Imputations, Inverse Hyperbolic Sines, and Impossible Values

Jack Fitzgerald*

25 April 2024

Abstract

Wolfowicz et al. (2023) find that more arrests and convictions for terrorism offenses decrease terrorism, more charges increase terrorism, and longer sentences do not deter terrorism in 28 European Union member states from 2006-2021. I assess the computational reproducibility of their study and find many data irregularities. The article's primary dependent variable – purportedly an inverse hyperbolic sine transformation of terrorist attack rates – takes on 292 different values when attack rates equal zero, and negatively correlates with attack rates. Many variables exhibit impossible values or undisclosed imputations, often masking a lack of reporting in the article's main data sources. I estimate that the authors have access to 57% fewer observations than claimed. Reproduction attempts produce estimates at least 77.7% smaller than the published estimates. Models reflecting the true degree of missing data produce estimates that are not statistically significantly different from zero for any independent variable of interest.

1 Introduction

Wolfowicz et al. (2023a), henceforth WEA23, use dynamic generalized method of moments (GMM) models and pseudo-Poisson maximum likelihood count data models to explore the effects of lagged sentence length and (rates of) arrests, charges, and convictions on terror attacks using a balanced country-year panel dataset of 28 European Union (EU) member states from 2006-2021. All terrorism-related variables in this dataset are derived from annual EU Terrorism Situation and Trend Reports (TE-SAT; Europol 2007-2022).[1] WEA23's main findings are as follows.

1. Higher (rates of) arrests for terrorism are associated with lower terrorist attacks. From their abstract: "… we find that increased probability of apprehension... demonstrate[s] an inverse relationship with terrorism offending…"

*Vrije Universiteit Amsterdam, School of Business and Economics, Department of Ethics, Governance, and Society, j.f.fitzgerald@vu.nl.
I thank Michael Wolfowicz and Gian Maria Campedelli for answering questions about their article, as well as Abel Brodeur and Derek Mikola for helpful comments and recommendations. For clarity, all tables and figures in WEA23 are numbered using their original numbering system, with Arabic numerals denoting tables and figures, and Roman numerals followed by letters denoting models. In contrast, all tables and figures in this paper are sorted by letter, with all models being termed 'columns' and numbered by Arabic numerals. This replication project has approval from the Ethics Review Board of the School of Business and Economics at Vrije Universiteit Amsterdam.

[1] From pg. 1880: "Our study draws on data from Europol's annual Terrorism Situation and Trends reports (TE-SAT) from 2006 to 2021, testing how arrests, charges, convictions and sentence length for terrorism offenses impact gross terrorism offending." For clarification, TE-SAT reports contain information about terrorism incidents in the prior years; e.g., the 2008 TE-SAT report (Europol 2008) reports data on terrorism-related variables for 2007 and 2006.

2. Higher (rates of) convictions for terrorism are associated with less terrorist attacks. From the abstract: "... we find that increased probability of... punishment demonstrate[s] an inverse relationship with terrorism offending..."

3. Higher (rates of) charges for terrorism are associated with fewer terrorist attacks. From their abstract: "... the rate of charged individuals is associated with a small increase in terrorism."

4. Increased sentence length for terrorism does not deter terrorism. From their abstract: "The results for sentence length are less clear but also indicate potential backlash effects." Their title also reads, "Arrests and convictions *but not sentence length* deter terrorism in 28 European Union member states" (emphasis added).

These claims are defended using a variety of model estimates, but WEA23's dynamic GMM models are in general preferenced in their discussion of results.[2] WEA23 appear to specifically preference the results from Model IIa in Table 1, whose results are reported to arise from a dynamic GMM model (pg. 1880).[3] WEA23 describe these primary estimates in the following way, where B refers to coefficients, reported alongside 95% confidence intervals and P values (see pg. 1880). Because all of WEA23's precision estimates are reported in confidence intervals rather than standard errors, and WEA23 do not provide sufficient replication code to calculate the standard errors directly (see Section 2.1), I compute standard errors by dividing the range of the reported confidence interval by 3.92 (i.e., 2×1.96) and rounding to three decimal places.

1. **Arrest rates**: "In terms of deterrent effects, a 1% increase in the arrest rate is associated with a 0.017% reduction in the terrorism rate ($B = 0.017$ [sic] $(-0.029, -0.006)$, $P = 0.004$)..." This implies an approximate standard error of 0.006.

    (a) The above quote is reported verbatim, but the authors' textually-reported value for B is a typo; the coefficient estimate in Table 1, Model IIa is in fact -0.017 (pg. 1882). This is in line with the authors' interpretation of the estimate as evidence of a negative association, as well as the strictly negative confidence interval bounds.

2. **Conviction rates**: "... for the conviction rate, a 1% increase is associated with an approximately 0.09% reduction [in the terrorism rate] ($B = -0.095$ $(-0.168, -0.023)$, $P = 0.010$)." This implies an approximate standard error of 0.037.

3. **Charge rates**: "A 1% increase in the rate of charged individuals was associated with a 0.16% increase in the terrorism rate ($B = 0.016$ $(0.006, 0.026)$, $P = 0.001$)..." This implies an approximate standard error

[2] When interpreting the magnitude of their estimates, WEA23 write (pg. 1882): "*Drawing on our GMM models*, our results for both likelihood of arrest and charge would be equivalent to correlations of $r = -0.13$, and for conviction rate, $r = -0.10$. With regard to sentence severity, the results would be equivalent to $r = 0.03$ and, combined with our divergent results from the count models, would provide a similar, non-significant pooled result." Emphasis added.

[3] When plotting the "Effects of deterrence variables on terrorism incident rates" in Figure 4 (pg. 1883), the authors specifically reference the "Average predictive margins... along with 95% confidence intervals... of the lagged (y_{t-1}) [sic] deterrence variables *based on GMM Model IIa* as reported in Table 1." Emphasis added.

of 0.003.

4. **Sentence length**: "... a one-unit increase in the average sentence length was associated with a 0.12% increase [in the terrorism rate] ($B = 0.012$ (0.005, 0.019), $P = 0.001$)." This implies an approximate standard error of 0.004.

With the exception of their findings for sentence length, WEA23 also generally assert that these findings are replicated by their count data models in Table 2.[4]

This report is prepared as part of an ongoing collaborative initiative to assess the reproducibility and replicability of findings published in *Nature Human Behaviour* (Brodeur et al. 2024). I assess the computational reproducibility of WEA23's findings and find that they are not reproducible, with all attempts at reproduction yielding estimates that are at least 77.7% smaller than the published estimates. I also conduct a robustness check that makes several reasonable adjustments, which produces no estimates that are statistically significantly different from zero. Section 2 details these reproducibility and robustness checks.

However, perhaps more importantly, I also uncover many irregularities in WEA23's data. I show that the primary dependent variable used in WEA23's estimates exhibits impossible values, and is in fact negatively correlated with terrorist attack rates. Each independent variable of interest is undefined in much of the replication dataset (and in some cases much of the underlying data), but this is masked with a series of undisclosed and incorrect imputations (see Sections 3.1 and 3.4-3.7). WEA23 thus have far less data than they claim. I estimate that 57% of WEA23's claimed observations are only possible to include in their regression models due to their undisclosed imputations. Section 3 covers these irregularities in depth. The replication repository containing the data and code necessary to replicate these analyses is available at doi.org/10.17605/OSF.IO/HCE6N.

2 Computational Reproducibility and Robustness

This section proceeds as follows. In Section 2.1, I provide a general overview of the computational reproducibility afforded by WEA23's replication data. In Section 2.2, I discuss WEA23's original estimates, the estimates from my computational reproductions, and the estimates from my robustness check.

2.1 Reproducibility Overview

My starting point for reproducing WEA23's findings is the replication repository associated with their article (Wolfowicz et al. 2023b). Table A documents the completeness of the repository. Raw data is not provided, and the analysis data is incomplete. Specifically, the dataset is missing information on human development indices (HDIs) and proportion of GDP spent on law enforcement (both of which are necessary to reproduce

[4]From pg. 1880: "These models were replicated using count models, where raw counts of terrorism events served as the dependent variable. The results are displayed in Table 2."

Models IIa and IIIa in Table 1).[5] There is no cleaning code contained in the dataset, and the analysis code is incomplete. The analysis code is largely sufficient for producing Figures 1-3 and Supplementary Figure 1, but no code is provided for any other table or figure. Of the six exhibits in the main article (Figures 1-4 and Tables 1-2), only three are reproducible using WEA23's replication code. In general, I find that the paper's findings are not computationally reproducible, either from raw or analysis data. I defer discussion of coding errors to Section 3.

2.2 Computational Reproduction and Robustness Check

For the purposes of this report, I reproduce the estimates from Model Ia in Table 1, as it is the only model in Table 1 that is reproducible with the data in WEA23's replication repository. Its estimates are also quite close to those of Model IIa and yield all of the same conclusions based on the sign and statistical significance of the coefficient estimates. In reproducing Model Ia, I run models of the following form:

$$\text{Att}_{i,t} = \alpha + \rho\text{Att}_{i,t-1} + \beta_1\text{Arr}_{i,t-1} + \beta_2\text{Char}_{i,t-1} + \beta_3\text{Conv}_{i,t-1} + \beta_4\text{Sent}_{i,t-1} + \psi 1[\max\{\text{Att}_i\} = 0] + \lambda_t + \epsilon_{i,t}. \quad (1)$$

Here $\text{Att}_{i,t}$ is a measure of attack rates, $\text{Arr}_{i,t}$ is a measure of arrest rates, $\text{Char}_{i,t}$ is a measure of charge rates, $\text{Conv}_{i,t}$ is a measure of conviction rates, and $\text{Sent}_{i,t}$ is a measure of average sentence length, where i and t respectively index the country and year. λ_t represents a time fixed effect, ψ reflects the coefficient on a "no-event dummy" that indicates whether country i experiences no terrorist attacks throughout the entire time horizon, and $\epsilon_{i,t}$ is an error term.[6] No other control variables are included in this model, reflecting WEA23's specification choices.[7]

Following WEA23,[8] I specify all models with the `xtabond2` command (Roodman 2009) in Stata v17, obtaining robust standard errors using the `twostep robust` option to implement a finite-sample correction for two-step covariance matrices (Windmeijer 2005).[9] In line with WEA23's model descriptions,[10] I specify $\text{Att}_{i,t-1}$ as endogenous using the `gmmstyle()` option and all other independent variables ($\text{Arr}_{i,t-1}$, $\text{Char}_{i,t-1}$, $\text{Conv}_{i,t-1}$, and $\text{Sent}_{i,t-1}$) as exogenous using the `ivstyle()` option. I do not specify $\psi 1\left[\text{Att}_{i,t} = 0\right]$ or λ_t before the options in `xtabond2` begin, specifying them only as exogenous variables in `ivstyle()`.

[5] WEA23 do cite the sources from which they obtain this raw data in citations 97 and 100.

[6] WEA23 also claim to specify a time trend variable (see pg. 1885), but I omit this variable in my specifications because it is perfectly collinear with λ_t.

[7] From pg. 1880: "*Model Ia (without controls)* and Model IIa (with controls) present the results in which arrests are treated as exogenous" (emphasis added).

[8] From pg. 1885: "We also implemented a two-step model with robust standard errors, using Windmeijer's finite-sample correction for the two-step covariance matrix. All analyses were conducted in Stata v.17 using the XTABOND2 command."

[9] See the help files for `xtabond2` (Roodman 2009).

[10] From pg. 1880: "Model Ia (without controls) and Model IIa (with controls) present the results in which *arrests are treated as exogenous*" (emphasis added). From pg. 1885: "The coefficient for the lagged value of the terrorism incident rate (α) is specified to measure the combined effects of both the short-run dynamics and omitted, time-varying factors that may be hidden within *the endogenous, lagged terrorism incident rates*" (emphasis added), and "... charge and conviction rates, as well as sentence lengths, are unlikely to be endogenous to incident rates and even to arrest rates."

Table B shows the results of my computational reproductions. Column 1 shows the original published estimates, which are copied directly from the paper. For reasons that I detail in Section 3.6, I perform two computational reproductions. Column 2 shows the results from a direct reproduction, with coefficients based on independent variables that are (largely) constructed according to WEA23's published formulas. Column 3 shows the results from an alternate reproduction, with coefficients based on alternative independent variables which I suspect that WEA23 used to produce their main estimates (see Section 3.6). Indeed, based on the sign and statistical significance of the coefficient estimates, though the estimates from Column 2 reflect the conclusions that can be obtained from the published estimates in Column 1 for only one out of four independent variables of interest (namely sentence length), Column 3's estimates reflect these conclusions for three out of four independent variables of interest (all *but* sentence length). However, both sets of estimates are still markedly different than the published estimates; the reproduction estimates are all at least 77.7% smaller in magnitude than the published estimates.

My robustness check model makes three key changes. First, I define all rate variables as linear per capita rates, specifically parameterizing all variables as rates of the underlying count variables for each terrorism-related rate variable per 100,000 people living in country i at time t. This change affects $\text{Att}_{i,t}$, $\text{Arr}_{i,t}$, $\text{Char}_{i,t}$, and $\text{Conv}_{i,t}$. I respectively parameterize the latter three as

$$\text{ArrestsPC}_{i,t} = \frac{\texttt{arrests}_{i,t}}{\texttt{population1}_{i,t}}$$
$$\text{ChargesPC}_{i,t} = \frac{\texttt{charges}_{i,t}}{\texttt{population1}_{i,t}}$$
$$\text{ConvictionsPC}_{i,t} = \frac{\texttt{convictionrate}_{i,t} \times \texttt{charges}_{i,t}}{\texttt{population1}_{i,t}}.$$

I parameterize $\text{Att}_{i,t}$ as $\text{AttackRate}_{i,t}$ in Column 4 (see Section 3.1). In Section 3.8, I detail how these changes increase data coverage. Second, I add country fixed effects, as do WEA23 in Models IIa and IIIa (see Equation 1, pg. 1885). Third and finally, I remove WEA23's undisclosed imputations for missing data (see Sections 3.5, 3.7, and 3.8). This robustness check produces estimates on the independent variables of interest that differ in magnitude from the published estimates by at least 20.1%.

Only one of the four coefficients in Column 4 has a P-value beneath 0.05, specifically that for $\text{Char}_{i,t}$ ($P = 0.043$). However, because WEA23 simultaneously test for four hypotheses at the same time, adjustments must be made for multiple hypothesis testing. As the smallest P-value in a set of four P-values of interest, both Bonferroni-Holm family-wise error rate corrections (Holm 1979) and Benjamini-Hochberg false discovery rate corrections (Benjamini & Hochberg 1995) would effectively multiply this P-value by four. The resulting multiple testing-corrected P-value would be equal to 0.171, rendering it far greater than any commonly-agreed standard for statistical significance. Therefore, none of the coefficients on the independent variables of interest in Column 4 is statistically significantly different from zero in my robustness check.

3 Data Irregularities

This section details issues with the data in WEA23's replication repository (Wolfowicz et al. 2023b). In Section 3.1, I document several issues with the primary dependent variable in WEA23's dynamic GMM models, `DVSin`. Though WEA23 claim that `DVSin` is a positive monotonic transformation of attack rates, I use a simple vertical line test to show that `DVSin` can not possibly be a function of attack rates, and show that `DVSin` is in fact negatively correlated with attack rates. In Section 3.2, I show that WEA23's parameteriztion of arrest rates is undefined in most of their dataset, and that even when it is defined, it takes on questionable values in the vast majority of observations. In Section 3.3, I show that charge rates are also undefined in most of WEA23's dataset, and also document that WEA23's underlying data sources do not contain data on charges. In Section 3.4, I show that though WEA23's stored variable for conviction rates is always defined, it often should not be. WEA23 appear to use an undisclosed imputation to construct conviction rates, which artificially makes conviction rates and terrorist attacks appear more correlated than they actually are, and is itself not an accurate computation of conviction rates. In Section 3.5, I note that the same is true of sentence length, and document that WEA23's underlying data sources do not contain data on sentence length for most observations.

In Section 3.6, I show that these aforementioned data issues raise doubts about the reported observation counts in Table 1, which implies that it is possible that either the observation counts are misreported or that WEA23 use undisclosed alternative variables in their model specifications. I provide evidence suggesting the latter, and show that these alternative variables also exhibit irregularities. In Section 3.7, I document that WEA23 claim to possess terrorism data on four different countries from the EU's TE-SAT reports in years during which these countries were not part of the EU, and thus did not provide data for the EU TE-SAT reports. When terrorism-related variables are unavailable for these country-years, it appears that WEA23 uniformly impute these values as zeroes. Finally, in Section 3.8, I show that the inescapable data coverage issues discussed in Sections 3.5 and 3.7 imply that if the truly missing observations in WEA23 were given their actual missing values rather than WEA23's undisclosed zero-value imputations, then WEA23 would possess far less data than they claim to have. WEA23's reported observation counts in Table 1 imply that their estimates are computed with 420 observations, but if their true data coverage is presented accurately, they can only estimate these models with at most 180 observations. This implies that WEA23 have 57% fewer observations than claimed.

3.1 Attacks, Attack Rates, and `DVSin`

The primary quoted results from WEA23's dynamic GMM models (see pg. 2 of this replication report) all make reference to effects on terrorism rates. WEA23 state that the terrorism rate used as their dependent variable in these models is in fact the inverse hyperbolic sine (IHS) of the per capita rate of terrorist attacks. From subsection "Dependent variable" (pg. 1884): "... we calculated terrorism incidents both as a raw count of the number of events and as a rate per 100,000 residents...[11] For the incident rate, we applied the IHS transformation..."

I follow WEA23's textual description and define the *attack rate* as

$$\text{AttackRate}_{i,t} = \frac{\text{Attacks}_{i,t}}{\text{Pop}_{i,t}}, \tag{2}$$

where $\text{Attacks}_{i,t}$ is the number of terrorist attacks in country i during year t (stored in the replication dataset as `attacks`) and $\text{Pop}_{i,t}$ is the population of country i in hundreds of thousands in year t (stored in the replication dataset as `population1`). I define the *IHS attack rate* as

$$\text{AttackRateIHS}_{i,t} = \sinh^{-1}(\text{AttackRate}_{i,t}), \tag{3}$$

where the IHS transformation of some variable X can be written as[12]

$$\sinh^{-1}(X) = \ln\left(X + \sqrt{X^2 + 1}\right). \tag{4}$$

Based on WEA23's replication code, the variable that purports to be $\text{AttackRateIHS}_{i,t}$ is stored in the replication dataset as `DVSin`.

`DVSin` takes on many impossible values. This problem is clearly visible in Figures 1 and 2, which I reproduce in Figures A and B (respectively) in this report for convenience and thoroughness using WEA23's exact replication code, confirming that `DVSin` is indeed the variable that WEA23 claim represents IHS attack rates.[13] As WEA23 acknowledge in the notes of Figure 3, nine countries in WEA23's data experience no terror attacks throughout the entirety of the dataset's time horizon (specifically Croatia, Cyprus, Estonia, Latvia, Luxembourg, Malta, Romania, Slovakia, and Slovenia). The attack rate for each of these nine countries is thus equal to zero for the entirety of the time horizon. A well-known property of the IHS transformation (Bellemare & Wichman 2020; Chen & Roth 2024) – one that WEA23 cite as an explicit motive for using the IHS transformation[14] – is that the IHS of zero is defined, and equals zero. Thus if `DVSin` is indeed the IHS attack rate, then `DVSin` should equal zero for the entirety of the time horizon for all nine of these countries. However, Figure 2 and Figure B show that in every one of these nine countries, the only values that `DVSin` takes on are strictly

[11]This description of the per capita attack rate contradicts the labels in Figures 1-2, as well as those of Supplementary Figure 1, which state that attack rates are measured in rates per 10,000 residents. However, it is easy to verify by inspection that the variable stored in WEA23's replication dataset for population, `population1` (which is itself used to construct Figures 1-2 and Supplementary Figure 1 in WEA23's replication code), is measured in hundreds of thousands rather than tens of thousands. For instance, Germany's value of `population1` for 2021 is 843.3909; Germany's population in 2021 is of course much closer to 84.39 million than 8.439 million.

[12]See Bellemare & Wichman (2020) and Chen & Roth (2024).

[13]Though the axes in the replicates of Figures 1 and 2 produced by WEA23's code take on different scales than those in the published versions of Figures 1-2, and the labels are not formatted the same, the patterns of the plotted values over time are virtually identical between the replicates and published versions of Figures 1-2.

[14]From pg. 1885: "Because the dependent and main independent variables include observations with zeroes, for the calculation of the event rate we used the IHS transformation, which is common in economics in situations where the presence of zeroes prohibits the calculation of the logarithm of the variable."

positive, and these values vary over time.

Impossible values affect more than just these nine countries; in fact, `DVSin` holds no clear relationship with either attack rates or IHS attack rates. WEA23's replication dataset contains data on $\text{Attacks}_{i,t}$ and $\text{Pop}_{i,t}$ (variables `attacks` and `population1` respectively), so attack rates and IHS attack rates can be manually calculated using Equations 2 and 3 respectively. Figure C plots the relationships between `DVSin` and both attack rates and IHS attack rates. Two features of Figure C immediately stand out. First, `DVSin` exhibits an inverse relationship with both attack rates and IHS attack rates. In fact, `DVSin` is negatively correlated with both attack rates ($r = -0.107, P = 0.024$) and IHS attack rates ($r = -0.108, P = 0.022$). Second, `DVSin` inexplicably maps (IHS) attack rates of zero onto many different values. Terrorist attacks are quite rare in WEA23's data, with 305 of the 420 country-year observations after 2006[15] (72.6%) experiencing zero attacks. For these 305 observations – all of whom share an (IHS) attack rate of zero – `DVSin` takes on 292 different values, all of which are strictly positive. This is particularly troubling, as it implies that `DVSin` fails a simple vertical line test, and thus that `DVSin` can not possibly be a function of attack rates. I contacted the authors for clarification about how DVSin was constructed through the Institute for Replication. Unfortunately, the authors could not respond to my queries at the moment.

3.2 Arrests and Arrest Rates

Per Equation 1 in WEA23 (pg. 1885), the *arrest rate* is defined as

$$\text{ArrestRate}_{i,t} = \frac{\text{Arrests}_{i,t}}{\text{Attacks}_{i,t}}. \tag{5}$$

Here $\text{Arrests}_{i,t}$ is the number of arrests for terrorism in country i during year t, stored in WEA23's replication data as `arrests`. WEA23 seem to interpret $\text{ArrestRate}_{i,t}$ as a probability of arrest.[16]

There are three reasons why $\text{ArrestRate}_{i,t}$ does not cleanly represent the probability of arrest. First, $\text{ArrestRate}_{i,t}$ is undefined whenever $\text{Attacks}_{i,t}$ is zero. Because terrorism is rare in WEA23's data, this issue arises frequently. Of the 420 pre-2021[17] country-year observations, 301 (71.7%) experience zero attacks, and thus have undefined values of $\text{ArrestRate}_{i,t}$.

Second, more than one person can commit a single terrorist attack. However, because WEA23's attacks data does not contain any information regarding how many perpetrators are responsible for each attack, $\text{Attacks}_{i,t}$ is not normalized to reflect group terror plots, which can lead to questionable values for the arrest rate. For

[15] This is the relevant sample for the purposes of evaluating WEA23's findings, as observations of the dependent variable from 2006 – the first year in the dataset's time horizon – are mechanically dropped from all dynamic models due to the lag specification.

[16] On pg. 1884: "Our study follows what could be referred to as a classic deterrence framework, in which the expected deterrent effect on terrorism activity is a function of the probability of arrest (P_A), the probability of charge conditional on arrest ($P_{C|A}$), the probability of conviction conditional on being charged ($P_{P|C}$) and the severity of punishment as expected prison sentence length (S)."

[17] This is the relevant sample for the purposes of evaluating WEA23's findings, as observations of the independent variables from 2021 – the last year in the dataset's time horizon – are mechanically dropped from WEA23's dynamic models due to the lag specification.

instance, suppose country i experiences exactly one terror attack in year t, that this attack is committed by a group of five people, and that all five individuals are arrested for the attack in that same year. By Equation 5, $\text{ArrestRate}_{i,t}$ would be equal to five, implying an arrest rate of 500%.

Third and finally, not all terror suspects are arrested in the same year as the attack that they are accused of committing; some take longer to find and arrest. This implies that $\text{Arrests}_{i,t}$ may be populated by suspects who are associated with terror attacks committed long before year t. It also implies that $\text{Arrests}_{i,t}$ can be positive even when $\text{Attacks}_{i,t}$ is zero. This occurs frequently in WEA23's data; of the 301 pre-2021 country-year observations that experience zero attacks, 134 (44.5%) have a positive value for `arrests`. This data feature can not be explained by terrorism attacks being prevented by arrests. As WEA23 note (pg. 1884), the country-year attack counts from the TE-SAT reports are unique because they comprise the sum of completed, failed, and foiled terror attacks (Europol 2007-2022). Thus if an arrest in year t foiled a terror attack that would have occurred in year t, this would have been included in the value of $\text{Attacks}_{i,t}$. This means that the frequent presence of positive `arrests` values in country-year observations with zero attacks almost certainly reflects temporal misattribution.

Due to the combination of these issues, most values of $\text{ArrestRate}_{i,t}$ in WEA23's data are undefined, and the vast majority of those that are defined are clearly not probability values. Of the 119 pre-2021 country-year observations with defined values of $\text{ArrestRate}_{i,t}$, 101 (84.8%) are strictly greater than one, implying an arrest rate of more than 100%. The largest value of $\text{ArrestRate}_{i,t}$ – belonging to Belgium in 2018 – is 166, implying that Belgium had a 16,600% arrest rate for terror attacks in 2018. WEA23 do not store a variable for $\text{ArrestRate}_{i,t}$ in their replication dataset, nor do they report descriptive information about this variable in the article.

3.3 Charges, `charges`, and Charge Rates

By Equation 1 in WEA23 (pg. 1885), the *charge rate* can be written as

$$\text{ChargeRate}_{i,t} = \frac{\text{Charges}_{i,t}}{\text{Arrests}_{i,t}}, \tag{6}$$

where the variable purporting to be $\text{Charges}_{i,t}$ is stored in WEA23's data as `charges`. $\text{ChargeRate}_{i,t}$ is thus undefined whenever $\text{Arrests}_{i,t} = 0$. This occurs frequently in WEA23's data; 174 of 420 pre-2021 country-year observations (41.4%) see zero arrests, and thus have undefined values of $\text{ChargeRate}_{i,t}$. WEA23 do not store a variable in their replication dataset for $\text{ChargeRate}_{i,t}$, and do not provide descriptive statistics on $\text{ChargeRate}_{i,t}$ for readers to observe this missing data issue.

However, I find a more serious issue concerning WEA23's `charges` variable: TE-SAT reports do not contain data on terrorism charges. Attempting to match `charges` to appendix data from the TE-SATs reveals that `charges` most closely represents the count of individuals who are *tried* for terrorist attacks in country i in year t. This variable is substantively different from $\text{Charges}_{i,t}$ because not all charged suspects face trial. Many terrorism charges are dropped (a point which WEA23 acknowledge explicitly),[18] and some are resolved *via* plea deals. This proxy also yields substantial temporal misattribution, as not all terrorism trials begin or conclude in the same year as their respective charges. WEA23 at no point disclose that they proxy terrorism charges with the number of individuals who face trial for terrorism charges.

3.4 Convictions and `convictionrate`

Per Equation 1 in WEA23 (pg. 1885), the *conviction rate* can be written as

$$\text{ConvictionRate}_{i,t} = \frac{\text{Convictions}_{i,t}}{\text{Charges}_{i,t}}. \tag{7}$$

WEA23's replication dataset does not contain a variable for $\text{Convictions}_{i,t}$. However, the dataset does contain variable `convictionrate`, which should in principle be $\text{ConvictionRate}_{i,t}$.

However, `convictionrate` is often defined when it should not be. Per Equation 7, $\text{ConvictionRate}_{i,t}$ is undefined whenever $\text{Charges}_{i,t} = 0$. Thus `convictionrate` should be frequently undefined, as `charges` is equal to zero for 219/420 pre-2021 observations (52.1%). However, `convictionrate` is defined for all 219 of these observations. Attempting to match `convictionrate` to data from the TE-SAT reports shows that $\texttt{convictionrate}_{i,t}$ most closely represents the proportion of terrorism verdicts in country i that were convictions, rather than acquittals, in year t. However, even this undisclosed imputation does not explain the preponderance of defined values when country-year observations exhibit zero verdicts for terrorism cases; for such country-years, the conviction proportion of verdicts is still undefined. Returning to the TE-SAT reports, WEA23 appear to have imputed all values of `convictionrate` to zero for country-years where no verdicts are issued in terrorism cases.

These undisclosed imputations are inaccurate for two reasons. First, the fact that a given country concludes no cases against terrorism suspects in a given year does not imply that this country would fail to convict 100% of suspects charged with terrorism offenses if some were charged in that year. Critically for WEA23's findings in Table 1, this imputation artificially inflates the magnitude of correlation between `convictionrate` and terror attacks. Intuitively, country-years where no terrorism verdicts are levied experience few terror attacks, and predominantly experience none at all. Of the 216 observations for which $\texttt{convictionrate}_{i,t-1} = 0$ and $\texttt{charges}_{i,t-1} = 0$, $\texttt{attacks}_{i,t} = 0$ in 205 (94.9%). Thus due to this imputation, $\texttt{convictionrate}_{i,t-1}$ and $\texttt{attacks}_{i,t}$ often artificially converge to zero at the same time, making the two variables appear more correlated than they actually are.

[18] From pgs. 1879-1880: "Additional issues with how the criminal justice systems in democratic and semi-democratic countries treat terrorism could lead to differential deterrent effects. For example, compared with ordinary crime, the rate of release without charge may be quite high. According to the United Kingdom's Home Office, some years have seen more than 50% of those arrested under the Terrorism Act (2000) released without charge. Our own review of these statistics shows that this rate increases commensurate with the volume of arrests."

Second, WEA23's undisclosed imputation of `convictionrate` is also not equal to the conviction rate that WEA23 claim to compute (i.e., from Equation 7) even when their proxy (i.e., the conviction proportion of verdicts) is defined. Not all terrorism trials that begin in year t also end in year t. Therefore, even though $\texttt{charges}_{i,t}$ represents the number of individuals that faced trial for terrorism in country i during year t (as discussed in Section 3.3), comparing published conviction counts from the TE-SAT reports with the data used to create variable $\texttt{charges}_{i,t}$ reveals clear temporal misattributions. To give an example from the 2009 TE-SAT report (Europol 2009),[19] Spain tried 141 individuals for terrorism charges in 2008, yet convicted 162 individuals on terrorism charges that same year. Per Equation 7, this would imply that Spain had a terrorism conviction rate above 114% in 2008. In fact, relying on the same TE-SAT report, Equation 7 would imply that four countries – Denmark, the Netherlands, Spain, and the United Kingdom (UK) – have conviction rates above 100% in 2008. These temporal misattributions are masked by WEA23's undisclosed proxy; though the proportion of terrorism verdicts that are convictions is mechanically constricted to range between zero and one, the conviction rate WEA23 claim to compute in Equation 7 lacks this constraint. If `convictionrate` was stored according to WEA23's disclosed formula, basic descriptive statistics of the variable would reveal these questionable values. However, even the `convictionrate` variable stored in WEA23's replication data shows some evidence of such misattribution, such as three pre-2021 observations with `charges` = 0 and `convictionrate = 1`, implying that these country-years exhibited terrorism conviction rates of 100% despite the fact that no suspects were tried on terrorism charges.

For the purposes of my reproductions, I recover conviction counts from WEA23's replication data using the Stata expression `gen convictions = convictionrate*charges`. Though `convictions` should in principle equal $\text{Convictions}_{i,t}$ if `convictionrate` is indeed calculated according to Equation 7, it is clear given the previous discussion in this section that `convictions` is not an accurate count of the number of terrorism convictions assessed in country i at time t. Another easy way to see this is by noting that `convictions` exhibits many impossible values. 83 of 420 pre-2021 observations (19.8%) exhibit non-integer values for `convictions`. This partially reflects coarse rounding, as `convictionrate` is bounded from zero to one and is stored with at most two decimal places. However, it also reflects the fact that the conviction counts in the TE-SAT reports are not a subset of `charges`, and confirms that it is impossible to obtain a completely accurate count of convictions from WEA23's replication data.

A simple exercise reveals other problems with `convictionrate`. If `convictionrate` is $\text{ConvictionRate}_{i,t}$, and is constructed (as claimed) according to Equation 7, then a variable `conviction_rate_calc` produced by the Stata expression `gen conviction_rate_calc = convictions/charges` should be identical to the variable `convictionrate`, given that `convictions` = `convictionrate` × `charges` by construction. I perform this exercise and find that though `conviction_rate_calc` holds identical values to `convictionrate` whenever `conviction_rate_calc` is defined, `conviction_rate_calc` is undefined for 219/420 pre-2021 observations

[19] Calculations in this paragraph specifically reference Figures 4-5 in Europol (2009).

(52.1%), while `convictionrate` is defined for all of these values. This confirms that `convictionrate` can not possibly have been constructed as WEA23 claim.

3.5 Sentence Length and `sentence`

WEA23 parameterize punishment severity as the average prison sentence given to terrorism convicts within country i in year t.[20] I term this variable $\text{Sentence}_{i,t}$. WEA23 store the variable purporting to reflect this measure under the name `sentence`.

Like `convictionrate` (see Section 3.4), `sentence` is also often defined when it should not be. The average sentence length for a given country is not defined if that country has not sentenced anyone for terror offenses. Accordingly, the TE-SAT only reports sentence length for a subset of EU member states (Europol 2007-2022), presumably those with imprisoned terrorism convicts. However, `sentence` is defined for all pre-2021 observations. WEA23 appear to impute all undefined values of $\text{Sentence}_{i,t}$ as zeroes. Resultantly, `sentence` is equal to zero for 240/420 (57.1%) of the pre-2021 observations in WEA23's data.

This undisclosed imputation is inaccurate. A country's lack of terrorist prisoners does not imply that all individuals convicted of terrorism-related offenses in that country would receive zero prison time for such offenses. For example, in WEA23's data, Luxembourg is assigned `sentence` values of zero for the entirety of the time horizon. This is not reflective of potential terrorists expecting to receive no punishment for terror attacks committed in Luxembourg, even if caught; it is reflective of the fact that Luxembourg experiences no terror attacks over the entire time horizon.

Similarly to the zero-value imputations for `convictionrate` (see Section 3.4), the zero-value imputations for `sentence` inflate the magnitude of correlation between sentence length and terror attacks. Countries lacking convicted terrorists to sentence naturally experience fewer attacks, and predominantly experience none at all. In 221 of 240 observations where $\texttt{sentence}_{i,t-1} = 0$ (92.1%), $\texttt{attacks}_{i,t} = 0$ as well. Like `convictionrate`, WEA23's zero-value imputations often cause $\texttt{sentence}_{i,t-1}$ and $\texttt{attacks}_{i,t}$ to converge to zero together by construction, making the two variables appear more correlated than they actually are.

3.6 Discrepancies in Table 1 and Other 'IHS' Variables

WEA23's Equation 1 (pg. 1885) implies that the dynamic GMM models which produce Table 1's estimates incorporate $\text{ArrestRate}_{i,t-1}$, $\text{ChargeRate}_{i,t-1}$, $\text{ConvictionRate}_{i,t-1}$, and $\text{Sentence}_{i,t-1}$ (i.e., the independent variables of interest). As discussed in Sections 3.2-3.5, these variables should be – and are – often undefined. If these variables are computed without WEA23's undisclosed imputations for `convictions` and `sentence`, then 327/420 pre-2021 observations (77.9%) contain missing values of at least one independent vari-

[20] On pg. 1884: "Our main independent variables were... (4) sentence severity (*average number of years*)..." (emphasis added).

able of interest.[21] Even if the undisclosed imputations for `convictions` and `sentence` are treated as genuine non-missing values, 308/420 pre-2021 observations (73.3%) still contain missing values of at least one independent variable of interest because either $\texttt{attacks}_{i,t-1}$ or $\texttt{arrests}_{i,t-1}$ is zero. WEA23 partially acknowledge this issue of zeroes in observations of the independent variables of interest.[22]

These facts imply that there are reporting discrepancies in Table 1. `xtabond2`, the Stata command that WEA23 cite as the source of their dynamic GMM estimates (see pg. 1885), will automatically drop observations that contain missing values in any specified variable. Table 1 reports 420 observations in each regression. However, because 308 observations contain missing values of either $\text{AttackRate}_{i,t}$ or $\text{ChargeRate}_{i,t}$, these 308 observations are mechanically dropped from any dynamic GMM model estimated using `xtabond2`, so the regression models WEA23 specify should have at most 112 observations. This is reflected in my direct reproduction estimates in Table B, Column 2. This further implies that it is possible that either the observation counts in Table 1 are misreported or that WEA23 produced the estimates in Table 1 with different variables than claimed.

There are in fact strong candidates for alternate rate variables that WEA23 may have used instead without disclosing. WEA23's replication dataset contains two variables – `CharSin` and `ConvictionSin` – that are titled in the same manner as `DVSin`, implying that they may respectively be IHS-transformed charge and conviction rates. There is also variable `NewArrest`, which appears to be an offshoot of arrest rates, the only such variable in WEA23's replication data. Critically, `CharSin`, `ConvictionSin`, and `NewArrest` all exhibit full data coverage with no undefined observations, and thus a model regressed with these three variables in place of $\text{ArrestRate}_{i,t}$, $\text{ChargeRate}_{i,t}$, and $\text{ConvictionRate}_{i,t}$ could in principle produce the same observation counts as are reported in Table 1.

Though WEA23 never explicitly state that they use IHS transformations for the arrest rate, charge rate, and conviction rate variables modelled in Table 1, there are four reasons to believe that they used these 'Sin' variables in place of actual rate variables beyond the aforementioned sample size discrepancy. First, as Columns 1-3 in Table B show, the conclusions that can be derived from a computational reproduction of Model Ia in Table 1 that uses `NewArrest`, `CharSin`, and `ConvictionSin` (i.e., the alternate reproduction) are much closer to the conclusions that can be derived from the published estimates than those that can be derived from the estimates in a reproduction that uses the true values of $\text{ArrestRate}_{i,t}$, $\text{ChargeRate}_{i,t}$, and $\text{ConvictionRate}_{i,t}$ (i.e., the direct reproduction). Based on the sign and statistical significance of the coefficient estimates, though conclusions arising from the published estimates align with those arising from only one out of four independent variables of interest in the direct reproduction, they align with the conclusions arising from three out of four independent variables of interest in the alternate reproduction. However, the coefficients on all independent variables of interest are at least 77.7% smaller than the published estimates in both reproductions.

[21] Such observations have zero values in at least one of $\texttt{attacks}_{i,t-1}$, $\texttt{arrests}_{i,t-1}$, $\texttt{charges}_{i,t-1}$, and $\texttt{sentence}_{i,t-1}$.

[22] From pg. 1885: "Because the dependent *and main independent variables* include observations with zeroes..." (emphasis added).

Second, WEA23's interpretations of the coefficients in Table 1 make much more sense if the rate variables are IHS-transformed. When interpreting the coefficients on arrest rates, conviction rates, and charge rates from Table 1, WEA23 interpret the coefficients in elasticity terms, describing the coefficients as the effects of "a 1% increase" in arrest, charge, and conviction rates; notably, they do not make the same claims about sentence length (see pg. 2 of this report).[23] This interpretation of the regression coefficients on the rate variables is not justified if the rate variables are incorporated into regression models linearly (without the use of a post-estimation suite that converts the regression estimates into elasticities,[24] which WEA23 do not claim to use). However, under the notion that coefficients on IHS-transformed variables can be viewed as (semi-)elasticities (or close approximations thereof; see Bellemare & Wichman 2020), WEA23's interpretation of the coefficients as associations with a 1% increase in the rate variable would be largely justified.[25]

Third, WEA23 allude to the IHS transformation being appropriate due to the values of the independent variables. WEA23 justify their use of the IHS transformation on pg. 1885: "Because the dependent *and main independent variables* include observations with zeroes, for the calculation of the event rate we used the IHS transformation, which is common in economics in situtations where the presence of zeroes prohibits the calculation of the logarithm of the variable" (emphasis added). The mention of independent variables taking on zero values as a justification to employ the IHS transformation does not make sense if only the dependent variable is IHS-transformed; regression results are of course estimable without observation drops if the dependent variable is IHS-transformed while the independent variables are left linear. However, this explanation does make sense under the assumption that the rate variables are also transformed to permit (approximate) elasticity estimates (as is implied by WEA23's interpretations discussed in the previous paragraph), in which case justifying the IHS transformation over a logarithmic transformation by pointing to zero values in the rate variables does make sense.

Fourth and finally, the pre-print version of WEA23 (Wolfowicz et al. 2023c) IHS-transforms the three independent rate variables of interest. This is clear from Equation 1 in Wolfowicz et al. (2023c), which provides their model specifications, as well as their textual description (pg. 5): "Our main independent variables were 1) the probability of arrest, calculated as the number of arrests per incident, 2) the probability of being charged, calculated as the number of charged cases per number of arrests, 3) the probability of convictions, calculated as the percentage of charges ending in conviction, and 4) the average prison sentence given, measured in years... *All of these variables were subsequently converted to their inverse hyperbolic sines as has been done in previous studies on terrorism in which many observation points are equal to zero...*" (emphasis added). It is plausible that this specification choice may have carried over from the pre-print to the published version, and that the one-sentence mention of the IHS transformation for the independent variables was removed in the

[23] The replication dataset also contains variable `SentenceSin`, but because of this explicit departure in coefficient interpretation when discussing sentence duration, I presume that `SentenceSin` is never incorporated in WEA23's GMM models and thus omit analysis of this variable.

[24] For example, consider `margins, eyex()` in Stata or the `marginaleffects` suite in R.

[25] See Chen & Roth (2024) for problems with such interpretations.

final version of the article. Though none of these facts in isolation definitively proves that WEA23 used 'Sin' rate variables in place of the true values of rate variables, they combine to provide convincing evidence of such specification choices.

If WEA23 used the 'Sin' variables when estimating the models in Table 1, then the serious data issues that affect `DVSin` (see Section 3.1) also impact several of the independent variables used in Table 1's models. Figure D plots the relationships between each 'Sin' variable and both the raw and IHS-transformed values of the underlying rate variable from which the respective 'Sin' variable is purportedly calculated. Each graph shows irregularities in the construction of the 'Sin' variables.

The top row of graphs in Figure D shows that `NewArrest` is inversely related with both $\text{ArrestRate}_{i,t}$ and its IHS transformation. In fact, NewArrest is substantially negatively correlated with both $\text{ArrestRate}_{i,t}$ ($r = -0.316, P = 0.000$) and its IHS transformation ($r = -0.726, P = 0.000$). It is also plainly visible that `NewArrest` fails the vertical line test when $\text{ArrestRate}_{i,t} = 0$. The seven observations in WEA23's data for which $\text{ArrestRate}_{i,t} = 0$ are assigned five different values of `NewArrest`, all of which are strictly positive. This vertical line test failure implies that `NewArrest` can not possibly be a function of $\text{ArrestRate}_{i,t}$ or its IHS transformation.

The middle row of graphs in Figure D shows that `CharSin` is positively related, albeit noisily, with both $\text{ChargeRate}_{i,t}$ and its IHS transformation. However, while not clearly visible in Figure D, `CharSin` also fails the vertical line test. The 83 observations with zeroes for $\text{ChargeRate}_{i,t}$ exhibit 21 different positive values of `CharSin`, all of which are strictly greater than zero. It is thus not possible that `CharSin` is a function of $\text{ChargeRate}_{i,t}$.

The bottom row of graphs in Figure D shows that `ConvictionSin` is perfectly positively correlated with `convictionrate`, but not with its IHS transformation. Though this implies that a formula for `ConvictionSin` can be reverse-engineered via OLS estimation, this is itself rather troubling. `ConvictionSin` should be more correlated with the IHS transformation of `convictionrate` than with the raw values of `convictionrate`, as it is in principle an IHS-transformed variable. The fact that `ConvictionSin` can be perfectly reconstructed by giving `convictionrate` a linear slope and adding a constant implies that `ConvictionSin` is not a genuine IHS transformation of `convictionrate`, as the IHS transformation is nonlinear (see Equation 4). This in turn implies that even if WEA23 used `ConvictionSin` instead of `convictionrate`, and even under the presumption that regression coefficients on IHS-transformed variables yield (approximate) (semi-)elasticity estimates, that WEA23's interpretation of the conviction rate coefficients in Table 1 as the effects of a 1% increase in conviction rates are not justified. `ConvictionSin` appears to just be a linearly rescaled version of `convictionrate`.

One additional data issue with the 'Sin' variables is not visible in Figure D: the 'Sin' variables map undefined values of their respective underlying variables to many different real values. Such mappings are naturally not displayed in Figure D, as the underlying variable values for such observations are missing. However, `NewArrest` takes on 37 different defined values for 321 observations in which $\text{Attacks}_{i,t} = 0$, and `CharSin` takes on 10 different values for 184 observations for which $\text{Arrests}_{i,t} = 0$. The only 'Sin' variable that does not exhibit such mappings is `ConvictionSin`. However, recall from Section 3.4 that this arises by construction because WEA23 impute conviction rates that should be undefined as zeroes without disclosure. Further, recall from Section 3.5 that they do the same for average sentence durations that should be undefined. Thus if WEA23 indeed use the 'Sin' variables to estimate Table 1's models, then WEA23 only manage to reach 420 observations in these models through simulating full data coverage by imputing hundreds of different real values to independent variables that are in fact undefined, doing so for roughly three quarters of their dataset.

3.7 Data Coverage and EU Membership

Though WEA23's article and replication data imply that they possess a balanced panel dataset on all 28 EU member states for the entirety of the time horizon, this is not an accurate characterization of the data coverage afforded to WEA23 by the TE-SAT reports. This is because a TE-SAT report published in year t contains no data on countries that are not part of the EU at time $t-1$. This affects four countries over the time horizon of WEA23's data. Bulgaria and Romania did not join the EU until 1 January 2007, Croatia did not join until 1 July 2013, and the UK formally departed from the EU on 1 February 2020.[26] The TE-SAT reports are missing data on at least some variables in all country-years in which country i was not part of the EU in year t (Europol 2006-2013; Europol 2021). However, WEA23 appear to incorrectly impute all of these missing values as zeroes in their replication dataset.

This issue makes WEA23's data at times quite misleading. For example, Figure 1 (replicated here in Figure A) implies that in the UK, terrorism went from being relatively frequent before 2021 to being completely eliminated in 2021. This can be easily misconstrued as a consequence of the COVID-19 pandemic; Figure A shows that terrorist attacks were less frequent in most countries from 2020 to 2021. However, the UK's 2021 pure zero value for attacks – as well as every other terrorism-related variable – does not reflect a genuine disappearance of terrorism or anti-terrorism enforcement. 169 people were arrested, 49 were charged, and 38 were convicted of terrorism-related offenses in the UK during 2021 (Allen, Burton, & Pratt 2022). The sharp declines in all terrorism-related variables for the UK in 2021 solely reflect WEA23's undisclosed imputations. These coordinated zeroes also bias the magnitudes of correlations between terrorism incidence and deterrence variables upward; because all terrorism variables 'go to zero' at the same time in these observations, their correlations are artificially inflated.

[26] See https://neighbourhood-enlargement.ec.europa.eu/enlargement-policy/6-27-members_en, accessed 18 April 2024.

3.8 Real vs. Stated Data Coverage

WEA23 report having a balanced panel dataset of 28 EU member states over 16 years. Given the one-lag specification implied by their models,[27] this implies that WEA23 have 420 observations to conduct the model estimations for Table 1, which is reflected in Table 1's observation counts. This same is (in principle) true for the models used to estimate Table 2.

However, given the many undisclosed imputations that mask data coverage issues, WEA23 in fact possess far fewer than 420 observations for their models, regardless of whether 'Sin' variables are used instead of regular rate variables in Table 1 (see Section 3.6). Some alternative specification choices can reduce the severity of data coverage issues. For instance, reparameterizing arrest rates, charge rates, and conviction rates as per capita rates can reduce data coverage problems that arise from 'dividing-by-zero' issues, as population counts are of course always strictly positive for all observations in the dataset. This is a key motivation for changing $\text{Att}_{i,t}$ in Table B, Column 4 from $\texttt{DVSin}_{i,t}$ to $\text{AttackRate}_{i,t}$, alongside the aforementioned irregularities with `DVSin` (see Section 3.1) and recent findings regarding the inappropriateness of IHS transformations for obtaining (semi-)elasticity estimates (Chen & Roth 2024).

However, there are two inescapable sources that decrease actual data coverage in WEA23's dataset. First, the TE-SAT does not report average sentence length for the majority of country-years, as most country-years experience no terrorism convictions. As discussed in Section 3.5, WEA23 incorrectly impute values of zero into `sentence` for these observations. Not a single TE-SAT throughout the entire time horizon reports that a country has an average sentence duration of zero years (Europol 2007-2022).[28] The 240 values of zero in `sentence` for pre-2021 observations are in fact undefined and should therefore be missing. Second, the TE-SAT does not report data on all variables for Bulgaria or Romania in 2006, for Croatia from 2006-2012, or for the UK in 2021, as these countries were not part of the EU in these years. As discussed in Section 3.7, WEA23 again incorrectly impute these missing values in the TE-SATs as zeroes without disclosure.

Together, these two unavoidable data coverage limitations imply that at least 240 observations which reportedly contribute to WEA23's estimates in Tables 1-2 exhibit undefined values of at least one variable of interest, and should thus be dropped from the estimation. Such observations either have zeroes imputed into `sentence` or represent country-years in which the country was not part of the EU in that year. This leaves at most 180 clean observations with real defined values for all dependent and independent variables of interest for these models. Thus 57% of WEA23's observations are only possible to include in their regression models due to their undisclosed imputations, and therefore WEA23 functionally have 57% fewer observations than

[27] Though Equations 1 and 2 in WEA23 imply that the independent variables of interest are regressed contemporaneously rather than in a first-lag specification, this is almost certainly a typo. If this is not a typo, then Equation 1 also implies that $\texttt{DVSin}_{i,t}$ is regressed on $\texttt{DVSin}_{i,t}$, while Equation 2 implies that $\text{Attacks}_{i,t}$ is regressed on $\text{Attacks}_{i,t}$. Because `xtabond2` is a GMM estimator, the model in Equation 1 will run, but the coefficient on $\texttt{DVSin}_{i,t}$ will converge to one while all other coefficients will converge to zero. The estimates in Table 1 and the note in Figure 4 thus provide sufficient assurance that the lag specifications given in Equations 1 and 2 in WEA23 are typos.

[28] Though a handful of country-years are assigned average sentences of ≤ 1 or < 1, this is still not zero.

claimed.

4 Conclusion

Governments spend hundreds of billions of dollars on anti-terrorism enforcement globally each year. The United States spent 15% of the more than $18 trillion dedicated to discretionary spending from 2002-2017 – a total that sums to $2.7 trillion – on counterterrorism efforts, including more than $175 billion in 2017 alone (Heeley et al. 2018). Given the sheer amount of money and lives at stake, research that aims to inform policymakers on the most (cost-)effective ways to combat terrorism through legal enforcement should arise from credible analyses of solid evidence.

My reproducibility and robustness analysis of WEA23 reveals severe issues. The main estimates in the article are not reproducible, and reasonable robustness checks render all estimates statistically insignificant. I also find irregularities in WEA23's replication data, which raise serious doubts about the credibility of the empirical results.

5 Tables

Replication Package Item	**Fully**	**Partial**	**No**
Raw data provided			✓
Analysis data provided		✓	
Cleaning code provided			✓
Analysis code provided		✓	
Reproducible from raw data			✓
Reproducible from analysis data			✓

This table summarizes the contents of WEA23's replication package (Wolfowicz et al. 2023b).

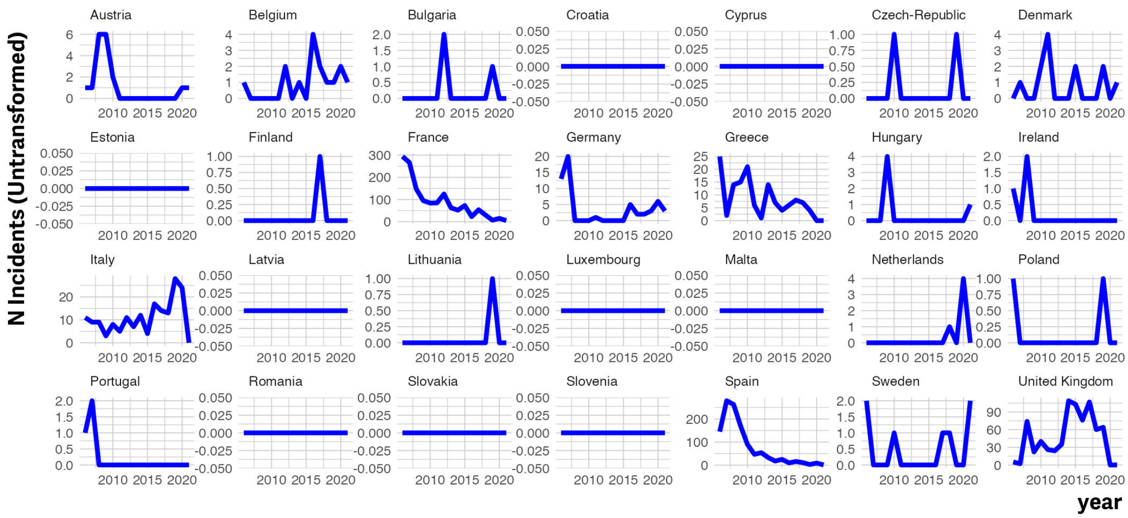
Table A: Reproducibility Summary

Estimate Type	(1) Published	(2) Direct Reproduction	(3) Alternate Reproduction	(4) Robustness Check
$\text{Att}_{i,t-1}$	0.929	0.62094	0.98572	0.76858
	(0.071)	(0.02890)	(0.00442)	(0.04143)
	[$P = 0.000$]	[$P = 0.000$]	[$P = 0.000$]	[$P = 0.000$]
$\text{Arr}_{i,t-1}$	-0.018	0.00001	-0.00297	-0.01069
	(0.006)	(0.00001)	(0.00067)	(0.01582)
	[$P = 0.002$]	[$P = 0.123$]	[$P = 0.000$]	[$P = 0.499$]
$\text{Char}_{i,t-1}$	0.016	-0.00010	0.00356	0.01922
	(0.003)	(0.00026)	(0.00073)	(0.00949)
	[$P = 0.001$]	[$P = 0.685$]	[$P = 0.000$]	[$P = 0.043$]
$\text{Conv}_{i,t-1}$	-0.113	-0.00185	-0.00372	-0.03513
	(0.045)	(0.00152)	(0.00158)	(0.03520)
	[$P = 0.012$]	[$P = 0.222$]	[$P = 0.018$]	[$P = 0.318$]
$\text{Sent}_{i,t-1}$	0.012	0.00022	-0.00001	0.00082
	(0.005)	(0.00011)	(0.00006)	(0.00068)
	[$P = 0.016$]	[$P = 0.047$]	[$P = 0.825$]	[$P = 0.232$]
N	420	112	420	180
Time FE	✓	✓	✓	✓
No-Event Dummy	✓	✓	✓	✓
Country FE				✓
Undisclosed Imputations	✓	✓	✓	
$\text{Att}_{i,t}$	$\texttt{DVSin}_{i,t}$	$\texttt{DVSin}_{i,t}$	$\texttt{DVSin}_{i,t}$	$\text{AttackRate}_{i,t}$
$\text{Arr}_{i,t}$	?	$\text{ArrestRate}_{i,t}$	$\texttt{NewArrest}_{i,t}$	$\text{ArrestsPC}_{i,t}$
$\text{Char}_{i,t}$	?	$\text{ChargeRate}_{i,t}$	$\texttt{CharSin}_{i,t}$	$\text{ChargesPC}_{i,t}$
$\text{Conv}_{i,t}$	?	$\texttt{convictionrate}_{i,t}$	$\texttt{ConvictionSin}_{i,t}$	$\text{ConvictionsPC}_{i,t}$
$\text{Sent}_{i,t}$	$\texttt{sentence}_{i,t}$	$\texttt{sentence}_{i,t}$	$\texttt{sentence}_{i,t}$	$\texttt{sentence}_{i,t}$

This table provides the regression estimates for the published estimates (Column 1), my computational reproductions (Columns 2-3) and my robustness check (Column 4) for Model Ia in Table 1, whose primary specification is given by Equation 1. Standard errors are stated in parentheses; in Column 1, I compute standard errors by dividing the range of the reported confidence interval in the published estimates by 3.92, while robust standard errors are computed for the remainder of the models using Windmeijer's (2005) finite-sample correction for two-step covariance matrices. P-values are reported in brackets. Regression estimates are accompanied by observation counts and indicators for whether time fixed effects, a no-event dummy ($1\left[\text{Att}_{i,t} - 0\right]$), and/or country fixed effects are included in the model. Column 4 is estimated with undisclosed imputations removed (see Sections 3.7 and 3.8). The bottom five rows of the table state the specific name of each variable used in the model, when known.

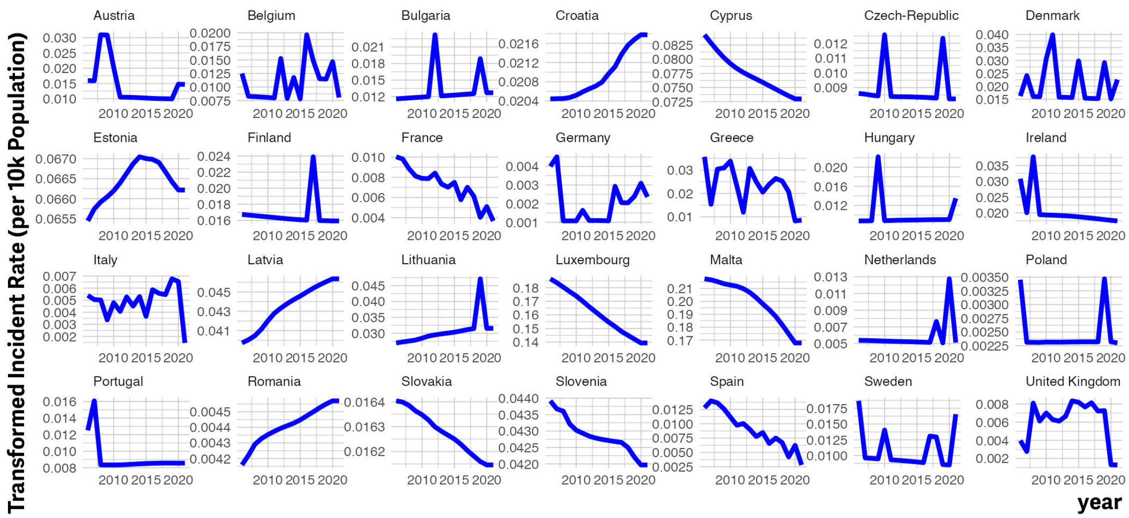
Table B: Published Estimates, Computational Reproductions, and Robustness Checks for Table 1, Model Ia

6 Figures



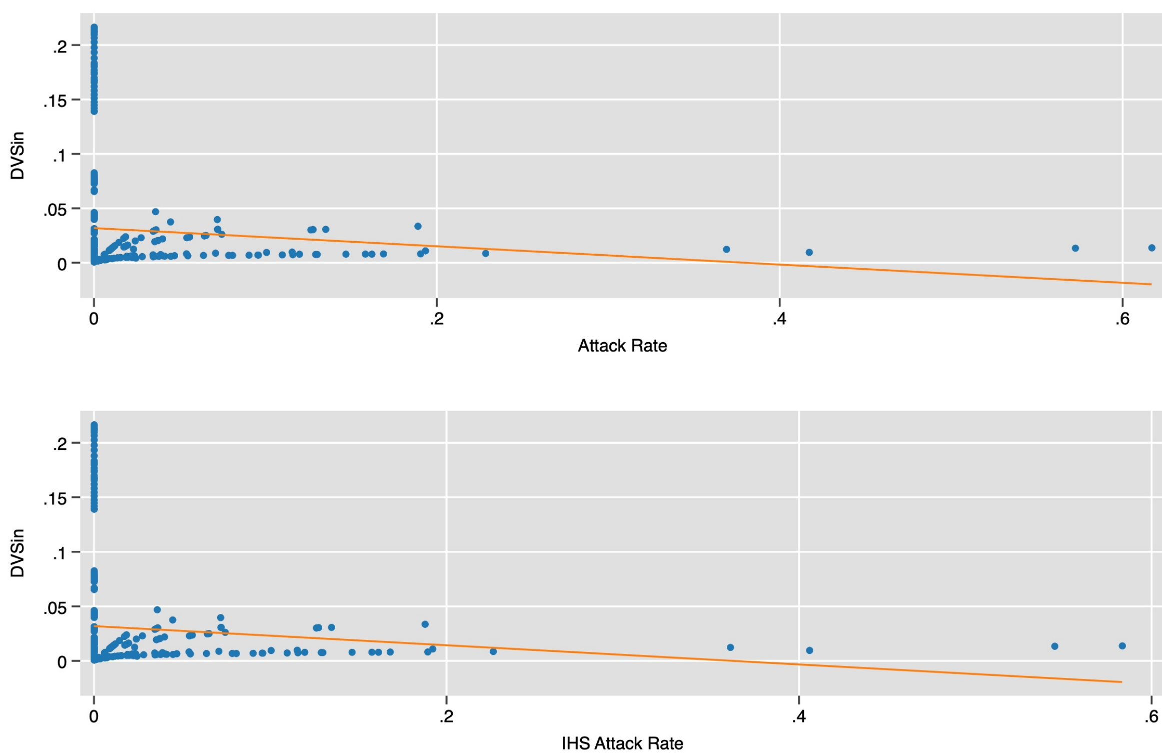
Note: This figure is exactly reproduced using WEA23's replication code for producing Figure 1 in their paper.

Figure A: Exact Reproduction of WEA23's Figure 1



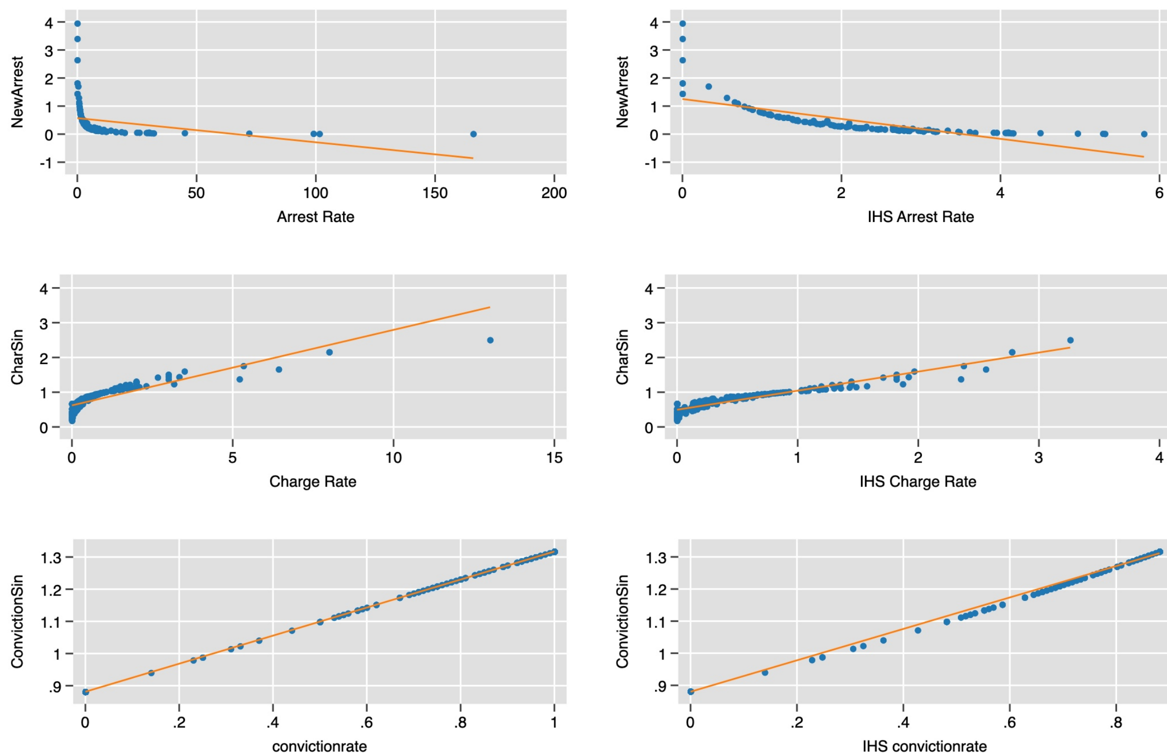
Note: This figure is exactly reproduced using WEA23's replication code for producing Figure 2.

Figure B: Exact Reproduction of WEA23's Figure 2



Note: The top and bottom graphs respectively display scatterplots and OLS lines of best fit between `DVSin` (as stored in WEA23's replication dataset) and both AttackRate$_{i,t}$ (calculated according to Equation 2) and AttackRateIHS$_{i,t}$ (calculated according to Equation 3).

Figure C: `DVSin`'s Relationships with Attack Rates and IHS Attack Rates



Note: The graphs display scatterplots and OLS lines of best fit. The top row plots the relationships between `NewArrest` and both the ArrestRate$_{i,t}$ and its IHS transformation. The middle row plots the relationships between `CharSin` and both ChargeRate$_{i,t}$ and its IHS transformation. The bottom row plots the relationships between `ConvictionSin` and both `convictionrate` and its IHS transformation.

Figure D: Relationships between 'Sin' Variables and Underlying Variables

References

Allen, Grahame, Matthew Burton, and Alison Pratt (2022). *Terrorism in Great Britain: The statistics*. Briefing. House of Commons Library. URL: https://researchbriefings.files.parliament.uk/documents/CBP-7613/CBP-7613.pdf (visited on 04/18/2024).

Bellemare, Marc F. and Casey J. Wichman (2020). "Elasticities and the inverse hyperbolic sine transformation". In: *Oxford Bulletin of Economics and Statistics* 82.1, pp. 50–61. DOI: 10.1111/obes.12325.

Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the false discovery rate: A practical and powerful approach to multiple testing". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 57.1, pp. 289–300. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

Brodeur, Abel et al. (2024). "Reproduction and replication at scale". In: *Nature Human Behaviour* 8.1, pp. 2–3. DOI: 10.1038/s41562-023-01807-2.

Chen, Jiafeng and Jonathan Roth (2024). "Logs with zeros? some problems and solutions". In: *The Quarterly Journal of Economics* 139.2, pp. 891–936. DOI: 10.1093/qje/qjad054.

Europol (2007). *EU Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/tesat2007_1.pdf (visited on 04/12/2024).

– (2008). *EU Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/tesat2008_1.pdf (visited on 04/12/2024).

– (2009). *EU Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/tesat2009_1.pdf (visited on 04/12/2024).

– (2010). *EU Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/tesat2010_1.pdf (visited on 04/12/2024).

– (2011). *EU Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/te-sat2011_0.pdf (visited on 04/12/2024).

– (2012). *EU Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/europoltsat.pdf (visited on 04/12/2024).

– (2013). *EU Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/europol_te-sat2013_lr_0.pdf (visited on 04/12/2024).

– (2014). *European Union Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/europol_tsat14_web_1%20%281%29.pdf (visited on 04/12/2024).

– (2015). *European Union Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/p_europol_tsat15_09jun15_low-rev.pdf (visited on 04/12/2024).

– (2016). *European Union Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/europol_tesat_2016.pdf (visited on 04/12/2024).

– (2017). *European Union Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/sites/default/files/documents/tesat2017.pdf (visited on 04/12/2024).

Europol (2018). *European Union Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/tesat_2018_1.pdf (visited on 04/12/2024).

– (2019). *European Union Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/tesat_2019_final.pdf (visited on 04/12/2024).

– (2020). *European Union Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/european_union_terrorism_situation_and_trend_report_te-sat_2020_0.pdf (visited on 04/12/2024).

– (2021). *European Union Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/tesat_2021_0.pdf (visited on 04/12/2024).

– (2022). *European Union Terrorism Situation and Trend Report*. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/europol_tsat14_web_1%20%281%29.pdf (visited on 04/12/2024).

Heeley, Laicie et al. (2018). *Counterterrorism Spending: Protecting America While Promoting Efficiencies and Accountability*. Report. Stimson Center. URL: https://www.stimson.org/wp-content/files/file-attachments/CT_Spending_Report_0.pdf (visited on 04/19/2024).

Holm, Sture (1979). "A Simple Sequentially Rejective Multiple Test Procedure". In: *Scandinavian Journal of Statistics* 6.2, pp. 65–70.

Roodman, David (2009). "How to do xtabond2: An introduction to difference and system GMM in stata". In: *The Stata Journal* 9.1, pp. 86–136. DOI: 10.1177/1536867x0900900106.

Windmeijer, Frank (2005). "A finite sample correction for the variance of linear efficient two-step GMM estimators". In: *Journal of Econometrics* 126.1, pp. 25–51. DOI: 10.1016/j.jeconom.2004.02.005.

Wolfowicz, Michael et al. (2023a). "Arrests and convictions but not sentence length deter terrorism in 28 European union member states". In: *Nature Human Behaviour* 7.11, pp. 1878–1889. DOI: 10.1038/s41562-023-01695-6.

– (2023b). *Data and code files for "Arrests and convictions but not sentence length deter terrorism in 28 EU-member states"*. Data set V1. Zenodo. DOI: 10.5281/zenodo.8196717. (Visited on 04/12/2024).

– (2023c). "Not so different after all: Increased arrests and convictions (but not sentence length) deter terrorism". In: *Research Square*. DOI: 10.21203/rs.3.rs-2504881.