular prejudices.

The great use of Government is as a restraint; and there is no restraint which it ought to put upon others, and upon itself too, rather than on the fury of speculating under circumstances of irritation. The number of idle tales spread about by the industry of faction, and by the zeal of foolish good-intention, and greedily devoured by the malignant credulity of mankind, tends infinitely to aggravate prejudices, which, in themselves, are more than sufficiently strong. In that state of affairs, and of the publick with relation to them, the first thing that Government owes to us, the people, is information; the next is timely coercion:—the one to guide our judgment; the other to regulate our tempers.

To provide for us in our necessities is not in the power of Government. It would be a vain presumption in statesmen to think they can do it. The people maintain them, and not they the people. It is in the power of Government to prevent much evil; it can do very little positive good in this, or perhaps in any thing else. It is not only so of the state and statesman, but of all the classes and descriptions of the Rich—they are the pensioners of the poor, and are maintained by their superfluity. They are under an absolute, hereditary, and indefeasible dependance on those who labour, and are miscalled the Poor.

The labouring people are only poor, because they are numerous. Numbers in their nature imply poverty. In a fair distribution among a vast multitude, none can have much. That class of dependant pensioners called the rich, is so extremely small, that if all their throats were cut, and a distribution made of all they consume in a year, it would not give a bit of bread and cheese for one night’s supper to those who labour, and who in reality feed both the pensioners and themselves.

But the throats of the rich ought not to be cut, nor their magazines plundered; because, in their persons they are trustees for those who labour, and their hoards are the banking-houses of these latter. Whether they mean it or not, they do, in effect, execute their trust—some with more, some with less fidelity and judgment. But on the whole, the duty is performed, and every thing returns, deducting some very trifling commission and discount, to the place from whence it arose.
When the poor rise to destroy the rich, they act as wisely for their own purposes as when they burn mills, and throw corn into the river, to make bread cheap.

When I say, that we of the people ought to be informed, inclusively I say, we ought not to be flattered: flattery is the reverse of instruction. The poor in that case would be rendered as improvident as the rich, which would not be at all good for them.

Nothing can be so base and so wicked as the political canting language, “The Labouring Poor.” Let compassion be shewn in action, the more the better, according to every man’s ability, but let there be no lamentation of their condition. It is no relief to their miserable circumstances; it is only an insult to their miserable understandings. It arises from a total want of charity, or a total want of thought. Want of one kind was never relieved by want of any other kind. Patience, labour, sobriety, frugality, and religion, should be recommended to them; all the rest is downright fraud. It is horrible to call them “The once happy labourer.”

Whether what may be called moral or philosophical happiness of the laborious classes is increased or not, I cannot say. The seat of that species of happiness is in the mind; and there are few data to ascertain the comparative state of the mind at any two periods. Philosophical happiness is to want little. Civil or vulgar happiness is to want much, and to enjoy much.

If the happiness of the animal man (which certainly goes somewhere towards the happiness of the rational man) be the object of our estimate, then I assert, without the least hesitation, that the condition of those who labour (in all descriptions of labour, and in all gradations of labour, from the highest to the lowest inclusively) is on the whole extremely meliorated, if more and better food is any standard of melioration. They work more, it is certain; but they have the advantage of their augmented labour; yet whether that increase of labour be on the whole a good or an evil, is a consideration that would lead us a great way, and is not for my present purpose. But as to the fact of the melioration of their diet, I shall enter into the detail of proof whenever I am called upon: in the mean time, the known difficulty of contenting them with any thing but bread made of the finest flour, and meat of the first quality, is proof sufficient.

I further assert, that even under all the hardships of the last year, the labouring people did, either out of their direct gains, or from charity, (which it seems is now an insult to them) in fact, fare better than they did, in seasons of common plenty, 50 or 60 years ago; or even at the period of my English observation, which is about 44 years. I even assert, that full as many in that class, as ever were known to do it before, continued to save money; and this I can prove, so far as my own information and experience extend.

It is not true that the rate of wages has not encreased with the nominal price of provisions. I allow it has not fluctuated with that price, nor ought it; and the
Squires of Norfolk had dined, when they gave it as their opinion, that it might or ought to rise and fall with the market of provisions. The rate of wages in truth has no direct relation to that price. Labour is a commodity like every other, and rises or falls according to the demand. This is in the nature of things; however, the nature of things has provided for their necessities. Wages have been twice raised in my time, and they bear a full proportion, or even a greater than formerly, to the medium of provision during the last bad cycle of twenty years. They bear a full proportion to the result of their labour. If we were wildly to attempt to force them beyond it, the stone which we had forced up the hill would only fall back upon them in a diminished demand, or, what indeed is the far lesser evil, an aggravated price of all the provisions, which are the result of their manual toil.

There is an implied contract, much stronger than any instrument or article of agreement, between the labourer in any occupation and his employer—that the labour, so far as that labour is concerned, shall be sufficient to pay to the employer a profit on his capital, and a compensation for his risk; in a word, that the labour shall produce an advantage equal to the payment. Whatever is above that, is a direct tax; and if the amount of that tax be left to the will and pleasure of another, it is an arbitrary tax.

If I understand it rightly, the tax proposed on the farming interest of this kingdom, is to be levied at what is called the discretion of justices of peace.

The questions arising on this scheme of arbitrary taxation are these—Whether it is better to leave all dealing, in which there is no force or fraud, collusion or combination, entirely to the persons mutually concerned in the matter contracted for; or to put the contract into the hands of those, who can have none, or a very remote interest in it, and little or no knowledge of the subject.

It might be imagined that there would be very little difficulty in solving this question; for what man, of any degree of reflection, can think, that a want of interest in any subject closely connected with a want of skill in it, qualifies a person to intermeddle in any the least affair; much less in affairs that vitally concern the agriculture of the kingdom, the first of all its concerns, and the foundation of all its prosperity in every other matter, by which that prosperity is produced?

The vulgar error on this subject arises from a total confusion in the very

---

1. Burke may have mistakenly written Norfolk when he meant Suffolk, where the Justices of the Peace recommended that the wages of laborers should be adjusted in proportion to the price of corn.
2. The reference is to the so-called Speenhamland system, which inspired Burke to write this memorandum to William Pitt. In 1782, Parliament had enacted Gilbert’s Act, which authorized local governments to grant allowances in aid of wages. Subsidizing the wages of the poor was not even then a new departure in English law. On this basis, in 1795 the magistrates of Berkshire, a county adjacent to Burke’s Buckinghamshire, met in the Pelican Inn in Speenhamland, and adopted a scheme to ensure laborers a living wage. A minimum wage was fixed, which varied with the price of corn; if the wages actually paid fell below that, they would be supplemented from the poor rates.
idea of things widely different in themselves;—those of convention, and those of judicature. When a contract is making, it is a matter of discretion and of interest between the parties. In that intercourse, and in what is to arise from it, the parties are the masters. If they are not completely so, they are not free, and therefore their contracts are void.

But this freedom has no farther extent, when the contract is made; then their discretionary powers expire, and a new order of things takes its origin. Then, and not till then, and on a difference between the parties, the office of the judge commences. He cannot dictate the contract. It is his business to see that it be enforced; provided that it is not contrary to pre-existing laws, or obtained by force or fraud. If he is in any way a maker or regulator of the contract, in so much he is disqualified from being a judge. But this sort of confused distribution of administrative and judicial characters, (of which we have already as much as is sufficient, and a little more) is not the only perplexity of notions and passions which trouble us in the present hour.

What is doing, supposes or pretends that the farmer and the labourer have opposite interests;—that the farmer oppresses the labourer; and that a gentleman called a justice of peace, is the protector of the latter, and a controul and restraint on the former; and this is a point I wish to examine in a manner a good deal different from that in which gentlemen proceed, who confide more in their abilities than is fit, and suppose them capable of more than any natural abilities, fed with no other than the provender furnished by their own private speculations, can accomplish. Legislative acts, attempting to regulate this part of economy, do, at least, as much as any other, require the exactest detail of circumstances, guided by the surest general principles that are necessary to direct experiment and enquiry, in order again from those details to elicit principles, firm and luminous general principles, to direct a practical legislative proceeding.

First, then, I deny that it is in this case, as in any other of necessary implication, that contracting parties should originally have had different interests. By accident it may be so undoubtedly at the outset; but then the contract is of the nature of a compromise; and compromise is founded on circumstances that suppose it the interest of the parties to be reconciled in some medium. The principle of compromise adopted, of consequence the interests cease to be different.

But in the case of the farmer and the labourer, their interests are always the same, and it is absolutely impossible that their free contracts can be onerous to either party. It is the interest of the farmer, that his work should be done with effect and celerity: and that cannot be, unless the labourer is well fed, and otherwise found with such necessaries of animal life, according to its habits, as may keep the body in full force, and the mind gay and cheerful. For of all the instruments of his trade, the labour of man (what the ancient writers have called the instrumentum vocale)
is that on which he is most to rely for the re-payment of his capital. The other two, the *semivocale* in the ancient classification, that is, the working stock of cattle, and the *instrumentum mutum*, such as carts, ploughs, spades, and so forth, though not all inconsiderable in themselves, are very much inferior in utility or in expense; and without a given portion of the first, are nothing at all. For in all things whatever, the mind is the most valuable and the most important; and in this scale the whole of agriculture is in a natural and just order; the beast is as an informing principle to the plough and cart; the labourer is as reason to the beast; and the farmer is as a thinking and presiding principle to the labourer. An attempt to break this chain of subordination in any part is equally absurd; but the absurdity is the most mischievous in practical operation, where it is the most easy, that is, where it is the most subject to an erroneous judgment.

It is plainly more the farmer’s interest that his men should thrive, than that his horses should be well fed, sleek, plump, and fit for use, or than that his waggon and ploughs should be strong, in good repair, and fit for service.

On the other hand, if the farmer ceases to profit of the labourer, and that his capital is not continually manured and fructified, it is impossible that he should continue that abundant nutriment, and cloathing, and lodging, proper for the protection of the instruments he employs.

It is therefore the first and fundamental interest of the labourer, that the farmer should have a full incoming profit on the product of his labour. The proposition is self-evident, and nothing but the malignity, perverseness, and ill-governed passions of mankind, and particularly the envy they bear to each other’s prosperity, could prevent their seeing and acknowledging it, with thankfulness to the benign and wise disposer of all things, who obliges men, whether they will or not, in pursuing their own selfish interests, to connect the general good with their own individual success.

But who are to judge what that profit and advantage ought to be? certainly no authority on earth. It is a matter of convention dictated by the reciprocal conveniences of the parties, and indeed by their reciprocal necessities.—But, if the farmer is excessively avaricious?—why so much the better—the more he desires to increase his gains, the more interested is he in the good condition of those, upon whose labour his gains must principally depend.

I shall be told by the zealots of the sect of regulation, that this may be true, and may be safely committed to the convention of the farmer and the labourer, when the latter is in the prime of his youth, and at the time of his health and vigour, and in ordinary times of abundance. But in calamitous seasons, under accidental illness, in declining life, and with the pressure of a numerous offspring, the future nourishers of the community but the present drains and blood-suckers of those who produce them, what is to be done? When a man cannot live and maintain his
family by the natural hire of his labour, ought it not to be raised by authority?

On this head I must be allowed to submit, what my opinions have ever been; and somewhat at large.

And, first, I premise that labour is, as I have already intimated, a commodity, and as such, an article of trade. If I am right in this notion, then labour must be subject to all the laws and principles of trade, and not to regulations foreign to them, and that may be totally inconsistent with those principles and those laws. When any commodity is carried to market, it is not the necessity of the vender, but the necessity of the purchaser that raises the price. The extreme want of the seller has rather (by the nature of things with which we shall in vain contend) the direct contrary operation. If the goods at market are beyond the demand, they fall in their value; if below it, they rise. The impossibility of the subsistence of a man, who carries his labour to a market, is totally beside the question in this way of viewing it. The only question is, what is it worth to the buyer?

But if authority comes in and forces the buyer to a price, who is this in the case (say) of a farmer, who buys the labour of ten or twelve labouring men, and three or four handycrafts, what is it, but to make an arbitrary division of his property among them?

The whole of his gains, I say it with the most certain conviction, never do amount any thing like in value to what he pays to his labourers and artificers; so that a very small advance upon what one man pays to many, may absorb the whole of what he possesses, and amount to an actual partition of all his substance among them. A perfect equality will indeed be produced;—that is to say, equal want, equal wretchedness, equal beggary, and on the part of the partitioners, a woeful, helpless, and desperate disappointment. Such is the event of all compulsory equalizations. They pull down what is above. They never raise what is below: and they depress high and low together beneath the level of what was originally the lowest.

If a commodity is raised by authority above what it will yield with a profit to the buyer, that commodity will be the less dealt in. If a second blundering interposition be used to correct the blunder of the first, and an attempt is made to force the purchase of the commodity (of labour for instance), then one of these two things must happen, either that the forced buyer is ruined, or the price of the product of the labour, in that proportion, is raised. Then the wheel turns round, and the evil complained of falls with aggravated weight on the complainant. The price of corn, which is the result of the expence of all the operations of husbandry, taken together, and for some time continued, will rise on the labourer, considered as a consumer. The very best will be, that he remains where he was. But if the price of the corn should not compensate the price of labour, what is far more to be feared, the most serious evil, the very destruction of agriculture itself, is to be apprehended.

Nothing is such an enemy to accuracy of judgment as a coarse discrimina-
tion; a want of such classification and distribution as the subject admits of. En-
crease the rate of wages to the labourer, say the regulators—as if labour was but one
thing and of one value. But this very broad generic term, labour, admits, at least, of
two or three specific descriptions: and these will suffice, at least, to let gentlemen
discern a little the necessity of proceeding with caution in their coercive guidance
of those whose existence depends upon the observance of still nicer distinctions
and sub-divisions, than commonly they resort to in forming their judgments on this
very enlarged part of economy.

The labourers in husbandry may be divided: 1st. into those who are able to
perform the full work of a man; that is, what can be done by a person from twenty-
one years of age to fifty. I know no husbandry work (mowing hardly excepted) that
is not equally within the power of all persons within those ages, the more advanced
fully compensating by knack and habit what they lose in activity. Unquestionably,
there is a good deal of difference between the value of one man’s labour and that of
another, from strength, dexterity, and honest application. But I am quite sure, from
my best observation, that any given five men will, in their total, afford a proportion
of labour equal to any other five within the periods of life I have stated; that is,
that among such five men there will be one possessing all the qualifications of a
good workman, one bad, and the other three middling, and approximating to the
first and the last. So that in so small a platoon as that of even five, you will find
the full complement of all that five men can earn. Taking five and five throughout
the kingdom, they are equal: therefore, an error with regard to the equalization of
their wages by those who employ five, as farmers do at the very least, cannot be
considerable.

2dly. Those who are able to work, but not the complete task of a day-
labourer. This class is infinitely diversified, but will aptly enough fall into principal
divisions. Men, from the decline, which after fifty becomes every year more sensi-
tible, to the period of debility and decrepitude, and the maladies that precede a
final dissolution. Women, whose employment on husbandry is but occasional, and
who differ more in effective labour one from another than men do, on account of
gestation, nursing, and domestic management, over and above the difference they
have in common with men in advancing, in stationary, and in declining life. Children,
who proceed on the reverse order, growing from less to greater utility, but with a
still greater disproportion of nutriment to labour than is found in the second of
these sub-divisions; as is visible to those who will give themselves the trouble of
examining into the interior economy of a poor-house.

This inferior classification is introduced to shew, that laws prescribing, or
magistrates exercising, a very stiff, and often inapplicable rule, or a blind and rash
discretion, never can provide the just proportions between earning and salary on
the one hand, and nutriment on the other: whereas interest, habit, and the tacit
convention, that arise from a thousand nameless circumstances, produce a *fact* that regulates without difficulty, what laws and magistrates cannot regulate at all. The first class of labour wants nothing to equalize it; it equalizes itself. The second and third are not capable of any equalization.

But what if the rate of hire to the labourer comes far short of his necessary subsistence, and the calamity of the time is so great as to threaten actual famine? Is the poor labourer to be abandoned to the flinty heart and griping hand of base self-interest, supported by the sword of law, especially when there is reason to suppose that the very avarice of farmers themselves has concurred with the errors of Government to bring famine on the land.

In that case, my opinion is this. Whenever it happens that a man can claim nothing according to the rules of commerce, and the principles of justice, he passes out of that department, and comes within the jurisdiction of mercy. In that province the magistrate has nothing at all to do: his interference is a violation of the property which it is his office to protect. Without all doubt, charity to the poor is a direct and obligatory duty upon all Christians, next in order after the payment of debts, full as strong, and by nature made infinitely more delightful to us. Pufendorf, and other casuists do not, I think, denominate it quite properly, when they call it a duty of imperfect obligation. But the manner, mode, time, choice of objects, and proportion, are left to private discretion; and perhaps, for that very reason it is performed with the greater satisfaction, because the discharge of it has more the appearance of freedom; recommending us besides very specially to the divine favour, as the exercise of a virtue most suitable to a being sensible of its own infirmity.

The cry of the people in cities and towns, though unfortunately (from a fear of their multitude and combination) the most regarded, ought, in *fact*, to be the *least* attended to upon this subject; for citizens are in a state of utter ignorance of the means by which they are to be fed, and they contribute little or nothing, except in an infinitely circuitous manner, to their own maintenance. They are truly “*Fruges consumere nati.*” They are to be heard with great respect and attention upon matters within their province, that is, on trades and manufactures; but on any thing that relates to agriculture, they are to be listened to with the same *reverence* which we pay to the dogmas of other ignorant and presumptuous men.

If any one were to tell them, that they were to give in an account of all the stock in their shops; that attempts would be made to limit their profits, or raise the price of the labouring manufacturers upon them, or recommend to Government, out of a capital from the publick revenues, to set up a shop of the same commodities, in order to rival them, and keep them to reasonable dealing, they would

---

3. “Born to consume the fruits [of the earth].” Horace *Epistles* 1.2.27.
very soon see the impudence, injustice, and oppression of such a course. They would not be mistaken; but they are of opinion, that agriculture is to be subject to other laws, and to be governed by other principles.

A greater and more ruinous mistake cannot be fallen into, than that the trades of agriculture and grazing can be conducted upon any other than the common principles of commerce; namely, that the producer should be permitted, and even expected, to look to all possible profit which, without fraud or violence, he can make; to turn plenty or scarcity to the best advantage he can; to keep back or to bring forward his commodities at his pleasure; to account to no one for his stock or for his gain. On any other terms he is the slave of the consumer; and that he should be so is of no benefit to the consumer. No slave was ever so beneficial to the master as a freeman that deals with him on an equal footing by convention, formed on the rules and principles of contending interests and compromised advantages. The consumer, if he were suffered, would in the end always be the dupe of his own tyranny and injustice. The landed gentleman is never to forget, that the farmer is his representative.

It is a perilous thing to try experiments on the farmer. The farmer’s capital (except in a few persons, and in a very few places) is far more feeble than commonly is imagined. The trade is a very poor trade; it is subject to great risks and losses. The capital, such as it is, is turned but once in the year; in some branches it requires three years before the money is paid. I believe never less than three in the turnip and grass-land course, which is the prevalent course on the more or less fertile, sandy and gravelly loams, and these compose the soil in the south and south-east of England, the best adapted, and perhaps the only ones that are adapted, to the turnip husbandry.

It is very rare that the most prosperous farmer, counting the value of his quick and dead stock, the interest of the money he turns, together with his own wages as a bailiff or overseer, ever does make twelve or fifteen per centum by the year on his capital. I speak of the prosperous. In most of the parts of England which have fallen within my observation, I have rarely known a farmer, who to his own trade has not added some other employment or traffic, that, after a course of the most unremitting parsimony and labour (such for the greater part is theirs), and persevering in his business for a long course of years, died worth more than paid his debts, leaving his posterity to continue in nearly the same equal conflict between industry and want, in which the last predecessor, and a long line of predecessors before him, lived and died.

Observe that I speak of the generality of farmers who have not more than from one hundred and fifty to three or four hundred acres. There are few in this part of the country within the former, or much beyond the latter, extent. Unquestionably in other places there are much larger. But, I am convinced, whatever
part of England be the theatre of his operations, a farmer who cultivates twelve hundred acres, which I consider as a large farm, though I know there are larger, cannot proceed, with any degree of safety and effect, with a smaller capital than ten thousand pounds; and that he cannot, in the ordinary course of culture, make more upon that great capital of ten thousand pounds, than twelve hundred a year.

As to the weaker capitals, an easy judgment may be formed by what very small errors they may be farther attenuated, enervated, rendered unproductive, and perhaps totally destroyed.

This constant precariousness and ultimate moderate limits of a farmer’s fortune, on the strongest capital, I press, not only on account of the hazardous speculations of the times, but because the excellent and most useful works of my friend, Mr. Arthur Young, tend to propagate that error (such I am very certain it is), of the largeness of a farmer’s profits. It is not that his account of the produce does often greatly exceed, but he by no means makes the proper allowance for accidents and losses. I might enter into a convincing detail, if other more troublesome and more necessary details were not before me.

This proposed discretionary tax on labour militates with the recommendations of the Board of Agriculture: they recommend a general use of the drill culture. I agree with the Board, that where the soil is not excessively heavy, or incumbered with large loose stones (which however is the case with much otherwise good land), that course is the best, and most productive, provided that the most accurate eye; the most vigilant superintendance; the most prompt activity, which has no such day as to-morrow in its calendar; the most steady foresight and pre-disposing order to have every body and every thing ready in its place, and prepared to take advantage of the fortunate fugitive moment in this coquetting climate of ours—provided, I say, all these combine to speed the plough, I admit its superiority over the old and general methods. But under procrastinating, improvident, ordinary husbandmen, who may neglect or let slip the few opportunities of sweetening and purifying their ground with perpetually renovated toil, and undissipated attention, nothing, when tried to any extent, can be worse, or more dangerous: the farm may be ruined, instead of having the soil enriched and sweetened by it.

But the excellence of the method on a proper soil, and conducted by an husbandman, of whom there are few, being readily granted, how, and on what conditions, is this culture obtained? Why, by a very great encrease of labour; by an augmentation of the third part, at least, of the hand-labour, to say nothing of the horses and machinery employed in ordinary tillage. Now, every man must be sensible how little becoming the gravity of Legislature it is to encourage a Board,

4. To drill is to sow seeds or seedlings along a shallow furrow.
which recommends to us, and upon very weighty reasons unquestionably, an 
enlargement of the capital we employ in the operations of the land, and then to pass 
an act which taxes that manual labour, already at a very high rate; thus compelling 
us to diminish the quantity of labour which in the vulgar course we actually employ.

What is true of the farmer is equally true of the middle man; whether the 
middle man acts as factor, 5 jobber, salesman, or speculator, in the markets of grain. 
These traders are to be left to their free course; and the more they make, and the 
richer they are, and the more largely they deal, the better both for the farmer and consumer, between whom they form a natural and most useful link of connection; though, by the machinations of the old evil counsellor, Envy, they are hated and 
maligned by both parties.

I hear that middle men are accused of monopoly. Without question, the 
monopoly of authority is, in every instance and in every degree, an evil; but the 
monopoly of capital is the contrary. It is a great benefit, and a benefit particularly 
to the poor. A tradesman who has but a hundred pound capital, which (say) he can 
turn but once a year, cannot live upon a profit of 10 per cent. because he cannot live 
upon ten pounds a year; but a man of ten thousand pounds capital can live and 
thrive upon 5 per cent. profit in the year, because he has five hundred pounds a year. 
The same proportion holds in turning it twice or thrice. These principles are plain 
and simple; and it is not our ignorance, so much as the levity, the envy, and the 
malignity of our nature, that hinders us from perceiving and yielding to them: but 
we are not to suffer our vices to usurp the place of our judgment.

The balance between consumption and production makes price. The market 
settles, and alone can settle, that price. Market is the meeting and conference of the consumer and producer, when they mutually discover each other’s wants. Nobody, I 
believe, has observed with any reflection what market is, without being astonished 
at the truth, the correctness, the celerity, the general equity, with which the balance 
of wants is settled. They who wish the destruction of that balance, and would 
fain by arbitrary regulation decree, that defective production should not be 
compensated by increased price, directly lay their axe to the root of production 
itself.

They may even in one year of such false policy, do mischiefs incalculable; 
because the trade of a farmer is, as I have before explained, one of the most 
precarious in its advantages, the most liable to losses, and the least profitable of any 
that is carried on. It requires ten times more of labour, of vigilance, of attention, 
of skill, and let me add, of good fortune also, to carry on the business of a farmer 
with success, than what belongs to any other trade. Seeing things in this light, I

---

5. One who acts for another as an agent, deputy, or representative; more narrowly, an agent who buys or 
sells for another; a commission merchant.
am far from presuming to censure the late circular instruction of Council to Lord Lieutenants— but I confess I do not clearly discern its object. I am greatly afraid that the enquiry will raise some alarm as a measure, leading to the French system of putting corn into requisition. For that was preceded by an inquisition somewhat similar in its principle, though, according to their mode, their principles are full of that violence, which here is not much to be feared. It goes on a principle directly opposite to mine: it presumes, that the market is no fair test of plenty or scarcity. It raises a suspicion, which may affect the tranquillity of the public mind, “that the farmer keeps back, and takes unfair advantages by delay;” on the part of the dealer, it gives rise obviously to a thousand nefarious speculations.

In case the return should on the whole prove favourable, is it meant to ground a measure for encouraging exportation and checking the import of corn? If it is not, what end can it answer? And, I believe, it is not.

This opinion may be fortified by a report gone abroad, that intentions are entertained of erecting public granaries, and that this enquiry is to give Government an advantage in its purchases.

I hear that such a measure has been proposed, and is under deliberation, that is, for Government to set up a granary in every market town, at the expense of the state, in order to extinguish the dealer, and to subject the farmer to the consumer, by securing corn to the latter at a certain and steady price.

If such a scheme is adopted, I should not like to answer for the safety of the granary, of the agents, or of the town itself, in which the granary was erected—the first storm of popular phrenzy would fall upon that granary.

So far in a political light.

In an economical light, I must observe, that the construction of such granaries throughout the kingdom, would be at an expense beyond all calculation. The keeping them up would be at a great charge. The management and attendance would require an army of agents, store-keepers, clerks, and servants. The capital to be employed in the purchase of grain would be enormous. The waste, decay, and corruption, would be a dreadful drawback on the whole dealing; and the dissatisfaction of the people, at having decayed, tainted, or corrupted corn sold to them, as must be the case, would be serious.

This climate (whatever others may be) is not favourable to granaries, where wheat is to be kept for any time. The best, and indeed the only good granary, is the rick-yard of the farmer, where the corn is preserved in its own straw, sweet, clean, wholesome, free from vermin and from insects, and comparatively at a trifle of expense. This, with the barn, enjoying many of the same advantages, have been

6. A circular letter sent by the Council through the Home Secretary to the Lords Lieutenant asking them to hold magistrates’ meetings in their counties to ascertain the produce of the recent harvest.
the sole granaries of England from the foundation of its agriculture to this day. All this is done at the expense of the undertaker, and at his sole risk. He contributes to Government; he receives nothing from it but protection; and to this he has a claim.

The moment that Government appears at market, all the principles of market will be subverted. I don’t know whether the farmer will suffer by it, as long as there is a tolerable market of competition; but I am sure that, in the first place, the trading government will speedily become a bankrupt, and the consumer in the end will suffer. If Government makes all its purchases at once, it will instantly raise the market upon itself. If it makes them by degrees, it must follow the course of the market. If it follows the course of the market, it will produce no effect, and the consumer may as well buy as he wants—therefore all the expense is incurred gratis.

But if the object of this scheme should be, what I suspect it is, to destroy the dealer, commonly called the middle man, and by incurring a voluntary loss to carry the baker to deal with Government, I am to tell them that they must set up another trade, that of a miller or a mealman, attended with a new train of expenses and risks. If in both these trades they should succeed, so as to exclude those who trade on natural and private capitals, then they will have a monopoly in their hands, which, under the appearance of a monopoly of capital, will, in reality, be a monopoly of authority, and will ruin whatever it touches. The agriculture of the kingdom cannot stand before it.

A little place like Geneva, of not more than from twenty-five to thirty thousand inhabitants, which has no territory, or next to none; which depends for its existence on the good-will of three neighbouring powers, and is of course continually in the state of something like a siege, or in the speculation of it, might find some resource in state granaries, and some revenue from the monopoly of what was sold to the keepers of public-houses. This is a policy for a state too small for agriculture. It is not (for instance) fit for so great a country as the Pope possesses, where, however, it is adopted and pursued in a greater extent, and with more strictness. Certain of the Pope’s territories, from whence the city of Rome is supplied, being obliged to furnish Rome and the granaries of his Holiness with corn at a certain price, that part of the papal territories is utterly ruined. That ruin may be traced with certainty to this sole cause, and it appears indubitably by a comparison of their state and condition with that of the other part of the ecclesiastical dominions not subjected to the same regulations, which are in circumstances highly flourishing.

The reformation of this evil system is in a manner impracticable; for, first, it does keep bread and all other provisions equally subject to the chamber of supply, at a pretty reasonable and regular price, in the city of Rome. This preserves quiet among the numerous poor, idle, and naturally mutinous people, of a very great capital. But the quiet of the town is purchased by the ruin of the country, and the
ultimate wretchedness of both. The next cause which renders this evil incurable, is, the jobs which have grown out of it, and which, in spite of all precautions, would grow out of such things, even under governments far more potent than the feeble authority of the Pope.

This example of Rome which has been derived from the most ancient times, and the most flourishing period of the Roman empire (but not of the Roman agriculture) may serve as a great caution to all Governments, not to attempt to feed the people out of the hands of the magistrates. If once they are habituated to it, though but for one half-year, they will never be satisfied to have it otherwise. And, having looked to Government for bread, on the very first scarcity they will turn and bite the hand that fed them. To avoid that evil, Government will redouble the causes of it; and then it will become inveterate and incurable.

I beseech the Government (which I take in the largest sense of the word, comprehending the two Houses of Parliament) seriously to consider that years of scarcity or plenty, do not come alternately or at short intervals, but in pretty long cycles and irregularly, and consequently that we cannot assure ourselves, if we take a wrong measure, from the temporary necessities of one season; but that the next, and probably more, will drive us to the continuance of it; so that in my opinion, there is no way of preventing this evil which goes to the destruction of all our agriculture, and of that part of our internal commerce which touches our agriculture the most nearly, as well as the safety and very being of Government, but manfully to resist the very first idea, speculative or practical, that it is within the competence of Government, taken as Government, or even of the rich, as rich, to supply to the poor, those necessaries which it has pleased the Divine Providence for a while to withhold from them. We, the people, ought to be made sensible, that it is not in breaking the laws of commerce, which are the laws of nature, and consequently the laws of God, that we are to place our hope of softening the Divine displeasure to remove any calamity under which we suffer, or which hangs over us.

So far as to the principles of general policy.

As to the state of things which is urged as a reason to deviate from them, these are the circumstances of the harvest of 1795 and 1794. With regard to the harvest of 1794, in relation to the noblest grain, wheat, it is allowed to have been somewhat short, but not excessively; and in quality, for the seven and twenty years, during which I have been a farmer, I never remember wheat to have been so good. The world were, however, deceived in their speculations upon it—the farmer as well as the dealer. Accordingly the price fluctuated beyond any thing I can remember; for, at one time of the year, I sold my wheat at 14l. a load, (I sold off all I had, as I thought this was a reasonable price), when at the end of the season, if I had then had any to sell, I might have got thirty guineas for the same sort of grain. I sold all that I had, as I said, at a comparatively low price, because I thought it a good
price, compared with what I thought the general produce of the harvest; but when I came to consider what my own total was, I found that the quantity had not answered my expectation. It must be remembered, that this year of produce, (the year 1794) short, but excellent, followed a year which was not extraordinary in production, nor of a superior quality, and left but little in store. At first this was not felt, because the harvest came in unusually early—earlier than common, by a full month.

The winter, at the end of 1794, and beginning of 1795, was more than usually unfavourable both to corn and grass, owing to the sudden relaxation of very rigorous frosts, followed by rains, which were again rapidly succeeded by frosts of still greater rigour than the first.

Much wheat was utterly destroyed. The clover grass suffered in many places. What I never observed before, the rye-grass, or coarse bent, suffered more than the clover. Even the meadow-grass in some places was killed to the very roots. In the spring, appearances were better than we expected. All the early sown grain recovered itself, and came up with great vigour; but that, which was late sown, was feeble, and did not promise to resist any blights, in the spring, which, however, with all its unpleasant vicissitudes passed off very well; and nothing looked better than the wheat at the time of blooming—but at that most critical time of all, a cold dry east wind, attended with very sharp frosts, longer and stronger than I recollect at that time of year, destroyed the flowers, and withered up, in an astonishing manner, the whole side of the ear next to the wind. At that time I brought to town some of the ears, for the purpose of shewing to my friends the operation of those unnatural frosts, and according to their extent I predicted a great scarcity. But such is the pleasure of agreeable prospects, that my opinion was little regarded.

On threshing, I found things as I expected—the ears not filled, some of the capsules quite empty, and several others containing only withered hungry grain, inferior to the appearance of rye. My best ears and grains were not fine; never had I grain of so low a quality—yet I sold one load for 21l. At the same time I bought my seed wheat (it was excellent) at 23l. Since then the price has risen, and I have sold about two load of the same sort at 23l. Such was the state of the market when I left home last Monday. Little remains in my barn. I hope some in the rick may be better; since it was earlier sown, as well as I can recollect. Some of my neighbours have better, some quite as bad, or even worse. I suspect it will be found, that wherever the blighting wind and those frosts at blooming time have prevailed, the produce of the wheat crop will turn out very indifferent. Those parts which have escaped, will, I can hardly doubt, have a reasonable produce.

As to the other grains, it is to be observed, as the wheat ripened very late, (on account, I conceive, of the blights) the barley got the start of it, and was ripe first. The crop was with me, and wherever my enquiry could reach, excellent; in some places far superior to mine.
The clover, which came up with the barley, was the finest I remember to have seen.

The turnips of this year are generally good.

The clover sown last year, where not totally destroyed, gave two good crops, or one crop and a plentiful feed; and, bating the loss of the rye-grass, I do not remember a better produce.

The meadow-grass yielded but a middling crop, and neither of the sown or natural grass was there in any farmer’s possession any remainder from the year worth taking into account. In most places, there was none at all.

The meadow-grass yielded but a middling crop, and neither of the sown or natural grass was there in any farmer’s possession any remainder from the year worth taking into account. In most places, there was none at all.

Oats with me were not in a quantity more considerable than in commonly good seasons; but I have never known them heavier, than they were in other places. The oat was not only an heavy, but an uncommonly abundant crop. My ground under pease did not exceed an acre, or thereabouts, but the crop was great indeed. I believe it is throughout the country exuberant.

It is however to be remarked, that as generally of all the grains, so particularly of the pease, there was not the smallest quantity in reserve.

The demand of the year must depend solely on its own produce; and the price of the spring-corn is not to be expected to fall very soon, or at any time very low.

Uxbridge is a great corn market. As I came through that town, I found that at the last market-day, barley was at forty shillings a quarter; oats there were literally none; and the innkeeper was obliged to send for them to London. I forgot to ask about pease. Potatoes were 5s. the bushel.

In the debate on this subject in the House, I am told that a leading member of great ability, little conversant in these matters, observed, that the general uniform dearness of butcher’s meat, butter, and cheese, could not be owing to a defective produce of wheat; and on this ground insinuated a suspicion of some unfair practice on the subject, that called for enquiry.

Unquestionably the mere deficiency of wheat could not cause the dearness of the other articles, which extends not only to the provisions he mentioned, but to every other without exception.

The cause is indeed so very plain and obvious, that the wonder is the other way. When a properly directed enquiry is made, the gentlemen who are amazed at the price of these commodities will find, that when hay is at six pound a load, as they must know it is, herbage, and for more than one year, must be scanty, and they will conclude, that if grass be scarce, beef, veal, mutton, butter, milk, and cheese, must be dear.

But to take up the matter somewhat more in detail—if the wheat harvest in 1794, excellent in quality, was defective in quantity, the barley harvest was in quality ordinary enough; and in quantity deficient. This was soon felt in the price of malt.

Another article of produce (beans) was not at all plentiful. The crop of pease
was wholly destroyed, so that several farmers pretty early gave up all hopes on that head, and cut the green haulm as fodder for the cattle, then perishing for want of food in that dry and burning summer. I myself came off better than most—I had about the fourth of a crop of pease.

It will be recollected, that, in a manner, all the bacon and pork consumed in this country, (the far largest consumption of meat out of towns) is, when growing, fed on grass, and on whey, or skimmed milk; and when fatting, partly on the latter. This is the case in the dairy countries, all of them great breeders and feeders of swine; but for the much greater part, and in all the corn countries, they are fattened on beans, barley meal, and pease. When the food of the animal is scarce, his flesh must be dear. This, one would suppose, would require no great penetration to discover.

This failure of so very large a supply of flesh in one species, naturally throws the whole demand of the consumer on the diminished supply of all kinds of flesh, and, indeed, on all the matters of human sustenance. Nor, in my opinion, are we to expect a greater cheapness in that article for this year, even though corn should grow cheaper, as it is to be hoped it will. The store swine, from the failure of subsistence last year, are now at an extravagant price. Pigs, at our fairs, have sold lately for fifty shillings, which, two years ago, would not have brought more than twenty.

As to sheep, none, I thought, were strangers to the general failure of the article of turnips last year; the early having been burned as they came up, by the great drought and heat; the late, and those of the early which had escaped, were destroyed by the chilling frosts of the winter, and the wet and severe weather of the spring. In many places a full fourth of the sheep or the lambs were lost, what remained of the lambs were poor and ill-fed, the ewes having had no milk. The calves came late, and they were generally an article, the want of which was as much to be dreaded as any other. So that article of food, formerly so abundant in the early part of the summer, particularly in London, and which in a great part supplied the place of mutton for near two months, did little less than totally fail.

All the productions of the earth link in with each other. All the sources of plenty, in all and every article, were dried or frozen up. The scarcity was not as gentlemen seem to suppose, in wheat only.

Another cause, and that not of inconsiderable operation, tended to produce a scarcity in flesh provision. It is one that on many accounts cannot be too much regretted, and, the rather, as it was the sole cause of scarcity in that article, which arose from the proceedings of men themselves. I mean the stop put to the distillery.

The hogs (and that would be sufficient) which were fed with the waste wash of that produce, did not demand the fourth part of the corn used by farmers in fattening them. The spirit was nearly so much clear gain to the nation. It is an odd
way of making flesh cheap, to stop or check the distillery.

The distillery in itself produces an immense article of trade almost all over the world, to Africa, to North America, and to various parts of Europe. It is of great use, next to food itself, to our fisheries and to our whole navigation. A great part of the distillery was carried on by damaged corn, unfit for bread, and by barley and malt of the lowest quality. These things could not be more unexceptionably employed. The domestic consumption of spirits, produced, without complaints, a very great revenue, applicable, if we pleased, in bounties to the bringing corn from other places, far beyond the value of that consumed in making it, or to the encouragement of its increased production at home.

As to what is said, in a physical and moral view, against the home consumption of spirits, experience has long since taught me very little to respect the declamations on that subject—whether the thunder of the laws, or the thunder of eloquence, “is hurled on gin,” always I am thunder-proof. The alembic, in my mind, has furnished to the world a far greater benefit and blessing, than if the opus maximum had been really found by chemistry, and, like Midas, we could turn every thing into gold.

Undoubtedly there may be a dangerous abuse in the excess of spirits; and at one time I am ready to believe the abuse was great. When spirits are cheap, the business of drunkenness is achieved with little time or labour; but that evil I consider to be wholly done away. Observation for the last forty years, and very particularly for the last thirty, has furnished me with ten instances of drunkenness from other causes, for one from this. Ardent spirit is a great medicine, often to remove distempers—much more frequently to prevent them, or to chase them away in their beginnings. It is not nutritive in any great degree. But, if not food, it greatly alleviates the want of it. It invigorates the stomach for the digestion of poor meagre diet, not easily alliable to the human constitution. Wine the poor cannot touch. Beer, as applied to many occasions, (as among seamen and fishermen for instance) will by no means do the business. Let me add, what wits inspired with champaign and claret, will turn into ridicule—it is a medicine for the mind. Under the pressure of the cares and sorrows of our mortal condition, men have at all times,

---

7. See Alexander Pope’s *Epilogue to the Satires*, Dialogue 1, lines 129–31, deploiring the presumptuousness of the lower classes in imitating the vices of their social superiors:

This, this, my friend, I cannot, must not hear;
Vice thus abused demands a nation’s care;
This calls the Church to deprecate our sin,
And hurls the thunder of the laws on gin.

8. An apparatus used in distilling spirits.
9. The greatest work or art, that of realizing alchemy’s dream of turning base metals into gold.
and in all countries, called in some physical aid to their moral consolations,—wine, beer, opium, brandy, or tobacco.

I consider therefore the stopping of the distillery, æconomically, financially, commercially, medicinally, and in some degree morally too, as a measure rather well meant than well considered. It is too precious a sacrifice to prejudice.

Gentlemen well know whether there be a scarcity of partridges, and whether that be an effect of hoarding and combination. All the tame race of birds live and die as the wild do.

As to the lesser articles, they are like the greater. They have followed the fortune of the season. Why are fowls dear? was not this the farmer’s or jobber’s fault. I sold from my yard to a jobber, six young and lean fowls, for four and twenty shillings; fowls, for which, two years ago, the same man would not have given a shilling a-piece.—He sold them afterwards at Uxbridge, and they were taken to London to receive the last hand.

As to the operation of the war in causing the scarcity of provisions, I understand that Mr. Pitt has given a particular answer to it—but I do not think it worth powder and shot.

I do not wonder the papers are so full of this sort of matter, but I am a little surprised it should be mentioned in parliament. Like all great state questions, peace and war may be discussed, and different opinions fairly formed, on political grounds, but on a question of the present price of provisions, when peace with the regicides is always uppermost, I can only say, that great is the love of it.

After all, have we not reason to be thankful to the giver of all good? In our history, and when “The labourer of England is said to have been once happy,” we find constantly, after certain intervals, a period of real famine; by which, a melancholy havock was made among the human race. The price of provisions fluctuated dreadfully, demonstrating a deficiency very different from the worst failures of the present moment. Never since I have known England, have I known more than a comparative scarcity. The price of wheat, taking a number of years together, has had no very considerable fluctuation, nor has it risen exceedingly until within this twelvemonth. Even now, I do not know of one man, woman, or child, that has perished from famine; fewer, if any, I believe, than in years of plenty, when such a thing may happen by accident. This is owing to a care and superintendance of the poor, far greater than any I remember.

The consideration of this ought to bind us all, rich and poor together, against those wicked writers of the newspapers, who would inflame the poor against their friends, guardians, patrons, and protectors. Not only very few (I have observed, that I know of none, though I live in a place as poor as most) have actually died of want, but we have seen no traces of those dreadful exterminating epidemics, which, in consequence of scanty and unwholesome food, in former times, not
unfrequently, wasted whole nations. Let us be saved from too much wisdom of our own, and we shall do tolerably well.

It is one of the finest problems in legislation, and what has often engaged my thoughts whilst I followed that profession, “What the State ought to take upon itself to direct by the public wisdom, and what it ought to leave, with as little interference as possible, to individual discretion.” Nothing, certainly, can be laid down on the subject that will not admit of exceptions, many permanent, some occasional. But the clearest line of distinction which I could draw, whilst I had my chalk to draw any line, was this: That the State ought to confine itself to what regards the State, or the creatures of the State, namely, the exterior establishment of its religion; its magistracy; its revenue; its military force by sea and land; the corporations that owe their existence to its fiat; in a word, to every thing that is truly and properly public, to the public peace, to the public safety, to the public order, to the public prosperity. In its preventive police it ought to be sparing of its efforts, and to employ means, rather few, unfrequent, and strong, than many, and frequent, and, of course, as they multiply their puny politic race, and dwindle, small and feeble. Statesmen who know themselves will, with the dignity which belongs to wisdom, proceed only in this the superior orb and first mover of their duty, steadily, vigilantly, severely, courageously: whatever remains will, in a manner, provide for itself. But as they descend from the state to a province, from a province to a parish, and from a parish to a private house, they go on accelerated in their fall. They cannot do the lower duty; and, in proportion as they try it, they will certainly fail in the higher. They ought to know the different departments of things; what belongs to laws, and what manners alone can regulate. To these, great politicians may give a leaning, but they cannot give a law.

Our Legislature has fallen into this fault as well as other governments; all have fallen into it more or less. The once mighty State, which was nearest to us locally, nearest to us in every way, and whose ruins threaten to fall upon our heads, is a strong instance of this error. 10 I can never quote France without a foreboding sigh—ΕΣΣΕΤΑΙ’ΗΜΑΡ! 11 Scipio said it to his recording Greek friend amidst the flames of the great rival of his country. 12 That state has fallen by the hands of

10. That of the fixing of prices by government.
11. The first words of a passage in Homer (Iliad 6.448–49), in which Hector tells his wife that he knows that Troy is doomed:

The day will come when sacred Troy will perish,
And Priam and his people shall be slain.

12. The Roman general Scipio, who had finally and fully conquered Carthage, repeated Hector’s words, “The day will come,” when his Greek friend, the historian Polybius, asked him why he wept when he saw Carthage in flames. He feared for Rome, too, says Polybius, “when he reflected on the fate of all things
the parricides of their country, called the Revolutionists, and Constitutionalists, of France, a species of traitors, of whose fury and atrocious wickedness nothing in the annals of the phrenzy and depravation of mankind had before furnished an example, and of whom I can never think or speak without a mixed sensation of disgust, of horror, and of detestation, not easy to be expressed. These nefarious monsters destroyed their country for what was good in it: for much good there was in the constitution of that noble monarchy, which, in all kinds, formed and nourished great men, and great patterns of virtue to the world. But though its enemies were not enemies to its faults, its faults furnished them with means for its destruction. My dear departed friend,13 whose loss is even greater to the public than to me, had often remarked, that the leading vice of the French monarchy (which he had well studied) was in good intention ill-directed, and a restless desire of governing too much. The hand of authority was seen in every thing, and in every place. All, therefore, that happened amiss in the course even of domestic affairs, was attributed to the Government; and, as it always happens in this kind of officious universal interference, what began in odious power, ended always, I may say without an exception, in contemptible imbecility. For this reason, as far as I can approve of any novelty, I thought well of the Provincial Administrations. Those, if the superior power had been severe, and vigilant, and vigorous, might have been of much use politically in removing government from many invidious details. But as every thing is good or bad, as it is related or combined, government being relaxed above as it was relaxed below, and the brains of the people growing more and more addle with every sort of visionary speculation, the shiftings of the scene in the provincial theatres became only preparatives to a revolution in the kingdom, and the popular actings there only the rehearsals of the terrible drama of the republic.

Tyranny and cruelty may make men justly wish the downfall of abused powers, but I believe that no government ever yet perished from any other direct cause than its own weakness. My opinion is against an over-doing of any sort of administration, and more especially against this most momentous of all meddling on the part of authority; the meddling with the subsistence of the people.

FINIS

human.” Histories 38.22.1–3.
13. His son Richard, who had died the year before, on August 2, 1794.
Edmund Burke (1729–1797) was an author, statesman, and publicist, born in Dublin. He was an MP in the House of Commons between 1766 and 1794. His writings include *A Vindication of Natural Society*, *A Philosophical Enquiry into the Origin of Our Ideas of the Sublime and Beautiful*, *Thoughts on the Cause of the Present Discontents*, *Reflections on the Revolution in France*, *Appeal from the New to the Old Whigs*, *Letters on a Regicide Peace*, and *Thoughts and Details on Scarcity*.

**Go to archive of Watchpad section**
**Go to March 2019 issue**

Discuss this article at Journaltalk:
https://journaltalk.net/articles/5988/