


# Characterizing and Assessing Temporal Heterogeneity: Introducing a Change Point Framework, with Applications on the Study of Democratization

Gudmund Horn Hermansen<sup>1,2</sup>, Carl Henrik Knutsen<sup>3</sup>  and Håvard Mokleiv Nygård<sup>2</sup>

<sup>1</sup>Department of Mathematics, University of Oslo, Oslo, Norway. Email: [gudmund.hermansen@gmail.com](mailto:gudmund.hermansen@gmail.com)

<sup>2</sup>Peace Research Institute Oslo, Oslo, Norway. Email: [havnyg@prio.org](mailto:havnyg@prio.org)

<sup>3</sup>Department of Political Science, University of Oslo, Oslo, Norway. Email: [c.h.knutsen@stv.uio.no](mailto:c.h.knutsen@stv.uio.no)

## Abstract

Various theories in political science point to temporal heterogeneity in relationships of interest. Yet, empirical research typically ignores such heterogeneity or employs fairly crude measures to evaluate it. Advances in models for change point detection offer opportunities to study temporal heterogeneity more carefully. We customize a recent such method for political science purposes, for instance so that it accommodates panel data, and provide an accompanying R-package. We evaluate the methodology, and how it behaves when different assumptions about the number and abrupt nature of change points are violated, by using simulated data. Importantly, the methodology allows us to evaluate changes to different quantities of interest (for various estimators). It also allows us to provide comprehensive estimates concerning uncertainty in the timing and size of changes. We illustrate the utility of this flexible change point methodology on two types of regression models (Probit and OLS) in two empirical applications. We first re-investigate the proposition by Albertus (2017) that labor-dependent agriculture had a more pronounced negative effect on democratic survival before the “third wave of democratization.” Next, we utilize data extending from the French revolution to the present, from V-Dem, to examine the time-variant nature of the income–democracy relationship.

*Keywords:* change point model, temporal heterogeneity, panel data, democracy, democratization

*Political Analysis* (2020)

DOI: [10.1017/pan.2020.39](https://doi.org/10.1017/pan.2020.39)

**Corresponding author**  
Carl Henrik Knutsen

**Edited by**  
Jeff Gill

© The Author(s), 2020. Published by Cambridge University Press on behalf of the Society for Political Methodology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial reuse or in order to create a derivative work.

## 1 Introduction

Time is fundamental to our understanding of many political processes. For instance, influential theories suggest that certain points in time corresponded with structural changes that altered the “data-generating process” behind episodes of democratization (Huntington 1991), or even that these changes altered causal relationships between factors such as economic development and democracy (Boix 2011). Proposed changes to data-generating processes or particular causal relationships are often tied to terms such as “critical junctures,” “structural changes,” or “turning points” (e.g., Tilly 1995; Pierson 2011). Consider the “End of History” thesis formulated by Fukuyama (1992). The end of the cold war supposedly represented the culmination of human history understood as the struggle between fundamentally opposing ideas for how human society should be organized; democracy remained as the only legitimate regime. Consequently, the underlying likelihood of democratic onset and democratic reversal, as well as their determinants, may have changed.

These considerations point to the importance of explicitly assessing *temporal* heterogeneity in empirical studies of democratization. Similar considerations can be done for other political science questions. Yet, attempts to explicitly model temporal heterogeneity by empirical researchers are infrequent. Researchers that do assess such heterogeneity typically do so via “statistical fixes” that are easy to implement, but which come with limitations. Some researchers limit the time frame of the study, for instance studying determinants of democracy only during the “third wave

of democratization” (Teorell 2010). Others employ longer time series, but add temporal dummies to their models. Yet others go further and evaluate possible changes to the influence of particular covariates using split-sample- or Chow tests, or out of sample analysis (Hegre *et al.* 2013).

Such methods provide fairly coarse instruments for studying temporal heterogeneity and hinge on strong assumptions. Split-sample or Chow tests are easy to implement and efficient tools for estimating the size of the change, *if* the timing of the change is known or can be determined with high certainty. Yet, this assumption is often violated in practice, and identifying the timing of the change may often be equally interesting to estimating its size. Moreover, temporal dummies or split samples that typically span several decades may shed little light on the specific timing of a transition, even when they indicate that a relationship has changed.<sup>1</sup> Another concern is that some changes represent sharp breaks whereas other changes happen over protracted periods of time. Extant methods are poorly equipped to distinguish between such different types of changes, or to provide reasonable estimates of uncertainty for exactly when a change occurred.

Change point methods represent an alternative and arguably better suited modeling framework for handling temporal heterogeneity. These models are inductive in nature. They identify systematic patterns in the data, which researchers can interpret after the fact. Several important change point models have already been introduced to the discipline, illustrating that change point methods include a versatile and powerful class of models. Yet, change point methods remain rare in applied research. Why is this so, given the supposedly strong demand for tools that can appropriately assess temporal heterogeneity? One reason, we believe, is that considerably more work, and technical expertise, is required to fit an appropriate change point model compared to simply running a split-sample regression. Available change point methods still lack some of the “functionality” that applied researchers demand. This goes especially for researchers dealing with time series–cross section data, which are common in comparative politics and international relations.

To alleviate these issues, we adapt and further develop the change point *framework* originally developed by Cunen, Hermansen, and Hjort (2018) for political science purposes. This flexible framework can be applied to different data structures and estimation techniques. In other words, we introduce a framework that can incorporate a broad range of statistical specifications, rather than one specific change point model. Our applications pertain to probit and Ordinary Least Squares (OLS) models run on panel data, but the framework can accommodate most standard models used by comparative politics and international relations scholars. We have a built an R package to accompany this article, which should enable other researchers to employ the framework in applied research.<sup>2</sup>

This framework has several benefits. First, it allows researchers to fully account for the uncertainty both in the estimated location *and* the size of the change point through the use of so-called confidence curves. Second, most existing models are geared towards finding a shift in the mean level of some parameter. This framework, in contrast, makes it easy to obtain inference for the change in any sufficiently smooth function of the model parameters at the location of the change point, meaning that it can assess changes in a variety of distribution properties. Third, while we use it to study *temporal* heterogeneity, the framework can handle heterogeneity according to any ordinal variable (e.g., income or population size). In one application, we show how the framework can be used to study changes that happen at different time points in different groups

- 1 For a simulation illustrating how Chow-tests may give misleading indications on change point location, see Figure A-1. One flexible alternative that could, in principle, allow for identifying change point location is a model with a full set of time-dummies interacted with the relevant covariate. Yet, this approach is often intractable; it requires estimation of numerous parameters and introduces issues related to multiple testing. It is unclear what criteria should be used to determine a change if interaction terms for consecutive periods before/after a presumed change fluctuate in size and significance.
- 2 The R package will be available at GitHub [gudmundhermansen/CDCPRegression](https://github.com/gudmundhermansen/CDCPRegression). The full replication materials for this paper can be found at Hermansen, Knutsen, and Nygard (2020) [<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/XR3IDV>].

of observations (world regions and countries). Incorporating such group-heterogeneity could be useful for studying various political science questions, such as how technological, institutional, or ideological changes spread from one region or country to others.

Below, we first review previous use of change point models in political science, before we describe and illustrate our framework. By using simulated data, we also discuss key issues pertaining to studying temporal heterogeneity (e.g., whether the change occurred as a crisp break or more gradually), and how our framework handles such issues. Next, we demonstrate the usefulness of the framework in two applications. Both focus on issues of time-variant determinants of democracy and draw on panel data, with unbalanced panels and country-year as the unit of analysis, but differ in other regards. One uses a categorical dependent variable (and probit estimator) and the other a continuous outcome (and OLS estimator). Specifically, we first re-investigate the proposition by Albertus (2017) that labor-dependent agriculture had a more pronounced negative effect on democratic survival before the “third wave of democratization.” Next, we use extensive data from Varieties of Democracy (V-Dem; Coppedge *et al.* 2017a, b) to more inductively investigate temporal heterogeneity in the income–democracy relationship.

## 2 Change Point Models in Political Science

Models that allow researchers to study changes over time have been widely known to political scientists at least since the seminal contribution by Beck (1983) on how to estimate structural changes in regression models. Park (2012) unifies much of this literature and develops a (Bayesian) framework in which researchers can accommodate time-varying effects in both random- and fixed effects specifications. While, for example, Mitchell, Gates, and Hegre (1999) use Kalman filter models to study democracy and interstate conflict, these techniques have not been widely used.

More recent methodological advances have, instead, focused on the use of change point detection models.<sup>3</sup> In certain regards, such models generalize the use of temporal dummies in the classical regression framework. While using temporal dummies assumes the presence of a change at an *a priori* prespecified point in time, change point models instead allow researchers to treat the change point as a quantity that one can draw inferences about. In an early application, Western and Kleykamp (2004) used Bayesian change point models that treat the point of structural change as a parameter to be estimated. Focusing on 1965–1992, they show a structural break in the process of Organisation for Economic Co-operation and Development (OECD) wage growth in 1976. Yet, political scientists often employ limited and categorical dependent variables (democracy vs. autocracy, war vs. peace, etc.), and Spirling (2007) shows how change point models can be used to study count, binary, and duration-type data.

A limitation of these earlier models was that they generally required researchers to assume the presence of at least one change point. Recently, Blackwell (2018) introduced a Bayesian change point model for count data that uses a so-called Dirichlet prior. Notably, these models allow the researcher to remain agnostic about the number, or presence, of change points, and rather estimate both the number and temporal location of the change point(s) from the data. These models, however, mostly deal with time series data, such as the monthly global number of terrorist attacks or campaign contributions to a candidate (Blackwell 2018).

Yet, several research questions in comparative politics and international relations call for the use of time series–cross section data. The “workhorse” model on many topics continues to be an OLS, or alternatively Logit or Probit, regression fitted on time series–cross section data, often including unit- and/or time fixed effects and clustered standard errors. Unfortunately, available change point models are difficult to employ on this type of data. Researchers familiar with

3 Beyond political science, change point methods surface in, for example, engineering, biology, ecology, finance, meteorology, and literature studies. We will not review this wider literature, but see Frigessi and Hjort (2002) for a broad introduction to a special journal issue on discontinuities.

Bayesian methods may be able to adapt an existing model to this data structure, but this requires a level of technical and methodological expertise well beyond what is standard among most political scientists.

The framework that we introduce is frequentist in nature. It was used by Cunen, Hjort, and Nygård (2020) to locate a change point in the power-law tail of a single time series of battle deaths. Here, we extend this framework to handle time series–cross section data and to accommodate standard features such as fixed effects and clustered standard errors. This framework is particularly powerful in that it allows researchers to provide a full accounting of uncertainty for change point characteristics for all relevant parameters as well as its flexibility in accommodating different types of data, dependent variables, and estimation techniques.

### 3 The Change Point Framework: Theory and Estimation

Our change point methodology draws on—but substantially adapts and customizes for political science purposes—the state-of-the-art techniques from statistics developed by Cunen et al. (2018). This methodology uses the confidence distributions framework (Schweder and Hjort 2016) to assess uncertainty associated with both the location and effect of a change point.

Consider the observations  $y_1, \dots, y_n$  from a parametric model  $f(y, \theta)$ , with the parameter  $\theta$  taking the value  $\theta_L$  for  $y_1, \dots, y_\tau$ , and a different value  $\theta_R$  for the observations  $y_{\tau+1}, \dots, y_n$ . In this case, the methodology allows for pinpointing and providing a full inference for a change point, and also for the corresponding change in any sufficiently smooth function of the model parameter,  $\theta$ . In the applications presented below,  $\tau$  represents a point in time (e.g., a year), but the methodology can be used to study other features that generate an ordering of the observations  $y_1, \dots, y_n$ , such as income levels or degree of democracy. The general setup is flexible, and allows for studying heterogeneity in relationships of interest across very different contexts and models. In Cunen et al. (2018), the methodology is used for various regression and time series models. We extend the focus to applications in a panel regression environment.

To illustrate the methodology, consider a simple regression model where we only expect to see a change point in the intercept:

$$y_{i,t} = \begin{cases} \beta_L + \sigma \epsilon_{i,t} & \text{if } t \leq \tau \\ \beta_R + \sigma \epsilon_{i,t} & \text{if } t \geq \tau + 1 \end{cases}, \tag{1}$$

and where  $\epsilon_{i,t}$  are i.i.d. for  $i = 1, \dots, N$  and  $t = 1, \dots, T$ . For the sake of simplicity, we assume that  $\epsilon_{i,t} \sim N(0, 1)$ . In this model,  $\theta_L = (\beta_L, \sigma)$  and  $\theta_R = (\beta_R, \sigma)$ . The standard deviation is fixed and the only part of the model that can change is the intercept,  $\beta$ . We note, more generally, that determining what specific parts of a model that is allowed to change is potentially an important aspect of the model specification and this choice requires careful consideration.

The statistical task at hand is first to estimate  $\tau$ , that is, the location of the change point, along with related measures of uncertainty. Secondly, we would also like to draw inferences for the parameter of interest. This could be, for example, the difference between the intercepts  $\mu = \mu(\theta_L, \theta_R) = \beta_L - \beta_R$  in Model (1), but in general it can be any sufficiently smooth function  $\mu(\theta_L, \theta_R)$  of the model parameters.

For Model (1), the likelihood is given by

$$\ell_n(\tau, \theta_L, \theta_R) = \ell_n(\tau, \beta_L, \beta_R, \sigma) = \sum_{t \leq \tau} \log f(y_{i,t}, \beta_L, \sigma) + \sum_{t \geq \tau+1} \log f(y_{i,t}, \beta_R, \sigma),$$

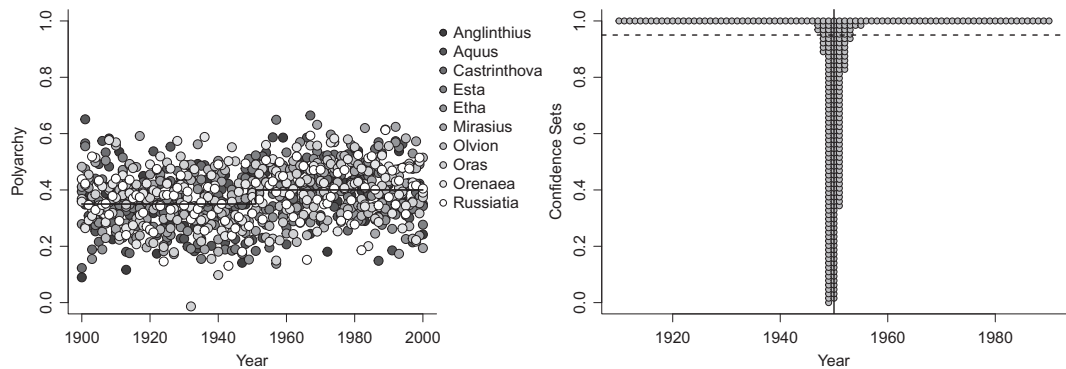
where  $f$  is the associated density. From this, we can compute the profile log-likelihood function

$$\ell_{\text{prof}}(\tau) = \max_{\beta_L, \beta_R, \sigma} \ell_n(\tau, \beta_L, \beta_R, \sigma) = \ell_n(\tau, \beta_L(\tau), \beta_R(\tau), \sigma(\tau)),$$

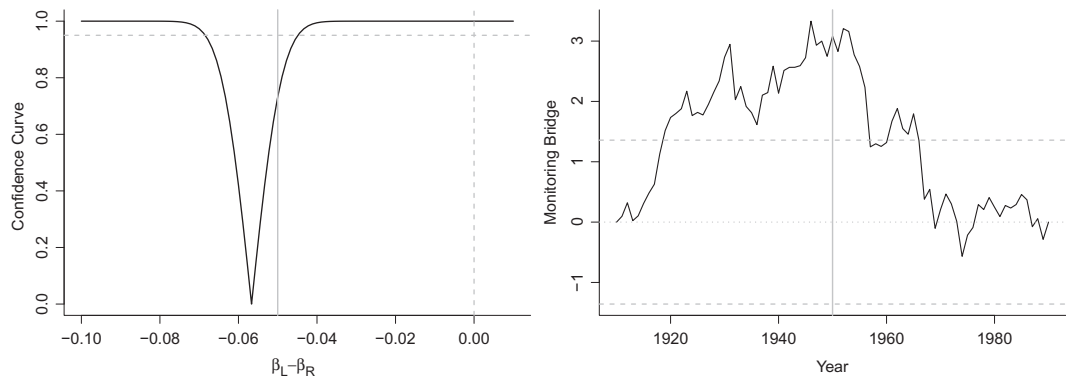
which is the maximization over  $\beta_L, \beta_R$ , and  $\sigma$  for a given  $\tau$ . The maximizer of  $\ell_{\text{prof}}(\tau)$ , resulting in a maximum likelihood estimate  $\hat{\tau}$ , also yields maximum likelihood estimators for the remaining parameters by  $\hat{\beta}_L = \hat{\beta}_L(\hat{\tau})$ ,  $\hat{\beta}_R = \hat{\beta}_R(\hat{\tau})$ , and  $\hat{\sigma} = \hat{\sigma}(\hat{\tau})$ .

The traditional way of reporting uncertainty in parameter estimates is by providing standard errors or, alternatively,  $t$ -values or confidence intervals. Here, we will instead build on recent work by Schweder and Hjort (2016) and use confidence distributions as a comprehensive tool to understand and report uncertainty, both the uncertainty for the location of  $\tau$  and the uncertainty associated with the parameters of interest  $\mu = \mu(\theta_L, \theta_R)$ .

Confidence distributions, and the closely related confidence curves (derived from the confidence distribution) are particularly useful for two reasons. First, they allow us to easily assess uncertainty at any confidence level—uncertainty can be read directly from the plotted confidence curve (see, e.g., Figure 1 or 2). Second, the general theory provides a powerful tool for combining “information” via confidence distributions or confidence curves across different data to assess uncertainty of more complex quantities of interest (see Schweder and Hjort, 2016). Here, we prefer the confidence curve as our main tool for summarizing inference. In brief, a (full) confidence curve—denoted by  $cc(\tau, y_{\text{obs}})$ , based on the observed dataset  $y_{\text{obs}}$ —has the following



**Figure 1.** Data simulated from Model 1 for 10 imaginary countries with a common true change in intercept from 0.35 to 0.40 at 1950–1951 (left panel), and corresponding confidence sets for the location of the change (right panel). The dashed line indicates the 95% confidence level.



**Figure 2.** Left panel: Confidence curve for the difference in intercept from Figure 1. Note that the confidence curve does not cross zero (dashed vertical line) for reasonable levels of confidence (the 95% confidence level is dashed horizontal line). Right panel: monitoring bridge for Model 1, based on the same observations as in Figure 1. The monitoring bridge plot does not tell us which part of the model that changes, only that there is evidence for some change. If the solid line crosses or comes close to one of the two dashed lines, this indicates that the assumption that the model stays unchanged (i.e., samples are homogeneous) across time does not hold. Here there is thus strong evidence of a change and our best guess (according to this method) is that it is located where the solid curve is maximized, which happens around 1947–1954.

interpretation: at the true change-point parameter,  $\tau$ , the set  $R(\alpha) = \{\tau : cc(\tau, Y) < \alpha\}$  has a probability (approximately) equal to  $\alpha$  with  $Y$  generated by the true model.<sup>4</sup>

To construct the confidence curve, we start with the deviance function, which is calculated based on the profile log-likelihood above. The deviance function is given by

$$D(\tau, Y) = 2\{\ell_{\text{prof}}(\widehat{\tau}) - \ell_{\text{prof}}(\tau)\}.$$

To obtain a confidence curve for  $\tau$  based on the deviance function, consider the estimated distribution of  $D(\tau, Y)$  at position  $\tau$ ,

$$K_{\tau}(x) = \Pr_{\tau, \widehat{\beta}_L, \widehat{\beta}_R, \widehat{\sigma}}(D(\tau, Y) < x).$$

Then we use a simulation procedure to construct the corresponding confidence sets by

$$cc(\tau, y_{\text{obs}}) = B^{-1} \sum_{b=1}^B I(D(\tau, Y_b^*) < D(\tau, y_{\text{obs}})) \tag{2}$$

for a large number,  $B$ , of simulated copies of datasets,  $Y_b^*$ , and where  $I(\cdot)$  is the indicator function.<sup>5</sup>

To illustrate how the methodology works, we will consider a few simple examples based on Model 1. In order to fix ideas, suppose that the outcome,  $y_{i,t}$ , is democracy, as measured by an index (let us call it “Polyarchy”) that ranges from 0 to 1, in a panel of imaginary countries, realized each year  $t$  from 1900 to 2000. Recall that Model 1 contains only an intercept (interpretable as mean Polyarchy score) and errors. But, we further assume that the intercept, for some reason, changes at  $\tau = 1950$ , so that  $\beta_L = 0.35$  and  $\beta_R = 0.40$ . We set the (i.i.d.) errors to:  $\sigma = 0.10$ .

Figure 1 (left panel) plots a simulated dataset for this panel of countries with the line marking the true evolution of  $\beta$ . The right panel shows the corresponding confidence set—the discrete version of the confidence curve arrived at by using the simulation method in Equation (2). For this particular dataset (i.e., this realization of the imaginary countries’ histories), the intercept change is sufficiently clear for the method to easily detect it; there is relatively little uncertainty regarding the year in which the change point is located. For the 95% confidence level—demarcated by the horizontal dashed line at 0.95—the confidence set includes the true change point (1950) and spans the years [1948, 1953]. This is indicated by the grey bars for these years crossing the dashed 0.95-line. If we were to be “more liberal” regarding the inference for when  $\tau$  occurred, and select a, say, 75% confidence level (construct a horizontal line from the  $y$ -axis at 0.75), we would have concluded that this confidence interval only covered [1949, 1951].

The confidence sets for  $\tau$  will *always* point to at least one location as the best guess, that is, where the confidence sets are closest to the  $\tau$ -axis. Consequently, we should *not* take this best guess to be correct without considering the associated uncertainty. If the model is sufficiently uncertain about whether *any change* has occurred, this will be reflected in the size of the confidence sets at (e.g.) the 95% level being very wide. Fortunately, also other pieces of information can further inform us about whether any parameter change has occurred at all, and if the change is substantively large enough to warrant further interest.

For some practical purposes, estimating the *size of the change* in a parameter—the difference  $\mu = \beta_L - \beta_R$  in our case—is more interesting than locating  $\tau$ . The method for constructing the confidence curve for the size of the change is based on a similar continuous parameter construction

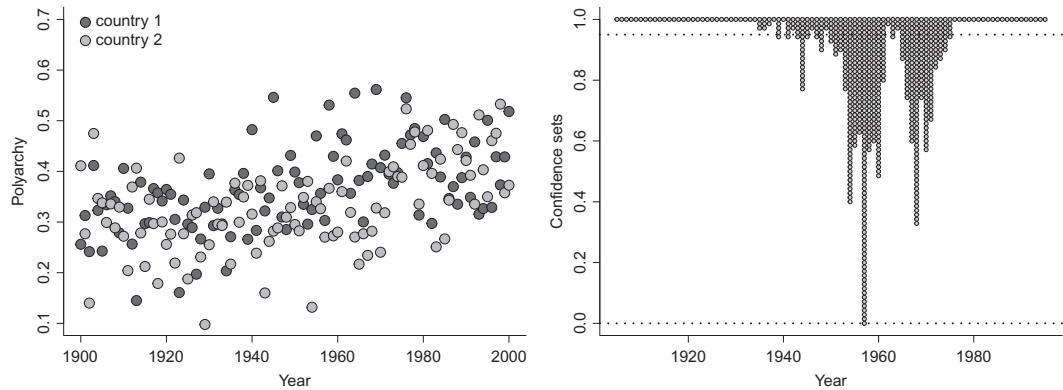
4 We write *approximately* in parenthesis since we sometimes use (large-sample) approximations derived from asymptotic theory and then this probability is only exact in the limit experiment.  
 5 For most standard parametric models, the Wilks theorem implies that  $K_{\tau}(x)$  is approximately the distribution function of a  $\chi^2$ . Here we rely on simulation, however, since there is no general Wilks theorem at play, since the parameter  $\tau$  is discrete.

as that of the discrete parameter version in Equation (2) (for details, see Cunen, Hermansen, and Hjort, 2018). The confidence curve for the degree of change is a useful tool for evaluating the likely substantial effect of the change point, which indirectly also informs about the probability of a significant and important change actually happening in the data. Here, the difference  $\mu = \beta_L - \beta_R$  should, for reasonable levels of confidence, not cross zero in order to be sufficiently interesting for further analysis. In other words, the estimated parameter change should not simultaneously be both positive and negative for reasonable levels of uncertainty. Figure 2, middle panel, shows that this is not the case for our simulated example on the intercept change for the Polyarchy model. The median estimate—as indicated by the minimum point for the solid line—is close to the true value of  $-0.05$ , and the 95% confidence interval for  $\mu = \beta_L - \beta_R$  extends from about  $-0.07$  to about  $-0.04$ .

The methodology just described assumes that there is an underlying change point. Therefore, the estimation and uncertainty is first and foremost related to *where*, and not *if*, there is a change in the underlying model. For most practical purposes, however, it makes good sense to investigate whether a change in the model is, indeed, reasonable to expect. If there are no true change points and the data generating process remains identical for the entire sample, this is typically reflected by very wide confidence sets, suggesting high uncertainty as to where the assumed change point is located. But, there are also other ways of assessing this question, including a so-called *monitoring bridge* plot (introduced in Hermansen, Hjort, and Kjesbu 2016). This is a visualization tool for investigating model homogeneity, and is based on the large-sample properties of the log-likelihood function under the assumption that the model is homogeneous across the sample. Figure 2, right panel, illustrates the tool for Model 1, for the simulated example. The plot indicates that ‘something’ happened between 1947 and 1954, since the solid line is maximized around these years and it also crosses one of the two dashed lines (here, the upper one). The latter property is what suggests that we cannot safely assume that the data-generating process is homogeneous across time.

The version of the methodology derived in Cunen et al. (2018) assumes that there is only one change point (although the methodology may be extended to accommodate multiple change points; see Cunen, Hermansen, and Hjort, 2018, section 10.3). Yet, for several real-world processes, changes could happen at different points in time, in different parts of the sample. One key strength of the confidence curves approach, however, is the possibility for combining multiple independent confidence curves. This, in turn, allows for greater flexibility in studying temporal heterogeneity. For example, for many political processes of change that happen across several countries, it is not plausible to assume that all countries experience the same change simultaneously—particular countries or regions may be ahead or behind others (see, e.g., our final empirical application on income and democracy, when separating world regions and individual countries). If so, both temporal dummies and split-sample designs may become laborious and impractical, unless we have clear prior specification of relevant subgroups of units. Confidence curves, in contrast, can deal with any ordering of groups or samples into coherent subgroups, for example similar countries and/or regions, where we expect to only see one change point per subgroup. And, the general methodology for confidence distributions is well suited to subsequently combine several, independent confidence curves into one combined confidence curve.

Figure 3, which represents a variation of Model (1), illustrates this point. It represents a dataset of two hypothetical countries that experience the same change in mean Polyarchy (+0.10), but at different years (1934–1935 and 1969–1970). We restrict this example to two countries instead of a panel of, say, 10 countries divided into two groups in order to ease visual interpretation and highlight key dynamics; extending this example to more than two countries is, however, straightforward. The left panel shows the simulated data points, whereas the right panel displays the confidence sets for  $\tau$  for the “naive,” combined model. This “naive” model looks for one change point in the pooled data, but the two countries actually have change points that are 35 years



**Figure 3.** Simulated data from Model 1 on two countries that experience a change in Polyarchy of the same amount (+0.10), but at different years 1934–1935 and 1969–1970 (left panel). The corresponding confidence sets are constructed by running the general method (2) for the combined dataset (right panel). Here, we do not get a clear answer to where the change point is located. The 95% confidence set includes almost all years from 1935 to 1975, with 1957 as the best guess.

apart. Indeed, the wide nature of the confidence set at the 95% level suggests that this model cannot clearly pin down a narrow time interval in which a change point occurred. Regarding the magnitude of the change (+0.10), however, we obtain much more reliable results if the analysis is done separately for the two countries, and then combined (right panel Appendix Figure A-2), than if the analysis is done simultaneously for the combined dataset (left panel Appendix Figure A-2).

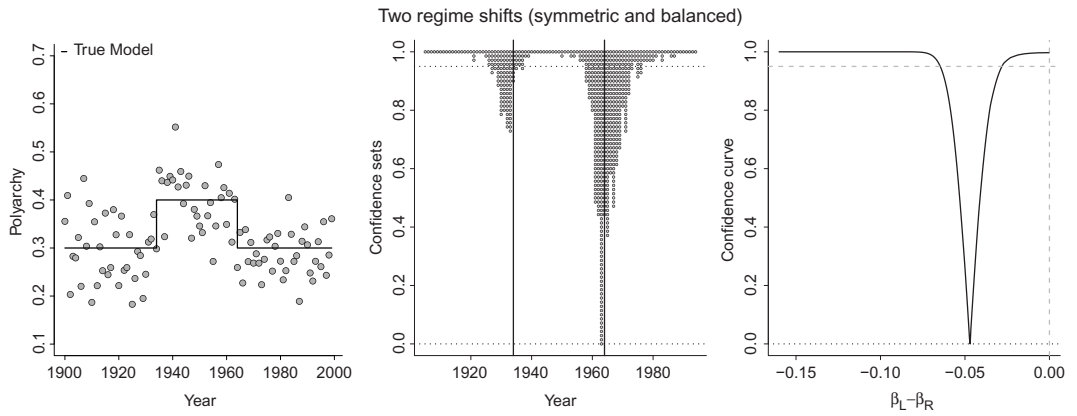
However, we could have multiple change points due to other mechanisms than different identifiable clusters of observations, such as countries or regions, experiencing changes at different points in time. For example, two or more structural breaks may occur across a longer time series in a particular  $X$ – $Y$  relationship, for a given set of units. As highlighted above, the original version of the framework developed in Cunen et al. (2018), and customized here, is mainly geared towards identifying one change point. Yet, it is also attuned to estimating the uncertainty about the location of that change point. Our simulations below show how the model behaves if there are actually multiple change points (that are jointly observed by all units).

To keep the illustrations and discussions as simple as possible, we restrict the discussion to simulated data from one country and situations with two change points; adding more countries with similar change points or having three or more change points would give more or less analogous discussions. We consider three different scenarios. Figure 4 captures a scenario with two identically sized change points with opposite signs. The scenario in Figure 5 is similar, but assumes that there is one larger (in terms of size of change) and one smaller change point. In Figure 6, we have two identically sized change points, where the two changes have similar signs.

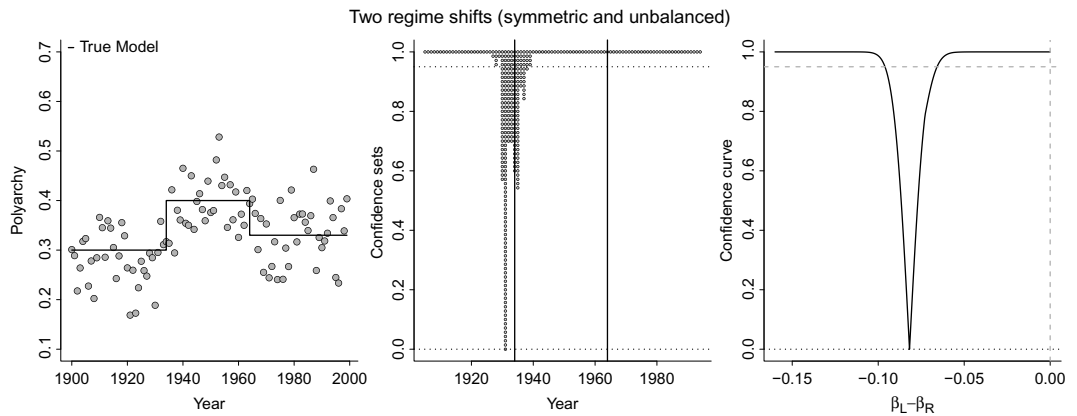
For the first scenario in Figure 4, with two equally sized change points and parameter-changes moving in opposite directions, the confidence sets tend to focus on one or both of the temporal locations, depending on the level of noise in the data. This is further illustrated by the left heatmap in Figure 7, displaying results from 100 simulations of this scenario. We tend to get indications of at least one change point, but with large uncertainty, and sometimes the confidence sets concentrate about equally on both change points.

For the second scenario with two imbalanced change points, Figure 5 (middle plot) illustrates that our method tends to focus mainly on the largest one, even when, in this case, the smaller parameter shift is about 70% the size of the larger one. Here, the level of noise is so large, compared to the size of the smaller change point, that the method often overlooks the smaller one. This is further illustrated by the middle heatmap of Figure 7. Hence, if our framework is applied to an empirical relationship of interest and detects a change point, this is not necessarily the *only* one. Instead, it may be the largest change point out of several.

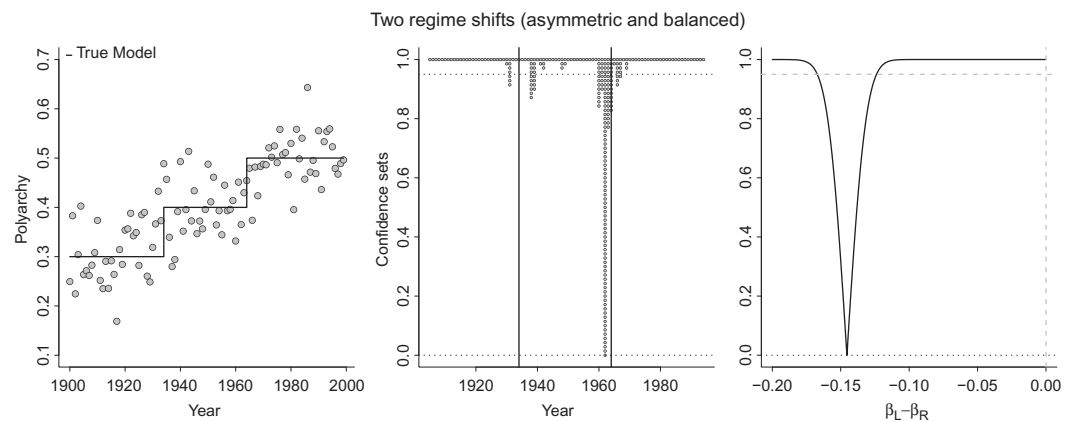




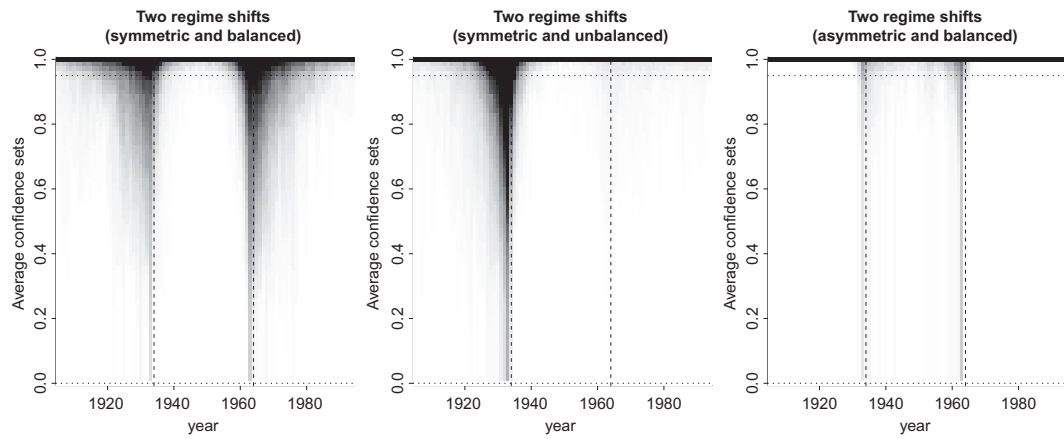
**Figure 4.** Data simulated with two similar change points at 1934 and 1964—change in mean from 0.3 to 0.4 and then back again to 0.3—under the same assumptions as in the above examples (left panel). The confidence sets (middle panel) indicates that there are two reasonable change point locations (concentrated on the two real change points). Yet, the method does not do a good job at estimating the degree of change in this scenario (best guess around  $-0,05$ ; right panel).



**Figure 5.** Data simulated with two change points; the change at 1934 is larger, of size 0.1 (from 0.3 to 0.4), than the change at 1964, which is of size 0.07 (from 0.4 to 0.33). Here, the method focuses on the largest change point.



**Figure 6.** Data simulated with two change points moving in the same direction. For this case, the method points to the leftrightmost change point. When running a larger number of simulations, we find that the method tends to put the estimated change point at or between the two true change points.



**Figure 7.** Heatmaps that aggregate and summarize the confidence sets from  $N = 100$  simulated datasets for models with two change points; as shown in Figures 4–6.

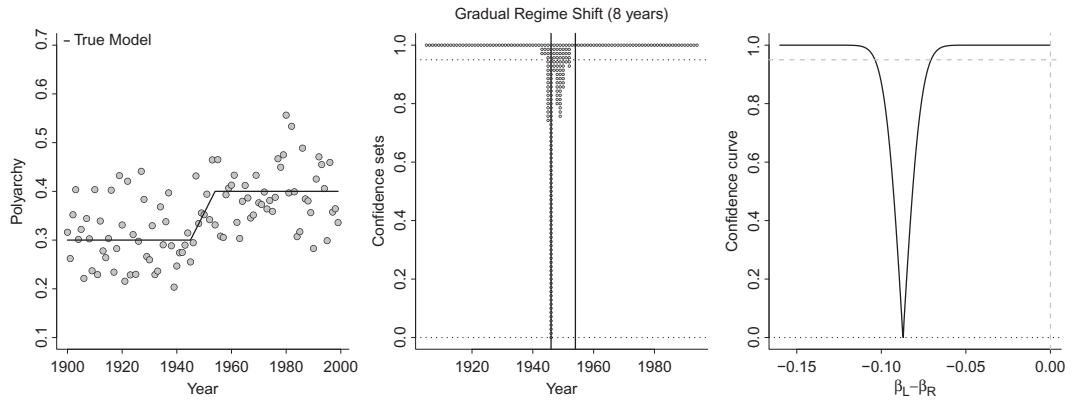
Third, Figure 6 assumes two identically sized change points with parameter shifts in the same direction. The right plot exemplifies that in this situation, the method tends to locate an estimated change point at, or close to, one of the actual change points (see also right heatmap of Figure 7).

Finally, changes may not always come as abrupt change points, but instead be gradual over several years. Concerning determinants of democratization, for example, changes generated by sudden shifts to the international system, such as the collapse of the Soviet Union, are likely abrupt, whereas changes generated by the diffusion of new ideas or technologies are likely gradual. Strictly speaking, our methodological framework is not constructed for gradual changes, nor are most other change point models, for that matter. However, as Figures 8 and 9 show, our model actually handles this type of mis-specification adequately and does a good job at determining the location of the change point and the corresponding degree of change. More specifically, we simulate data where the change in the parameter value (from 0.3 to 0.4 on Polyarchy) is a gradual change—here assumed to be linear—over 8 years, from 1946 to 1954. In Appendix Figure A-3, we make similar assumptions, but now assume the change occurs over 16 years, from 1942 to 1958.

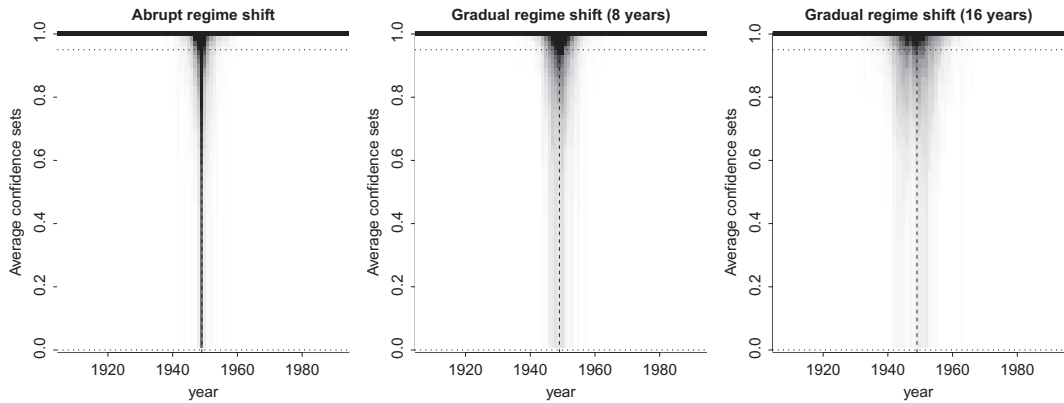
To provide a clearer picture of the more general performance of the method under these conditions, Figure 9 reports simulations from  $N = 100$  datasets, displaying heat-maps of the relevant confidence sets, for the two scenarios plus the benchmark case where the change in parameter value is abrupt (in 1949). The confidence sets tend to be wider and cover more years for the gradual than abrupt changes. The confidence sets are (as expected) wider for the 16 year than the 8-year scenario. Nonetheless, our evaluation is that the framework is useful for locating the change point in all these scenarios. While our set-up is, strictly speaking, not constructed for scenarios of gradual changes, the simulations indicate that it may still be used even where we anticipate changes to represent “change intervals” rather than “change points,” especially if intervals are short.

Before turning to applications on real-world data, we discuss a limitation: precisely identifying change points very early (or late) in a time series is difficult, due to limited information before (after) the change occurred. One practical solution is to omit the early and late years when estimating change point location.<sup>6</sup> We do this in our applications to ensure at least 4–10 data points for each estimated parameter, and conduct sensitivity analysis to assess the stability of results with respect to the selected range of years. Alternatively, one may restrict the complexity of the model and limit the number of parameters to be estimated. Hence, our framework will have more limited

<sup>6</sup> For unbalanced panels, this may even imply that certain units with very short time series are dropped altogether.



**Figure 8.** Data simulated with gradual changing regime shift over 8 years (from 1946 to 1954). Compared to a baseline case of an abrupt change in one year, the confidence set is somewhat wider.



**Figure 9.** Heatmaps that aggregate and summarize the confidence sets from  $N = 100$  simulated datasets, first with a normal abrupt change point and then for the two set-ups with a gradual change across, respectively, 8 and 16 year intervals (from Figure 8 and Appendix Figure A-3).

applicability—it will, for example, be difficult to estimate numerous categorical variables such as unit-fixed effects—and produce less precise results when time series are very short. We note that the use of bootstrapping in the final analysis may make our setup more stable and provide more reliable uncertainty estimates in small samples (compared to large-sample approximations).

#### 4 Application I: Labor-Dependent Agriculture and Democratic Survival

The point of departure for our first application is the recent study by Albertus (2017) on the production structure of the economy and democracy. Processes of urbanization and industrialization have often been considered key drivers of democratization, notably because they expand and strengthen two social groups with strong incentives to fight for democracy, the urban middle classes (Lipset 1959) and industrial workers (Rueschemeyer, Stephens, and Stephens 1992). In contrast, rural economies are widely presumed to be conducive to autocracy. Albertus points out that a negative association with democracy should mainly be anticipated in societies where agricultural production depends on reservoirs of cheap labor, and where these laborers do not own their own farmland, but work for large-scale land-owners. But, this relationship should have become weaker in recent decades, according to Albertus. He highlights three relevant changes—increased financial globalization, observed expropriation of land and land-reforms in several autocracies, and increased prevalence of civil war in rural areas—that were in motion (before and around the start of the “third wave of democratization” (1974).

Albertus proceeds to test for a heterogeneous relationship between his measure of labor-dependent agriculture and democratization and democratic survival. He employs the dichoto-

mous DD regime measure from Cheibub, Gandhi, and Vreeland (2010) and a dynamic probit specification. In brief, Albertus finds a nonrobust link between labor-dependent agriculture and democratization, but a negative relationship with democratic survival. Yet, when splitting his post-WWII sample in 1974, Albertus re-covers the robust relationship with democratic survival only in the pre-third wave sample. This corroborate the notion that labor-dependent agriculture is no longer as “bad for democracy” as it once was.

Yet, Albertus’ discussion on the three particular changes contributing to this shift makes it very clear that 1974 should *not* unequivocally be expected to be a crisp break-point. Indeed, Albertus notes that “[a]ll of these factors had begun to operate by the time of the third wave of democracy began with Portugal’s Carnation Revolution in 1974, and some had been operating even before” (p. 258). Thus, it is not clear why we should consider 1974 as the natural break point. We note that Albertus (2017), who is acutely aware of this issue, provides separate tests to study the mechanisms. He also carefully assesses the robustness of the results to alternative years for splitting the sample (indeed, the labor dependent agriculture coefficient on democratic duration is the most sizeable for the early time period when splitting the sample by 1969, see p. 261). Given the multiple mechanisms, 1974 is not a worse year to split the sample than, for example, 1972 or 1976, when using these conventional methods for assessing temporal heterogeneity. But, when employing our change point methodology, we are no longer forced to make this choice of change point, *a priori*.

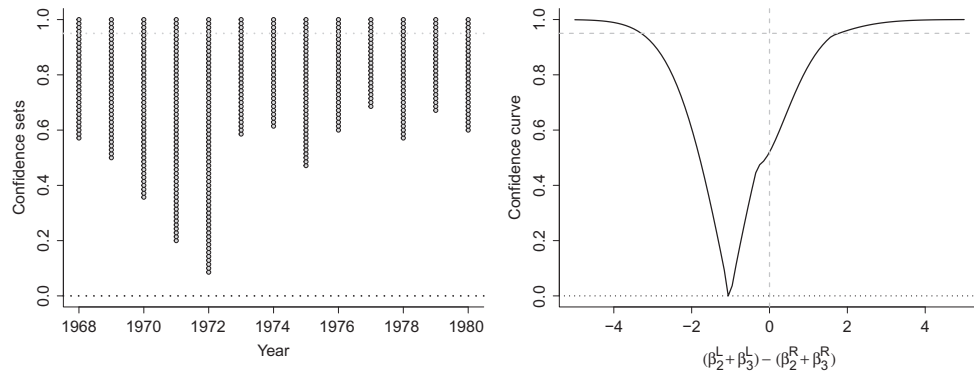
We employ the change point set-up described in the previous section. We follow Albertus (2017) in estimating a dynamic probit specification—more specifically his benchmark Model 3, Table 1—where  $D$  is the dummy variable capturing democracy,  $L$  is labor-dependent agriculture,  $\mathbf{X}$  is the above-listed vector of controls (which does not include country- and year-fixed effects), and  $j$  denotes country and  $t$  denotes year:

$$\Pr\{D_{j,t} = 1 \mid L_{j,t-1}, D_{j,t-1}, X_{j,t-1}, \beta, \beta_X, \beta_{XD}\} = \Phi(\beta_0 + \beta_1 L_{j,t-1} + \beta_2 D_{j,t-1} + \beta_3 L_{j,t-1} D_{j,t-1} + \beta_X X_{j,t-1} + \beta_{XD} X_{j,t-1} D_{j,t-1}). \quad (3)$$

Albertus’ study is not focused on identifying a break point in the overall regression model, but rather assessing a specific set of parameters, namely the estimated effect of labor repressive agriculture ( $\beta_2$ ) and this variable’s interaction with the lagged regime measure ( $\beta_3$ ). To focus more specifically on this, we use the change point methodology described above to probe for changes in  $\beta_2 + \beta_3$ —which can be interpreted as the relationship between labor-dependent agriculture and democratic duration/survival—while letting the others parameters stay constant over time.

The left panel of Figure 10 shows confidence sets for the location of the break point  $\tau$ . We focus only on 1968–1980, which is a reasonable approximation of the broader time period in which we would expect to see a change if the argument in Albertus (2017) is correct. A clear and crisp break point, for example, in 1974, would have been represented by the confidence sets, the gray dots, centering on this year, and not being spread across other years. Our method pinpoints 1972 as the most likely year for a change. However, by reading off the confidence sets for conventional levels of confidence (95% is indicated by the dashed line) we cannot reject the hypothesis that all years in the 1968–80 interval are *equally* likely candidates for the change point. We stress that one should not interpret this as implying that there *ipso facto* has been a change in the relationship between labor dependent agriculture and democratic survival, and that the change occurred somewhere between 1968 and 1980. The high level of uncertainty simply reflects that the method does not put much stock in a change happening in any of the particular years. Another plausible conclusion is thus that the method is indicating a situation of no change point.

This latter interpretation is further strengthened by the confidence curve for the difference between  $\beta_2 + \beta_3$  before and after the potential change point. Figure 10 (right panel) shows that,



**Figure 10.** Confidence sets, focus parameters from the Albertus model, representing change in the estimated coefficient of labor-dependent agriculture on democratic survival.

for all reasonable confidence intervals, the estimated change in the relationship between labor-dependent agriculture and democratic survival covers zero. The 95% confidence interval, for example, covers a change in  $\beta_2 + \beta_3$  from about  $-3.5$  to about  $+1.5$ , even if the point estimate for the difference is about  $-1.1$  (where the confidence curve in Figure 10 touches the x-axis). Hence, our results do not warrant a clear conclusion that the relationship has changed during this period of time.

In sum, when using our methodology for identifying a change point in the relationship between labor-dependent agriculture and democratic survival, we find little support for the specific hypothesis of a change point occurring in 1974. There is simply too much uncertainty associated with the potential change point to draw any strong inferences on when—or even whether—it occurred. When combined with a null-hypothesis of a constant relationship, a strict interpretation of our exercise would lead us to conclude that the relationship has *not* changed at all. This is, however, a premature conclusion. One plausible alternative explanation is that the (lack of) results may be driven by several issues with the underlying data:

The dataset used by Albertus (2017) has a considerable amount of missing data, which means that even if the time series on the surface is fairly long, the amount of available information is limited, especially at the beginning of the time series. Moreover, democratic onsets and breakdowns—as registered by the dichotomous DD regime measure from Cheibub et al. (2010)—are rare phenomena. Researchers using these data thus quickly run into degrees of freedom issues when estimating models with as many parameters as Albertus’ model. Below, we rely on data material that alleviate these issues, with longer time series, less missing data, and a continuous democracy measure with more frequent changes. This enables more precise estimation of change points.

## 5 Application II: Income and Democracy

The relationship between economic development and democracy is probably the most widely theorized and tested relationship in the democracy literature. Lipset (1959), in his seminal study, proposed that higher income levels increase the chances of countries becoming and staying democratic. Yet, recent studies have found mixed evidence (Acemoglu 2008). Empirical studies extending the time series back into the 19th century do, however, tend to find a stronger positive relationship between income and democratization (Boix 2011), and also a clearer relationship with democracy levels, even when accounting for country-fixed effects (Knutsen et al. 2019). The latter observations may be suggestive of temporal heterogeneity, which Boix (2011) theorizes and studies more carefully, for example, by using split-sample analysis. Boix argues that the number and the regime type of hegemonic actors, internationally, have varied across modern history and that these developments have strongly influenced the income–democracy link.

We re-assess the temporal heterogeneity of the income–democracy relationship by employing an OLS model on a graded democracy measure—complementing the probit regression above on a dichotomous measure, and thus displaying the flexibility of the change point set-up. We employ V-Dem’s core electoral democracy measure, Polyarchy, which extends from 1789 to 2018 (Teorell *et al.* 2019). Polyarchy relies on numerous indicators (mostly expert-coded, and scores are adjusted to ensure comparability across space and time by an item response theory (IRT) measurement model; see Marquardt and Pemstein 2018; Pemstein *et al.* 2020). The measure and is constructed to capture the democracy concept of Dahl (1971), and the theoretical range is 0–1 (0.01–0.95 in the data). The data on income, or more specifically Ln Gross Domestic Product (GDP) per capita, are from Fariss *et al.* (2017).

We estimate an OLS model with country-fixed effects ( $\phi_i$ ) and a third order polynomial for time trends ( $\theta_t = \beta_3 \text{year}_t + \beta_4 \text{year}_t^2 + \beta_5 \text{year}_t^3$ ). The country-fixed effects should alleviate concerns that time-invariant country-specific factors will bias the relationship, but we limit the addition of other covariates in order to mitigate issues of post-treatment bias and listwise deletion.<sup>7</sup> In the final analysis and inference, errors are clustered by country to account for panel-level autocorrelation:

$$\text{Polyarchy}_{i,t+1} = \beta_0 + \beta_1 \text{GDPpc}_{i,t} + \beta_2 \text{Polyarchy}_{i,t} + \phi_i + \theta_{t+1} + \epsilon_{i,t+1}. \quad (4)$$

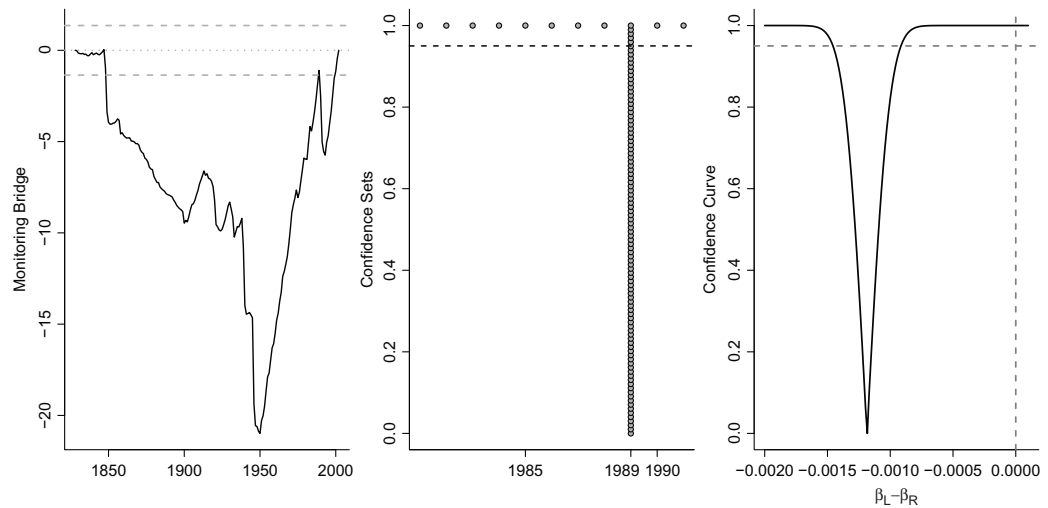
We use the same tools as above to probe for change points in this more parsimonious model. We initially include all polities with available data, globally, across the 1789–2015 time span. We focus on the years 1828–2002, and “shave off” the early and late parts of the sample where investigating change points is, by default, very difficult to do in a credible manner. Before presenting our analysis, we note one caveat: the income–democracy relationship may be influenced by the former affecting the latter (see, e.g., Acemoglu *et al.* 2008). Our results could thus partly pick up conditional correlation patterns coming from “reverse causality,” and we should not conclude unequivocally that identified change points reflect changes in a causal effect of income on democracy.<sup>8</sup>

Nonetheless, results are presented in Figure 11. The leftmost plot displays the monitoring bridge. The solid line crosses, and goes far beyond, the lower dashed line, providing evidence of temporal heterogeneity in the data-generating process. The middle panel pertains to the more specific question of when the GDP per capita coefficient displays a likely change point. The grey dots—falling well below the 95% confidence line—centers on one particular year, 1989. This was the year the Berlin Wall fell, and when democratizing changes started in several other Eastern European countries, where autocratic regimes had stayed in power with Soviet Union support, despite their relatively industrialized and developed economies (see also Boix, 2011). Right after 1989, the Soviet Union disintegrated (in 1991) and the Cold War ended, removing the structural conditions and lifelines of support that had helped keep many autocratic regimes (in both rich and poor countries in different regions) in power. 1989 is thus a plausible change point for the income–democracy relationship, and the confidence sets do not point to any other plausible candidate years for a structural change in this relationship.

Lastly, the right panel of Figure 11 pertains to the change in the magnitude of the income coefficient. This coefficient is interpreted as predicted change from  $t$  to  $t + 1$  on the 0–1 Polyarchy Index when Ln GDP per capita increases by one unit in  $t$ . The best estimate of  $\beta_{\text{GDP}}^L - \beta_{\text{GDP}}^R$  is

7 Including both country-fixed effects and a lagged dependent variable introduces the well-known (attenuating) Nickell bias, but this bias is negligible for as long time series as ours.

8 One possible solution would be to identify valid instruments of income and run our change-point framework on a 2SLS rather than OLS model. In general, drawing reliable inferences from 2SLS may require more customization of our framework. Yet, even without further customization, if the first and second stages of the 2SLS are preformed independently to the left and then the right of the potential change-points, everything else will be as in the standard OLS case. In general, as long as the observations are approximately i.i.d. to the left and to the right of the change-point, the general methodology should provide reliable inference for any statistical model.



**Figure 11.** A global aggregated model on Polyarchy. Does the model change over time? (Monitoring bridge, left plot). When does the relationship between GDP per capita and Polyarchy change? (Confidence sets, middle plot). What is the estimated change in the relationship? (Confidence curves for change GDP per capita coefficient; right plot).

around  $-0.001$ . Hence, the estimated relationship between income and democracy has become larger over time ( $\beta_{\text{GDP}}^L - \beta_{\text{GDP}}^R < 0 \implies \beta_{\text{GDP}}^R > \beta_{\text{GDP}}^L$ ). But, the estimate is also indicative of a very small change, albeit a statistically significant one; the 95% confidence interval for  $\mu = \beta_{\text{GDP}}^L - \beta_{\text{GDP}}^R$  does not cover zero. One plausible reason for why the estimated change is so small, is the presence of multiple change points, which could come at different points in time and vary in size, across different regions. We elaborate on this more complex scenario in the next section.

## 5.1 Geographically Specific Temporal Heterogeneity

Both the frequency of democratization episodes *and* the (perceived) drivers of regime change have differed substantially across regions of the world (see Haerpfer *et al.* 2019). The assumption that every region, or for that matter country, should experience the same shift in the income–democracy relationship, at the exact same time, is thus a strong one. While our focus is on assessing and understanding temporal heterogeneity, various types of geographic heterogeneity is also important to applied researchers. Hence, we briefly display and discuss how our framework can incorporate such additional heterogeneity.

Before presenting these results, we note that a specialized literature already exists, with customized models that allow researchers to explicitly account for group-based heterogeneity and estimate what are the relevant groups directly from the data (e.g., Bonhomme and Manresa 2015; Ando and Bai 2016). Yet, these models are not designed to deal with all the aspects of temporal heterogeneity addressed by our framework, and specific models, such as the one developed by Bonhomme and Manresa (2015), also become computationally challenging as the size of the dataset increases. Our change point framework is flexible enough to incorporate group-based heterogeneity, for example, through the use of confidence curves that can be combined (into different groups) across multiple, distinct estimations (for more details, see Schweder and Hjort 2016). For example, we may disaggregate the global sample from the previous sections and run individual estimations for all countries, as long as they have sufficiently long time series, before combining countries into groups according to prior knowledge or some criterion (estimated sign or size of change points, timing of change points, etc.).<sup>9</sup> We include such analysis on individual

<sup>9</sup> The latter approach is more problematic, since data are used twice and the second part of the analysis is dependent on the first. This may introduce a bias to the final inference, and is related to so-called “postselection inference.” In general,

countries in the appendix. Below, we employ the somewhat stronger assumption that change-points are uniform for countries that belong to a prespecified region of the world (but may differ across regions).

We use the eight-fold regional classification by Miller (2015), which is a modified version of the regional categorization by Hadenius and Teorell (2007), in order to divide the world into subsamples. Next, we re-run the OLS model on Polyarchy detailed above, with country-year as unit of analysis, on each region. The regions are Eastern Europe and the (post-)Soviet space (1), Latin America (2), Middle East and North Africa (3), Sub-Saharan Africa (4), Western Europe and North America (plus Australia and New Zealand) (5), East Asia (6), South-East Asia (7), and South Asia (8). We focus on three of these regions, with results for remaining regions being plotted in the Appendix. Figure 12 shows diagnostics plots for Eastern Europe and the Soviet space, Middle East and North Africa (MENA), and Western Europe and North America. The monitoring bridges (left column of Figure 12) provide substantial evidence that structural changes in the “data-generating process” behind democracy occur, at different points in time, for each region; the curves cross the dashed lines at least once.

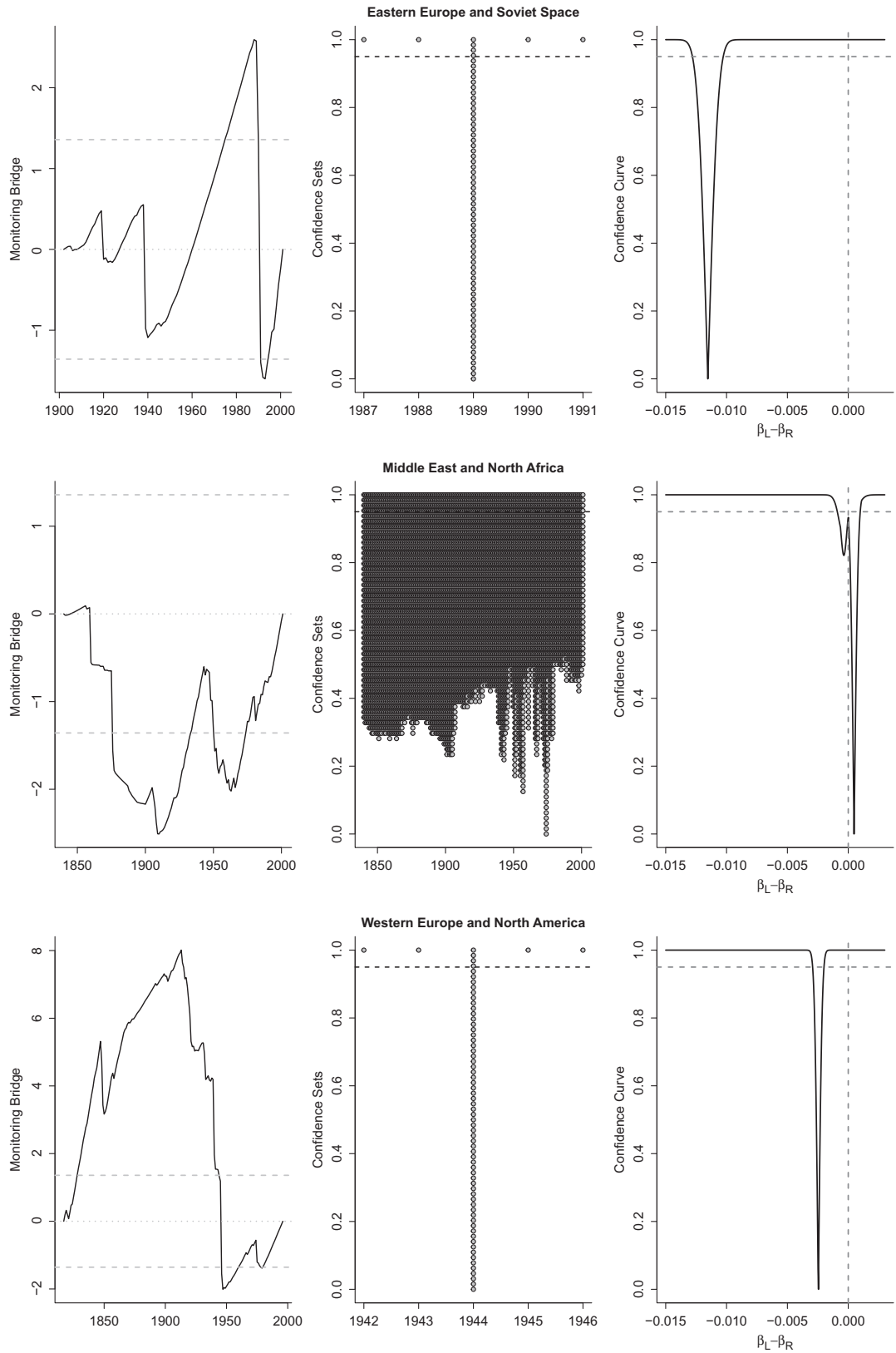
However, we are here primarily interested in the relationship between income and democracy, rather than the overall regression model. When we focus on this relationship, we find distinctly estimated change-point years in different regions (middle column). For Eastern Europe and the Soviet space (top row), we find a change point in 1989, the year of revolutions in Eastern Europe. Indeed, 1989 is not only the maximum likelihood estimate, it is also the only year in which the method places any confidence in a potential change point. For Western Europe and North America (bottom row), there is clear evidence that the change point occurred earlier, in 1944, towards the end of WWII and Allied victory. For MENA (middle row), our methodology does not locate any unique point in time in which the relationship changed. The method reports a maximum likelihood estimate, namely 1974, but the 95% confidence interval covers *all* years included in the study. Whereas the income–democracy relationship has changed in some regions, our results suggest that such a change may not have occurred in MENA.

The confidence curves in the right columns of Figure 12 indicate the change in the coefficient on income—that is, the size of  $\mu = \beta_{\text{GDP}}^L - \beta_{\text{GDP}}^R$ —for the different regions. For Eastern Europe and the Soviet space as well as Western Europe and North America the estimated change is negative—indicating that the development–democracy relationship has become more pronounced after the change—and does not overlap 0 at the 95% confidence level. For Eastern Europe in 1989, the estimated change in the income coefficient (–.011) is much larger than what we estimated for the global analysis (–.001). Also for Western Europe and North America, the estimated change (–.002) is larger than globally, though less pronounced than in Eastern Europe. For MENA, in contrast, the maximum likelihood estimate is essentially 0 and there is no statistically significant pattern to discern.

One plausible interpretation of these results is that the identified change points mark junctures at which income became a relatively more important factor in affecting regime developments, compared to region-specific factors that dominated up until that point. For Western Europe, WWII and Nazi occupation of many countries may have dominated the income effect in explaining regime development. For Eastern Europe, 1989 marks the end of the Cold War. One interpretation, along the lines discussed in Boix (2011), is that Soviet influence, and the larger dynamic of the US versus USSR competition, washed out any effect of income on the level of democracy in this region during the Cold War, and kept countries, both rich and poor alike, autocratic. This suppression of the potential effect of income, however, ends with the collapse of the Soviet Union, as—to put it

one key question is whether using the data to select the groups introduces a more substantial error to the main inference than potential errors from estimating all countries together.





**Figure 12.** Regressions on Polyarchy, with country-year as unit of analysis, subsampled by region: Change point investigation for Eastern Europe and Soviet space (top row), Middle East and North Africa (middle row), and Western Europe and North America (bottom row). Monitoring bridges (left column), confidence sets (middle column, where we have chosen years where there was something to see) and confidence curves (right column).

simply—both rich and poor countries are allowed to democratize without external intervention, but rich countries are more susceptible to do so.

Finally, we note that the assumption that change points are identical for countries in the same region is, of course, a strong one. Both the timing and size of change points, across units, may follow a variety of patterns. Hence, we point readers to the appendix, where we conduct and discuss an even more fine-grained analysis with change points estimated separately for each individual country. Yet, also this analysis displays a strong clustering of change points for Eastern European countries around 1989, corroborating at least one aspect of the regional analysis.

## 6 Concluding Discussion

We have introduced and discussed a novel approach, building on the framework developed in Cunen et al. (2018), for detecting, describing and drawing inferences about change points in statistical relationships. We have used this approach—both in a more deductive fashion to test for a specific, hypothesized change point, and in a more inductive fashion to ‘let the data speak’ on where likely change points are—in two empirical applications on the study of democracy: First, we replicated the recent study by Albertus (2017). When doing so and using our change point methodology, we show that the hypothesized shift in the relationship between labor-intensive agriculture and democratic breakdown at the beginning of the “third wave of democratization” is associated with much more uncertainty than conventional approaches suggest. Second, we use new and extensive time series data from V-Dem, going back to the French Revolution, to re-evaluate the relationship between income and democracy. This study indicates that, globally, the most important change point—corresponding with an increased strength in the link between income and democracy—occurred relatively late, around the end of the Cold War. Yet, disaggregated analysis focusing on specific regions shows that this strengthening of the income–democracy relationship occurred only in some regions, and then at different points in time.

The approach to modeling change points that we have taken has several notable benefits, which should make it suitable to a range of empirical questions in political science (and related disciplines). We have described and illustrated these benefits in the paper, both by using simulations and the two empirical applications, but let us briefly summarize them here:

First, it is a very flexible approach, statistically, as it can be fitted to different types of data and estimators. We illustrated the approach by employing it to panel data, and using OLS and probit models. It can also be used to infer about changes in different parts of the statistical model, both concerning particular (combinations of) parameters but also the overall data-generating process.

Second, the framework can be applied to a number of relevant real world scenarios that political scientists may face. Notably, while the framework is originally developed for identifying one, crisp, change–point—and thus certainly has its limitations—our simulations reveal that it works adequately well and is still useful even in some cases where these conditions are only approximately true. These include situations when changes occur gradually over a (limited) time interval, as well as situations where there are several change points of different magnitudes, where our approach will then often detect the most important one. In other words, our framework is fairly robust against certain types of model mis-specifications that are presumably common in real-world political science applications.

Third, the use of confidence distributions theory and, in particular, confidence curves allow us to give a more comprehensive assessment of uncertainty pertaining to inferences about change points, including their temporal location *and* the size of the change. This is an important benefit, as many existing approaches to detecting changes over time could lead to over-confident conclusions about the timing and nature of structural breaks in relationships of interest to political scientists.

Finally, the framework is accessible to empirical researchers. Alongside this article, we provide an R-package that will allow others to conduct the same type of assessments and tests that we have done in our applications on democracy for various relationships of interest.

## Funding

This work was supported by Research Council Norway (grant numbers 240505, 275400).

## Acknowledgments

We thank Nils Lid Hjort, Dan Pemstein, Magnus B. Rasmussen, Sebastian Ziaja, Alexander Batur, three anonymous reviewers and the Editor of Political Analysis, as well as participants at the 2018 Historical V-Dem Workshop in Oslo, 2018 V-Dem Annual Conference in Gothenburg, and 2019 Annual EPSA meeting in Belfast for very valuable comments and suggestions. Thanks also to Michael Albertus for generously sharing his replication data.

## Data Availability Statement

The replication materials for this paper can be found at Hermansen *et al.* (2020) or <https://doi.org/10.7910/DVN/XR3IDV>.

## Conflict of Interest

There is no conflict of interest to disclose.

## Supplementary Material

For supplementary material accompanying this paper, please visit <https://doi.org/10.1017/pan.2020.39>.

## Bibliography

- Acemoglu, D. 2008. *Introduction to Modern Economic Growth*. Princeton, NJ: Princeton University Press.
- Acemoglu, D., S. Johnson, J. A. Robinson, and P. Yared. 2008. "Income and Democracy." *American Economic Review* 98(3):808–842.
- Albertus, M. 2017. "Landowners and Democracy: The Social Origins of Democracy Reconsidered." *World Politics* 69(2):233–276.
- Ando, T., and J. Bai. 2016. "Panel Data Models with Grouped Factor Structure Under Unknown Group Membership." *Journal of Applied Econometrics* 31(1):163–191.
- Beck, N. 1983. "Time-Varying Parameter Regression Models." *American Journal of Political Science* 27(3):557–600.
- Blackwell, M. 2018. "Game Changers: Detecting Shifts in Overdispersed Count Data." *Political Analysis* 26(2):230–239.
- Boix, C. 2011. "Democracy, Development, and the International System." *American Political Science Review* 105(4):809–828.
- Bonhomme, S., and E. Manresa. 2015. "Grouped Patterns of Heterogeneity in Panel Data." *Econometrica* 83(3):1147–1184.
- Cheibub, J., J. Gandhi, and J. Vreeland. 2010. "Democracy and Dictatorship Revisited." *Public Choice* 143(1–2):67–101.
- Coppedge, M. *et al.* 2017a. V-Dem country-year dataset v7.1.
- Coppedge, M. *et al.* 2017b. V-Dem v.7, codebook.
- Cunén, C., G. Hermansen, and N. L. Hjort. 2018. "Confidence Distributions for Change-Points and Regime Shifts." *Journal of Statistical Planning and Inference* 195(1):14–34.
- Cunén, C., N. L. Hjort, and H. M. Nygård. 2020. "Statistical Sightings of Better Angels: Analysing the Distribution of Battle-Deaths in Interstate Conflict Over Time." *Journal of Peace Research* 57(2):221–234.
- Dahl, R. A. 1971. *Polyarchy: Political Participation and Opposition*. New Haven, CT: Yale University Press.
- Fariss, C. J., C. D. Crabtree, T. Anders, Z. M. Jones, F. J. Linder, and J. N. Markowitz. 2017. "Latent Estimation of GDP, GDP Per Capita, and Population from Historic and Contemporary Sources." Working paper.
- Frigessi, A., and N. L. Hjort. 2002. "Statistical Models and Methods for Discontinuous Phenomena." *Journal of Nonparametric Statistics* 14(1–2):1–5.
- Fukuyama, F. 1992. *The End of History and the Last Man*. New York: Free Press.
- Hadenius, A., and J. Teorell. 2007. "Pathways from Authoritarianism." *Journal of Democracy* 18(1):143–156.

- Haerpfer, C., P. Bernhagen, C. Welzel, and R. F. Inglehart (Eds.) 2019. *Democratization*, 2nd edn. Oxford: Oxford University Press.
- Hegre, H., J. Karlsen, H. M. Nygård, H. Strand, and H. Urdal. 2013. "Predicting Armed Conflict 2010–2050." *International Studies Quarterly* 55(2):250–270.
- Hermansen, G. H., N. L. Hjort, and O. S. Kjesbu. 2016. "Recent Advances in Statistical Methodology Applied to the Hjord Liver Index Time Series (1859–2012) and Associated Influential Factors." *Canadian Journal of Fisheries and Aquatic Sciences* 73(2):279–295.
- Hermansen, G. H., C. H. Knutsen, and H. M. Nygard. 2020. "Replication Data for: Characterizing and assessing temporal heterogeneity: Introducing a change point framework, with applications on the study of democratization." <https://doi.org/10.7910/DVN/XR3IDV>, Harvard Dataverse, V1.
- Huntington, S. P. 1991. *The Third Wave: Democratization in the Late Twentieth Century*. Norman, OK: University of Oklahoma Press.
- Knutsen, C. H. et al. 2019. "Economic Development and Democracy: An Electoral Connection." *European Journal of Political Research* 58(1):292–314.
- Lipset, S. M. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *American Political Science Review* 53(1):69–105.
- Marquardt, K., and D. Pemstein. 2018. "IRT Models for Expert-Coded Panel Data." *Political Analysis* 26(4):431–456.
- Miller, M. K. 2015. "Democratic Pieces: Autocratic Elections and Democratic Development since 1815." *British Journal of Political Science* 45(3):501–530.
- Mitchell, S. M., S. Gates, and H. Hegre. 1999. "Evolution in Democracy-War Dynamics." *Journal of Conflict Resolution* 43(6):771–792.
- Park, J. H. 2012. "A Unified Method for Dynamic and Cross-Sectional Heterogeneity: Introducing Hidden Markov Panel Models." *American Journal of Political Science* 56(4):1040–1054.
- Pemstein, D. et al. 2020. "The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data." V-Dem Working Paper no. 21, 5th Edition. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3595962](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3595962).
- Pierson, P. 2011. *Politics In Time: History, Institutions and Social Analysis*. Princeton, NJ: Princeton University Press.
- Rueschemeyer, D., E. H. Stephens, and J. D. Stephens. 1992. *Capitalist Development and Democracy*. Chicago, IL: University of Chicago Press.
- Schweder, T., and N. L. Hjort. 2016. *Confidence, Likelihood, Probability*. Cambridge: Cambridge University Press.
- Spirling, A. 2007. "Bayesian Approaches for Limited Dependent Variable Change Point Problems." *Political Analysis* 15(4):387–405.
- Teorell, J. 2010. *Determinants of Democracy: Explaining Regime Change in the World, 1972–2006*. Cambridge: Cambridge University Press.
- Teorell, J., M. Coppedge, S. I. Lindberg, and S.-E. Skaaning. 2019. "Measuring Polyarchy Across the Globe, 1900–2017." *Studies in Comparative International Development* 54(1):71–95.
- Tilly, C. 1995. "To Explain Political Processes." *American Journal of Sociology* 100(6):1594–1610.
- Western, B., and M. Kleykamp. 2004. "A Bayesian Change Point Model for Historical Time Series Analysis." *Political Analysis* 12(4):354–374.