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

Gudmund Horn Hermansen
Carl Henrik Knutsen
Håvard Mokleiv Nygård

November 2019

Working Paper
SERIES 2019:93
THE VARIETIES OF DEMOCRACY INSTITUTE

UNIVERSITY OF GOTHENBURG
DEPT OF POLITICAL SCIENCE

Varieties of Democracy (V-Dem) is a new approach to conceptualization and measurement of democracy. The headquarters – the V-Dem Institute – is based at the University of Gothenburg with 17 staff. The project includes a worldwide team with six Principal Investigators, 14 Project Managers, 30 Regional Managers, 170 Country Coordinators, Research Assistants, and 3,000 Country Experts. The V-Dem project is one of the largest ever social science research-oriented data collection programs.

Please address comments and/or queries for information to:

V-Dem Institute

Department of Political Science

University of Gothenburg

Sprängkullsgatan 19, PO Box 711

SE 40530 Gothenburg

Sweden

E-mail: contact@v-dem.net

V-Dem Working Papers are available in electronic format at www.v-dem.net.

Characterizing and assessing temporal heterogeneity: Introducing a change point framework, with applications on the study of democratization.*

Gudmund Horn Hermansen[1,3], Carl Henrik Knutsen[2], and Håvard Mokleiv Nygård[3]

[1]Department of Mathematics, University of Oslo
[2]Department of Political Science, University of Oslo
[3]Peace Research Institute Oslo

November 28, 2019

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. It also allows us to provide comprehensive and nuanced estimates concerning uncertainty in the timing and size of changes. We illustrate the utility of the 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 outcomes 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.

*Authors listed in alphabetical order. We thank Nils Lid Hjort, Dan Pemstein, Magnus B. Rasmussen, Sebastian Ziaja, Alexander Baturo, 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. The research has been funded by the Research Council Norway, 'Young Research Talent' grant, pnr 240505 and 'Young Research Talent' grant, pnr 275400.

1 Introduction

Time is fundamental to our understanding of many political processes. For example, influential theories suggest that certain points in time corresponded with structural changes that altered the "data-generating process" behind episodes of democratization (Huntington 1991). Some theories even suggest that these changes altered causal relationships between factors such as economic development and democratic outcomes (Boix 2011b). 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., Pierson 2011; Tilly 1995). Consider, for instance, 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; liberalism won and 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.

The above 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 . Nonetheless, attempts to explicitly model temporal heterogeneity by empirical researchers are few and far between. The researchers that do assess temporal heterogeneity typically do so via one of several 'statistical fixes' that are easy to implement, but which come with limitations. Some researchers limit the time frame of the study, for instance studying the determinants of democracy only during the 'Third Wave of Democratization' (Teorell 2010). Other researchers employ longer time series, but then typically 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, more recently, out of sample analysis (Hegre et al. 2013).

Still, such methods provide fairly coarse instruments for studying temporal heterogeneity. To name one concern, temporal dummies or split samples that typically span at least a couple of decades shed little light on the specific timing of a transition. 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 specific change took place.

Change point methods represent an alternative and arguably better suited modelling framework for handling temporal heterogeneity. These models are inductive in nature. They identify systematic patterns in the data and then researchers can interpret these patterns after the fact. A number of change point methods have already been introduced to the discipline. These let researchers treat the structural change, or the change point, as a parameter to be estimated (Western and Kleykamp 2004), and allow for studying change points in count, binary, and duration-type data (Spirling 2007). More recently, Blackwell (2018) developed a Bayesian change point model for count data that uses a Dirichlet prior, which allows researchers to remain agnostic about the number of change points in a time series.

These are important contributions – together they show 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 theories that postulate temporal heterogeneity? The reason, we believe, is straightforward: considerably more work, as well as technical expertise, is required to fit an appropriate change point model compared to simply running a regression on a split sample. Available change point methods still lack some of the 'functionality' that applied researchers need, in practice, 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 develop the change point framework originally developed by Cunen, Hermansen, and Hjort (2018) for political science purposes. This is a flexible framework that can be applied to different data structures, models, and estimation techniques, thus giving applied researchers all the 'functionality' they are used to. Importantly, the framework allows researchers to fully account for the uncertainty both in the estimated location of the change point and the degree of the change. While we here use it to study *temporal* heterogeneity, the framework is flexible and can handle heterogeneity according to any ordinal variable (e.g., income or population size). We have a developed an R package that accompanies this article, which should make it relatively easy to employ the methods described here.[1]

We discuss how this framework can overcome, or at least mitigate, issues pertaining to studying temporal heterogenity. By using simulated data, we discuss fundamental, but largely neglected, issues relating, e.g., to whether the change occurred as a crisp break or more gradually over time. Next, we demonstrate the usefulness of the framework in two applications. Both are focused on issues of time-variant determinants of democracy and both draw on panel data. But the applications differ in other relevant regards. For example, 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 outcomes before the 'third wave of democratization'. Next, we use extensive data from the Varieties of Democracy (V-Dem; Coppedge et al. 2017; Coppedge et al. 2017b) dataset to more inductively investigate temporal heterogeneity in the income–democracy relationship.

2 Change point models in political science

Models that allow researchers to study how something changes over time have been 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 a large part of this literature and develops a (Bayesian) framework in which researchers can accommodate time-varying effects in both random- and fixed effects specifications. These techniques, however, have not been widely used by empirical researchers in political science. One exception is Mitchell, Gates, and Hegre (1999), who use Kalman filter models to study the relationship between democracy and interstate conflict, and find that the pacifying effect of democracy on interstate war has increased over time.

More recent methodological advances have, instead, focused on the use of change point detection models. In certain regards, such models generalize the use of temporal dummies in the classical regression framework. While the use of temporal dummies assumes the presence of a change at an *a priori* pre-specified point in time, change point models instead allow researchers to treat the

[1] Link to package:

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 structural change, or the change point, as a parameter to be estimated. Focusing on the 1965–1992 period, they show that a structural break in the process of wage growth happened in 1976. Research in comparative politics and international relations often employs 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 the researcher to assume the presence of at least one change point. Recently, Blackwell (2018), has introduced a Bayesian change point model for count data that uses a so-called Dirichlet prior. This is similar to the model developed by Fox et al. (2011) for studying speaker diarization in an audio file. The strength of these models is that they allow the researcher to remain agnostic about the number, or presence, of change points. Instead, one can estimate both the number and temporal location of the change point(s) from 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, for example covering numerous countries across time, with each country constituting one time series. The 'workhorse' model on many topics in these fields continues to be an OLS, or alternatively Logit or Probit, regression fitted on time series–cross section data, invariably including country- 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 Bayesian methods may be able adapt an existing model to this data structure, but this requires a level of technical and methodological expertise well beyond what is standard among researchers in these sub-fields.

The model that we introduce is frequentist in nature. It is also easily able to handle time series–cross section data, allows for the inclusion of standard features such as fixed effects and clustered standard errors, and, importantly, can do so while still providing full estimates of uncertainty for all parameters in the change point model. In the following section, we detail a specification that can accommodate different types of dependent variables and estimation techniques.

3 The change point method: estimation and uncertainty

Our change point methods draws on – but substantially adapts and customizes for political science purposes – the state-of-the-art techniques from statistics developed by Cunen, Hermansen, and Hjort (2018). In addition to allowing for a formal test of the presence of a change point, this methodology allows for a full assessment of the uncertainty of the change point through confidence distributions (Schweder and Hjort 2016); details on this follows below.

Consider the observations $y_\mathrm{i} = y_1, \dots, y_n$ from a parametric model $f(y, \theta)$, with the parameter θ taking the value θ_L for $y_1, ..., y_\tau$, and a different value θ_R for the observations $y_{\tau+1}, ..., y_n$. In this case, the methodology allows for pinpointing and providing a full inference for a change point τ for the parameter θ. In most applications – including the ones presented below – τ represents time (e.g. year). Yet, the methodology can also be used to study, as τ, other features that generate an ordering of the observations $y_1, \dots, y_n$, such as income level or degree of democracy. This flexibility opens up for studying heterogeneity in relationships of interest across very different contexts. In Cunen, Hermansen, and Hjort (2018), the methodology is used to analyse and find break points in time series models. Here we extend it to be used in a (panel) regression environment.

To illustrate the methodology, take a simplified regression model where τ is related to time (e.g. year) and where we only expect to see a change point in the intercept. Consider the model:

$$y_i = \begin{cases} \beta_L + \sigma\epsilon_i \text{ if } i \leq \tau \\ \beta_R + \sigma\epsilon_i \text{ if } i \geq \tau + 1 \end{cases}, \tag{1}$$

where ϵ_i are i.i.d. For the sake of simplicity, we assume that $\epsilon_i \sim \mathrm{N}(0, 1)$. In this model, $\theta_L = (\beta_L, \sigma)$ and $\theta_R = (\beta_R, \sigma)$. The standard deviation is here assumed to be the same before and after τ, and the only part of the model that is allowed to change is the intercept, β. 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 in some cases careful consideration. The statistical task at hand is first to estimate τ, i.e., the existence and location of the change point, along with related measures of uncertainty. And then obtain inference for the parameter of interest, which for simplicity is here assumed to be $\mu = \beta_L - \beta_R$.

For the model (1), the likelihood is given by

$$\ell_n(\tau, \theta_L, \theta_R) = \ell_n(\tau, \beta_L, \beta_R, \sigma) = \sum_{i \le \tau} \log f(y_i, \beta_L, \sigma) + \sum_{i \ge \tau+1} \log f(y_i, \beta_R, \sigma),$$

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 maximisation over β_L, β_R and σ for a given τ. The maximizer of $\ell_{\text{prof}}(\tau)$, resulting in a maximum likelihood estimate $\widehat{\tau}$, also yields maximum likelihood estimators for the remaining parameters by $\widehat{\beta}_L = \widehat{\beta}_L(\widehat{\tau})$, $\widehat{\beta}_R = \widehat{\beta}_R(\widehat{\tau})$, and $\widehat{\sigma} = \widehat{\sigma}(\widehat{\tau})$.

The next step is to assess the uncertainty of these estimates. The traditional way of reporting uncertainty of 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 the associated uncertainty. These tools are used to assess both the uncertainty for the location of τ and the uncertainty associated with the parameters of interest, the so-called focus parameter; i.e., $\mu = \mu(\theta_L, \theta_R)$. This, parameters of interest, could be, for example, as simple as the difference between the intercepts β_L and β_R in the model (1), or any other sufficiently smooth function $\mu(\theta_L, \theta_R)$ of the model parameters.

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. Indeed, the extent of uncertainty can be read directly from the plotted confidence curve; see e.g. Figure 1 or 2. Second, the general theory provide a powerful and flexible tool for combining "information" via confidence distributions or confidence curves across different models to assess uncertainty of more complex quantities of interest.[2] Here, we prefer the confidence

[2]See Schweder and Hjort (2016) for a detailed account of confidence distributions.

curve as our main tool for summarizing inference. In brief, a (full) confidence curve – which we will denote by $cc(\tau, y_{\text{obs}})$, based on the observed dataset y_{obs} – has the following interpretation: at the true change-point parameter, τ, the set $R(\alpha) = \{\tau : cc(\tau, Y) < \alpha\}$ must have probability (approximately) equal to α with Y generated by the true model.

Here, 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 τ based on the deviance function, consider the estimated distribution of $D(\tau, Y)$ at position τ,

$$K_\tau(x) = \text{Pr}_{\tau,\widehat{\beta}_L,\widehat{\beta}_R,\widehat{\sigma}}(D(\tau, Y) < x)$$

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

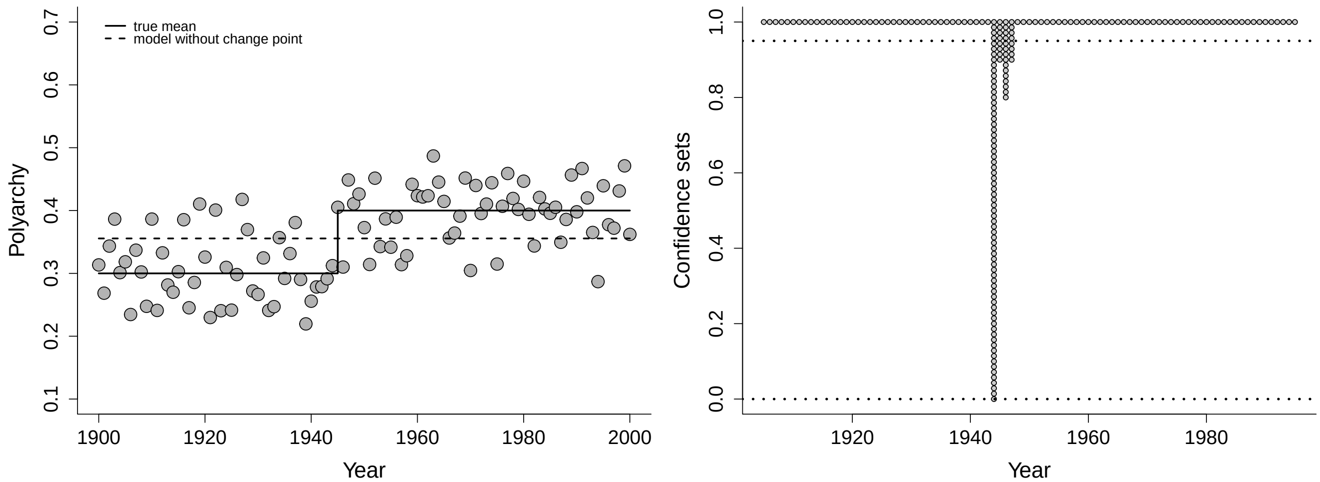
$$cc(\tau, y_{\text{obs}}) = B^{-1} \sum_{j=1}^{B} I(D(\tau, Y_j^*) < D(\tau, y_{\text{obs}})) \tag{2}$$

for a large number, B, of simulated copies of datasets, Y^*, and where $I(\cdot)$ is the indicator function.[3]

To illustrate how the methodology works, and instead of going into details on how to construct confidence curves by simulation using Equation 2 (see Cunen, Hermansen, and Hjort 2018), we will consider a few simple examples based on model (1). In order to fix ideas, suppose that the outcome in this model, y_i, is democracy, as measured by an index (let us call it 'Polyarchy') that ranges from 0-1, in an imaginary country, realized each year$_i$ from 1900 to 2000. Recall that the simple 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 $\tau = 1944$, so that $\beta_L = 0.30$ and $\beta_R = 0.40$. We set the (i.i.d.) errors to: $\sigma = 0.06$.)

[3]For most standard parametric models, the Wilks theorem implies that $K_\tau(x)$ is approximately the distribution function of a χ^2_1. Here we rely on simulation, however, since there is no general Wilks theorem at play.

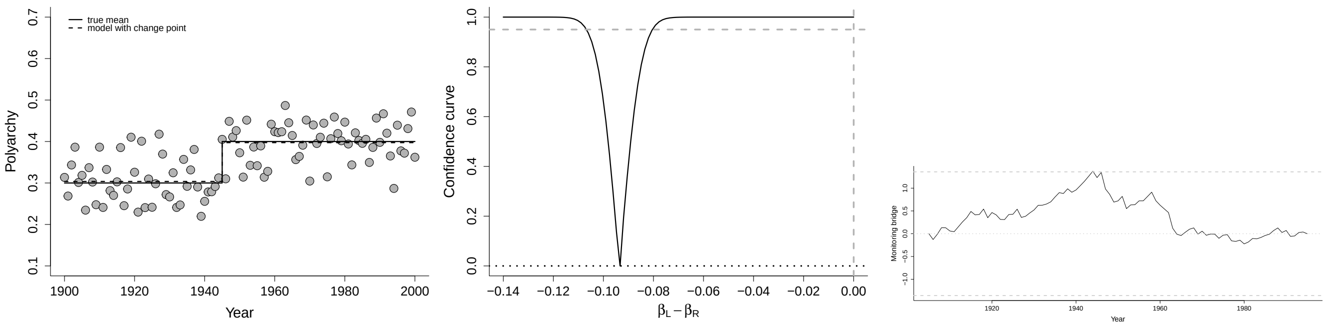
Figure 1: Data simulated from the simple model (2) with a true change in intercept from 0.3 to 0.4 at 1944-1945 (left panel). The corresponding confidence sets for the location of the change (right panel). The dashed line indicate the 95% confidence level.



In the left panel of Figure 1, a simulated dataset for this imaginary country is plotted against the true evolution of β. The corresponding confidence set, which is the discrete version of the confidence curve arrived at by using the simulation method in Equation 2, is shown in the right panel. For this particular dataset (i.e., this realization of the imaginary country's history), the change in the intercept is sufficiently clear for the method to detect it; there is relatively little uncertainty regarding the year in which the change point is located. For the 95 percent confidence level – demarcated by the horizontal dashed line at 0.95 – the confidence set includes the years $[1944, 1947]$. 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 τ occurred, and select a 75 percent confidence level (construct a horizontal line from the y-axis at 0.75), we would have concluded that this confidence interval only covers $\tau = 1944$, the true value of the change point.

The confidence sets for τ will always point to *at least* one location as the best guess – where the confidence sets are closest to the τ-axis – for where a change point is located. 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, there are other pieces of information that 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.

Figure 2: Simulated data (left panel) and confidence curve (middle panel) for the difference in intercept from Figure 1. Note that the confidence curve does not cross zero for any reasonable levels of confidence (i.e. around the 95% confidence level (dashed line)). Also, we display the monitoring bridge (right panel) for the model in (1) based on the same observations as in Figure 1. If the solid line crosses or comes close to one of the two dashed lines (as happens here around 1944-45), this indicates that the assumption that the model stays unchanged (i.e. that the samples are homogeneous) across time does not hold. The monitoring bridge plot does not tell us which part of the model that changes, only that there is evidence for some change.



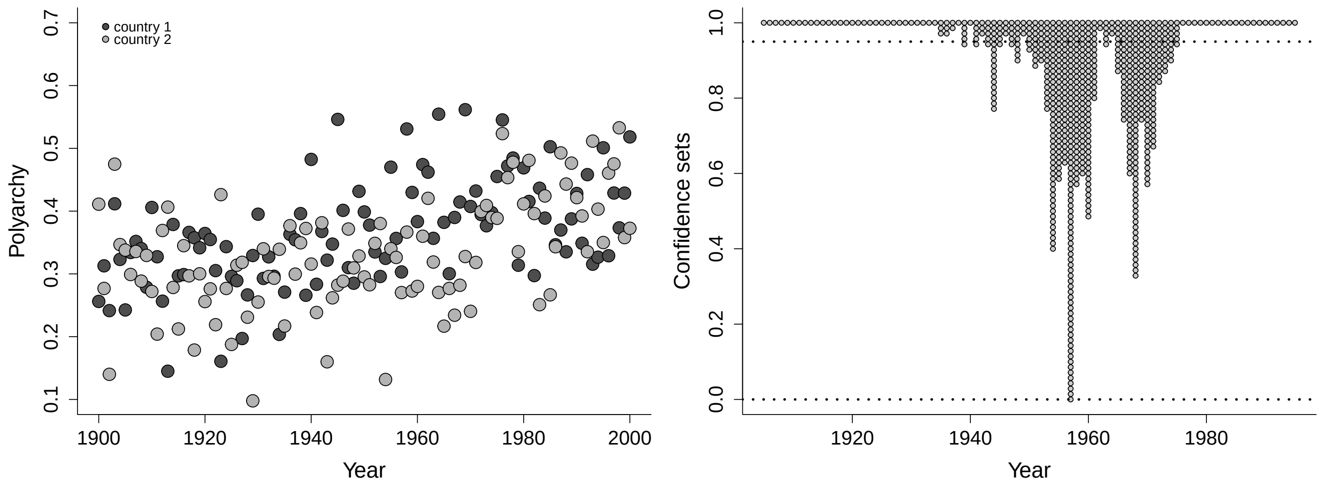
For many practical purposes, estimating the *size of the change* in a parameter – i.e., the difference $\mu = \beta_L - \beta_R$ in our case – is often equally, or perhaps more, interesting than locating τ. The method for constructing the confidence curve for the size of the change is based on a similar continuous parameter construction as that of the discrete parameter version in equation (2); see Cunen, Hermansen, and Hjort (2018) for details. The confidence curve for the degree of change, or any other parameter of interest, is a useful tool for gaining additional insight into the likelihood that a change point has occurred. This difference 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 -0.10, and the 95% confidence interval for $\mu = \beta_L - \beta_R$ extends from about -0.08 to about -0.11.

The methodology just described assumes that there is an underlying change point in the observed sequence. 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 it is reasonable to expect that there is, indeed, a change in the underlying model. If, in reality, there are no change points and the data generating process remains the same for the entire sample, this is typically reflected by very wide confidence sets, suggesting that there is high uncertainty as to where the assumed change point is located.

But, there are also other ways of assessing this question. One useful method is the so-called *monitoring bridge*. This is a visualisation 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 the model in (1), as specified for the hypothetical Polyarchy example. The plot indicates that there is something happening between 1944 and 1945. Around these years, the solid line approaches one of the two dashed lines (in this case the upper one), suggesting that we cannot safely assume that the data-generating process for Polyarchy is homogeneous across time.

The version of the methodology derived in Cunen, Hermansen, and Hjort (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). While this could be the case for some real-world processes of interest, for many other processes, changes could happen at different points in time, in different parts of the sample (e.g., in different countries). One key strength of the confidence curves approach, however, is the possibility for combining multiple independent such 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 the different world regions, further below). If so, both temporal dummies and split-sample designs will come up short, unless we have clear prior specification of relevant sub-groups of units. Confidence curves, in contrast, can deal with any ordering of groups or samples into coherent sub-groups, for example similar countries and/or regions, where we expect to only see one change point per sub-group. Adding to this, the general methodology for confidence distribution is well suited to subsequently combine several, independent confidence curves into one combined confidence curve.

Figure 3: Extending the simple model (1) on simulated data from 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 a best guess at 1957.



This point is illustrated in Figure 3, which is an extension of the model in (1) to a dataset consisting of two countries that experience the same change in mean Polyarchy (+0.10), but at different years (1934–1935 and 1969–1970). The simulated data points are represented by the scatter-plots in the leftmost panel, whereas the rightmost panel displays the confidence sets for τ the 'naive', combined model. The latter looks for one change point in the pooled data, where the two countries actually have change points that are 35 years apart. 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. If the analysis is done separately for the two countries, however, and then combined (as done for the rightmost panel of Appendix Figure A-1, estimating the magnitude of the change), we obtain much more reliable results than if the analysis is done simultaneously for the combined dataset (as done for the leftmost panel of Appendix Figure A-1).

However, different identifiable clusters of observations, such as countries or regions, experiencing different change points is just one way in which we, in a larger sample, could experience multiple change points. 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, Hermansen, and Hjort (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 in simulations when we assume that there are actually multiple change points (that are jointly observed by all units).

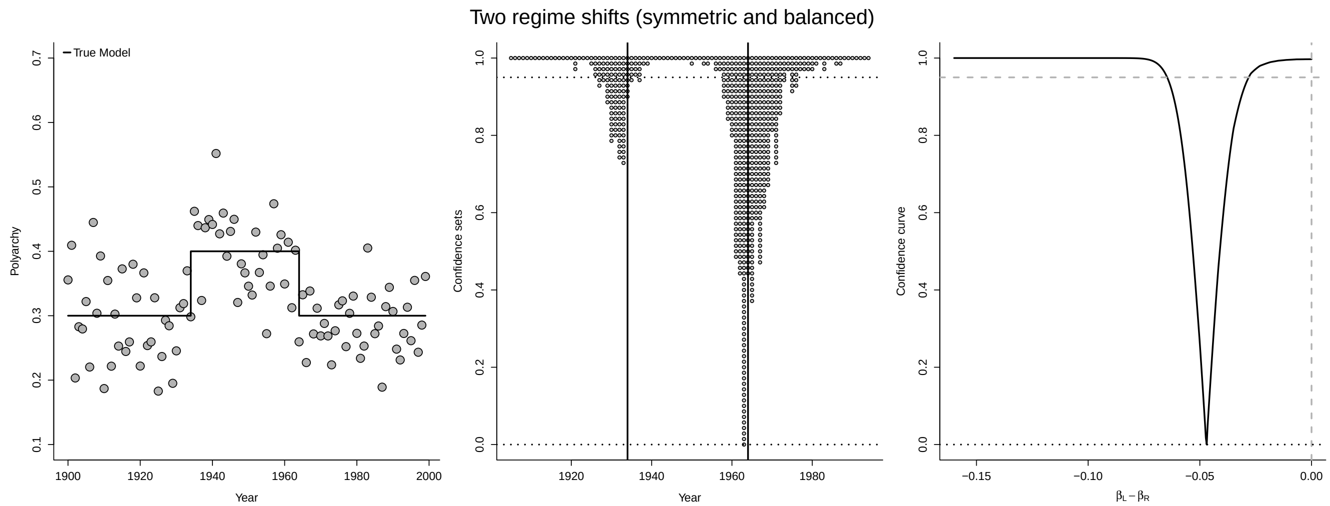


Figure 4: Data simulated with two similar change points at 1934 and 1964—change in mean from 0.3 to 0.4 and then back again to 0.3—under the same assumptions as in the above examples. We note that the method here focuses on the two real change points, indicating that there are two reasonable locations. We further note that the method does not do a good job at estimating the degree of change in this particular scenario.

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

The first scenario is depicted in Figure 4. In this scenario, 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 leftmost heatmap in Figure 7, displaying results from 100 simulations of this scenario. In other words, we tend to get indications of at least one change point, but with a

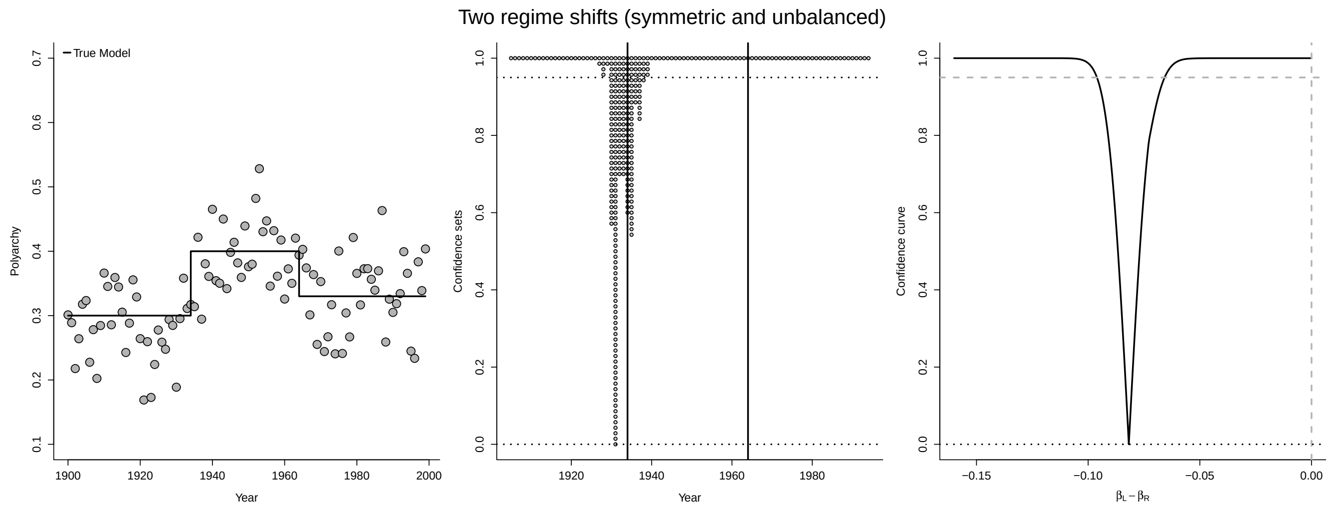


Figure 5: Data simulated with two change points; the change at 1934 is larger, of size 0.1 (from 0.3 to 0.4), than the change at 1964, which is of size 0.07 (from 0.4 to 0.33). Here, the method focuses on the largest change point.

fair amount of uncertainty, and sometimes the confidence sets concentrate about equally on both change points.

For the second case with two imbalanced change points, the middle plot of Figure 5 illustrates that our method tends to focus mainly on the largest of these change points, even when, in this case, the smaller parameter shift is about 70% the size of the larger one. In this scenario, 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 6 depicts the final scenario, assuming two identically sized change points with parameter shifts in the same direction. The rightmost plot of Figure 6 exemplifies that in this situation, the method tends to locate an estimated change point at, or close to, an actual change point (see also rightmost heatmap of Figure 7).

Finally, it is not always reasonable to assume that a change point manifests itself as an abrupt change. Instead, the change my be gradual over several years. When it comes to 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

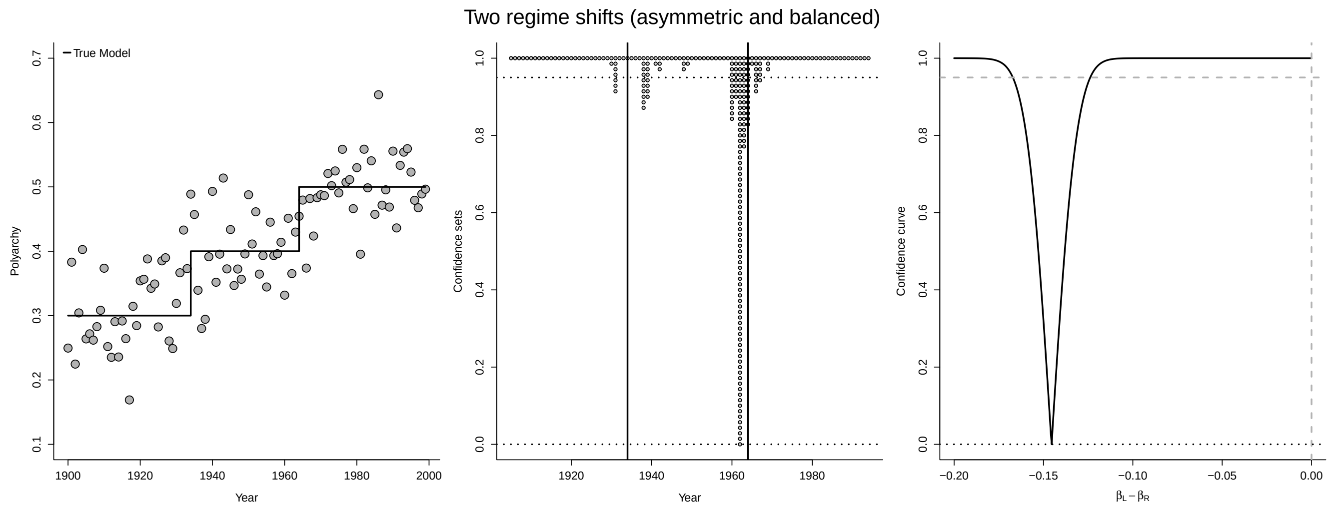


Figure 6: Data simulated with two change points moving in the same direction. For this case, the method points to the leftmost change point. When running a larger number of simulations, we find that the method tends to put the estimated change point at or between the two true change points.

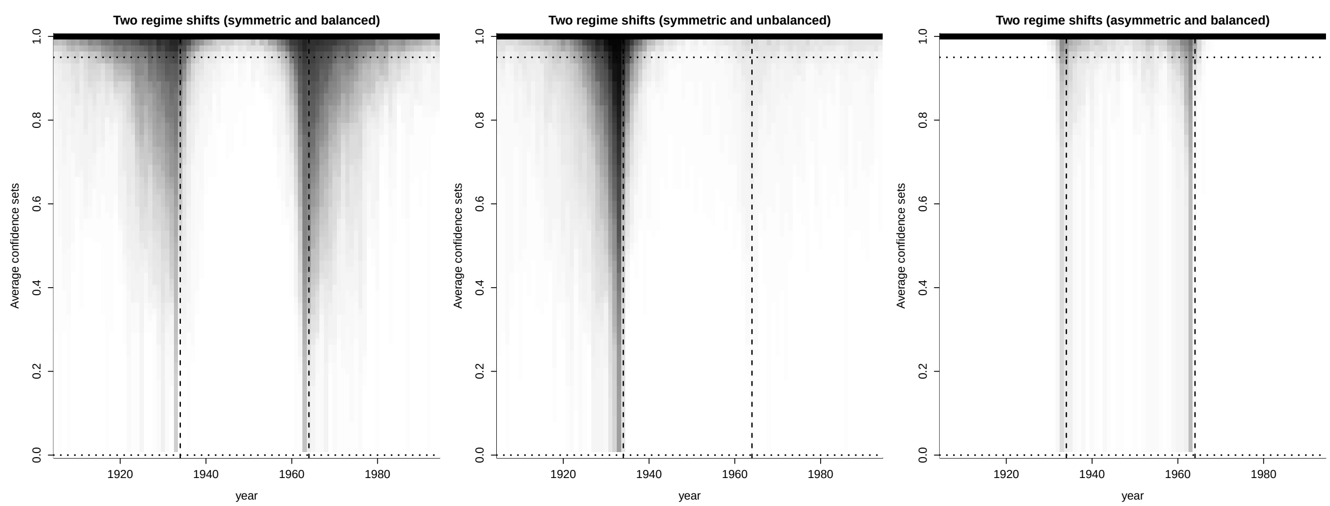


Figure 7: Heatmaps that aggregate and summarise the confidence sets from $N = 100$ simulated datasets for models with two change points; as shown in Figure 4, 5 and 6

of new ideas or technologies are likely gradual. Strictly speaking, our methodological framework is not constructed for gradual changes, nor are 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 in the parameter. More specifically, we simulate data where the change in the parameter value (from 0.3 to 0.4 on Polyarchy, as in prior examples) is a gradual change—here assumed to be a linear process—over 8 years, from 1946 to 1954. In Appendix Figure A-2, we make similar

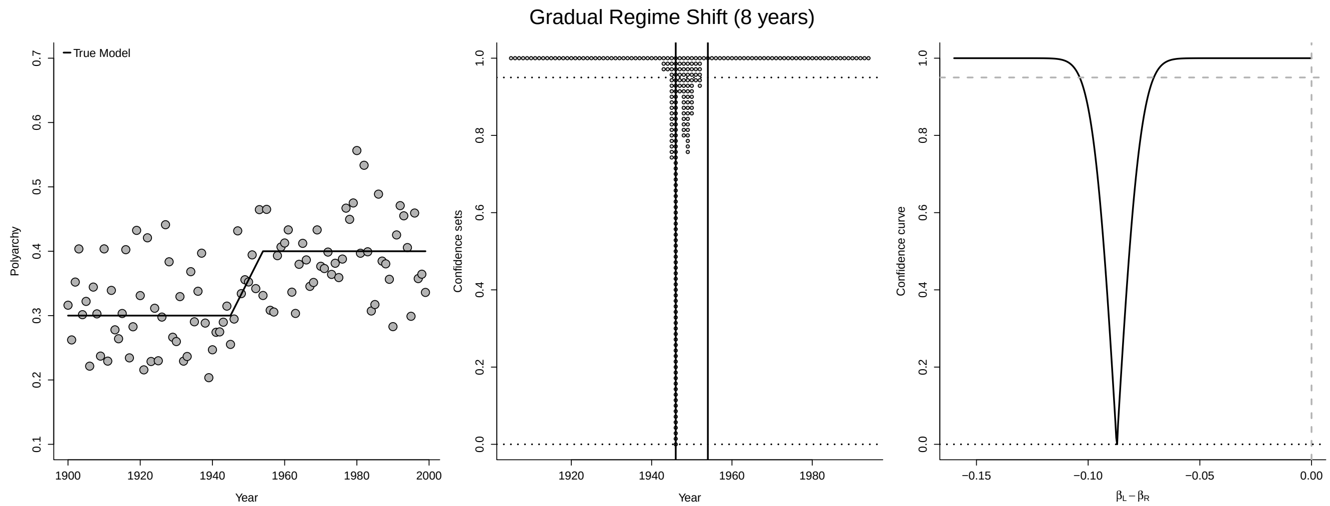


Figure 8: Data simulated with a gradual changing regime shift over 8 years (from 1946 to 1954). Compared to a baseline case of an abrupt change in one year, the confidence sets are somewhat wider.

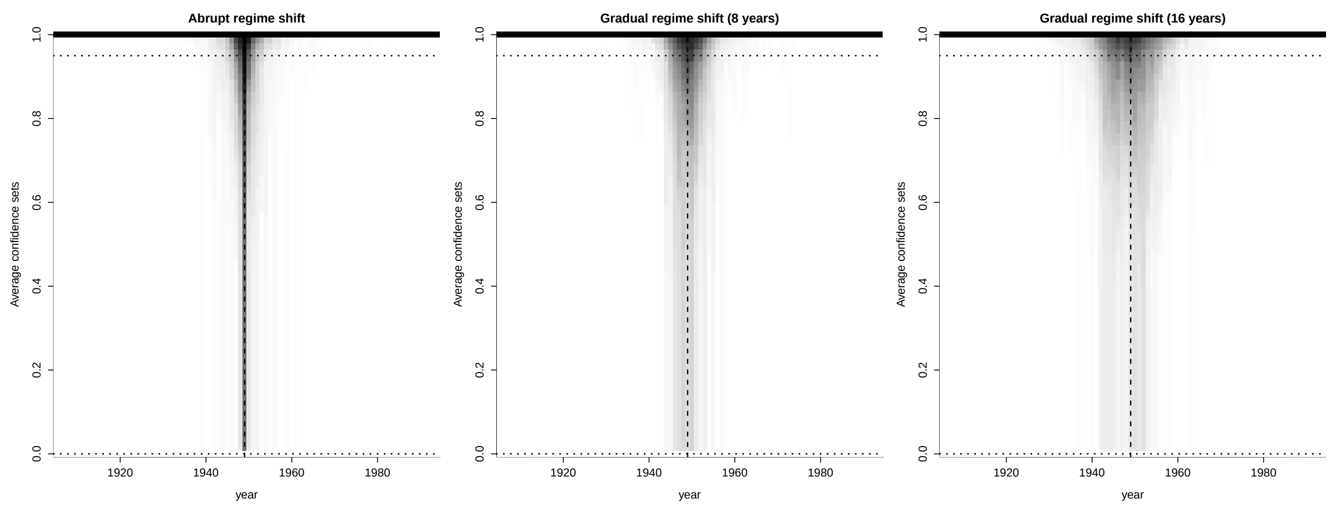


Figure 9: Heatmaps that aggregate and summarise the confidence sets from $N = 100$ simulated datasets, first with a normal abrupt regime shift (i.e. change point) and then for the two set-ups with a gradual change across, respectively, 8 and 16 year intervals (from Figure 8 and Appendix Figure A-2).

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). With the gradual changes, the confidence sets tend to be wider and cover more years than for the abrupt change. The confidence sets are (as expected) wider for the 16 year-scenario 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 not very long. This makes sense, considering that the methodology is designed to find the optimal change point for dividing data into a right and left model.

4 Application I: The changing role of labor-dependent agriculture for democratic survival

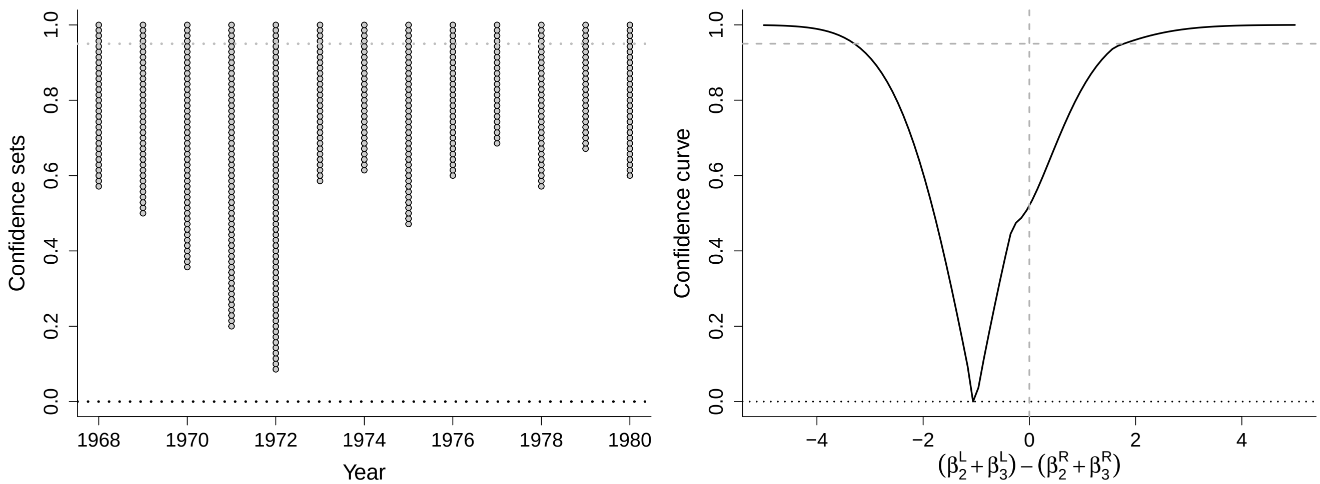
We now turn from a discussion of the methodology and use of simulated data towards applications on real-world data. The point of departure for our first application is the recent study by Albertus (2017). Albertus (2017) contributes to the large literature on economic determinants of democracy, focusing in particular on how specific production structures (with accompanying social class compositions) affect regime change. 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 (e.g., Lipset 1959; Ansell and Samuels 2014) and industrial workers (Rueschemeyer, Stephens, and Stephens 1992; Collier 1999). In contrast, rural economies – especially those with high inequalities in land ownership and that have labor-intensive forms of production – are widely presumed to be conducive to autocracy (see, e.g., Albertus 2017; Ansell and Samuels 2014). Landowners have assets that are hard to move, and thereby easy to tax. This gives them strong incentives to fight off democratization attempts in order to avoid progressive redistribution. Albertus points out that a negative association with democracy is mostly 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. He then discusses why the relationship should have become weaker in more recent decades. More specifically, Albertus highlights three mechanisms pertaining to increased financial globalization, observed expropriation of land and land-reforms in several autocracies, and increased prevalence of civil war in rural areas.

Albertus proceeds to test for a heterogeneous relationship between his measure of labor-dependent agriculture – constructed to capture the percentage share of families that work in agriculture but without owning the land they are working on – and democratization and democratic survival. He employs the dichotomous DD regime measure from Cheibub, Gandhi, and Vreeland (2010) and a dynamic probit specification. In brief, Albertus finds a non-robust link between labor-dependent agriculture and democratization, but a negative relationship with democratic survival. Yet, when splitting his post-WWII sample at the start-year of the 'Third Wave of Democratization' (1974), Albertus re-covers the robust relationship with democratic survival only in the pre-Third Wave sample. When combined with the higher positive coefficient on democratization in the Third Wave sample, this corroborates the notion that the relationship has changed, and that labor-dependent agriculture is no longer as 'bad for democracy' as it once was.

Yet, Albertus' discussion on the particular mechanisms contributing to this shift makes it very clear that 1974 should *not* unequivocally be expected to be a clear 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 – despite Huntington (1991) labelling it the start of the 'Third Wave' – 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, e.g., 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.

Thus, we employ the change point set-up described in the previous section, and follow Albertus (2017) in estimating a dynamic probit specification, where D is the dummy variable capturing democracy, L is labor-dependent agriculture, $\mathbf{X}$ is the above-listed vector of controls, j denotes

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



country, and t denotes year:

$$\begin{aligned}\Pr\{D_{j,t}=1 \mid L_{j,t-1}, D_{j,t-1}, X_{j,t-1}, \boldsymbol{\beta}, \boldsymbol{\beta}_X, \boldsymbol{\beta}_{XD}\} = \Phi(&\beta_0+\beta_1 L_{i,t-1}\\ &+\beta_2 D_{j,t-1}+\beta_3 L_{j,t-1}D_{j,t-1}\\ &+\boldsymbol{\beta}_X\mathbf{X}_{j,t-1}+\boldsymbol{\beta}_{XD}\mathbf{X}_{j,t-1}D_{j,t-1})\end{aligned} \tag{3}$$

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 (β_2) and this variable's interaction with the lagged regime measure (β_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. This test is shown in Figure 10.

The leftmost panel of Figure 10 shows confidence sets for the location of the break point τ. 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, e.g. in 1974, would have been represented by the confidence sets, the gray dots, centering on this year, and there would not be dots spread across the rest of the time series. Our method pinpoints 1972 as the most likely year for a change. However, by reading off the confidence sets for conventional levels of confidence – the 95% level is indicated by the grey dashed line – we cannot reject 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. But, the high level of uncertainty reflects that the method does not put much stock in a change happening in any of the particular years, and another plausible conclusion is thus that the method is pointing towards a finding 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, as displayed in the right panel of Figure 10. This plot shows that, 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 (the point where the confidence curve in Figure 10 touches the x-axis). Hence, our results do not warrant a clear conclusion on the relationship between labor-dependent agriculture and democratic survival having 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 draw the conclusion that the relationship between labor repressive agriculture and democracy has *not* changed at all. This is, however, a premature (and somewhat unsatisfying) 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) contains 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. This is true both at the start and end of the time series. Moreover, democratic onsets and breakdowns – as registered by dichotomous regime measures such as DD 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. In our next application, we rely on data material that alleviate these issues, and thus enable more precise estimation of change points. These data have longer time series and much less missing data. Additionally, we turn to estimating a continuous measure of democracy, which allows for more frequent changes in the dependent variable, and we estimate a more parsimonious model, which also contributes to increased degrees of freedom.

5 Application II: Income and democracy

The relationship between democracy and economic development is probably the most widely theorized and tested relationship in the democracy literature (see, e.g., Munck 2018). Notably, Lipset (1959), in his seminal study, proposed that higher income levels increase the chances of countries becoming and staying democratic. Contesting this proposed relationship, Przeworski and Limongi (1997) find a strong link between income and democratic survival, but not between income and democratization episodes. Yet, empirical studies extending the time series back into the 19th century tend to find a stronger positive relationship between income and democratization (Boix and Stokes 2003), and also a clearer relationship with democracy levels, even when accounting for country-fixed effects (Knutsen et al. 2019a). The latter observations may be suggestive of temporal heterogeneity, which Boix (2011a) theorizes and studies more carefully, e.g. by using split-sample analysis. Boix argues that the number and regime type of hegemonic actors, internationally, have varied across modern history *and* 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 analysis above on a dichotomous measure, and thus displaying the flexibility of the change point set-up. We include country-fixed effects to alleviate concerns that time-invariant country-specific factors will bias the income–

democracy relationship (see Acemoglu et al. 2008). Further, we restrict the addition of other covariates in order to mitigate issues of post-treatment bias (and maximize degrees of freedom). Finally, following Knutsen et al. (2019a), we use data from V-Dem. More specifically, we employ V-Dem's core electoral democracy measure, Polyarchy, which extends from 1789–2018 Teorell et al. (2019). Polyarchy is constructed to capture the democracy concept of Dahl (1971), and the theoretical range is 0–1 (0.01–0.95 in the data).

As discussed by Knutsen et al. (2019b), the graded nature of Polyarchy means that it captures regime changes not captured by binary democracy measures, such as the gradual liberalization of many countries prior to the 1848 revolutions. Further, the fact that it includes suffrage, in particular, means that it does not yield (artificially) high 19th century scores for countries such as the United States or Costa Rica (as does, for example, Polity2 from Marshall, Gurr, and Jaggers 2013), long prior to the enfranchisement of women or other large population groups, such as slaves in the Southern states in the US. Polyarchy may also be better attuned to capture nuances since many of the indicators draw on coding from several country-experts (and scores are subsequently adjusted to ensure comparability across space and time by the V-Dem IRT meaurement model; see Pemstein et al. 2018; Marquardt and Pemstein 2018). Hence, Polyarchy allows us to capture changes and trends also in the 'degree of democracy' among early democratizers, across modern history since the French Revolution.

The data on income, or more specifically Ln GDP per capita, are from Fariss et al. (2017), who run a dynamic latent trait model on several GDP and population datasets to provide GDP estimates. We use Fariss et al.'s estimates benchmarked in the extensive time series from the Maddison project. The latent model estimation mitigates various kinds of measurement error and extends the time series and mitigates missing values by imputation.

We run our change point set-up on the following OLS specification, where ϕ_i represent country-fixed effects, and $\theta_t = \beta_3 \text{year}_t + \beta_4 \text{year}_t^2 + \beta_5 \text{year}_t^3$ represent third order polynomial for time trends.[4] In the final analysis and inference the errors are clustered by country to account for panel-level autocorrelation:

[4] In the estimation we had to normalize year by subtracting $\text{year}_0 = 1789$ and scale by a constant so that the estimation should not break down.

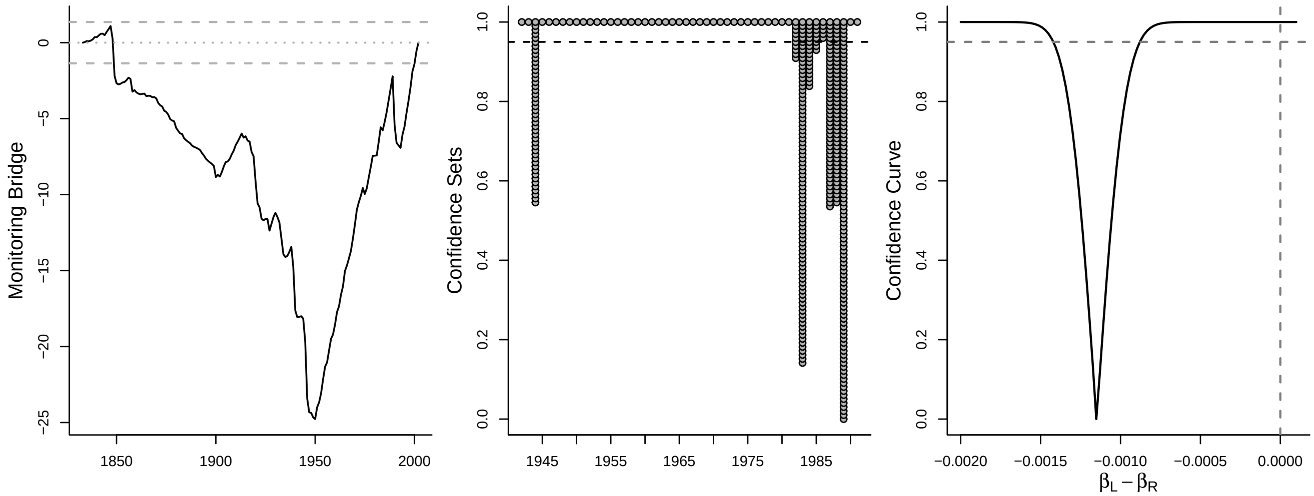
$$\text{Polyarchy}_{i,t+1} = \beta_0 + \beta_1 \text{GDPpc}_{i,t} + \beta_2 \text{Polyarchy}_{i,t} + \phi_i + \theta_t + \epsilon_{i,t} \tag{4}$$

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 entire 1789–2015 time span. We focus on the years 1833–2002, 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. The 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 behind Polyarchy.

We next investigate the more specific question of when the GDP per capita coefficient displays a likely change point in the middle panel of Figure 11. There are some indications of a change point occurring at the end of WWII. But, the strong clustering of grey dots – falling well below the dashed 95% confidence line – towards the end of the 1980s gives clear evidence for a change point in this decade. 1989 is marked as the maximum likelihood point-estimate for when the change point occurred, but the confidence sets also point to the years prior to (but not after) 1989 as candidate years for a structural change in the income–democracy relationship. As our simulation exercise in Section 3 suggests, a pattern of several adjacent years highlighted as possible change points could reflect that the method is capturing a gradual change in the relationship which unfolds over several years. But, as we will discuss in the next section, the pattern might also reflect that change points occur in different years for different parts of the sample, i.e., for different world regions.

We further investigate the change in the magnitude of the regression coefficient on income – interpreted as the predicted change from t to $t+1$ on the 0–1 Polyarchy Index when Ln GDP per capita increases by one unit in t – in the right panel of Figure 11. The best estimate of $\beta^L_{\text{GDP}} - \beta^R_{\text{GDP}}$ is around -0.001. In other words, the estimated relationship between income and democracy has become larger over time ($\beta^L_{\text{GDP}} - \beta^R_{\text{GDP}} < 0 \Longrightarrow \beta^R_{\text{GDP}} > \beta^L_{\text{GDP}}$). But, the estimate is also indicative of a very small change, albeit a statistically significant one; the 95% confidence interval for $\mu = \beta^L_{\text{GDP}} - \beta^R_{\text{GDP}}$ 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

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).



in time and vary in size, across different regions. We elaborate on this more complex scenario in the next section.

5.1 Regionally specific temporal heterogeneity

Extant work on democratization waves has highlighted that the frequency of democratization episodes *and* the (perceived) drivers of regime change have differed substantially across regions of the world (see, e.g., Haerpfer et al. 2019). The assumption that every region should experience the same shift in the income–democracy relationship, at the exact same point in time, is thus a strong one. We will now relax it by allowing change points to differ across world regions.

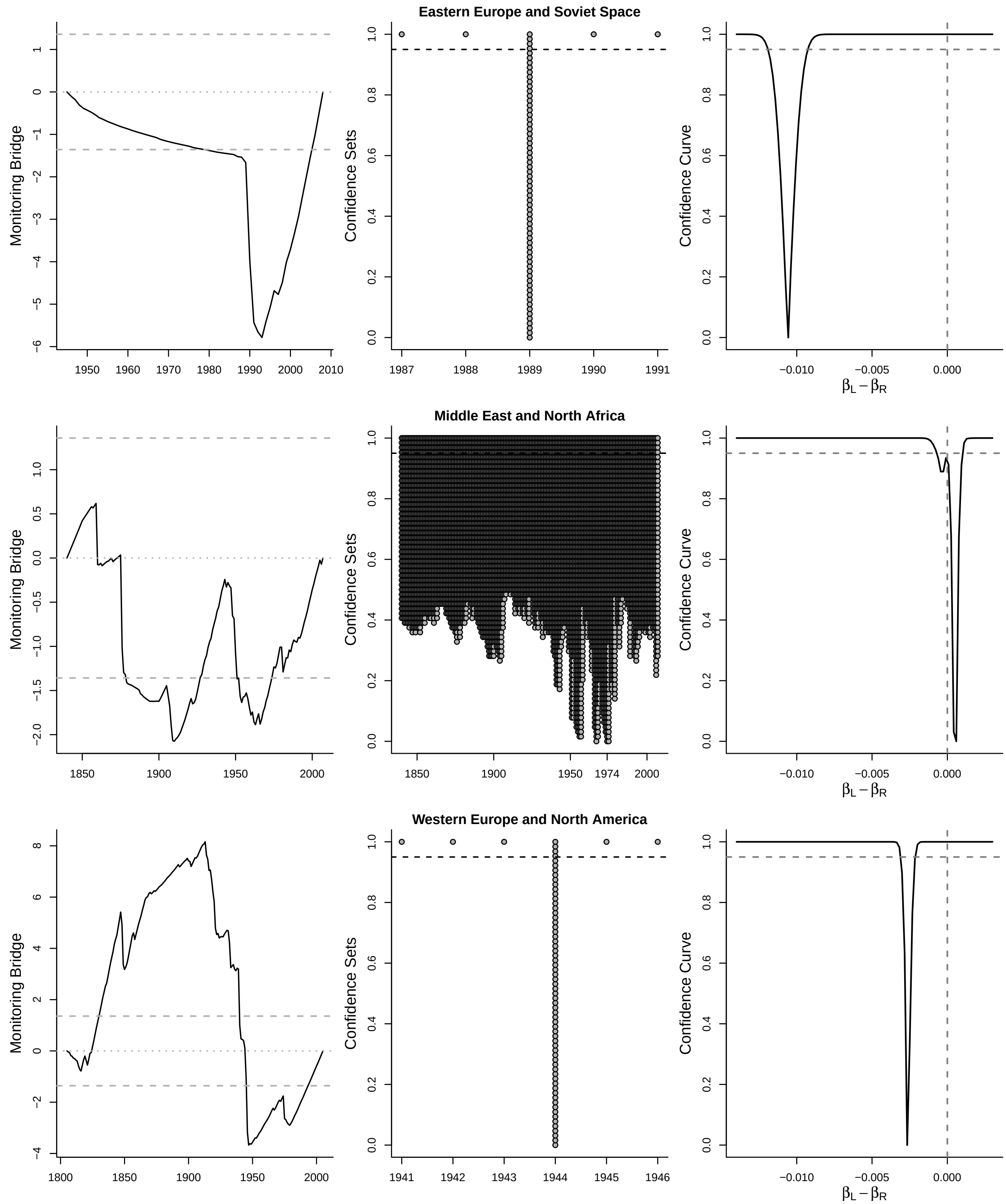
We use the eight-fold regional classification by Miller (2015) to divide up the world, and re-run the OLS model on Polyarchy detailed above on each region. The regions are Eastern Europe and the (post-)Soviet space (1), Latin America (2), Middle East and North Africa (3), Sub-Saharan Africa (4), Western Europe, North America, Australia and New Zealand (5), East Asia (6), South-East Asia (7), and South Asia (8). In the interest of space, we focus on three of these regions, with results for remaining regions being plotted in the Online Appendix. Hence, 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 displayed in the leftmost plots of Figure 12 provide quite substantial evidence that structural changes in the "data-generating process" behind democracy occur, at different points in time, for each of the regions; 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 distinct estimated change point years in different regions (middle plots Figure 12). Some of the estimated change points, we surmise, have considerable face validity and are easy to tie to prominent political processes; note that we have selected the range of years where there is confidence above zero for a potential change.

To be more specific, for Eastern Europe and the Soviet space (top row), we find a change point in 1989, which is the year the Berlin Wall came down. Indeed, 1989 is not only the maximum likelihood point-estimate, it is also the only year in which the method places any confidence as 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 Middle East and North Africa (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 democracy–development relationship has clearly changed in some world regions, these results suggest that such a change may not have occurred in MENA.

Finally, the confidence curves reported in the rightmost columns of Figure 12 indicate the change in the coefficient on income – i.e., the size of $\mu = \beta^{L}_{\mathrm{GDP}} - \beta^{R}_{\mathrm{GDP}}$ – for the different regions. For Eastern Europe and 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 (-.01) is much larger than what we estimated for the global analysis (-.001). Also for Western Europe and North America, the estimated change (-.005) is larger than at the global level, though less pronounced than in Eastern Europe. For

Figure 12: Regressions on Polyarchy, 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). As above, the rows show, the monitoring bridge (left plot), the confidence sets (middle plot, where we have chosen years where there was something to see) and the confidence curves (right plot)



MENA, in contrast, the maximum likelihood estimate for the change 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. Focusing on Eastern Europe, 1989 marks the end of the Cold War and the end of the influence of the Soviet Union. One interpretation, along the lines discussed in Boix and Stokes (2003), is that Soviet influence, and the larger dynamic of the US vs USSR competition, washed out any effect of income on the level of democracy in this region, 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 simply – both rich and poor countries are allowed to democratize without external intervention, but rich countries are more susceptible to do so.

6 Concluding discussion

We have introduced and discussed a novel approach, building on the framework developed in Cunen, Hermansen, and Hjort (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 the most 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. Further, 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 in this paper 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, and it can be used to study a change in a wide array of different parameters. We have illustrated the approach by employing it to panel data, and using OLS or probit models.

Second, the approach is also flexible in the sense that we can look for and infer about changes in different parts of the statistical model, both concerning particular parameters but also whether the overall data-generating process has changed.

Third, we have discussed how 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 in some cases even if these conditions are only approximately true. These include situations when changes occur gradually over a (limited) time interval rather than at one point in time, 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.

Fourth, and perhaps most importantly, the use of confidence distributions theory and in particular the confidence curves allows us to give a more comprehensive assessment of the uncertainty pertaining to our inferences about change points, including their temporal location *and* the size of the change. This is critical for political scientists and others, as many extant approaches to detecting changes over time could lead to over-confident conclusions about the timing and nature of structural breaks in relationships of interest.

In sum, we have showed how a flexible change point framework can be employed to assess key questions on the time-varying determinants of democracy, leading us to question some established findings and providing more nuance to others. Alongside this article, we also provide an R-package that will allow political scientists and others who are interested in studying temporal heterogeneity to conduct the same type of assessments and tests on various other relationships.

References

Acemoglu, Daron, Simon Johnson, James A. Robinson, and Pierre Yared (2008). "Income and Democracy". *American Economic Review* 98.3, pp. 808–842.

Albertus, Michael (2017). "Landowners and Democracy: The Social Origins of Democracy Reconsidered". *World Politics* 69.2, pp. 233–276.

Ansell, Ben W and David J Samuels (2014). *Inequality and Democratization: An Elite-competition Approach*. Cambridge: Cambridge University Press.

Beck, Nathaniel (1983). "Time–varying parameter regression models". *American Journal of Political Science* 27.3, pp. 557–600.

Blackwell, Mathew (2018). "Game Changers: Detecting Shifts in Overdispersed Count Data". *Political Analysis* 26.2, pp. 230–239.

Boix, Carles (2011a). "Democracy, Development, and the International System". *American Political Science Review* 105.4, pp. 809–828.

— (2011b). "Democracy, development, and the international system". *American Political Science Review* 105.4, pp. 809–828.

Boix, Carles and Susan C. Stokes (2003). "Endogenous Democratization". *World Politics* 55.4, pp. 517–549.

Cheibub, Jose, Jennifer Gandhi, and James Vreeland (2010). "Democracy and dictatorship revisited". *Public Choice* 143.1–2 (1), pp. 67–101.

Collier, Ruth Berins (1999). *Paths Towards Democracy: The Working Class and Elites in Western Europe and South America*. Cambridge: Cambridge University Press.

Coppedge, Michael, John Gerring, Staffan I Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, M Steven Fish, Adam Glynn, Allen Hicken, Carl Henrik Knutsen, Joshua Krusell, Anna Lührmann, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Moa Olin, Pamela Paxton, Daniel Pemstein, Josefine Pernes, Constanza Sanhueza Petrarca, Johannes von R¨mer, Laura Saxer, Brigitte Seim, Rachel Sigman, Jeffrey Staton, Natalia Stepanova, and Steven Wilson (2017). *V-Dem Country-Year Dataset v7.1.*

Coppedge, Michael, John Gerring, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, Steven Fish, Adam Glynn, Allen Hicken, Carl Henrik Knutsen, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Pamela Paxton, Daniel Pemstein, Laura Saxer, Brigitte Seim, Rachel Sigman, and Jeffreyn Staton (2017b). *Varieties of Democracy (V-Dem) Codebook v7.* Varieties of Democracy (V-Dem) Project.

Cunen, Céline, Gudmund Hermansen, and Nils Lid Hjort (2018). "Confidence distributions for change-points and regime shifts". *Journal of Statistical Planning and Inference.*

Dahl, Robert A. (1971). *Polyarchy: Political Participation and Opposition.* New Haven: Yale University Press.

Fariss, Christopher J., Charles D. Crabtree, Therese Anders, Zachary M. Jones, Fridolin J. Linder, and Jonathan N. Markowitz (2017). "Latent Estimation of GDP, GDP per capita, and Population from Historic and Contemporary Sources". Working paper.

Fox, Emily B, Erik B Sudderth, Michael I Jordan, Alan S Willsky, et al. (2011). "A sticky HDP-HMM with application to speaker diarization". *The Annals of Applied Statistics* 5.2A, pp. 1020–1056.

Fukuyama, Francis (1992). *The End of History and the Last Man.* New York: Free Press.

Haerpfer, Christian, Patrick Bernhagen, Christian Welzel, and Ronald F. Inglehart, eds. (2019). *Democratization, 2nd edition.* Oxford: Oxford University Press.

Hegre, Håvard, Joakim Karlsen, Håvard Mokleiv Nygård, Håvard Strand, and Henrik Urdal (2013). "Predicting Armed Conflict 2010–2050". *International Studies Quarterly* 55.2, pp. 250–270.

Huntington, Samuel P. (1991). *The Third Wave: Democratization in the Late Twentieth Century.* Norman, OK: University of Oklahoma Press.

Knutsen, Carl Henrik, John Gerring, Svend-Erik Skaaning, Jan Teorell, Matthew Maguire, Michael Coppedge, and Staffan I. Lindberg (2019a). "Economic Development and Democracy: An Electoral Connection". *European Journal of Political Research* 58.1, pp. 292–314.

Knutsen, Carl Henrik, Jan Teorell, Tore Wig, Agnes Cornell, John Gerring, Haakon Gjerløw, Svend-Erik Skaaning, Daniel Ziblatt, Kyle Marquardt, Dan Pemstein, and Brigitte Seim (2019b). "Introducing the Historical Varieties of Democracy Dataset: Patterns and Determinants of Democratization in the Long 19th Century". *Journal of Peace Research* 56.3, pp. 440–451.

Lipset, Seymour Martin (1959). "Some Social Requisites of Democracy: Economic Development and Political Legitimacy". *American Political Science Review* 53.1, pp. 69–105.

Marquardt, Kyle and Daniel Pemstein (2018). "IRT Models for Expert-Coded Panel Data". *Political Analysis* 26.4, pp. 431–456.

Marshall, Monty G., Ted Robert Gurr, and Keith Jaggers (2013). *POLITY IV PROJECT: Political Regime Characteristics and Transitions, 1800–2013*. Center for Systemic Peace.

Miller, Michael K (2015). "Democratic Pieces: Autocratic Elections and Democratic Development Since 1815". *British Journal of Political Science* 45.3, pp. 501–530.

Mitchell, Sara McLaughlin, Scott Gates, and Håvard Hegre (1999). "Evolution in Democracy-War Dynamics". *Journal of Conflict Resolution* 43.6, pp. 771–792.

Munck, Gerardo L. (2018). *The Quest for Knowledge About Societies: How Advances in the Social Sciences Have Been Made*. Book manuscript, University of Southern California.

Park, Jong Hee (2012). "A Unified Method for Dynamic and Cross-Sectional Heterogeneity: Introducing Hidden Markov Panel Models". *American Journal of Political Science* 56.4, pp. 1040–1054.

Pemstein, Dan, Kyle L Marquardt, Eitan Tzelgov, Yi-ting Wang, Joshua Krussel, and Farhad. 2018. Miri (2018). *The V-Dem measurement model: Latent variable analysis for cross-national and cross-temporal expert-coded data*. Gothenburg: V-Dem Working Paper 21.

Pierson, Paul (2011). *Politics In Time: History, institutions and social analysis*. Princeton: Princeton University Press.

Przeworski, Adam and Fernando Limongi (1997). "Modernization: Theory and Facts". *World Politics* 49.2, pp. 155–183.

Rueschemeyer, Dietrich, Evelyne Huber Stephens, and John D. Stephens (1992). *Capitalist Development and Democracy*. Chicago: University of Chicago Press.

Schweder, Tore and Nils Lid Hjort (2016). *Confidence, likelihood, probability*. Cambridge: Cambridge University Press.

Spirling, Arthur (2007). "Bayesian approaches for limited dependent variable change point problems". *Political Analysis* 15.4, pp. 387–405.

Teorell, Jan (2010). *Determinants of Democracy: Explaining Regime Change in the World, 1972–2006*. Cambridge: Cambridge University Press.

Teorell, Jan, Michael Coppedge, Staffan I. Lindberg, and Svend-Erik Skaaning (2019). "Measuring Polyarchy Across the Globe, 1900–2017". *Studies in Comparative International Development* 54.1, pp. 71–95.

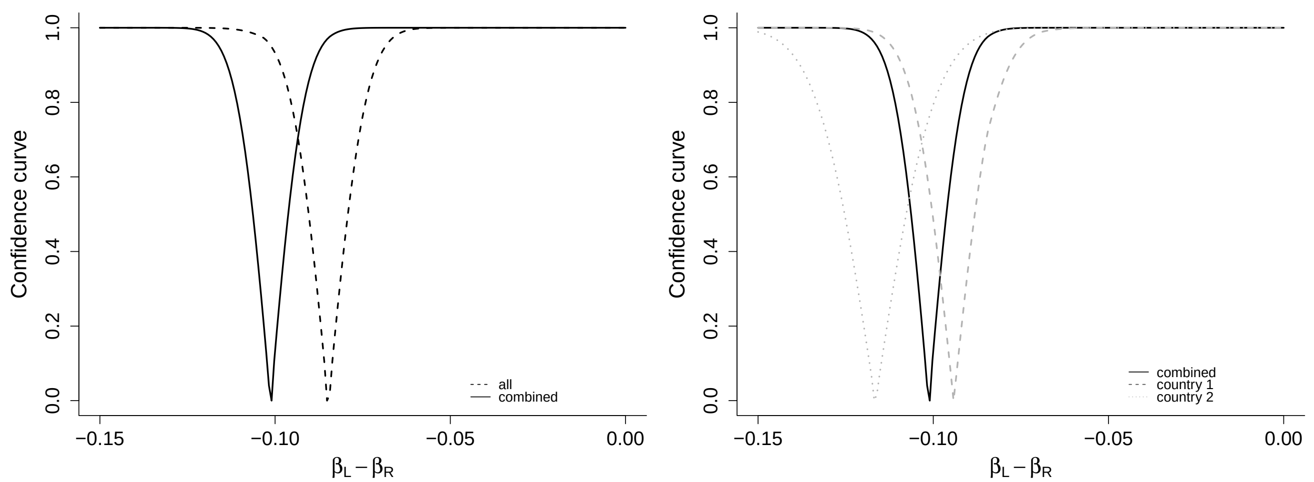
Tilly, Charles (1995). "To explain political processes". *American Journal of Sociology* 100.6, pp. 1594–1610.

Western, Bruce and Meredith Kleykamp (2004). "A Bayesian Change Point Model for Historical Time Series Analysis". *Political Analysis* 12.4, pp. 354–374.

A-1 Online Appendix

This section contains additional results that are discussed in the paper, but not reported in figures as well as results for regions not discussed above. We first present such additional results for different simulation exercises in Figures A-1 (separate change points, two different countries/parts of sample) and A-2 (gradual change over 16-year interval). Finally, we present additional results from Application II for the five regions not included in the discussion in the paper.

Figure A-1: Estimates of $\beta_L - \beta_R$. The leftmost panel displays a continuation of the 'naive' model in Figure 3, (mistakenly) assuming a joint change point for the two simulated countries as well as results for when the sample is first split into the two countries, the model is estimated, and the results are then combined. The rightmost panel takes the latter model as point of departure and shows the individual confidence curves for the two countries and the combined curve.



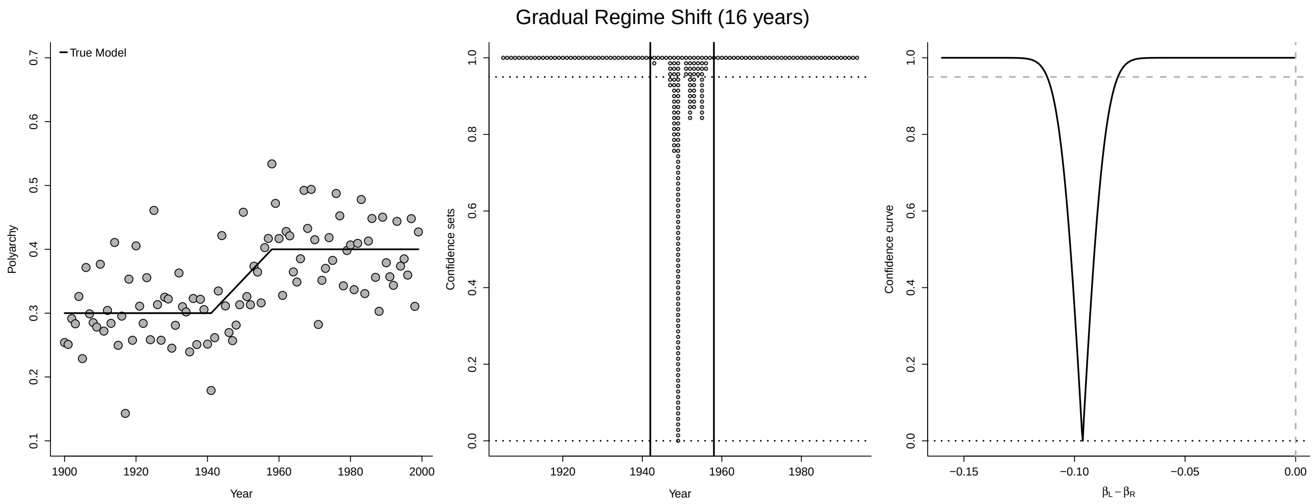


Figure A-2: Data simulated with a gradual and linear change over 16 years (from 1942 to 1958). Compared to the simulation with an 8-year change in Figure 8, the confidence sets are somewhat wider, which can be seen more clearly in the simulation study in Figure 9.

Figure A-3: Regressions on Polyarchy, by region, for regions not included in main paper. 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 regression coefficient; right plot).

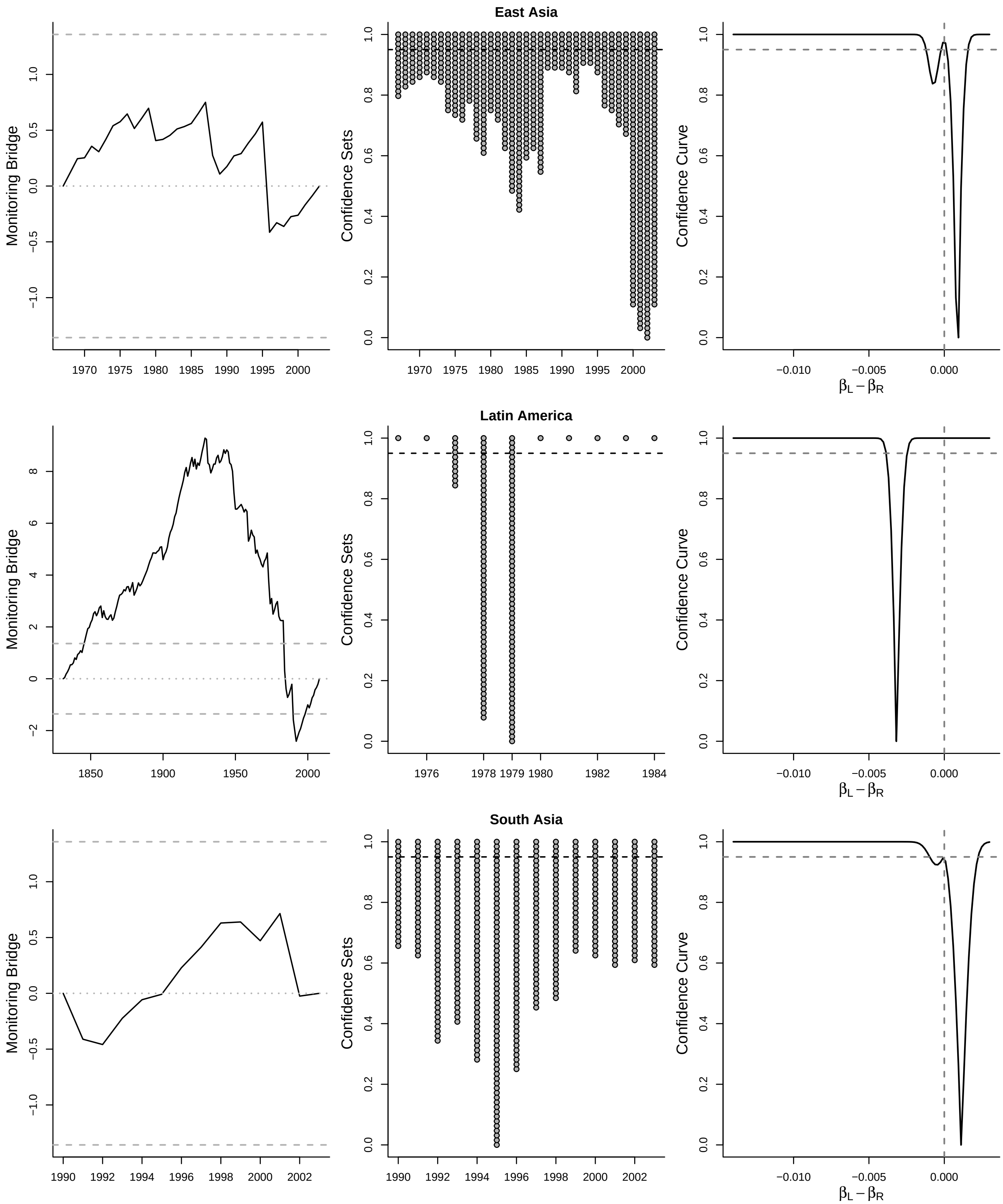


Figure A-4: Regressions on Polyarchy, by region, for regions not included in main paper. Does the model change over time? (Monitoring bridge, left plot). When does the relationship between GDP per capita and Polyarchy change? (Confidence sets, middle plot). What is the estimated change in the relationship? (Confidence curves for change regression coefficient; right plot).