Statistical Sightings of Better Angels:
Analysing the Distribution of Battle Deaths
in Interstate Conflict over Time

June 2019

Céline Cunen¹, Nils Lid Hjort¹, and Håvard Mokleiv Nygård²

¹Department of Mathematics, University of Oslo
²Peace Research Institute of Oslo (PRIO)

Abstract
Have great wars become less violent over time, and is there something we might identify as the long peace? We investigate statistical versions of such questions, by examining the number of battle deaths in the Correlates of War dataset, with 95 interstate wars from 1816 to 2007. Previous research has found this series of wars to be stationary, with no apparent change over time. We develop a framework to find and assess a change-point in this battle deaths series. Our change-point methodology takes into consideration the power-law distribution of the data, models the full battle death distribution, as opposed to focusing merely on the extreme tail, and evaluates the uncertainty in the estimation. Using this framework, we find evidence that the series has not been as stationary as past research has indicated. Our statistical sightings of better angels indicate that 1950 represents the most likely change-point in the battle deaths series – the point in time where the battle deaths distribution changed for the better.

Key words: battle deaths, change-point, confidence curves, interstate wars, Korean War, power law tails.

1 Introduction
Is the world becoming more peaceful? The question is both deceptively simple and quite controversial. Authors such as Gat (2006), Goldstein (2011), and Pinker (2011) have argued that the world is becoming steadily more peaceful, and a multidimensional quilt of research has contributed pieces of layers with similar stories and conclusions. Part of these arguments concern wars and armed conflicts, and there, the concept of the long peace (Gaddis 1989) has gained the weight of repeated respectful use, to signal the relatively few large interstate wars in the time after the 2nd World War (WW2). The more or less implicit change-point of war history in these arguments has been that since 1945 the world has changed.

¹See for instance the collection of review articles in the 50th Anniversary issue of the Journal of Peace Research (Volume 51, Issue 1).
While the empirical pattern constituting the long peace is not in itself disputed, some recent investigations have questioned whether the pattern can be said to constitute a statistically established trend; see for instance Cirillo & Taleb (2016); Clauset (2017, 2018). Could this long period of relative peace simply be a random occurrence in an otherwise homogeneous war-generating process, or does it represent a significant change, a trend towards peace? Cirillo & Taleb (2016) and Clauset (2017, 2018) answer the last question negatively: they find that the long peace is not a sufficiently unusual pattern when considering the variability inherent in long-term datasets of historical wars. The question investigated by these authors is essentially statistical in nature, and we follow in the same vein. We approach a similar question, with similar data, but with somewhat different statistical tools.

We see our contribution as two-fold. First, we introduce a set of statistical methods to the peace research community, some of them new. We have attempted to make the presentation of the methods accessible to most peace researchers, and have strived to push technical details to the appendix. Second, we present new results and conclusions, that partly challenge previous works, and that may generate hypotheses that can form the basis of future investigations. We will present evidence that a sequence of war sizes from the last two centuries is not entirely homogeneous, contrary to previously mentioned works by Cirillo & Taleb (2016) and Clauset (2017, 2018). In this sequence of observations, we find that the point of maximal change is in 1950, i.e. corresponding to the Korean war. Thus we differ from parts of the literature by not focussing exclusively on WW2 as the potential point of change, but by applying change-point methodology to investigate distributional changes in a time-series of wars. We also investigate the role of covariates, in particular democracy.

Our article is structured as follows. In Section 2 we draw on the existing literature to sharpen the question we will be considering. We also present the data we will use, and discuss the overall analysis framework. Then, we present the relevant statistical methods in more detail in Section 3. In Section 4 we give our main results: first we perform a homogeneity test, as this indicates non-homogeneity we go forward with change-point methodology, and crucially also present the degree of change. Finally, we investigate the effect of democracy. We discuss our findings in Section 5. There we examine the robustness of our approach to various choices, its relationship with previous works and also consider some potential theoretical mechanisms.

2 Modelling wars

Efforts to uncover trends in armed conflict have a long history and date back at least to the seminal contributions of Lewis Fry Richardson (1948, 1960). Richardson assembled datasets of historical wars, and sought to uncover long-term patterns by statistical modelling of various quantities, for instance the time between wars and also the number of fatalities in each war. We will consider datasets of that type, specifically the Correlates of War (CoW) interstate conflict dataset (Sarkees & Wayman, 2010), see Figure 2.1, which we discuss in a bit more detail below. For now, consider a general war dataset consisting of

\[(x_i, z_i) \text{ for } i = 1, \ldots, n,\]  

(2.1)

for a number n of historical wars, where \(x_i\) is the onset time of war i and \(z_i\) the number of fatalities; henceforth we will call \(z_i\) the size of war i. Richardson’s analyses of historical wars led him to two important statistical insights:
(i) the between-war times $d_i = x_i - x_{i-1}$ can be modelled as independent and identically distributed (i.i.d.), following a simple exponential distribution;

(ii) the war sizes $z_i$ can be modelled as i.i.d. with a power-law distribution.

Figure 2.1: War sizes and onset times for the 95 wars in the CoW data; here the war sizes $z_i$ are on the log10 scale.

Both the time between wars and the size of each war are relevant for investigating whether the world has become more peaceful. A peaceful world could be characterised by fewer wars (i.e. longer time between each war), smaller wars, or both. Potential trends in the number of interstate wars have been studied by for instance Harrison & Wolf (2012), Gleditsch & Pickering (2014), Cirillo & Taleb (2016), Braumoeller (2017) and Clauset (2018). Harrison & Wolf (2012) claim that interstate wars have become more frequent over time, while Gleditsch & Pickering (2014) criticise their approach and claim that wars are in fact becoming less frequent. Clauset (2018) finds that the time between wars in the CoW data is adequately modelled by a simple exponential distribution, a finding that supports insight (i) of Richardson above. Clauset (2018) takes this finding as an indication of a lack of trend in the war timings data. In the appendix (Section A) we provide a short investigation of the between-war waiting times $d_i$ in the CoW dataset and find that the observed waiting times are more consistent with an exponential-gamma mixture model than with a simple exponential model. This indicates that the waiting times in the CoW dataset are more variable than expected under an exponential model, but does not point towards any particular time-trend. While we consider this finding interesting and worthy of attention in future modelling of war sequences, we will leave the waiting times aside for the rest of the article and focus our attention on the war sizes.

Richardson’s second insight has possibly received even more attention than the first one. Power laws are a particular class of probability distributions, with

$$P(Z > z) \propto z^{-\theta} \text{ for all large } z,$$

and a positive parameter $\theta$. This means that the probability of observing an event, in our case a war, of size larger than $z$ is inversely proportional to $z$ raised to $\theta$. If $\theta$ is large this probability quickly decreases with $z$, but if $\theta$ is smaller $P(Z > z)$ can stay considerable even for large $z$. 

3
This last characteristic is sometimes referred to as the ‘fat-tailed’ property and entails a nonnegligible probability of observing truly enormous events. Often the power law distribution is only appropriate for observations larger than some threshold $z_0$, a point we will return to in Section 3.2.

Richardson’s insights concerning power laws are discussed by Pinker (2011) in his international best-seller *The Better Angels of Our Nature*. There, he argues that violence in a wide sense, including crime, torture, animal cruelty – and war, has declined. Interestingly, power laws also form the basis of empirical investigations that challenge Pinker’s conclusions about the decline of war and the long peace. In Cederman et al. (2011) a sequence of 118 war sizes from 1495 till 1997 is modelled with power law distributions. The authors find a shift in the power law parameter in 1789, indicating larger wars after that year compared to the period before. Cirillo & Taleb (2016) build their own database of war deaths from year 1 to the present. They use statistical models with power law tails and find that their dataset is well enough described by a single, stationary model. Clauset (2017, 2018) examines the CoW data discussed below and argues that it is still too early to confidently assert, from history and data alone, that the long peace is safely in place. Clauset (2017, 2018) models the size of interstate wars with power laws, and finds that he cannot reject the null hypothesis of no change. Indeed, he argues that the current trend would have to persist for 150 years until we could statistically claim that the world had become more peaceful.

Now we have decided on a quantity of interest, war sizes, and have found a class of appropriate statistical distributions to model this quantity. Still, there is a major question to resolve: should we normalise the war sizes by population size or should we consider the absolute number of fatalities instead? Here, normalisation refers to dividing the number of fatalities by the population size, typically the world population. Pinker (2011) forms most of his arguments around relative quantities, such as deaths per 100,000. Falk & Hildebolt (2017) criticise this normalisation choice because they claim that the risk of dying in battle is negatively related to the size of the population. Clauset (2017, 2018) discusses the choice of normalisation in some length, and decides to analyse the absolute numbers. The choice of normalisation in fact translates into different questions: are we interested in making claims about the absolute sizes of wars? Or in the risk of dying in wars? And in the latter case, with respect to which segment of the population should this risk be defined? All these questions are valid and interesting, but naturally the answers to one of them will not be directly relevant for the others. We have chosen to consider the absolute numbers. For the proponents of the long peace theory this is a conservative choice since normalising by world population inflates the size of ancient wars compared to more recent wars.

Further, there is a choice between different datasets. Naturally, we would prefer a dataset stretching as far as possible back in time, with measurements of high quality and constructed with careful and precise definitions. The previously mentioned study by Cederman et al. (2011) combines data from Levy (1983), the CoW project (Singer & Small 1994) and the PRIO/UCDP Armed Conflict Database (ACD) (Gleditsch et al., 2002). The dataset has a long timespan, but is unfortunately limited to wars involving ‘major powers’. Some datasets distinguish between inter- and intrastate wars, see Sarkees et al. (2003) and Lacina et al. (2006) for discussions on the appropriate analysis of these different types of wars. The quality of the reported battle deaths number can also be an issue. Even for recent wars involving developed countries the estimates

---

Clauset & Gleditsch (2018) provide a longer and more holistic overview of these and other issues pertaining to the study of trends in conflict.
of the number of battle deaths can be contested. The Falklands war, for instance, is included in the CoW interstate wars dataset with 1001 battle deaths, even though the actual number is most likely closer to 900 (Reiter et al., 2016). See also Obermeyer et al. (2008) and Spagat et al. (2009) for opposing views on the appropriate method for measuring battle deaths. We have used the Correlates of War (CoW) interstate conflict dataset (Sarkees & Wayman, 2010). This dataset contains onset dates $x_i$ and the number of battle deaths $z_i$ for all interstate wars with more than 1000 battle deaths in the period 1816 to 2007; comprising a total of 95 wars. The dates $x_i$ range from 1823.27 (the Franco-Spanish war) to 2003.22 (invasion of Iraq). Figure 2.1 displays these data, with $z_i$ on the log10-scale. The choice of the CoW dataset is motivated by its widespread use (Clauset, 2017, 2018; Fagan et al., 2018; Spagat & Weezel, 2018), which enables comparisons with other approaches. Also, the CoW dataset is considered to be of good quality, despite the issues mentioned above.

Finally, there are several different statistical frameworks for assessing whether a certain sequence of observations, war sizes in our case, supports a trend, or not. The possible options include regression models with respect to time, homogeneity tests and change-point analyses. We have not investigated regression models as these would impose too much of a constraint on the type of change present (also a quick look at the data clearly indicates that there is no simple linear time trend in the CoW data).

Homogeneity tests are a general class of methods which aim at testing a null hypothesis of stationarity, i.e. to test whether the observed sequence is consistent with a single, stationary statistical model or whether there is sufficient deviation from the model as to indicate that there has been a change. Most of the results in Clauset (2017, 2018) are based on tests of homogeneity, where Clauset does not find sufficient evidence to reject the null hypothesis of no change. Tests of homogeneity seem attractive because they can potentially discover many types of deviations from the stationary model. However for partly the same reason, they can often have low power in discovering actual changes. There are many homogeneity tests to choose between, which differ in for instance the assumptions made, the choice of test statistic and the choice of alternative hypothesis; see Hjort & Koning (2002), Cunen, Hermansen & Hjort (2018) for partial reviews and methods. We present a general homogeneity test in Section 3.1.

If the null hypothesis of homogeneity is rejected, there may be reasons to believe that the data are inconsistent with a completely stationary model. The rejection of the hypothesis does not necessarily give any indication on where the change took place, nor what type of changes the data support. Change-point analysis is a framework for investigating a certain type of ‘trend’: an abrupt change in the distribution of the data, with particular emphasis on where the change took place. There is a long tradition in social and political science for studying shifts in history, and for examining conditions for the potential for shifts; see e.g. Tilly (1995), and also Marx (1871), Spengler (1918), and for instance Beck (1983), Mitchell, Gates & Hegre (1999), Western & Kleykamp (2004), Spiriling (2007) and Blackwell (2018). Change-point methods have been applied to sequences of war sizes in Cederman et al. (2011), and very recently in Fagan et al. (2018). We will return to these two contributions in the discussion.
3 Methods

We construct a nonparametric homogeneity test which we present in Section 3.1. Since this test indicates non-homogeneity (see results in Section 4.1), we proceed with our change-point framework. First, we consider parametric models for the war sizes in Section 3.2, before presenting our change-point method in Section 3.3. In Section 3.4 we explain the inclusion of covariates.

3.1 Testing constancy over time

Suppose a sequence of observations $y_1, \ldots, y_n$ is registered over time, and that one wishes to query the null hypothesis $H_0$ that the distribution generating the sequence has remained constant, against the alternative that somewhere a change has taken place. Assume $\mu$ is a parameter of particular interest, like the median or standard deviation, with $\hat{\mu}_{a,b}$ the estimate of this quantity based on the stretch of data $y_a, \ldots, y_b$. For each candidate position $\tau$, inside a relevant pre-defined interval of time $[c,d]$, consider the relative difference in estimated $\mu$, to the left and to the right, via

$$H_n(\tau) = \frac{\hat{\mu}_L - \hat{\mu}_R}{(\hat{\kappa}_L^2/\tau + \hat{\kappa}_R^2/(n-\tau))^{1/2}}$$

for $\tau = c, c+1, \ldots, d-1, d$. (3.1)

Here $\hat{\mu}_L = \hat{\mu}_{1,\tau}$ and $\hat{\mu}_R = \hat{\mu}_{\tau+1,n}$, along with $\hat{\kappa}_L$ and $\hat{\kappa}_R$ being estimates of the relevant standard deviations, to the left and to the right, in the usual setup where $\hat{\mu}_{a,b}$ is approximately normal with variance of the form $\kappa^2/(b-a+1)$. The function $H_n(\tau)$ can be plotted for all potential $\tau$ values, and also provides natural test statistics for $H_0$, for instance $H_n,\text{max} = \max_{c \leq \tau \leq d} |H_n(\tau)|$, along with one-sided versions. The null hypothesis of homogeneity is rejected if $H_n(\tau)$ takes values sufficiently far from zero. In addition, the plot of $H_n(\tau)$ will indicate the position $\hat{\tau}$ at which the plot is farthest away from zero, which may serve as an estimate of the change-point (but from an entirely different perspective than the change-point method we discuss in Section 3.3).

Importantly, the $H_n$ plot may be utilised for the one-sided case where a change is assumed to have a given direction, on a priori grounds, thus yielding bigger detection power than with a two-sided version. Also, the method works for nonparametrically defined $\mu$. In order to find the p-value for the test, one needs to work out the distribution of the $H_n$ process. We present these derivations in Section B.1 of the appendix. There we also investigate a different homogeneity test based on a weighted Kolmogorov-Smirnov statistic, see Section B.2.

3.2 Models with power law tails

In order to use our change-point method we need a parametric model for the war sizes, $z_i$. As discussed in Section 2 we want to use a model with power law behaviour. One general option is to use the power law distribution directly, see (2.2). For most datasets, the power law distribution will not fit well for the entire dataset, but only for observations larger than a certain threshold, i.e. $z_i \geq z_0$ has a density proportional to $z_i^{-(\theta+1)}$. Then, one needs to estimate both the parameter $\theta$ and the tail-index threshold $z_0$. We investigate this approach in Section E.1 of the appendix; related approaches are used in Clauset (2017, 2018). This model is simple to use, but does not directly utilise the observations below the threshold $z_0$ and may therefore entail some loss of information compared to the next option. In the following, we will refer to this model as the ‘simple power law’ model.

Another option is to model the entire dataset, which only has wars of sizes 1001 and more (see appendix Section D), with a distribution that fulfils the power law requirement in the tails.
Generally speaking, the distribution function $F(z)$ for the $z_i$ is said to have power law tails, with power index $b$, if $z^b \{1 - F(z)\}$ tends to a positive constant as $z$ increases. One such model is the inverse Burr distribution, which also goes by the name of the Dagum distribution, taking

$$F(z; \mu, \alpha, \theta) = P\{Z \leq z\} = \left[\frac{(z - 1001)/\mu}{((z - 1001)/\mu)^\theta + 1}\right]^\alpha \quad \text{for } z \geq 1001,$$  

(3.2)

with parameters $(\mu, \alpha, \theta)$ to be estimated from the 95 wars. When $z$ increases we have $F(z) \approx 1 - \alpha (\mu/z)\theta$; thus $\theta$ plays the role of the power index, similarly to its namesake in the simple power law distribution above.

There are several other distributions with power law tails. The choice of distributions should ideally not influence the reported results to a great extent, as long as the chosen model has a reasonably good fit to the data. In the appendix, we examine goodness-of-fit, some model selection with the focussed information criterion, and also report results using other parametric models; see Section E.

### 3.3 Change-point methods

When faced with a sequence of observations, change-point methodology is used to search for when the point of maximal distributional change occurs. More formally, we have observations $z_1, \ldots, z_n$ from some parametric model, say $f(z, \gamma)$, where $\gamma$ is of dimension $p$. Assume that there is a change-point $\tau$ in the sequence, with parameter $\gamma_L$ for $i \leq \tau$ and $\gamma_R$ for $i \geq \tau + 1$. The aim of a change-point analysis is to estimate $\tau$ and, importantly, to assess the uncertainty around it. Subsequently, one should also assess the degree of change associated with the change-point, in order to investigate the magnitude and direction of the change, and to assess whether the change we have discovered is significant, in the sense of having any practical importance.

There are many ways in which to search for a change-point in a sequence of data; see Frigessi & Hjort (2002) for a broad introduction to a special journal issue on discontinuities in statistics. Here we employ change-point machinery developed in Cunen, Hermansen & Hjort (2018), both for spotting a potential change-point and, crucially, for assessing its uncertainty. To assess uncertainty and present our result, we use confidence curves, see Schweder & Hjort (2016). The confidence curves can be understood as graphical generalisations of confidence intervals. They present the uncertainty at all levels of confidence, instead of just a single confidence interval at some arbitrary level of confidence (typically 95%). See Section 4 for more on the interpretation of confidence curves.

In Section C of the appendix we provide a short technical overview of the change-point method we have used. The version of the method used here only allows for a single change-point in the sequence of data. Importantly, the method involves maximum likelihood estimators of the model parameters, $\hat{\gamma}_L$ to the left and $\hat{\gamma}_R$ to the right, and of the change-point parameter $\hat{\tau}$. The confidence curve $cc(\tau)$ is based on the deviance function and its construction requires computer simulations. Ideally, the results presented here should not be too sensitive to the choice among various change-point methods. The chosen method is easy to use, highly flexible, and relies on a natural extension of general likelihood theory to change-point parameters. It can be used in connection with any parametric model for the data and allows for changes in one, some, or all of the model parameters $\gamma_L$ and $\gamma_R$. Thus, it allows the user to discover more complex changes than simple jumps in the mean level (which parts of the change-point literature are constrained to). The framework we
The change-point method of Cunen, Hermansen & Hjort (2018) also allows us to construct confidence curves for the degree of change associated with the change-point. The degree of change is a one-dimensional parameter, called $\rho$, defined as a function of the model parameters on both sides of $\tau$, and meant to capture the size and direction of the change. Usually it will be in the form of a ratio or a difference; here we will study the ratio between quantiles of war sizes on each side of $\tau$. Confidence curves for the degree of change, $cc(\rho)$, are displayed in the result section. Importantly, $cc(\rho)$ takes into account the uncertainty in the change-point position. The confidence curves for the degree of change can therefore be considered an implicit homogeneity test. The change-point method described here always gives a point estimate for the change-point position, but if the degree of change analysis indicates that the magnitude of the change is very small, or highly uncertain, there is no reason to argue that there really has been a shift in distribution. Conversely, if the degree of change analysis indicates a change of large and significant magnitude, one may put faith in the existence of a change.

In our analysis, we will use the change-point method briefly discussed here along with the inverse Burr model described in the previous section. In addition to the choice of distribution, the modeller also needs to decide on which parameters of the distribution should be allowed to be (potentially) influenced by the change-point. For the model (3.2), we allow $\theta$ and $\mu$ to change, but assume the same $\alpha$ across the change-point. We then end up with a total of six parameters to estimate: the change-point $\tau$, along with $(\alpha, \mu_L, \theta_L, \mu_R, \theta_R)$.

### 3.4 Covariates

The change-point method above is sufficiently general to support the inclusion of covariates influencing the model parameters, for example democracy scores, as we will see. For simplicity of presentation, we will present the inclusion of a single covariate to the inverse Burr model described above; in the appendix we give a more general treatment (Section G).

Assume that we have covariate information $w_i$ for each war. In this illustration, the covariate is the mean democracy score of the countries involved in each war, measured the year before the war started. To measure democracy, we utilise the Polity index from the Polity IV dataset (Marshall & Jaggers, 2003). The Polity index scores regimes on a $-10$ to $10$ scale, where $-10$ are the most autocratic regimes and $10$ the most democratic. The covariate will be negative when a war involves mostly autocratic regimes, and large and positive if a war involves only democracies. Here, we will let the covariate influence the scale parameter $\mu$ of the inverse Burr:

$$
\mu_{L,i} = \mu_{L,0} \exp(\beta_L w_i) \quad \text{and} \quad \mu_{R,i} = \mu_{R,0} \exp(\beta_R w_i).
$$

Note that some of the wars have missing democracy scores. We remove these observations and end up with 90 wars for this analysis. The full model has now become moderately complex, with parameters $\theta_L, \mu_{L,0}, \beta_L$ to the left, $\theta_R, \mu_{R,0}, \beta_R$ to the right, a common $\alpha$, in addition to the change-point $\tau$.

When introducing covariates in this change-point model, there are some issues to consider. First, one can either assume that the covariate effect has changed across the change-point, or that it has remained constant (so $\beta_L = \beta_R$). This choice might depend on prior knowledge, or be
decided on based on some model selection criteria. Secondly, one must be aware that inclusion of covariates might alter the change-point inference (compared to a model without covariates).

4 Results

4.1 Testing constancy

For the sequence of log-battle-deaths \( y_i = \log z_i \) for \( i = 1, \ldots, n = 95 \), we may compute, display, and analyse \( H_n \) plots of \( 3.1 \) for any relevant choice of focus parameter \( \mu \). Figure 4.1 displays \( H_n \) plots for the median \( F^{-1}(0.50) \) and upper quartile \( F^{-1}(0.75) \), with maxima 1.621 and 2.675, respectively. When looking at the median level we cannot reject the null hypothesis of homogeneity at any ordinary level. For the upper quartile, however, the maximum of 2.675 corresponds to a p-value of 0.034. This p-value is computed using the theory from Section 3.1 with a one-sided version of the test statistic, since we judge it a priori clear that the battle death distribution has not gone up after WW2. In order to compute the test-statistic, we also need to choose a time range, we use \( c = 1934 \) and \( d = 1987 \).

The p-values, for monitoring the no-change hypothesis with respect to quantiles, become even smaller for higher quantiles than 0.75, and is e.g. 0.009 for \( q = 0.80 \). Thus the battle-death distribution has clearly not remained constant over time. More specifically, plots such as those in Figure 4.1 reveal that there are changes in the upper parts of the distribution, but not necessarily in the lower parts. Also, the max of \( H_n \), for the case of the 0.75 quantile, is attained for the start of the Korean war, 1950.483.

Figure 4.1: The relative change \( H_n \) plot of \( 3.1 \), for the median \( F^{-1}(0.50) \) (red broken curve) and the upper quartile \( F^{-1}(0.75) \) (black full curve). The two horizontal curves give the 5% significance thresholds. The lower one indicates the point-wise threshold, while the upper gives the threshold for max\(_{c_0 \leq s \leq d_0} H(s) \), with time window corresponding to all wars between 1934 and 1987.
4.2 Change-point results

Our change-point method provides the maximum likelihood estimate for the change-point at $\hat{\tau} = 1950.483$. Thus, the point of maximal change in the parameters of the inverse Burr model is found between the 60 wars up to and including the Korean war on the one side and the 35 wars following the Korean war on the other side.

The full uncertainty around the point estimate is given by the confidence curve in Figure 4.2. The potential change-point values are on the horizontal axis, while the degree of confidence is on the vertical axis. The confidence curve hits zero at the point estimate (1950), and we can read off confidence intervals at all levels. Note that these intervals can consist of disjoint parts. Clearly there is some uncertainty in the change-point position; we see that the 95% confidence interval, indicated by the red horizontal line in the figure, encompasses the whole range of possible change-point values. The 80% interval encompasses only 30 war-onset-times however, most of them from 1939 to 1992, but with ‘gaps’. Note that the analysis places considerable confidence on three onset-war-times in the dataset in addition to the point estimate, especially 1965.103, the Vietnam war, 1939.669, i.e. WW2, and 1982.236, the Falkland war.

![Confidence curve for the change-point using the inverse Burr model](image)

Figure 4.2: Confidence curve for the change-point using the inverse Burr model [4.2], pointing to the Korean war 1950. The red dashed line corresponds to the 0.95 confidence level.

For the inverse Burr model [4.2], the estimated parameters are: $\hat{\alpha} = 0.499$, $\hat{\mu}_L = 43887$, $\hat{\theta}_L = 0.702$, $\hat{\mu}_R = 10940$, $\hat{\theta}_R = 1.022$ (see appendix Section D). We assess the direction and magnitude of the potential change by computing confidence curves for the degree of change. We examine the ratio between certain quantiles before and after the estimated change-point, $\rho_1 = \phi_{0.50,L}/\phi_{0.50,R}$ and $\rho_2 = \phi_{0.75,L}/\phi_{0.75,R}$, with $L$ and $R$ again referring to the parameters to the left and to the right of the change-point. When the bigger wars are of primary interest, the ratio $\rho_2$ of the upper quartiles would be more relevant to assess than the ratio $\rho_1$ of medians. With the inverse Burr we
have the following expression for the $100q\%$ quantile,

$$
\phi_q = 1001 + \mu \left( \frac{q^{1/\alpha}}{1 - q^{1/\alpha}} \right)^{1/\theta}.
$$

Here we use $q = 0.50$ and $q = 0.75$ to estimate the medians and the upper quartiles, respectively. Note that the number 1001 here simply serves to bring the quantiles back to the battle death scale. The point estimates via the inverse Burr are $\hat{\rho}_1 = 2.15$ and $\hat{\rho}_2 = 4.25$. The fitted median decrease from 10129 battle deaths pre 1950 to 4721 after the change-point. The upper quartile decreases from 63545 to 14943 battle deaths.

The left panel of Figure 4.3 gives the confidence curves for the two degree of change parameters described above. These are computed with the simulation based method described in Section C of the appendix. The confidence curves reveal that the ratio between upper quartiles is significantly larger than 1 on the 95% level, whereas the ratio of medians is larger than 1 only at somewhat lower confidence levels. Thus, the upper quartiles on each side of the potential change-point are significantly different on a 5% level. This analysis is not conditional on a given change-point value, but takes into account the full uncertainty in the change-point position.

4.3 Covariate results

We include the democracy covariate and allows the effect of democracy to change across the change-point. The inclusion of the covariate changes the point estimate of the change-point somewhat,
from 1950.483 to 1967.431 (the Six Day war). The Korean war is still given high confidence and we have therefore performed follow-up analysis taking the 1950.483 change-point as given. When it comes to parameters $\theta_L, \mu_L, 0, \theta_R, \mu_R, 0, \alpha$, estimates with precision correspond roughly to those found in Section 4.2. The most interesting parameters, in this context, are $\beta_L$ (estimate $-0.007$, 90% interval $[-0.202, 0.187]$) and $\beta_R$ (estimate $-0.163$, 90% interval $[-0.308, -0.018]$). The estimated $\beta_L$ is close to zero and its confidence interval covers zero, while the interval for $\beta_R$ indicates that the scale parameter decreases as the mean democracy score increases. The changing effect of democracy is reflected in Figure 4.4, showing the fitted median as a function of mean democracy on both sides of the change-point. Before 1950 the median number of battle deaths is almost constant across democracy scores, while after 1950 the median number of battle deaths decrease sharply with increasing democracy.

Figure 4.4: Regression with inverse Burr model and mean democracy as covariate. The left plot is on the z scale, while the right plot is on the log z scale. Each of the 95 wars is represented by a point. The lines are the fitted medians; in red before 1950 and in black after.

5 Discussion and concluding remarks

While some recent contributions argue that there is no clear evidence of change, neither in the sizes nor the times between interstate wars since 1816, we find evidence that a change in the distribution of war sizes has taken place, and that it may have happened in the years after WW2, rather than in 1945 which is the assumed change-point within the current literature. Specifically, the change in the parameters of this distribution manifests itself in smaller wars in the period after the change-point. In addition to enriching the long peace debate by generating hypotheses concerning the long-term characteristics of interstate wars, we have also introduced models and methods to the peace research literature.
Our claim rests upon two distinct analyses. First, we presented a nonparametric test of homogeneity. The test suggests that the sequence of war sizes has not been homogeneous when considering the higher quantiles of the war size distribution; see the results for the upper quartiles in Section 4.1. With this test the null hypothesis of no change is rejected at the 5% level. Second, we have conducted a change-point analysis. Here, we needed a parametric model for the data, and we found suitable models among the class of models with power law tails. The results from the change-point analysis are open to interpretation. On one hand there is considerable uncertainty in the change-point position: the 95% interval for \( \tau \) covers the entire range of possible change-point positions. Some readers will thus interpret Figure 4.2 as favouring the ‘no-change’ hypothesis. On the other hand, the figure also indicates that all the most likely candidates for the change-point positions are found either at or after WW2. Moreover, the degree of change analysis shows a significant decrease in battle deaths after the change, at least when considering the upper quartiles. On the whole, our analyses support a decrease in battle deaths at some point in the time-span we are considering. The exact position of the shift remains somewhat uncertain, but the most likely candidate is the Korean war.

We have also introduced the use of covariates – pointing towards further modelling efforts including mechanisms and explanations. In the rest of this section, we will discuss our findings on various levels. First, we will take a critical look at our approach and report on some robustness checks we have conducted. Then we will explore connections between our contribution and related papers, both in terms of methods and results. Finally, we will discuss our findings in light of the general peace research literature, and in particular consider some theoretical explanations.

5.1 Robustness of our approach

Statistical analyses require a series of assumptions and some level of abstraction to get from a real world question to a statistical question. Here, we return to some of the choices we discussed in Section 2 and attempt to assess their influence on our results.

For our statistical modelling we have been guided by previous works using power law distributions. There have been a few attempts to give a theoretical justification to the power law behaviour of war sizes, see for instance Cederman (2003), but for most authors, including Richardson, the power law models have been used as essentially descriptive models, i.e. as ‘lower dimensional representations’ allowing us to assess potential regularities given the inherent variation in the data. In that case, it is particularly important that the model fits well to the data, i.e. that the distribution of war sizes according to the model is close to the actually observed war size distribution. We have therefore conducted various goodness of fit evaluations, for instance the log-log plot in Figure 4.3. We see that the data in general have a good fit to the inverse Burr models on each side of the change-point. The clearest deviation from the model is found for the very largest wars, especially among those taking place after 1950. The three largest wars in this period have more battle deaths than expected under the model. This particular aspect of the data was not successfully accounted for by any of the models we considered, see the corresponding Figures in Section E of the appendix, and would necessitate a more complex model than those considered so far. We have also conducted some goodness of fit tests. On both sides of 1950, the observed data were consistent with having been generated by the fitted inverse Burr distributions (\( p_L = 0.64 \) and \( p_R = 0.23 \), see details in Section F in the appendix).

Several models within the class of distributions with power law tails provide adequate fit to
the data. In order to investigate the sensitivity of our results to the modelling assumptions, we present results for similar change-point analyses assuming two different models for the data in the appendix Section E, the simple power law distribution and the inverse Pareto distribution. The inverse Pareto, like the inverse Burr, models the full sequence of 95 war sizes, and we obtained very similar results to the ones presented in Section 4, the same point estimate for the change-point, \( \hat{\tau} = 1950.483 \), and similar looking confidence curves for both \( \tau \) and the parameters representing the degree of change. This is not surprising since the inverse Pareto distribution is just a simplification of the inverse Burr. With the simple power law model the results were somewhat different. Here, we needed to set the tail-index threshold \( z_0 \), and we used \( z_0 = 7061 \), see details in Section E.2. The subsequent change-point analysis then makes use of only the 51 wars larger than \( z_0 \). Using this model we found \( \hat{\tau} = 1965.103 \) as the point-estimate for the change, corresponding to the Vietnam war. We provide the full confidence curve in Section E.3, and it displays more uncertainty than we saw with the two other models (i.e. wider confidence intervals). In particular, the degree of change analysis indicates that the change was non-significant, in contrast with the analyses with the inverse Burr and inverse Pareto models. The increased uncertainty is related to the reduced sample size.

The different estimated change-points, for the full battle deaths distribution and only the large wars (the simple power law analysis), underscores an important aspect inherent to any change-point exercise. What constitutes a change-point when analysing some aspects of the available data will not necessarily be recognised as a change-point when examining other relevant data. Thus it should not be seen as a paradox that the Vietnam war in 1965 can be a change-point for the extreme tail of the battle death distribution, whereas perhaps the Korean war in 1950 is more of a change-point when examining more complex models involving the full battle death distribution.

Some readers might question our choice of using a change-point framework at all. As mentioned in Section 2, change-point methods assume a very particular form of change, an abrupt shift in the distribution generating the data. In the case of our change-point method, we have in addition assumed that only a single such shift takes place. Is it realistic to assume that the long peace emerged in that way? Hardly, but a single change-point model could be considered a reasonable approximation to various other patterns, for instance to more gradual changes. We are inclined to interpret the change-points we identify here as the culmination of a process that has unfolded over some time. This could apply to several of the mechanisms discussed in Section 5.3.

5.2 Connections to other analyses

There are several recent contributions with clear connections to our paper. Many of these also analyse the CoW interstate conflict dataset (Clauset 2017, 2018, Spagat & Weezel 2018, Fagan et al. 2018), while Cederman et al. (2011) and Cirillo & Taleb (2016) use datasets with a longer time span (from year 1494, and year 1, respectively). Cirillo & Taleb (2016) and Spagat & Weezel (2018) normalise the war sizes with respect to world population, while Clauset (2017, 2018) and Cederman et al. (2011) analyse the absolute numbers. Fagan et al. (2018) conduct analyses of both absolute and relative numbers. As expected, analyses using relative war sizes find a clearer decline of war than those focusing on absolute numbers.

Parametric models within the class of distributions with power law tails are used in Cederman et al. (2011), Cirillo & Taleb (2016), Clauset (2017, 2018), while Fagan et al. (2018) and Spagat & Weezel (2018) use nonparametric approaches. Clauset (2017, 2018) also investigated a certain
semiparametric model. The papers also differ in their choice of framework for investigating potential trends. Cirillo & Taleb (2016) and Clauset (2017, 2018) use types of homogeneity tests. Spagat & Weezel (2018) test for differences in the probability of observing wars of a certain size across specific potential years-of-change, namely 1945 and 1950. Initially, Cederman et al. (2011) also investigate a single, specific year-of-change, 1789, but the authors proceed by searching for a change-point along the full sequence of wars. Their approach differs from ours: they do not make use of a formal change-point method and their method does not provide any measures of uncertainty. Fagan et al. (2018) use a formal change-point method based on work by Killick et al. (2012) and Haynes et al. (2017), but their approach has several differences from ours. Their methodology relies on an algorithm which introduces distributional changes in the data sequence when the change-point leads to a sufficiently large increase in the fit to the data. The fit is measured by some cost function, which the user has to define, along with some penalty function (against introducing unnecessary change-points). In contrast, our change-point method treats the change-point as a parameter in the model and we therefore analyse it in a parallel manner as we would ordinary model parameters. Our method also allows investigating the magnitude and direction of the change, which Fagan et al. (2018) do not provide. On the other hand the method in Fagan et al. (2018) naturally allows for multiple change-points, while we have only investigated the introduction of a single potential change-point.

As mentioned in Section 2, Cirillo & Taleb (2016) and Clauset (2017, 2018) test a null hypothesis of stationarity, and do not find sufficient statistical evidence to reject it. Cederman et al. (2011) find a shift towards larger wars in 1789, while Spagat & Weezel (2018) find a shift towards smaller wars after 1950. Fagan et al. (2018) find multiple change-points in the sequence of wars between 1816 and 2007, notably around 1950. How can all these results be reconciled with each other? First of all it is important to realise that they do not necessarily stand in stark opposition to each other. The studies differ in the timespan considered and in the specific research question they treat, through their choices of for instance normalisation and statistical framework. Also, as usual, non-rejections do not imply that the null hypothesis is true. Further, the homogeneity tests used in Cirillo & Taleb (2016) and Clauset (2017, 2018) differ from the one we use in Section 4.1. The test in Clauset (2017, 2018) investigates whether the observed dataset as a whole is sufficiently different from simulated data from a stationary model. Our test focuses on specific aspects of the distribution of the data, specifically the upper quartile for instance. This sharper focus likely increases the statistical power. This focus is also shared by the degree of change investigations in Section 4.2 where we study changes in the medians and upper quartiles.

There is no clear consensus among the studies mentioned here, but neither is there any strong incompatibility, despite the differences in methodology. Each should be considered as providing some evidence to the full picture, which remains to be fully understood. In further work, we hope to draw on these studies and devote energy into the development of more realistic models for the underlying processes behind the war characteristics we observe, incorporating explicit theoretical mechanisms.

5.3 Mechanisms

So far, we have not discussed the mechanisms that may underlie the patterns our analysis has revealed. In this, our exercise is similar to the path-breaking work of Richardson (1960) and the aforementioned papers by Clauset (2018, 2017), Cirillo & Taleb (2016), Fagan et al. (2018), which
mainly focus on modelling battle deaths and uncovering potential trends.\footnote{This does not mean that the authors make no attempt at explaining the patterns they uncover, Clauset\cite{clauset2017} in particular does discuss this.} A full investigation of mechanisms is beyond the scope of this article. Nonetheless, we will discuss a set of plausible mechanisms that could help explain the change-point our analysis revealed. We base this discussion on existing theoretical work.

There exists a large literature attempting to explain the production of wars at the systems level, in addition to aforementioned theories by for instance, Pinker\cite{pinker2011} or Gat\cite{gat2006}, or indeed by Clauset\cite{clauset2017,clauset2018}.\footnote{The forthcoming book ‘Only the Dead: The Persistence of War in the Modern Age’ by Bear Braumoeller appears relevant to our topic, but unfortunately comes too late for us to engage with.} Of particular relevance are the contributions by Cederman and co-authors. Cederman\cite{cederman2003} builds an agent-based model for war and state formation that re-produces the power law distribution of war. He argues that ‘technological change and contextually activated decision-making go a long way toward explaining why power laws emerge in geopolitical systems’ (Cederman, 2003, 147). As mentioned above, Cederman et al.\cite{cederman2011} find a change-point in 1789, with a subsequent increase in the severity of war. They discuss potential explanations driving the shift, and argue that it was driven by a revolution in the technology of statecraft, especially in the ability of states to extract resources and organise their militaries.

Our analysis identified 1950 and the Korean war as the most likely change-point in the distribution of battle deaths in international wars. A change-point in the period around and following the Korean war fits well with the thesis developed by Pinker\cite{pinker2011} Ch. 5). Here the mechanism underlying the change-point would be the cultural, political, and moral shift that took place across especially the Western world. War went from being an appropriate part of statecraft, ‘the continuation of policy by other means’ (Clausewitz\cite{clausewitz1989}, to something inappropriate or even evil (Mueller\cite{mueller1989}). This shift began in the post-Korean war world, and is particularly associated with the Vietnam war period. As informal evidence for the argument Pinker\cite{pinker2011} lists a multitude of songs and movies from that period with clear and explicit anti-war themes, themes that were much less present in earlier periods.

In addition to this norms-based mechanism, we consider two other mechanisms particularly plausible. The first centers around the development of nuclear weapons. When the USSR conducted their first atomic weapons test in 1949, the two superpowers, the US and the USSR, created the basis by which war could escalate to a point where the world would face total annihilation. The development of the system of mutual assured destruction led all key actors to fear that low intensity conflict could escalate into thermonuclear war (Kahn\cite{kahn1965}). This restraining effect could operate as a mechanism reducing the intensity of international conflicts. We could label this the ‘George Orwell Mechanism’. In his essay ‘You and the Atomic Bomb’ (October 1945), Orwell predicted that power would be consolidated in the hands of the superpowers due to the atomic bomb, and that these two would perpetually threaten atomic war against each other, without actually risking it. As a result, large-scale wars would end and instead we would see the rise of a new form of smaller wars. The restraining effect of nuclear weapons could by itself be an important mechanism, but this mechanism may have been further strengthened by the system of international governance, and especially the United Nations, which was developed to help defuse conflicts before they escalated out of control (Goldstein\cite{goldstein2011}).

A second mechanism centers around the role of democracy. Democracies very rarely go to war against each other, a tendency labelled the democratic peace (see e.g. Maoz & Russett\cite{maoz1993}).
Moreover, Mitchell, Gates & Hegre (1999) show that the relationship between democracy and war has become more pronounced over time, indicating that democracy could be particularly useful for studying change-points in the history of interstate wars. In Section 4.3 we do indeed find an increasingly pacifying effect of democracy, though this analysis is only indicative, and the results should be treated with caution. In the period before 1950, democracy seems to have no effect on the number of battle-deaths. After 1950, however, the wars between more democratic countries have become much less violent. The increasing effect of democracy on conflict coupled with a simultaneous increase in the number of democracies in the world could translate into a more peaceful world in the aggregate.

Acknowledgements

We are grateful to the NRC funded five-year project FocuStat, and to several of its participants for active discussions, in particular Gudmund Hermansen and Emil Stoltenberg, and to the NRC funded Young Research Talent project MiCE. In addition we appreciate comments from Aaron Clauset, Jens Kristoffer Haug, and Steven Pinker. We are also very grateful for the detailed and constructive comments from three anonymous reviewers.

References


