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

May 2018

Céline Cunen[1], Nils Lid Hjort[1], and Håvard Mokleiv Nygård[2]

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

Abstract

Has the world of great wars become less violent over time, and is there, in fact, something we might identify as The long peace? We investigate statistical versions of such questions, by examining the Correlates of War dataset on 95 interstate wars, from 1823 to 2003, pertaining in particular to the number of battle deaths. The data have been analysed in depth in Clauset (2017, 2018), where he concludes that the series of wars have been amazingly stationary, with no apparent change. We complement and finesse his analyses in several ways, and reach alternative and more optimistic conclusions. Our statistical methods involve finding and assessing change-points; modelling the full battle fatalities distribution, as opposed to focusing merely on the extreme tail; factoring in the level of democracy as covariates; and focused model selection. Our statistical sightings of better angels indicate that 1965 or 1950, depending on whether we focus on only the most severe wars or study the full distribution, represent the most likely game changer – 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, tail index, Vietnam War.

1 Introduction and summary

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 work has in various ways contributed pieces of layers with similar stories, messages, interpretations, and conclusions.[1] Thus the concept of The long peace (Gaddis, 1989) has gained the weight of repeated respectful

[1]The literature is too large to review here; see for instance the collection of review articles in the 50th Anniversary issue of the *Journal of Peace Research* (Volume 51, Issue 1).

use, over several decades, to signal the Cold War but also the absence of the more terrible grand-scale 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.

The long peace is not without its critiques. Cirillo & Taleb (2016) and Clauset (2017, 2018) have argued that it is still too early to confidently assert, from history and data alone, that The long peace is safely in place. Clauset (2017) re-iterates, using more appropriate and formal procedures, that interstate wars are power law distributed and that the waiting times between wars follow a Poisson process.[2] Both these characteristics are important when addressing the question of whether the world has become more peaceful. Quantities that follow a power law distribution generally have very fat tails. This has broad implications for the conclusions we may draw from data. Moreover, Clauset (2017) and Cirillo & Taleb (2016) both claim that it is still to early to conclude that the recent dearth in major interstate conflicts is part of a larger trend of a more peaceful world. Indeed, Clauset (2017) argues that the current trend would have to persist for as much as 100 to 150 more years before we could make such a conclusion.

We add to this debate in two ways. First, we implement a better and more appropriate methodology for identifying break points or change-points, the point in time at which 'things change', for the data on interstate wars. We use this methodology to query if and when a change in the history of wars has happened, and find evidence that The long peace started around 1965. As we describe in detail below, the tail index parameter, describing the sizes of the biggest wars, undergoes a change for the better, meaning much less intense wars, in 1965. Certain other statistical sightings of better angels are also pointed to. Second, whereas the parametric models in Clauset (2017) only use the most intense wars, those having incurred more than 7061 battle deaths, we study parametric models for the full distribution of interstate wars. Wars are, thankfully, rare events and utilising data on all wars, not just the most extreme, allows us to draw more informed conclusions about The long peace. For this more extensive analysis we find evidence of a break point happening earlier than 1965.

Our article is structured as follows. In Section 2 we offer both a general discussion and a brief technical description of the change-point methodology we will use. Then, we investigate potential change-points in the CoW dataset, first for the large wars only, in Section 3, and then for the full battle death distribution in Section 4. Within these sections, we offer fresh statistical insights concerning several issues: identifying the threshold value in power law models (Section 3.1), assessing the degree of change associated with a potential change-point (Sections 3.4 and 4.3), and importantly, finding appropriate models for the full battle-death distribution (Section 4). Further, we present yet other potential extensions in two directions; focused model selection via the FIC (in Section 5), and methods for extending previous parts of the theory to include potentially influential covariates (in Section 6).

While most of our efforts concern the war sizes, i.e. the number of battle-deaths in each war, we analyse the between-war-times in Section 7. The methods we describe can be used for various similar war-and-conflict datasets, for instance the UCDP/PRIO Armed Conflict Database (Gleditsch et al., 2002), also when these are more complex in character. Further interpretations of our analyses are offered in Section 8, with some concluding remarks in Section 9, some of which point to further research issues.

[2]A pattern first demonstrated by Richardson (1960); see also Cederman (2003) for a theory of why severe wars are power-law distributed.

2 Change-point analysis: motivation and methodology

When faced with a sequence of observations, change-point methodology is used to search for *when* the point of maximal change occurs and, importantly, to assess the uncertainty around it. We will start with a short overview of change-point methods in social sciences, then we give a short technical overview of the specific change-point method we will apply.

2.1 Change-points in social and political sciences

There is a long tradition in social and political science for studying shifts in history, and for examining conditions for the potential for shifts. Change-points or phase shifts – often labelled 'periods', 'critical junctures', 'structural change', or 'turning points' in social science – are critical for understanding a variety of political processes (see e.g. Tilly (1995), and also Marx (1871), Spengler (1918)). Other terms are used for similar phenomena in other fields, such as 'regime shift', 'tipping point', etc.; see Cunen, Hermansen & Hjort (2018a). Change-point analysis is a formal framework for studying the question 'has there been a change, and, in that case, when', and for investigating the size and direction of such a change.

Models that allow researchers to study if a specific parameter has changed over time have been utilised for some time. Beck (1983), for instance, shows how to estimate structural changes in regression models. Mitchell, Gates & Hegre (1999) draw on this and use Kalman filter models to study the relationship between democracy and interstate conflict. They find that the pacifying effect of democracy on interstate war has increased over time. Only more recently, however, have formal change-point models been introduced to the discipline. Western & Kleykamp (2004) develop a Bayesian change-point model and show that a structural break in the process of wage growth happened in 1976. Spirling (2007) introduces similar models for count, binary, and duration-type data. Blackwell (2018) develops a change-point model for overdispersed count data and uses this to study shifts in terrorist violence. These models all make use of the Bayesian framework to analyse change-points. The framework we develop here is frequentist in nature and thus does not necessitate the use of prior distributions for parameters.

There are by necessity many ways in which to search for a change-point in a sequence of data, even when the search is narrowed down via a plausible statistical model, where the question becomes that of change in its model parameter; see Frigessi & Hjort (2002) for a broad introduction to a special journal issue on discontinuities in statistics (modelling, purposes, applications). Here we employ change-point machinery developed in Cunen, Hermansen & Hjort (2018a), both for spotting a potential break-point and, crucially, for assessing its uncertainty. To assess uncertainty we use confidence curves, see Schweder & Hjort (2016); Hjort & Schweder (2018). The confidence curves allow us to assess the uncertainty at all levels of confidence and to investigate the extent to which the uncertainty is symmetric across different levels of confidence.

Change-point questions can be posed and analysed for several aspects related to the history of interstate wars, depending on which parts of the combined available information sources are brought to the table. For each such set or subset of the data we may model and assess distributions, and then search for change-points. (i) We might study the full distribution of battle deaths, across all 95 wars, cf. Section 4. (ii) We may concentrate on only the larger wars, for which the power-law behaviour has set in. The emphasis in the literature has been on these extreme tails (cf. Clauset's work) so we give this primary attention in Section 3. (iii) Covariate information might also be factored in, as in Section 6, which leads to more refined models and then new searches for regime shifts and associated influential factors.

2.2 Change-points for general models

Suppose we study observations $z_1, \ldots, z_n$ from some parametric model, say $f(z, \gamma)$, where γ is of dimension p. Assume that there is a break-point τ in the sequence, with parameter γ_L for $i \le \tau$ and γ_R for $i \ge \tau + 1$. The profiled log-likelihood function becomes

$$\ell_{\rm prof}(\tau) = \max_{\gamma_L, \gamma_R} \Big\{ \sum_{i \le \tau} \log f(z_i, \gamma) + \sum_{i \ge \tau+1} \log f(z_i, \gamma) \Big\} = \sum_{i \le \tau} \log f(z_i, \widehat\gamma_L) + \sum_{i \ge \tau+1} \log f(z_i, \widehat\gamma_R),$$

involving maximum likelihood (ML) estimators $\widehat\gamma_L$ to the left and $\widehat\gamma_R$ to the right. The ML estimate $\widehat\tau$ can be read off from plotting the profile function. We also need the deviance function

$$D(\tau, z) = 2\{\ell_{\rm prof}(\widehat\tau) - \ell_{\rm prof}(\tau)\},$$

now defined for a general $z = (z_1, \ldots, z_n)$ sequence. For the observed data $z_{\rm obs}$, the deviance $D(\tau, z_{\rm obs})$ is zero for $\widehat\tau$ and bigger than zero elsewhere.

One of several related methods developed in Cunen, Hermansen & Hjort (2018a) starts from the well-defined probability function

$$\mathrm{cc}_0(\tau; \gamma_L, \gamma_R) = P_\tau\{D(\tau, Z) < D(\tau, z_{\rm obs})\},$$

with Z denoting a random sequence drawn from γ_L to the left and γ_R to the right of τ. The recipe then plugs in the ML estimates of these left and right parameter vectors, leading to $\mathrm{cc}(\tau) = \mathrm{cc}_0(\tau; \widehat\gamma_L, \widehat\gamma_R)$. The actual computation of the confidence curve requires computer simulations from the estimated models, to the left and the right of each τ. There are certain modifications available, see again Cunen, Hermansen & Hjort (2018a). This general method will be used for the full battle deaths distribution in Section 4, and also, briefly, when covariates become part of the information modelling, in Section 6. For the simple power law model in Section 3, the method simplifies somewhat, as we will see. Specifically, in that case, the change-point inference is completely exact, while in the general case the method is dependent on how well the model parameter γ is estimated on each side of the change.

We underline that the general statistical methodology briefly described in this section is broadly applicable, not merely for finding a jump in level of an observed process, but for changes of less direct parameters in the data generating mechanisms.

3 Modelling and analysing the large wars

Clauset (2017, 2018) argues that when it comes to two fundamental aspects of major wars, namely the interwar-periods and the number of battle deaths, the history is essentially a stationary one, over the past two hundred years. We take this finding, and the dataset used in that analysis, as our point of departure. The specific dataset under scrutiny is that of

$$(x_i, z_i) \quad \text{for } i = 1, \ldots, n, \tag{3.1}$$

for the $n = 95$ interstate wars carefully catalogued in the Correlates of War (CoW) database (Sarkees & Wayman, 2010), where x_i is the onset time and z_i the (estimated) number of battlefield fatalities. We record the onset times in years, months, and days; thus the x_i range from 1823.27 (the Franco-Spanish war) to 2003.22 (invasion of Iraq). Also, each $z_i \geq 1000$ for this dataset. Figure 3.1 displays these data, with z_i on the log-scale. The horizontal dashed line, corresponding to the threshold $z_0 = 7061$, indicates that z_i above that level behave according to a power law with density proportional to $1/z^\alpha$ for a certain parameter α. The 7061 is Clauset's estimated threshold value; we discuss details and alternatives in Section 3.1.

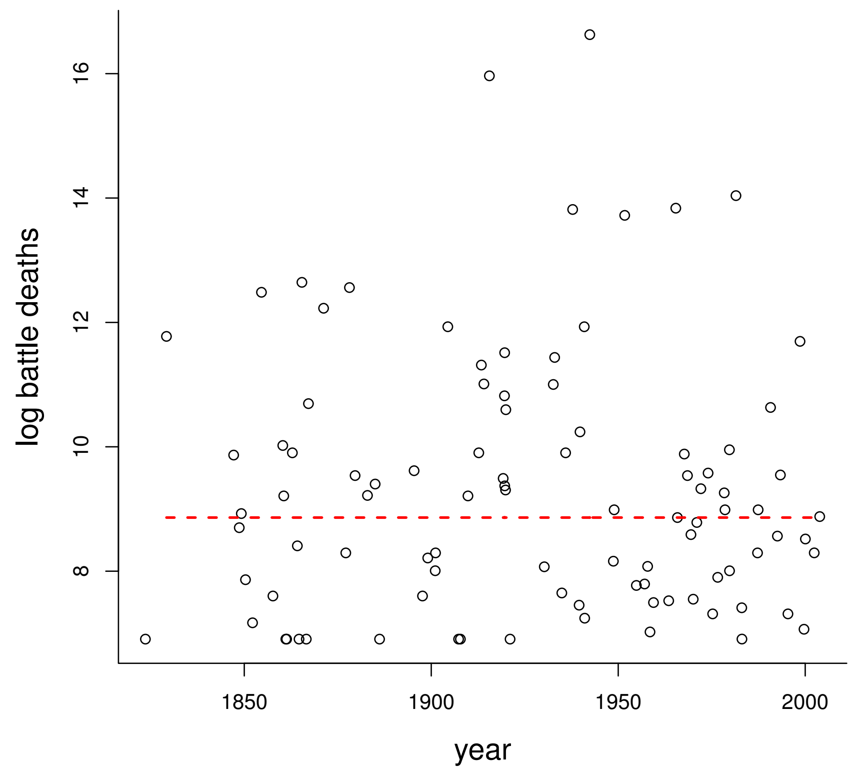

Figure 3.1: Onset times x_i and the logarithm $y_i = \log z_i$ of the number of battle deaths are shown, for 95 wars, from 1823 to 2003. The horizontal line indicates the threshold $\log 7061$ above which exponentiality is claimed.

Clauset (2017, 2018) analyses these data, and does not find sufficient reasons to contradict the basic stationary view that

(i) the between-times $d_i = x_i - x_{i-1}$ are independent and identically distributed (i.i.d.), following a simple exponential distribution;

(ii) the war sizes z_i are i.i.d. with a power-law tail behaviour, and independent of the d_i.

The violence level distribution for great wars has at least approximately a power-law behaviour, for several types of measures. The necessity of selecting a threshold entails a potential loss of information, since we only make use of the information in the wars with more than z_0 battle deaths, i.e. the large wars.

We note at the outset that statistical modelling and analysis in some ways become easier when passing from z_i to the log-scale of $y_i = \log z_i$. That the z_i follow a density proportional to $1/z^\alpha$ above a certain threshold z_0, where $\alpha > 1$, is equivalent to the y_i having an exponential tail; the density is then $\theta \exp\{-\theta(y - y_0)\}$ above $y_0 = \log z_0$, with $\theta = \alpha - 1$ positive. Discussions for (z_0, α) are hence equivalent to such for (y_0, θ). Results can then be transformed back to and interpreted for the original battle field deaths scale.

Below we start with the selection of the tail-index threshold. Specifically, the methods for threshold selection need to take into account the existence of a potential change-point. Then in Section 3.2 we describe how the general change-point method in Section 2.2 pans out with the exponential model. The results of this analysis are presented in Section 3.3. In Section 3.4 we analyse the magnitude of the degree of change across the change-point.

3.1 Setting the tail-index threshold: Methodology

A critical question when dealing with power laws is how to set a threshold z_0, or, equivalently, the $y_0 = \log z_0$. One may in several ways operationalise the idea that y_0 ought to be pushed incrementally upwards, until the remaining $v_i = y_i - y_0$ above that threshold, i.e. all conflicts above a certain battle deaths threshold, pass natural tests for exponentiality.

Using a Kolmogorov–Smirnov (KS) type test for this purpose, one may for each z_0 compute

$$K(z_0) = \sqrt{m} \max |F_{\rm emp}(v) - F(v, \widehat{\theta})| = \sqrt{m} \max_{i \le m} |F_m(v_{(i)}) - \{1 - \exp(-\widehat{\theta} v_{(i)})\}|.$$

Here $F_{\rm emp}$ is the empirical distribution function of the m data points, the inverse sample mean $\widehat{\theta} = 1/\bar{v}$ is the ML estimate, and with $v_{(i)}$ the ordered sample. With some extra efforts, via simulations or a large-sample approximation, one may compute the p-value

$$p(z_0) = P_0\{K^*(z_0) \ge K_{\rm obs}(z_0)\},$$

with $K^*(z_0)$ denoting a random variable constructed as above from purely exponential data. As long as $p(z_0)$ is small, exponentiality is not trusted, and the threshold is now increased until $p(z_0)$ reaches a satisfactory high value, or its maximum. This is a version of the method used in Clauset (2018, App. A), but here formulated on our log scale. The $\sqrt{m}$ factor is not necessary in this argument, but is used here since $K^*(z_0)$ then has a well-defined limit distribution which can be appealed to instead of more laborious simulations.

This recipe works when the sample of y_i is homogeneous, with the same tail index. When we below work to find a change-point, a more refined argument is required, based on there being a θ_L value to the left of the unknown break-point τ and another value θ_R to the right. Our recipe is as follows. For given z_0, we first find the ML estimate $\widehat{\tau}$ for the break-point, the one maximising the profiled log-likelihood function

$$\ell_{\rm prof}(\tau) = \tau \log \widehat{\theta}_L(\tau) + (m - \tau) \log \widehat{\theta}_R(\tau). \tag{3.2}$$

Here $\widehat{\theta}_L(\tau) = 1/\bar{v}_L$ and $\widehat{\theta}_R(\tau) = 1/\bar{v}_R$ are the ML estimates to the left and to the right. Then, for this value $\widehat{\tau}$, compute the KS statistics

$$K_L(z_0) = \sqrt{\tau} \max |F_{L,\rm emp}(t) - F(t, \widehat{\theta}_L)| \quad \text{and} \quad K_R(z_0) = \sqrt{m - \tau} \max |F_{R,\rm emp}(t) - F(t, \widehat{\theta}_R)|$$

to the left and to the right, along with the consequent p-values

$$p_L(z_0) = P_0\{K_L^*(z_0) \ge K_{L,\rm obs}(z_0)\} \quad \text{and} \quad p_R(z_0) = P_0\{K_R^*(z_0) \ge K_{R,\rm obs}(z_0)\}. \tag{3.3}$$

When one or both is small, we must reject the idea that the distributions to the left and to the right have reached exponentiality. When both have reached an adequately high level, we trust exponentiality on both sides. We set the threshold to be the first z_0 where such a level has been reached, or, more conservatively, to maximise $p(z_0) = \min\{p_L(z_0), p_R(z_0)\}$.

Applying these methods to the CoW dataset leads to Figure 3.2. The $p(z_0)$ curve is maximised at position 7147 (with 50 wars above or at that level), and this is almost the same as Clauset's 7061 (with 51 wars above). Also the threshold value $z_0 = 6000$ (with 53 wars above) is judged to be satisfactory for the tail-index behaviour purpose. We have worked with mild variations of the recipe above, using distances between distributions other than the KS one, and report that they lead to very similar results.

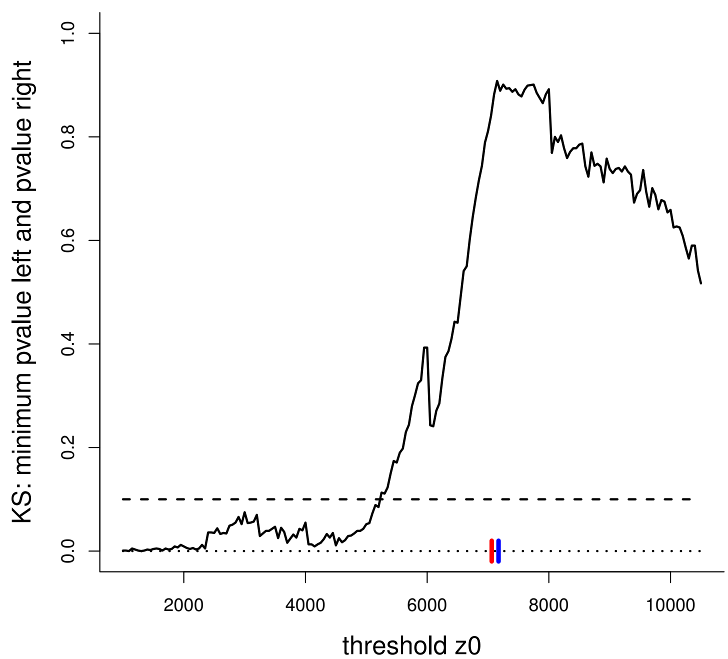

Figure 3.2: The minimum curve of the $p_L(z_0)$ and $p_R(z_0)$, given in (3.3), for selecting the cutoff point z_0. The maximum is at 7173, just to the right of Clauset's 7061. The threshold $z_0 = 6000$ is also satisfactory.

The choice of power-law threshold has separate importance, for this and similar datasets on war, conflict, violence; see e.g. Cirillo & Taleb (2016). Setting the z_0 also has repercussions for the following change-point analysis.

3.2 Change-point methodology for a simple exponential model

The analysis above leaves us with a set of conflicts above a certain battle deaths threshold for which the power law behaviour holds. The next step is to analyse if this series of interstate conflicts is homogenous over time, or if it at some point changes. If the series does change, and the change is for the better in terms of less lethal wars, this could be indicative of the onset of The long peace.

We have seen that power-law behaviour for the battle fatalities z_i corresponds to exponential distributions for $v_i = \log z_i - \log z_0$ for wars over a threshold z_0; cf. Figure 3.1. We therefore start describing a method for finding a change-point for the special case of data $v_1, \ldots, v_m$ coming from the exponential model, with one parameter θ_L to the left of τ and another θ_R to the right. As in Section 2.2, the starting point is the deviance function $D(\tau, v) = 2\{\ell_{\rm prof}(\widehat\tau) - \ell_{\rm prof}(\tau)\}$, with the profile function $\ell_{\rm prof}$ as in (3.2), and with v being the full data sequence. The change-point ML estimate $\widehat\tau$ is the maximiser of this profile function, and hence the point where the deviance reaches its minimum point of zero. We then form the confidence curve

$$\mathrm{cc}(\tau) = P_\tau\{D(\tau, V) < D(\tau, v_{\rm obs}) \,|\, \bar v_{L,{\rm obs}}(\tau), \bar v_{R,{\rm obs}}(\tau)\}, \tag{3.4}$$

for all candidate breakpoints τ. Here $v_{\rm obs}$ is the observed sequence, along with average values $\bar v_{L,{\rm obs}}(\tau)$ and $\bar v_{R,{\rm obs}}(\tau)$ to the left and right of the τ under scrutiny, and V denotes a random sequence from the relevant distribution of v given these observed left and right averages.

The first point to note is that these are well-defined probabilities which may be computed and displayed, for each τ, via the distributions of the left and right pieces V_L and V_R, from the conditional distributions given the observed averages, and where distributions do not depend on the (unknown) parameter values θ_L and θ_R. This is a consequence of these averages being sufficient, in the statistical sense, under the model used. In fact we may show that

$$(V_1, \ldots, V_\tau) \,|\, \bar v_{L,{\rm obs}} \quad \text{has the same distribution as} \quad \tau \bar v_{L,{\rm obs}}(\tau)(w_1, \ldots, w_\tau),$$

where the $(w_1, \ldots, w_\tau)$ has a Dirichlet distribution with parameter $(1, \ldots, 1)$, and similarly for the right part. The operational consequence is that we may calculate $\mathrm{cc}(\tau)$ via computer simulations from a known model, with no extra parameters. See Siegmund (1988); Cunen, Hermansen & Hjort (2018a) for more general constructions.

The second point relates to the properties and interpretation of the $\mathrm{cc}(\tau)$. It is a confidence curve, with the property that the confidence set $R(\alpha) = \{\tau \colon \mathrm{cc}(\tau) \le \alpha\}$ has probability precisely equal to α, for all such confidence levels α. For applications of the present type we prefer plotting such confidence curves as function of time x rather than the running label τ; thus we present $\mathrm{cc}(x_0)$ below, where x_0 signifies the change-point on the calendar scale.

3.3 For the most severe wars, the likely change-point is 1965

We now apply the change-point methodology above to the interstate wars in the CoW dataset that are above the battle deaths threshold. For the sake of simplicity and ease of comparison, we present the results using the same tail-index threshold as in Clauset (2018), $z_0 = 7061$, and the corresponding 51 wars. For our analysis we exclude the three first and the three last starting dates from the set of potential change-point values, since the log-likelihood profile may have erratic behaviour close to the edges. The ML estimate for the change-point is the 37th war (i.e. $\widehat\tau = 37$, in the formalism above), corresponding to the coded onset-war-time of $\widehat x_0 = 1965.103$. This is the Vietnam War, or more precisely what is coded as its Phase Two, with the attack on Camp Holloway, 7-Feb-1965. Thus, the point of clearest change in the tail exponent parameter is found to be between the 37 wars up to and including the Vietnam War on the one side and the 14 wars following the Vietnam War on the other side.

In this analysis we have not specified any direction for the change, and a priori the identified change-point could signal an increase in battle deaths. However, the exponential parameters to the left and right of the change-point are estimated to 0.451 and 0.928. These estimates correspond to a median of 32880 deaths for wars in the pre-and-up-to Vietnam period and 14906 for wars in the post-Vietnam period.

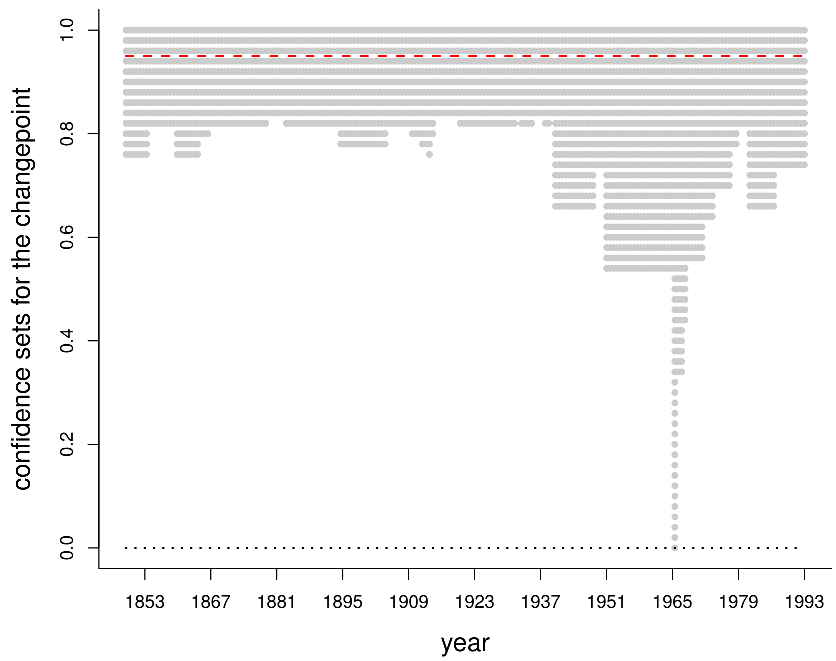

Figure 3.3: Confidence curve for the change-point, pointing to 1965. Here $z_0 = 7061$ has been used for the tail-index threshold. The red dashed line corresponds to the 0.95 confidence level.

The uncertainty around the point estimate is given by the confidence curve in Figure 3.3. 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 (1965), and we can read off confidence sets at all levels. These can be interpreted in the same manner as the more commonly used confidence intervals. Clearly there is considerable uncertainty; we see that around the 80% level the confidence set covers the whole set of potential change-point values. At the same time, however, the curve rapidly narrows as the confidence level decreases, and e.g. the 60% set covers only five wars (Korean War Oct 1950, Vietnam Feb 1965, 2nd Kashmir Aug 1965, Six-Day Israel Jun 1967, 2nd Laotian Phase 2 Jan 1968).

There is a potential risk that change-point analyses of this type rely too much on the selected cutoff threshold z_0. It is reassuring, then, to learn from examination that the same Vietnam War Feb 1965 is pointed to for a broad range of reasonable threshold values, actually from 4481 to 14000, thus lending credit to 1965 being the the game changer. In addition to these all pointing to 1965, the confidence curves are also broadly similar, for a range of such z_0. These mainly differ in the width of the higher level confidence sets. For several choices of z_0 below the Clauset value of 7061 the 80% and even 90% sets are quite narrow. This reflects having more data on board and also a better fit to the underlying model structure.

3.4 The degree of change

In a change-point setting one is typically interested in the degree of change, in addition to the position of the change-point. In the present context we may enquire about the intensity ratio $\rho = \theta_R/\theta_L$, for example. Confidence curves $\mathrm{cc}(\rho)$ can be constructed, from different sets of starting assumptions, and two such curves are displayed in Figure 3.4.

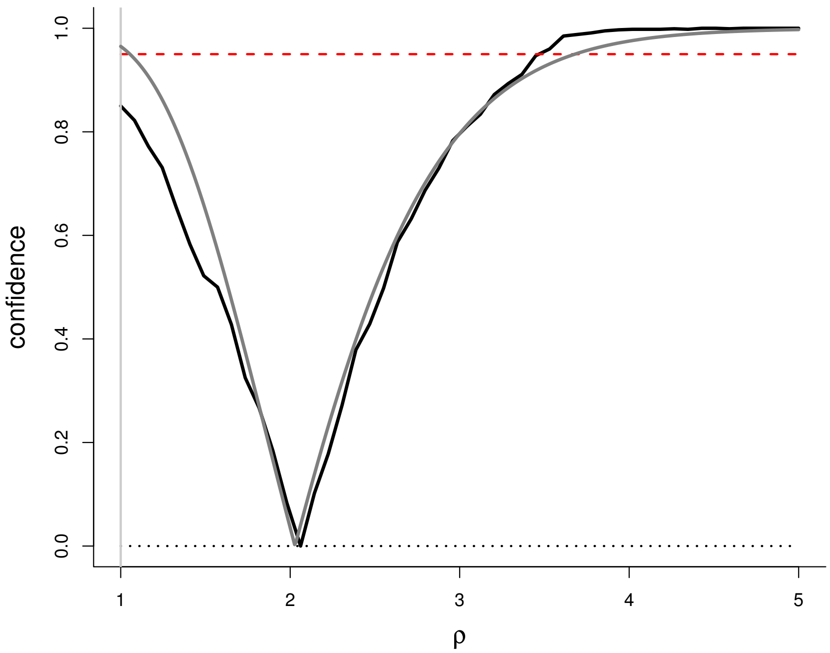

Figure 3.4: Confidence curves for $\rho = \theta_R/\theta_L$, with 7061 as the tail-index threshold. The narrow one is conditional on the 1965 hypothesis, the broader one is conservative and has no such assumption. The red dashed line corresponds to the 0.95 confidence level.

The smooth and more narrow $\mathrm{cc}_{1965}(\rho)$ is the easier statistical construction, with computations carried out under the Vietnam hypothesis that the change-point has been agreed upon and is 1965. The 95% interval is $[1.056, 3.667]$. The more wiggly and statistically conservative $\mathrm{cc}(\rho)$ emerges via methods of Cunen, Hermansen & Hjort (2018a), assuming there is a breakpoint, but not claiming to know in advance that it is at 1965 (or any other timepoint). Its construction involves maximisation over the model parameters and the change-point τ, leading to the profiled log-likelihood function $\ell_{\mathrm{prof}}(\rho)$. In its turn this leads to the full confidence curve, expressed via probabilities which can be computed via simulations. The ML estimate is $\widehat{\theta}_R/\widehat{\theta}_L = 2.059$. This indicates that the median level of fatalities has gone down with a factor of about 2.20, since 1965. Crucially, also the overall variability has gone down, and, in particular, the higher quantiles (e.g. with a factor of 13.80 for the 90% quantiles).

The uncertainty around the change-point estimate expressed in Figure 3.3 translates more uncertainty around the ratio ρ, expressed via the conservative $\mathrm{cc}(\rho)$, compared to that of $\mathrm{cc}_{1965}(\rho)$. The confidence intervals from $\mathrm{cc}(\rho)$, from around the 80% level, cover a relatively large range of

ρ values. Importantly, they also cover $\rho = 1$, implying that the hypothesis of no change cannot be rejected at common levels of significance. Again, for other choices of tail-index threshold the resulting intervals are narrower (even leading to rejection of no change hypothesis on the 10% level).

4 Modelling the full battle deaths distribution

The emphasis in Clauset's and several other analyses is on the tail of the statistical distribution of the battle deaths z_i, and the associated tail index parameter (see also Cirillo & Taleb, 2016). We depart from this, and instead attempt to model the full distribution, i.e. for the full range $z_i \geq 1000$, thus including in our analysis all conflicts in the CoW dataset. This potentially leads to better inference also for the wars in the tail, as we may exploit data for all wars, not merely those above the threshold value.

We will start by presenting some candidate models for the full battle deaths dataset and discuss some of their properties. Then we will use the general method of Section 2.2 to investigate a potential change-point in the sequence of wars, along with its uncertainty. Finally, we will also suggest some quantities for measuring the degree of change, and present confidence curves for these.

4.1 Parametric models with fat tails

The task is to model the nonnegative data $z_i^* = z_i - 1001$, and there is a long list of plausible such.[3] A suitable model should have a good fit to the data, as measured by relevant goodness-of-fit criteria. We may examine the degree of fit of the fitted cumulative or quantile distribution functions to the empirical ones, in various ways. Models may also be compared via model selection criteria, such as the AIC (the Akaike Information Criterion), or the FIC (the Focused Information Criterion); cf. Claeskens & Hjort (2008), and Section 5 below. In addition, in the present context, we also want to ensure that a model exhibits the appropriate power tail behaviour. Generally speaking, the distribution function $F(z)$ for the z_i is said to have the fat tail property, with power index b, if $z^b\{1 - F(z)\}$ tends to a positive constant as z increases. It also means that the density is approximately proportional to $1/z^{b+1}$ for large z. In this, we follow the existing literature on the distribution of interstate wars. This literature, starting with Richardson (1960), has showed that the tail of the distribution follows a power law. This issue is also discussed by Pinker (2011) and in detail by Cirillo & Taleb (2016). Note that we still know comparatively little about *why* interstate wars are power law distributed (Cederman, 2003).

We have investigated a generous list of potential models for the full battle deaths data, including the so-called log-logistic, the Burr, the log-gamma, the paralogistic, the inverse paralogistic, along with a few models we invented ourselves. When choosing between these distributions, we have been guided by plausibility, the fat tail property, and the AIC. Note however, that the use of AIC in a change-point context is not straightforward, since the change-point parameter itself is discrete, and outside the usual setup of regular statistical models. For simplicity we have chosen

[3] In the CoW database, 9 of the wars are listed as having had exactly 1000 battle deaths. These are clearly meant as estimates, however, in contrast to e.g. the number 1001 given for the Falkland War. To avoid certain artificial issues with the statistical estimation procedures, we have changed the battle deaths estimates for these 9 wars so that they are recorded as having $1002, 1003, \ldots, 1010$ battle deaths instead. Furthermore, in order to avoid having an observation exactly equal to 0 after we subtract 1001, we add 0.01 to the war with the minimum number of battle deaths.

to look at $\mathrm{AIC}_L + \mathrm{AIC}_R$, with L and R indicating analyses to the left and to the right of the ML position for the change-point.

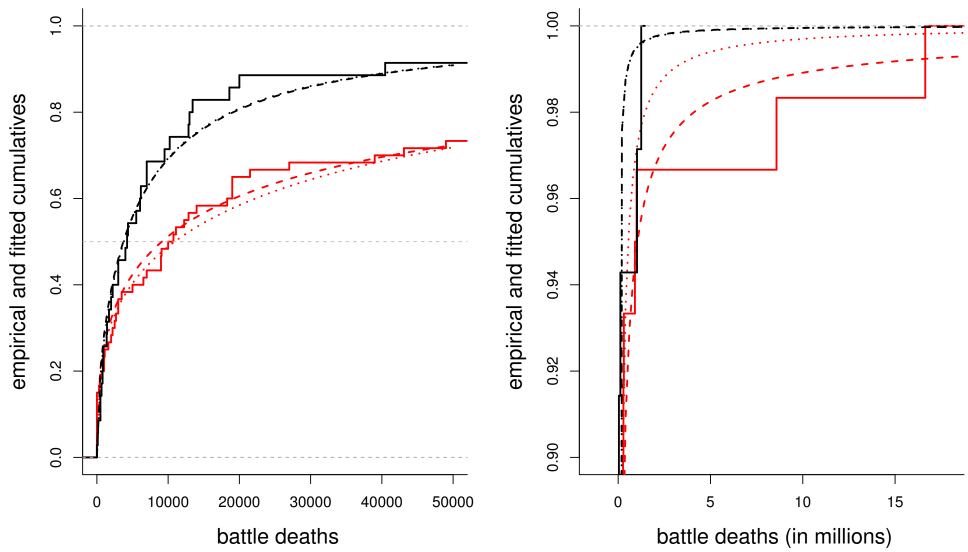

Figure 4.1: Empirical and fitted cumulative distribution functions for the battle deaths across 95 wars, for the inverse Burr and the inverse Patero. The red curves correspond to wars before 1950, the black to wars after 1950. The dashed lines are for the inverse Burr model; the dotted for the inverse Pareto. The two black curves are virtually identical. In the left panel we see the parts of distribution functions for wars up to 50000 battle deaths; in the right panel we focus on the very large wars.

Two of the models we have investigated will be given a more complete treatment, as these did best according to our examination and selection criteria, among various competitors. These are *the inverse Burr distribution*, which also goes by the name of the Dagum distribution, and the *inverse Pareto distribution*. In particular, these models were favoured by the AIC; with the inverse Pareto as the very best among those we investigated. Both models also fit the data well when considering the fitted and empirical cumulative distribution functions; see Figure 4.1. Importantly, they both have fat tails, but with somewhat different tail properties, as we will see. The inverse Burr takes

$$F(z;\mu,\alpha,\theta) = P\{Z \le z\} = \left[\frac{\{(z-1001)/\mu\}^\theta}{\{(z-1001)/\mu\}^\theta + 1}\right]^\alpha \quad \text{for } z \ge 1001, \tag{4.1}$$

with parameters (μ,α,θ) to be estimated from the 95 wars. When z increases we have $F(z) \approx 1-\alpha(\mu/z)^\theta$; thus θ plays the role of the power index, similarly to its namesake in Section 3.2. For the inverse Pareto we have

$$G(z;\mu,\alpha,\theta) = P\{Z \le z\} = \left[\frac{\{(z-1001)/\mu\}}{\{(z-1001)/\mu\} + 1}\right]^\alpha \quad \text{for } z \ge 1001, \tag{4.2}$$

corresponding to setting $\theta = 1$ in the inverse Burr. For large wars, $G(z) \approx 1-\alpha(\mu/z)$, revealing that the power index for the inverse Pareto is constrained to have the value 1. This distribution accordingly has a less flexible tail behaviour compared to the inverse Burr, and indeed to the general power law distribution in Section 3.2.

Figure 4.1 presents the empirical cumulative distribution functions (CDFs) for wars before and after 1950 (see change-point analysis in the next section). The fitted CDFs for the inverse Burr and inverse Pareto are also given. We observe that for the smaller wars both distributions fit well and are very similar. In fact for the wars after 1950, the two fitted distributions seem exactly equal; this happens here since the θ_R estimate for the inverse Burr is 1.021, i.e. very close to 1. The differences between the two distributions is largest for the most severe wars (see the right panel). We observe that the inverse Burr seems to fit these wars (before 1950) better than the inverse Pareto. Also for the highest quantiles, the inverse Pareto gives less difference between the fitted CDFs across the change-point than with inverse Burr.

Accordingly, in the following we rely on the inverse Burr distribution when we model the full distribution of wars. As in Section 3.2, our framework involves a single change-point. The two models described above can be used with the general change-point method described in Section 2.2, in order to identify the most likely change-point and the degree of change. In addition to the choice of distribution, the modeller also needs to decide on which parameters of the distribution should be allowed to be influenced by the change-point. The change-point method we use here is highly flexible and allows potential changes in one, some, or all of the model parameters. For the inverse Burr, AIC indicated that the best option was to allow θ and μ to change, but to assume the same α across the change-point. We then end up with a total of six parameters to estimate in the case of a change-point situation with inverse Burr: the change-point τ, along with $(\alpha, \mu_L, \theta_L, \mu_R, \theta_R)$.

4.2 For the full distribution of wars, the likely change-point is 1950

We now use the methodology from Section 2.2 to search for a change-point in the full CoW series of interstate conflicts. The specific task is to use the parametric modelling, employing both the inverse Burr (4.1) and the inverse Pareto (4.2), for the full distribution of wars.

In our analyses we have excluded the ten first and the ten last starting dates from the set of potential change-point values, since we had several more parameters to estimate than for the power law or exponential model in Section 3.2. The ML estimate for the change-point is $\widehat{x}_0 = 1950.483$.[4] With the full battle deaths data, therefore, we find the Korean War, rather than the Vietnam War, at the most likely game changer. Note that the model gives more confidence to 1965.103, however, compared to many other onset-war-times in the dataset (see Figure 3.3). Moreover, the model assigns some confidence to 1939.669, i.e. WW2, and we also see a marked drop around 1980.

The full confidence curve for the inverse Burr model in Figure 4.2 points towards four different periods as potential change-points, but indicates at the same time less uncertainty in the higher level confidence sets than in Figure 3.3. While the 80% confidence set for the power law model encompassed close to the whole range of possible change-point values, the confidence of same level for the inverse Burr encompasses 30 war-onset-times; most of them from 1939 to 1992, but with 'gaps' (as we see in the confidence curve). This demonstrates the increase in power from using all available data and not just the tail of the distribution.

For the inverse Burr, the estimated parameters are as follows (with approximate 95% intervals): $\widehat{\alpha} = 0.499$ (with $[0.244, 0.751]$), $\widehat{\mu}_L = 43887$ (with $[12058, 159736]$), $\widehat{\theta}_L = 0.702$ (with $[0.455, 0.948]$), $\widehat{\mu}_R = 10940$ (with $[4115, 29087]$), $\widehat{\theta}_R = 1.022$ (with $[0.644, 1.399]$). These estimates correspond to a change in the fitted median, from 9118 to 3719 battle deaths, a drop in battle deaths by a factor of 2.45. Modelling the full battle deaths distribution enables us to make more precise estimates of

[4] The inverse Pareto discussed above yields the same change-points.

Figure 4.2: Confidence curve for the change-point, using the full set of 95 wars, and the inverse Burr model (4.1), pointing to the Korean War 1950. The red dashed line corresponds to the 0.95 confidence level.

the likely change-point. These models also point to a somewhat sharper decline in battle deaths pre and and post the change-point than the model only including the most severe wars.

4.3 The degree of change: ratio of quantiles

For the multi-parameter models we have used in this section, the definition of the degree of change is more ambiguous than for the one-parameter exponential model in Section 3. One possibility is to study the ratio between certain quantiles before and after the estimated change-point. We may for instance consider $\rho_1 = \phi_{0.50,L}/\phi_{0.50,R}$ 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 ρ_2 of the upper quartiles would be more relevant to assess than the ratio ρ_1 of medians. With the inverse inverse Burr we have the following expressions for the $100q\%$ quantiles,

$$\phi_{q,\mathrm{ip}} = \mu \frac{q^{1/\alpha}}{1-q^{1/\alpha}} \quad \text{and} \quad \phi_{q,\mathrm{ib}} = \mu \Big(\frac{q^{1/\alpha}}{1-q^{1/\alpha}}\Big)^{1/\theta}. \tag{4.3}$$

Here we use $q = 0.50$ and $q = 0.75$ to estimate the medians and the upper quartiles, respectively. The point estimates via the inverse Burr are $\widehat{\rho}_1 = 2.45$ and $\widehat{\rho}_2 = 4.32$. The upper quartiles decrease from 62544 pre 1950 to 13942 after the change-point, and thus experience a decline of an even larger magnitude of than the medians.

Figure 4.3 gives the full confidence curves. These are computed with the same simulation based method as the conservative $\mathrm{cc}(\rho)$ in Section 3.4, taking into account the uncertainty in the change-point position. 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.

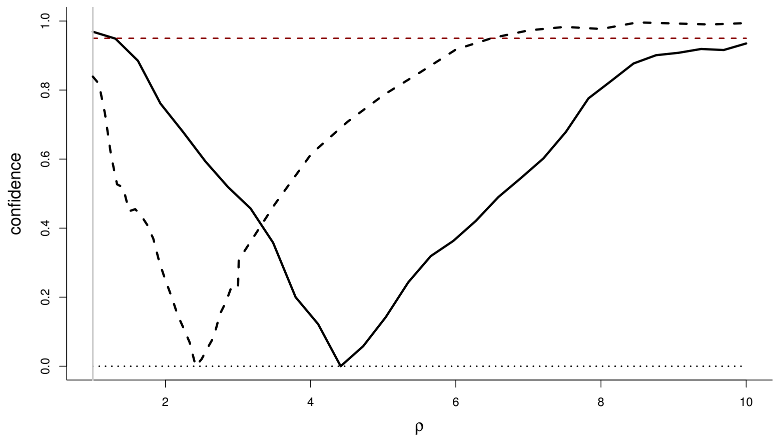

Figure 4.3: Confidence curves for the degree of change, using the full set of 95 wars and the inverse Burr model. The dashed curve belongs to the ratio of medians while the fully drawn one is for the ratio of 75% quantiles.

5 Focused model selection

In the context of analysing the CoW battle deaths data there are several avenues for further sophistication. In the present section we present one such suggestion, concerning potentially fruitful (but not very well known) methods for model selection. In the previous section we have considered different models, and then partly assessed their adequacy, and compared them, via the AIC. Different model selection methods have different strengths and weaknesses, and are suitable for different purposes. Most methods, like the AIC, aim at identifying the model with maximal fit to the data, while including a penalty for complexity. Sometimes our aims may be different, however; in certain contexts one is primarily interested in using models where a specific quantity, which we will call the focus parameter, is estimated with the best precision. The focused information criteria (FIC; see Claeskens & Hjort, 2008; Jullum & Hjort, 2017) offer a natural model selection approach in situations where we have a focus parameter in mind. Thus some models are good for estimating lower quantiles while others might be better for estimating higher quantiles, etc.

The focus parameter depends on the objective of the study. Here we are interested in the larger question of whether the world has become more peaceful. For purposes of a concrete illustration we let the focus parameter below be the median number of battle deaths. In our case, the critical model selection question is which of the inverse Burr and inverse Pareto models perform best in estimating this quantity. Other foci might be considered as well.

The FIC approach consists in ranking models according to their abilities to produce precise estimates of the focus parameter, say ϕ. Such a parameter needs to have a precise interpretation across models. Consider a candidate model, say M, with parameter vector γ_M and ML estimator $\widehat{\gamma}_M$. We require ϕ to be expressed as $\phi(\gamma_M)$, a smooth function of that model's parameter, leading to the estimate $\widehat{\phi}_M = \phi(\widehat{\gamma}_M)$. The precision of these estimators $\widehat{\phi}_M$ is measured by their mean squared errors,

$$\mathrm{mse}_M = \mathrm{E}\,(\widehat{\phi}_M - \phi)^2 = \mathrm{Var}\,\widehat{\phi}_M + (\mathrm{E}\,\widehat{\phi}_M - \phi)^2,$$

which conveniently splits into a variance and squared bias part. The FIC scores, by which we will compare the candidate models, are the estimated mse of the estimator associated with each model, i.e.

$$\mathrm{fic}_M = \widehat{\mathrm{mse}}_M, \quad \text{the estimated } \mathrm{mse}_M.$$

For the present purposes we will use a version of the FIC approach from Jullum & Hjort (2017). That framework consists in comparing the performance of some parametric models of interest with respect to a nonparametric alternative. Here, the nonparametric model merely consists of estimating the focus parameter, the median, with its empirical estimate, $\widehat{\phi}_{\mathrm{np}}$. Nonparametric models can be attractive because the user is spared from making the assumptions underlying parametric models; as we will see, this usually comes with the price of higher variance, however. For this illustration we will compare the nonparametric alternative with the two candidate models we discussed previously, the inverse Pareto and the inverse Burr, having the expressions for the focus parameter given in (4.3).

Jullum & Hjort (2017) provide formulae for computing the FIC scores for each candidate model. Again, the FIC score is the estimated mean-squared error related to the estimator from each model,

$$\begin{aligned}
\mathrm{fic}_{\mathrm{np}} &= \widehat{\mathrm{mse}}(\widehat{\phi}_{\mathrm{np}}) = \widehat{\mathrm{Var}}\,\widehat{\phi}_{\mathrm{np}}, \\
\mathrm{fic}_{\mathrm{ip}} &= \widehat{\mathrm{mse}}(\widehat{\phi}_{\mathrm{ip}}) = \widehat{\mathrm{Var}}\,\widehat{\phi}_{\mathrm{ip}} + \widehat{\mathrm{bsq}}(\widehat{\phi}_{\mathrm{ip}}), \\
\mathrm{fic}_{\mathrm{ib}} &= \widehat{\mathrm{mse}}(\widehat{\phi}_{\mathrm{ib}}) = \widehat{\mathrm{Var}}\,\widehat{\phi}_{\mathrm{ib}} + \widehat{\mathrm{bsq}}(\widehat{\phi}_{\mathrm{ib}}),
\end{aligned}$$

where $\mathrm{bsq} = (\mathrm{E}\,\widehat{\phi} - \phi)^2$ is the squared bias associated with the estimator from the parametric model under scrutiny. The mse of the nonparametric estimator has no bias term because this estimator is converging to the correct quantity as the sample size increases.

In the change-point setting of this article, we are a bit outside the standard FIC theory. Since $\widehat{\tau}$ has a faster convergence rate to the true value, compared to what is the case for the regular ML estimators $\widehat{\gamma}$ for the parametric models, we will take $\widehat{\tau}$ as given, and calculate separate FIC scores to the left and to the right of the given change-point. We will thus consider two estimators for each model, $\widehat{\gamma}_L$ and $\widehat{\gamma}_R$, and present fic_L and fic_R separately on each side of the point-estimate, 1950.483.

The results are summarised in the form of two FIC plots, one for the median to the left of the change-point and one for the right side (Figure 5.1). A FIC plot is a convenient graphical summary for model selection with FIC; on the horizontal axis it displays the square root of the FIC scores and on the vertical axis the estimated median from the three models under consideration. On both sides of the change-point, the FIC considers the inverse Burr to be the best model for estimating the median, with the inverse Pareto on a close second place. On the right side, these two models produce very similar estimates, while on the left side the estimated medians are considerably different. Despite having no bias, the nonparametric alternative got a higher FIC score than the parametric alternatives, due to higher variance. The figure also illustrates, once again, the considerable decline in median battle-deaths before and after 1950.

We note that there is potentially even more room for the use of FIC in the more complex world of models with covariates, as for the section below. There the number of candidate models grows quickly and it becomes both more difficult and important to screen, compare and rank them, for relevance and efficacy. In the present context it would be important not only to spot

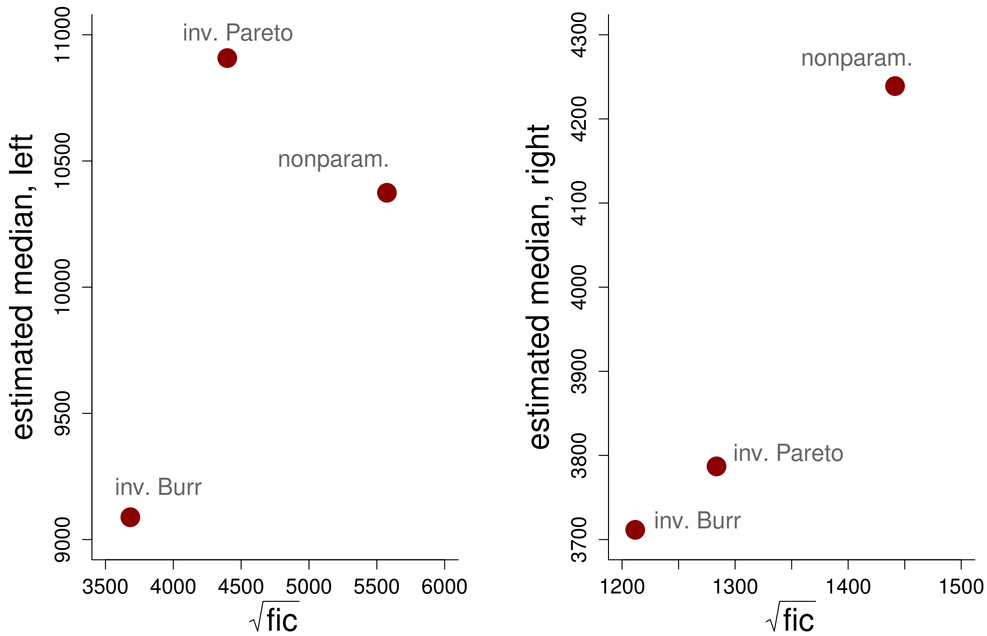

Figure 5.1: FIC plots for the median on the left side of the change-point, and on the right side. Note that the scale of both axes are different in the two panels, reflecting both that the fatality parameter is much lower after the change-point than before, and that its precision is higher.

a change-point, but to understand which underlying mechanisms have been at work. Methods recently developed in Jullum & Hjort (2017) and Cunen, Walløe & Hjort (2018b) can be applied to create FIC schemes to sort through relevant models for such issues.

6 Covariates: Democracy scores for the CoW dataset

So far we have not considered covariates that may inform our search for change-points in the history of interstate wars. Of course, in conflict research there is a rich literature on the causes of interstate conflicts, and parts of this literature has identified tendencies that could inform our search for better angels. Indeed, Pinker (2011) himself is mainly interested in explaining *why* there has been a decline in violence. It is outside the scope of this article to fully delve into these issues. Instead, we focus on one specific covariate; democracy. A large literature has shown that democracies very rarely go to war against each other, a tendency labelled the democratic peace (see e.g Maoz & Russett, 1993). 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 to study change-points in the history of interstate wars.

Our change-point methods are sufficiently general to support the inclusion of covariates such as democracy. Assume the model we work with has a parametric form $f(z\,|\,\gamma)$, with γ of length say p, and that there is covariate information, say w_i, of length say q, for each war. With such a setup we fit the model $f(z_i\,|\,\gamma_{L,i})$ for wars $i \le \tau$ and $f(z_i\,|\,\gamma_{R,i})$ for wars $i \ge \tau+1$, and with the change-point τ as one of the unknown parameters. Then we can investigate the effect of covariates on the model parameters in the following general way. First, consider

$$\gamma_{L,i} = \gamma_{L,0}\rho_{L,i} \quad \text{for } i \le \tau, \quad \gamma_{R,i} = \gamma_{R,0}\rho_{R,i} \quad \text{for } i \ge \tau + 1, \tag{6.1}$$

where $\rho_{L,i}$ and $\rho_{R,i}$ are p-dimensional modification factors, for the left and right parts of the sequence, to be seen as varying around an average level 1. Thus $\gamma_{L,0}$ and $\gamma_{R,0}$ are seen as the average levels of γ, to the left and to the right of the change-point. Second, we model the $\rho_{L,i}$ and $\rho_{R,i}$ as influenced by the covariates, in a log-linear fashion. This can be accomplished in various ways, e.g. by taking

$$\rho_{L,j,i} = \exp(w_i^{\mathrm{t}}\beta_L) \quad \text{and} \quad \rho_{R,j,i} = \exp(w_i^{\mathrm{t}}\beta_R) \tag{6.2}$$

for say one of the parameters components at work in (6.1), here marked by the index j among the p, and the other $\rho_{L,k,i}$ and $\rho_{R,k,i}$ set to 1, i.e. not influenced by the w_i, for $k \neq j$. Thus β_L and β_R are q-dimensional regression parameters.

There are several issues to consider here. First, one can either assume that the trend 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). Ideally, one should therefore re-do the whole change-point analysis. Alternatively, one can assume that the change-point is given (but this should be clearly stated).

We illustrate the inclusion of covariates using the inverse Burr model from Section 4. To this model we add information on how democratic the parties involved in a particular war were 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. As our predictor variable we use the mean democracy score (w) of the countries involved in each war. This variable will be negative when a war involves mostly autocratic regimes, and large and positive if a war involves only democracies. We let the predictor influence the scale parameter μ, assuming that the effect of democracy shifts at the change-point:

$$\mu_{L,i} = \mu_{L,0}\exp(\beta_L w_i) \quad \text{and} \quad \mu_{R,i} = \mu_{R,0}\exp(\beta_R w_i). \tag{6.3}$$

Note that some of the wars have missing democracy scores. We remove these observations and ended up with 90 wars for this analysis.

The full model is becoming 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 commong α, in addition to the change-point τ. ML and change-point analysis can be carried out as per the general methods of Section 2. The inclusion of the democracy 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 assuming that the 1950.483 change-point is 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 β_L (estimate -0.007, 95% interval $[-0.202, 0.187]$) and β_R (estimate -0.163, 95% interval $[-0.308, -0.018]$). The estimated β_L is close to zero and its confidence interval covers zero, while the interval for β_R indicates that the scale parameter decreases as the mean democracy score increases. For states with Polity average index 5 of more, the number of battle deaths decreases with a factor of two or more after 1950 compared with states with an average index of 0 or less. The changing effect of democracy is also reflected in Figure 6.1, showing the fitted median as a function of mean democracy on both sides of the change-point.

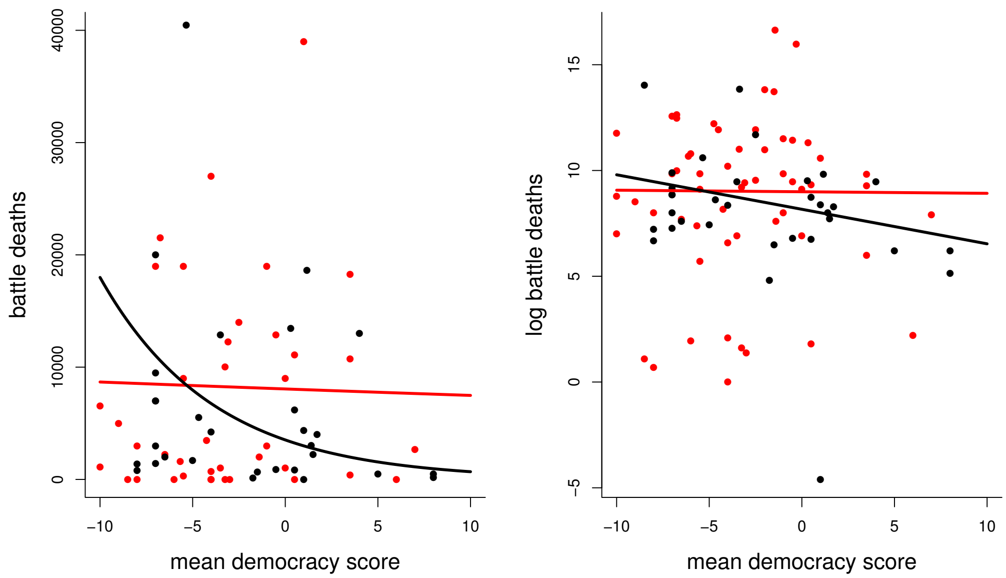

Figure 6.1: Regression with inverse Burr model, mean democracy as covariate and unequal effect of the covariate across the change-point. The left plot is on the z scale, while the right plot is on the $\log z$ scale (to be able to see all the wars in the same plot). Each of the 95 wars is represented by a point. The lines are the fitted medians; in red before 1950 and in black after.

This analysis lends some support to the already mentioned study by Mitchell, Gates & Hegre (1999) who argue that the role of democracy has become more pronounced over time. 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.

7 The between-war waiting times

Since Richardson (1948, 1960) a standard assumption in peace research has been that the between-war time periods $d_i = x_i - x_{i-1}$ follow the Expo(λ) distribution, for the relevant parameter λ, as with a constant intensity Poisson process. Poisson processes are interesting in this regard particularly because they have no statistical memory. The implication of this is that the number of wars that start in a period depends only on the length of that period, and not on the history of armed conflict up until that period. This assumption, although routinely used, see for instance Pinker (2011) and Clauset (2018, 2017) has seldom been the subject of much scrutiny. There are, however, several relevant extensions and refinements concerning this aspect of the CoW dataset.

7.1 Methodology

A more refined model, which can take on board that this intensity is not necessarily constant over time, emerges by placing a mixing distribution over the λ, leading to $f(d) = \int_0^\infty \lambda \exp(-\lambda d)\pi(\lambda)\,\mathrm{d}\lambda$. With $\pi(\lambda)$ a Gamma density with parameters (a, b), i.e. proportional to $\lambda^{a-1}\exp(-b\lambda)$, the result is

$$f(d) = \frac{b^a}{\Gamma(a)}\frac{\Gamma(a+1)}{(b+d)^{a+1}} = \frac{a}{b}\frac{1}{(1+d/b)^{a+1}} \quad \text{for } d > 0.$$

7.2 The inter-war periods for the CoW dataset

Fitting the simple Poisson process type model to the d_i leads to $\widehat{\lambda} = 0.522$. This would indicate that the probability of having a 15 year or more period without interstate wars, as we are blessed with since 2003, is $\exp(-15\,\widehat{\lambda}) = 0.0004$.

The two-parameter model pointed to above, where the λ is seen not as constant, but coming from a Gamma distribution, gives better fit, however. The ML estimates are $(7.395, 12.182)$, with mean 0.607 and standard deviation 0.223. The increase in twice-log-likelihood, when passing from the one- to the two-parameter model, say $D = 2(\ell_{\max} - \ell_{0,\max})$, is found to be 2.822. Its approximate null distribution is not the usual χ^2_1, since the one-parameter model does not correspond to an inner point in the two-parameter model. Its limit distribution is rather that of $\max(0, N)^2$, where N is standard normal; see Claeskens & Hjort (2008, Ch. 10.2). The p-value for testing the simple exponential is then found to be 0.046, indicating that the pure Poisson process view is too simple. This has implications for the broader study of armed conflict and calls into question this fundamental assumption. A more realistic estimate of the 15 year hiatus is hence $1/(1 + 15/b)^a$, estimated at 0.003.

If the stationary world hypothesis is taken at face value, this calculation would be sufficient to shoot it down; it would be extremely unlikely to experience a war-free period of 15 years or more. To some extent it could be interjected that the reason for this lies partly with the definitions underpinning the CoW dataset, where the drastic wars of Libya (2011) and Syria (since 2011) are classified as civil wars. We also note that the PRIO/UCDP Armed Conflict Database (ACD) (Gleditsch et al., 2002) has recorded interstate conflicts since 2003, but these have not reached the 1000 battle deaths criterion needed to be included in the CoW database.[5] We also note that the wars in Libya and Syria both have substantial international involvement, indeed they are both classified as internationalized internal conflicts in the ACD.

Having started with the CoW data (x_i, z_i) of (3.1), we have examined the z_i in Sections 3.1, 3.2, 4, and now the between-times $d_i = x_i - x_{i-1}$. Joint modelling of (d_i, z_i) can also be attempted, but plotting e.g. $(d_i, \log z_i)$ reveals no additional structure, with a correlation very close to zero.

8 Discussion

The work presented here has come about as a broad and fine-tuned statistical response to the interesting and stimulating contributions by Clauset. While Clauset argues that there is no clear evidence of change, neither in the sizes nor the times between interstate wars since 1823, we find evidence that a change in the distribution of war sizes has indeed taken place, and that it has

[5] The ACD counts five such wars.

happened in the years *after* WW2. Specifically, the parameters in this distribution changed in such a way as to generate smaller wars in the period after the change-point.

Inspired by Clauset (2018) and the CoW dataset we have treated a considerable number of statistical topics, which we trust to be useful in many contexts, also outside the peace research literature. Within our multifaceted contribution, we add to the 'long peace' debate in two main directions: (i) a flexible method for detecting change-points, assessing their uncertainty and the degree of change; and (ii) parametric modelling of the full battle death distribution.

One aspect of Clauset's investigations concerns testing whether 1939 can be seen as a change-point in the tail distribution of battle deaths (and his answer is negative). We have instead sought to find the time point of clearest change in the CoW dataset, without any prior beliefs related to WW2, and have landed in *1965 for the large wars*, and *1950 for the full sequence of wars*. Change-point methodology allows the possibility of making new discoveries, by generating new hypotheses, which in their turn may lead to further research and insights. Although not overwhelmingly significant in the traditional sense, we found these change-point estimates to be sufficiently precise to be worthy of further investigations. The analysis of the large wars with a simple exponential model gave the first indication that if there has been an abrupt change in the mechanism behind interstate wars it may not have happened in relation to WW2, but in the following years.

The different estimated change-points, for the full battle deaths distribution and only the large wars, 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 factoring in or 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 game-changer when examining more complex models involving the full battle death distribution. The change-point analyses by necessity point to a specific point in time in identifying when things change. We note, however, that this does not by any means necessarily indicate that the change that did happen occurred suddenly overnight. In some specific cases this may be the case; cf. the fall of the Berlin Wall or the collapse of the Soviet Union, which may have been change-points in specific circumstances. We are more inclined to interpret the change-points we identify here as the culmination of a process that has unfolded over some time. This could, for instance, be a process of a change in norms surrounding the use of violence, as argued by Pinker (2011), or it could be the result of changes in international governance as argued by Goldstein (2011). The notion of 'a change' depends on what we choose to observe and to study.

Further, we find that modelling the full sequence of battle deaths has many advantages compared to the simple exponential model we used in Section 3. Naturally, there is a gain in power from the increase in the number of observations. This gain manifests itself in narrower confidence curves for the change-point with the inverse Burr model compared to the exponential model. The inverse Burr model allots high confidence to several potential change-points, all of them later than or equal to 1939. Modelling the full sequence of war sizes also relieves us from having to determine a tail-index threshold, which despite Clauset's and our own efforts remain somewhat arbitrary. In addition, the more complex models we used in Section 4.3 offer a deeper and more nuanced impression of the war size distribution than can be obtained from just modelling the tail. Compare for instance the fitted inverse Burr model in Section 4.2 with the fitted exponential model in 3.3. Both models exhibit an increase in their power index parameter θ at their respective change-points, but the inverse Burr model provides additional information, with a considerable decrease in its scale parameter.

One of the more striking statements in Clauset (2018) is the claim that the current pattern of peace would need to endure for another 100 to 150 years for the long peace to be considered statistically significant. There are many ways to translate that statement into a precise statistical question within our framework. One possibility is to study the confidence curves for the degree of change in Section 4.3 and investigate how many years of new wars we would need to generate until the curve no longer contains 1 at some required level of confidence. We can generate new war timings from either one of the two distributions considered in Section 7 and then, independently, draw battle deaths from the fitted inverse Burr model to the right of the change-point. This procedure gives us a sequence of future wars and we can study for how many years we need to run this procedure until 'significance' (naturally we need to repeat the whole procedure many times). Again, there are choices to be made. If one considers the large wars to be the most relevant, we see that the ratio of 75% quantiles is already significantly different from 1 at the 95% level. Thus, the answer in that case is *zero years*. If one prefers to look at the ratio of medians, the situation is different: the confidence curve contains 1 from approximately the 80% level, and the 'time until significance' will be some way into the future. These type of analyses are somewhat non-standard and yield interesting revelations: when we study the ratio of medians in this way, the median to the right will benefit from the new observations and end up being estimated without errors. The median to the left however will never have more observations than the 60 observations it currently has and will therefore keep its current uncertainty. In statistical terms, there is a clear difference between assessing the median ratio $\rho = \mathrm{med}_L/\mathrm{med}_R$ associated with the two statistical models, and $\rho^* = \mathrm{med}_{L,\mathrm{obs}}/\mathrm{med}_R$, with $\mathrm{med}_{L,\mathrm{obs}}$ set to its actually observed number (which is 10374). We can ascertain here and now that $\rho^* > 1$ with 90 % confidence and no more need for waiting; the statistically more ambitious question pertaining to ρ has to do with being certain that the two statistical machines really are different with respect to their medians.

These investigations may inspire a short examination of a fundamental question: why do we model the past? Certainly, the battle deaths for the 95 wars between 1823 and 2003 are known, more or less without errors. In that sense, it is meaningless to discuss let us say the error associated with the median number of battle deaths from 1823 to 1950, since this number is not random (anymore), but known and fixed. However, as do the other authors in the field, we nonetheless work with statistical models, the viewpoint being that the observed battle deaths are realisations from an underlying distribution. How meaningful is this approach? Statistical methods in cases like these must be understood as tools that allow us to investigate different questions using 'lower-dimensional representations' of the data (as parametric models can be considered to be). These tools may enable us to identify and understand patterns, to pin-point influential factors, and build methods for predicting the future. These insights again help us plan actions to change our world in wished-for directions.

While we do not believe that the statistical models treated here really represent the mechanism generating war timings and sizes, the methods allow us to assess whether one potential pattern is more likely than another, given the inherent variation in the data. There are therefore sensible reasons for using statistical methods in the way that we and other authors do, but it is important not to forget the underlying model assumptions (and limitations). In further work, we hope to devote energy into the development of more realistic models for the underlying processes behind the data we observe.

While our analyses signal a decrease in war sizes after 1950, we are naturally careful with claiming that there has been a change in the general conflict-generating process. There are naturally many competing hypotheses that could explain the decrease in battle deaths in interstate wars (and these things should be investigated): changes in the types of conflicts? changes in definitions? changes in types of casualties (Fazal, 2014)?

It is beyond the scope of this article to provide much explanation for the two change-points we have identified. We hope that future studies will focus on this issue. Several hypotheses related to, for instance, the role and importance of the United Nations and the wider international architecture built up after WW2 to govern relations between states; the expansion of democracies world wide; the tightening of trade and commercial ties between countries; and the changing norms relating to the use of violence, all seem like plausible and worthwhile candidates to consider.

9 Concluding remarks

Here we offer a brief list of further statistical issues, some of which point to further research themes.

A: Predictions. Based on the tail-index behaviour of $z \geq z_0$ for extreme wars, we may predict the fatality numbers for future wars. With $\widehat{\theta}_R$ being the ML estimate for the $n_R = 14$ wars from 1965 and upwards, and the assumption that $z_{\rm new} \geq z_0$ follows the same post-Vietnam distribution in the future, we have $\widehat{\theta}_R\, y_{\rm new}$ follows an F distribution with degrees of freedom $(2, 2n_R)$, with $y_{\rm new} = \log z_{\rm new} - \log z_0$. This leads to the prediction confidence curve ${\rm cc}(z_{\rm new}) = F_{2,2n_R}(\widehat{\theta}_R y_{\rm new})$, portrayed to the left in Figure 9.1. Note that it is naturally rather skewed to the right. In the right panel we have given the median and the 0.10 and 0.90 prediction quantiles of this predictive distribution, corresponding to z levels 7913, 15189, 104751.

B: Per capita normalisation. Clauset found that the distribution of battle deaths has been relatively stationary, from 1823 to 2033, cf. our Figure 3.1. Our work above paints a somewhat more optimistic picture, with a change-point in 1965 and an expected level of fatalities lowered by a factor of two. These statements relate to the absolute number of battle fatalities. As soon as these figures are normalised, in a per capita manner, e.g. via world population numbers from 1823 to 2003, the picture is of course a different and more optimistic one. See Figure 9.2, where estimates ${\rm wp}_i$ for the world population for way year i, in billions, is plotted in, and where there is a downward trend over time in these relative $z_i/{\rm wp}_i$ numbers. We also mention the possibility of analysing say z_i/l_i, with l_i the length of war i (as measured in months or other time unit).

C: Extreme value statistics. There is a body of literature concerning the distributions for extreme counts, like the battle counts z_i. Relevant methods are surveyed in Embrechts, Klüppelberg & Mikosch (1997), along with certain types of application. Such methods might lead to more relevant models for the tails of the z_i distributions, with consequences also for predictions. See also Cirillo & Taleb (2016). To illustrate, consider $v_i = \log z_i - \log z_0$ for the 51 wars with $z_i \geq z_0 = 7061$. Extreme value theory suggests these ought to follow a distribution with cumulative function $G(v, \theta, a) = 1 - (1 - a\theta v)^{1/a}$, and where the $a \to 0$ case corresponds to $1 - \exp(-\theta v)$, which is equivalent to what we have used in analyses above. Fitting the data to the (θ, a) model leads to the small value $\widehat{a} = 0.072$, which is seen not to be significantly positive. This type of refinement might however be fruitful for other datasets.

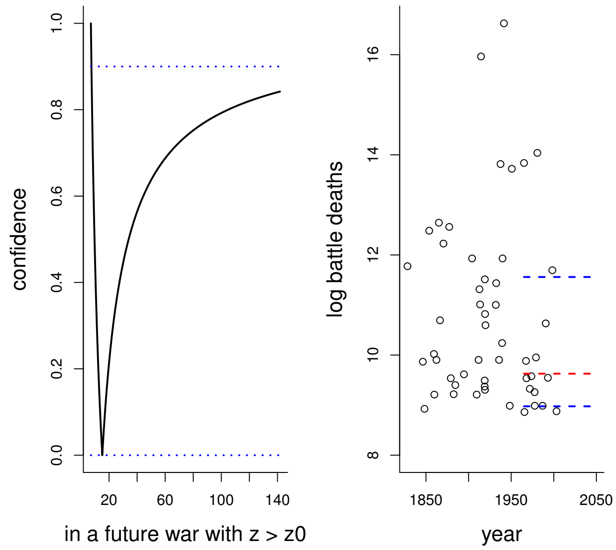

Figure 9.1: Left: prediction confidence curve for the number of battle deaths (in thousands), in a future war with $z \geq z_0 = 7061$. The median predicted number for such a war is 15189. Right: on the log scale, the red line indicates this median number, with the two blue lines indicating the 0.10 and 0.90 prediction quantiles, 7913 and 104751.

D: Monitoring plots for no-change stretches of data. When a model is worked with, for CoW type data, perhaps also with covariates, it is useful to have statistical monitoring methods to help sort out which stretches over time display no change, i.e. a stationary behaviour. Two types of such methods, along with monitoring plots, are in Hjort & Koning (2002) and in Cunen, Hermansen & Hjort (2018a). We have used these in the present context, and found that the (x_i, z_i) data have been sufficiently stationary from 1823 to 1965, and from 1965 to the present. Such plots can also be used to help identify breakpoints; see Hjort (2018).

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.

References

Beck, N. (1983). Time–varying parameter regression models. *American Journal of Political Science* **27**, 557–600.

Blackwell, M. (2018). Game changers: Detecting shifts in overdispersed count data. *Political Analysis* , 1–10.

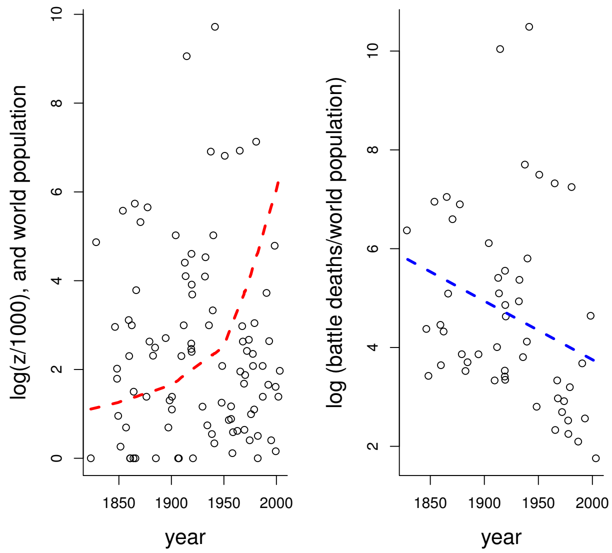

Figure 9.2: Left: the battle deaths $z_i/1000$, on the log scale, along with world population figures (red slanted), in billions. Right: the normalised battle deaths $z_i = \mathrm{wp}_i$, with a downward trend.

Cederman, L.-E. (2003). Modeling the size of wars: from billiard balls to sandpiles. *American Political Science Review* **97**, 135–150.

Cirillo, P. & Taleb, N. N. (2016). On the statistical properties and tail risk of violent conflicts. *Physica A: Statistical Mechanics and its Applications* **452**, 29–45.

Claeskens, G. & Hjort, N. L. (2008). *Model Selection and Model Averaging*. Cambridge: Cambridge University Press.

Clauset, A. (2017). The enduring threat of a large interstate war. Tech. rep., One Earth Foundation.

Clauset, A. (2018). Trends and fluctuations in the severity of interstate wars. *Science Advances* **4**, xx–xx.

Cunen, C., Hermansen, G. & Hjort, N. L. (2018a). Confidence distributions for change-points and regime-shifts. *Journal of Statistical Planning and Inference* **195**, 14–34.

Cunen, C., Walløe, L. & Hjort, N. L. (2018b). Focused model selection for linear mixed models, with an application to whale ecology. *Submitted for publication* **xx**, xx–xx.

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Berlin: Springer-Verlag.

Fazal, T. M. (2014). Dead wrong?: Battle deaths, military medicine, and exaggerated reports of war's demise. *International Security* **39**, 95–125.

Frigessi, A. & Hjort, N. L. (2002). Statistical models and methods for discontinuous phenomena [editorial overview, for a Special Issue of the journal, with Frigessi and Hjort as guest editors]. *Journal of Nonparametric Statistics* **14**, 1–6.

Gaddis, J. L. (1989). *The Long Peace: Inquiries Into the History of the Cold War*. Oxford: Oxford University Press.

Gat, A. (2006). *War in Human Civilization*. Oxford: Oxford University Press.

Gleditsch, N. P., Wallensteen, P., Eriksson, M., Sollenberg, M. & Strand, H. (2002). Armed conflict 1946-2001: A new dataset. *Journal of peace research* **39**, 615–637.

Goldstein, J. S. (2011). *Winning the War on War: The Decline of Armed Conflict Worldwide*. New York: Penguin Press.

Hjort, N. L. (2018). Towards a More Peaceful World [insert '!' or '?' here]. Tech. rep., FocuStat Blog at the University of Oslo.

Hjort, N. L. & Koning, A. (2002). Tests for constancy of model parameters over time. *Journal of Nonparametric Statistics* **14**, 113–132.

Hjort, N. L. & Schweder, T. (2018). Confidence distributions and related themes: Editorial overview, for a Special Issue of the journal, with Hjort and Schweder as guest editors. *Journal of Statistical Planning and Inference* **195**, 1–13.

Jullum, M. & Hjort, N. L. (2017). Parametric of nonparametric: The FIC approach. *Statitica Sinica* **27**, 951–981.

Maoz, Z. & Russett, B. (1993). Normative and structural causes of democratic peace, 1946–1986. *American Political Science Review* **87**, 624–638.

Marshall, M. G. & Jaggers, K. (2003). Polity IV project: Political regime characteristics and transitions, 1800–2003.

Marx, K. (1871). *The Civil War in France*. London: International Working-Men's Association.

Mitchell, S. M., Gates, S. & Hegre, H. (1999). Evolution in democracy-war dynamics. *Journal of Conflict Resolution* **43**, 771–792.

Pinker, S. (2011). *The Better Angels of Our Nature: Why Violence Has Declined*. Toronto: Viking Books.

Richardson, L. F. (1948). Variation of the frequency of fatal quarrels with magnitude. *Journal of the American Statistical Association* **43**, 523–546.

Richardson, L. F. (1960). *Statistics of Deadly Quarrels*. Los Angeles: Boxwood Press.

Sarkees, M. R. & Wayman, F. W. (2010). *Resort to war: a data guide to inter-state, extra-state, intra-state, and non-state wars, 1816-2007*. Washington, DC: CQ Press.

Schweder, T. & Hjort, N. L. (2016). *Confidence, Likelihood, Probability: Statistical Inference with Confidence Distributions*. Cambridge: Cambridge University Press.

Siegmund, D. (1988). Confidence sets in change-point problems. *International Statistical Review* **56**, 31–48.

Spengler, O. (1918). *Der Untergang des Abenlandes. Umrisse einer Morphologie der Weltgeschichte*. Wien: Braumüller.

Spirling, A. (2007). Bayesian approaches for limited dependent variable change point problems. *Political Analysis* **15**, 387–405.

Tilly, C. (1995). To explain political processes. *American Journal of Sociology* **100**, 1594–1610.

Western, B. & Kleykamp, M. (2004). A Bayesian change point model for historical time series analysis. *Political Analysis* **12**, 354–374.