UiT The Arctic University of Norway

School of Business and Economics

The Efficiency of Coercion

The modality and efficiency of coercive measures

Chris Rune Andersen

Master's thesis in Economics, SOK3901, June 2022

Foreword

This masters thesis marks the end of my time as a student at the school of business and economics at the University of Tromsø – The Arctic University of Norway. It is somewhat bittersweet, as I credit most of who I am today to this place and this program, both the bachelor and master program, and to see this chapter of my life close feels somewhat surreal.

Firstly, I would like to express my immense gratitude to my mentor, Øivind Devik Schøyen. His engagement and drive for the subject is hard to overstate, and the process of writing this thesis would have been severely hampered if not for his willingness to guide me, but also his insistence that I needed to understand certain things myself. I believe that genuinely looking forward to supervisory "classes" is a rare thing, but somehow the odds was beaten here. The possibility of passing ideas back and forth was both fun, engaging, and highly instructive, and resulted in this task not being nearly as exhausting as I anticipated.

I would also like to thank the rest of the faculty at HHT, for their seemingly unending willingness to help when it was needed and for providing a climate where the prospect of asking professors for help did not give me heart palpitations.

Secondly, I would like to thank my parents for their support. Both for their supportive attitude, but also their willingness to not let me starve while I hunched over books and computer screens for the better part of a decade. Them pushing me to go for a higher education is the best version of paternalism I have experienced.

Last, but certainly not least, I would like to extend large amounts of gratitude to my classmates. Both the "veterans" that came over from the bachelor program with me, and the "newcomers" that I met at the start of the Master program, it has been a genuine pleasure to meet and befriend them all. From hanging around the office discussing benign and silly ideas, to collectively smashing our heads into the desk when faced with what seemed like impossible tasks, there is little doubt in my mind that most of my good memories from these years are because of them. Thank you

Abstract:

Authoritarian regimes around the world employ coercive measures to attain legitimacy. It would then follow that more coercion equals more legitimacy, which in turn means more utility for a given regime. This is however not always the case, as not all authoritarian regimes, even relatively similar regimes, use the same levels of coercion. This paper supposes that this relative difference can be explained by an intrinsic response to coercion, called coercion resistance, which creates certain intervals of coercion that are not efficient in attaining legitimacy. We used kernel density estimation to ascertain that the coercion levels, defined by data on human rights abuses, restrictions on civil liberties, etc., are bimodal. The KDE was tested for bi- and multimodality using the Silverman Test and Dip test, which strengthened the assertion that coercion usage had a bimodal distribution, although not for all time periods. The total number of countries employing higher levels of coercion saw a heavy decline as the data approached the present, as the distribution had become unimodal at the datasets end in 2014, with the single mode located close to zero coercion usage. The attempt to identify specific intervals which were efficient or inefficient resulted in ambiguous answers, but the theory of their existence was not rejected. The movement and characteristics of certain intervals appeared to imply efficiency/inefficiency, but their classifications as such could not be done conclusively.

Keywords:

Coercion, Kernel Density Estimation, Authoritarianism, Democracy, Coercion Resentment

Contents

Figure List

1 Introduction

What is an authoritarian state? The common understanding is that it is a government headed by a person/group with near unfettered power, with little to no limit on implementing policy and, more relevant to this paper, coercive measures. While this understanding is far from explicitly wrong, the concept of uninhibited power is in general not grounded in reality. The reason North Korea is often the first example of authoritarianism one thinks of is how extreme their coercive measures are (Byman & Lind, 2010), but it raises the question: why does not every authoritarian regime do the same? If a regime has the capacity to coerce its population into complete subservience, why would it choose not to? This paper seeks to ascertain if there are intervals of coercion levels that a legitimacy maximizing authority should (and do) avoid, a theory proposed in the paper "What Limits the Efficacy of Coercion?" (Schøyen, 2020).

---

2 Theoretical Overview

Why would authoritarian regimes self-impose restrictions on their use of coercion? One would assume that any rational authority would choose to make decisions that maximize their utility, and this paper will primarily look at the decision-making process when choosing a coercion level. Coercion is defined here as the act of imposing extrinsic incentives on certain activities to dissuade participation in said activity. Such activities include the general methods of exhibiting and sowing dissent, such as protests, civil disobedience, or just openly criticizing the regime, but also the very act of holding values and opinions not endorsed by the regime. Examples range from attending religious services as a religious minority, adhering to cultural norms not endemic to the culture of the government, or just the appearance of belonging to a group which the state does not approve of, to name a few. This paper will focus on activities that have a clear ideological fundament, like religion or political ideology, and not on any visible ethnic/genetic traits. Using coercion to ensure a higher proportion of the population internalizes the same values as the regime builds legitimacy, which is always positive for the regime. In the short term, the strategy for attaining legitimacy manifests itself as governing in accordance with the values of the majority. The long-term strategy for attaining legitimacy is however to shape the populations values to better suit the regime itself, where the use of coercion functions as the means to do so. As such, the use of coercion will increase the regimes utility.

However, not all regimes coerce equally, and some regimes does not even coerce equally within their own borders (Froese, 2008). This implies that there is something obfuscating the picture, that there is some opposing "force" that incentivizes a regime to not always use the maximum level of coercion. This effect, referred to as coercion-resentment (Schøyen, 2020), is an amalgamation of multiple psychological effects, and is an intrinsic response to coercion which makes some coercion levels ineffective for achieving their goals. The theory that outlines this effect is found in the paper "What Limits The Efficacy of Coercion?" (Schøyen, 2020), and this paper is at its most fundamental level a test of this theory. As such, the hypotheses of the paper will be:

- There are effective and ineffective intervals of coercion levels
- The distribution of coercion levels is multimodal

A preliminary conjecture is that one efficient interval is located near/at the lowest coercion level possible, while another is located somewhat close to, but not completely at, the highest coercion level possible.

2.1 Theoretical Model

As mentioned above, this paper seeks to test one of the theories proposed in Schøyens paper, whose paper forms the basis for the principles of exclusion discussed below. The part of the theory that is directly relevant to this paper states, in very broad terms, that:

If a regime:

1) wishes to maximize legitimacy
2) Is capable of implementing any coercion level

And:

3) The coercion resentment response is non-linear

Then it follows that there are inefficient intervals of coercion[1]. Proof provided in (Schøyen, 2020).

If the coercion resentment function is linear, it would always be in the regimes best interest to always employ the maximum level of coercion. This is of course not the case, as not *every* authoritarian state holds their populace at gun point at all times. An alternative is a convex or

[1] Schøyen shows the basic premise in an overlapping generations model (Bisin and Verdier, 2001). The micro-foundation of the theory outlined in the paper considers the choices of parents socializing their children into either the state identity or the non-state identity, denoted as *s* and *ns* respectively. Identity is defined as the values and ideologies internalized by an agent, and the state and non-state identities are assumed to be mutually exclusive. The proportion of people that has internalized the non-state identity is denoted as q_t, while the proportion of people that has internalized the state identity is denoted as $(1 - q_t)$. The state chooses some level of coercion π, which creates incentives to socialize children into the state identity. The state is incentivized to minimize q_t, as it increases their legitimacy. The use of coercion π also produces the aforementioned coercion resentment, denoted as $C(\pi)$. The "behavior" of this $C(\pi)$ is what underpins the part of the theory this paper is meant to test.

concave function, which would mean that there are unique optimal points. While certainly not impossible, a convex/concave function would poorly account for the variation in coercion levels seen not just between countries, but the variation observed within countries. Lastly, the (current) most probable suspect is a non-linear $C(\pi)$. The coercion resentment increases with coercion to a point, but after a certain level the intrinsic response is diminished by the increased cost of resisting the higher coercion level. This idea is what gives way to the hypothesis of a multimodal distribution of coercion, when assuming rational regimes and actors that adjust their coercion levels to account for this effect.

---

3 Data

The paper will use data found in the "Authoritarian Regimes Dataset", which was compiled by (Hadenius & Teorell, 2007) in their paper "Pathways from Authoritarianism". The dataset categorizes all countries (N=195) in the time period 1972-2014 into 19 different classifications, where some are mutually exclusive while others are not. As an example, democracy excludes military one-party classification, but Monarchy does not exclude a one-party regime. For the purposes of this paper, multiple countries will be excluded based of their classification, as they do not fall within the scope.

The dataset itself combines indexes from both Polity (Polity IV Project, 2014) and Freedom House to make its classifications. The Freedom House index awards a score on a 0 to 10 scale for a country's political rights and civil liberties, and the dataset uses an average of these scores. Freedom House is a non-profit human rights advocacy and research organization that was founded in 1941 with the stated goal of "working to defend human rights and promote democratic change". The Polity index categorizes a country from strongly autocratic to strongly democratic, on a scale from -10 to 10 scale, which the dataset has revised to a 0 to 10 scale as well The points are awarded on the basis of constraints on executive authority (or lack thereof), institutionalized qualities of the governing authority, and so forth.

Finally, the dataset has created an imputed average of both indices, which "outperforms all rival indices of democracy" (Hadenius & Teorell, 2005). For all intents and purposes, this will be the most likely candidate when choosing which countries to exclude and include. It functions on the same 0 to 10 scale, where any country scored above 7 is defined as democracy. Some countries have been treated as continuations of previous versions of themselves, such as Germany absorbing West Germany, Vietnam absorbing North Vietnam, and so forth. Some countries did however not receive this treatment due to large governmental shifts, such as The Russian Federation not succeeding the USSR, but rather being treated as a new country from its inception.

3.1 Exclusions

As mentioned, some countries in the dataset will need to be excluded, as their style of governance either serve no purpose for the papers analysis, or they have associated factors that obfuscate the behavior that we are trying to analyze. Firstly, the principles for exclusion is as follows:

- Only authoritarian governments will be considered
- Only authoritarian regimes that use coercion will be considered
- The coercion must be used for the purpose of maximizing legitimacy
    - The reasons for any usage of coercion are seldom explicit, which means that this principle may be vague. At best this might be rectified, at worst it will require using an assumption and an associated acknowledgement of the issue.
- The regimes in question need to be bureaucratically capable of imposing any coercion level
    - In effect, this means that a regime that restricts its use of coercion due to some barrier will be excluded. If a regime cannot impose its desired coercion level due to state capacity or legislative measures, any notion of self-imposed restriction is void, which is not useful to this analysis.
- Only coercion that has its origin in the ruling regime itself is within the scope
    - This principle is added to remove coercion that is result of mob rule or other similar anarchical movements.
- Only coercion that is targeted at an ideological fundament will be considered
    - As mentioned in the introduction, this paper will not consider coercion towards groups based on any biological or genetic characteristics. This is because such characteristics are not transmissible, and as such there is no additional legitimacy to be gained, which breaks with the fundamental premise,

The dataset uses the following classifications:

1 Limited Multiparty

2 Partyless

3 No-Party

4 Military

5 Military No-Party

6 Military Multiparty

7 Military One-party

8 One-Party

9 Other

16 One-Party Monarchy

17 Monarchy

18 Rebel Regime

19 Civil War

20 Occupation

21 Theocracy

22 Transitional Regime

23 No-Party Monarchy

24 Multiparty Monarchy

100 Democracy

3.1.1 Excluded categories

This section will discuss the removal of entire classifications, not necessarily any of the specific countries contained within a classification.

Firstly, any country classified as democracy will be removed, as they do not explicitly use coercion to increase legitimacy. Secondly, transitional regimes. The code book for the dataset defines transitional regimes as temporary regimes whose only purpose is to carry out the transition into another regime/government. For such a regime, there is little reason to assume

their state capacity exceeds the necessary threshold to efficiently coerce, as well as attaining legitimacy for a regime that will invariably cease to exist is probably counterproductive.

Furthermore, countries that have recently transitioned, particularly from democracy, will also be excluded, as their potential restriction on coercion usage have a much higher probability of being a result of lack of state capacity than the self-imposed restraint this paper wants to analyze. Lastly, three other classifications will be excluded: rebel regime, occupation, and civil war. “Occupation” and “civil war” refer to countries where the parts or the entire territory is not controlled by the official government. Rebel regimes are countries where rebel forces have ousted the previous regime through military means, but the regime has not yet been reconstituted as a new governing body. These categories will also be omitted due to potential lack of state capacity.

3.1.2 Methodological Issues of Measuring Coercion

Before venturing into the statistical methods that will be employed, there are certain issues that needs to be at the very least acknowledged. Firstly, and most importantly, it is the issue of the scoring system in the data. The answer to the question of “how authoritarian is country x?” is beholden to a myriad of biases, from the ideological stance of those who compile the data to the process of gathering it. This is mentioned not to discredit the institutions and organizations that did this incredibly important work, but to underscore that the type of data being used is not necessarily completely rigid under sufficient scrutiny.

Furthermore, the use of “level of authoritarianism” as a proxy for coercion is not necessarily representative on a 1:1 basis. A given military one-party regime could be less authoritarian than some autocratic theocracy based on the scores, but it does not necessarily mean their methods are less coercive. For example, a country that does not restrict religious rights, but is liberal with the use of capital punishment, will by the rules of the dataset be less authoritarian than a given counterpart, even though it is easy to construe the threat of the death penalty as coercion. This issue is compounded with the issue mentioned in the principles-section, that the “mission statement” of any coercive strategy is unlikely to be stated outright. In the end, all of this means the use of the specific data this paper will employ is not necessarily perfect.

However, for the specific hypotheses this paper will test, the data is in all probability the best available. Any attempts to directly measure coercion will invariably run into the same problems of obfuscated and covert motivations, while not providing observations on the same scale as the dataset at hand. More concrete measurements, such as the number of prisoners of conscience, severity of police brutality or the mortality rate of activists and journalists, do provide more measurable data, but it would still be incomplete. The dataset's comprehensiveness outpaces its relative imperfections, which in turn makes it uniquely suited to tackle this very important question. This quality of imperfect but still useful also translates to the paper itself. The question of why an authoritarian state chooses to restrain itself in the use of coercion will not be definitively answered here, but a very important question will be raised, and a theory which could answer it will be tested.

While not necessarily an assumption that is completely rigorous, henceforth this paper will implicitly assume that the combined Polity and Freedom House scores correlate with coercion usage, and will use the scores as a metric for coercion levels.

---

4 Statistical Method

To quickly reiterate, the main hypothesis of this paper states that there are efficient and inefficient intervals of coercion levels. This hypothesis implies that the data has a multimodal distribution, meaning there are two or more peaks, which the first part of the paper will be dedicated to testing. The null hypothesis is that the distribution is uniform, that the efficiency of coercion is the same regardless of level, as this is the maximum entropy probability distribution. The maximum entropy distribution is, in broad terms, the distribution that minimizes assumptions when no additional information is supplied, which is the uniform distribution when there is a finite time horizon. If the null hypothesis is rejected, another test will be employed to test for the exact number of modes, which will be explained in more detail below (Silverman's test subchapter).

Furthermore, this paper will check if the included countries move towards these peaks (modes) over a certain time-period. If a country finds themselves in an inefficient interval, the main supposition of this paper is that a rational regime will change their coercion levels to an efficient interval, meaning that there exist internal steady state equilibria. This could explain the question raised in the introduction, why North Korea is so different from many other authoritarian states. If the threshold for inefficiency exists between a given position and the maximum coercion level, it would disincentivize crossing this boundary.

The methods that will be used will attempt to discern the number of modes in the dataset. Seeing as the data is ordinal as result of the use of the Likert scale, the best solution is using a non-parametric statistical test, namely Kernel density estimation. Using the kernel density estimate for univariate observations $x = (x_1, x_2, \dots x_N)$, which is defined by:

$$\hat{f}(x) = (nh)^{-1} \sum_i K\left(\frac{x_i - x}{h}\right)$$

The *n* is, as in most cases, the number of observations, while the *h* is the bin width (also called smoothing parameter or bandwidth). The bin width is the frequency of observations. Graphically a bin is just the rectangle on a histogram, and the bin width is the size of this bin.

Mathematically, the bin width is the equally sized sub-intervals that the complete data interval is made up of. A smaller bin width will generally allow for a more accurate representation of the data, as there are more counts of the relevant variable. As an example, splitting up observations of some often-occurring phenomena, like millimeters of rainfall, from a per-hour basis to a per-minute basis, will yield a more accurate representation as it does not need to be averaged over such a large group. A general graphical outlining the difference looks like this:

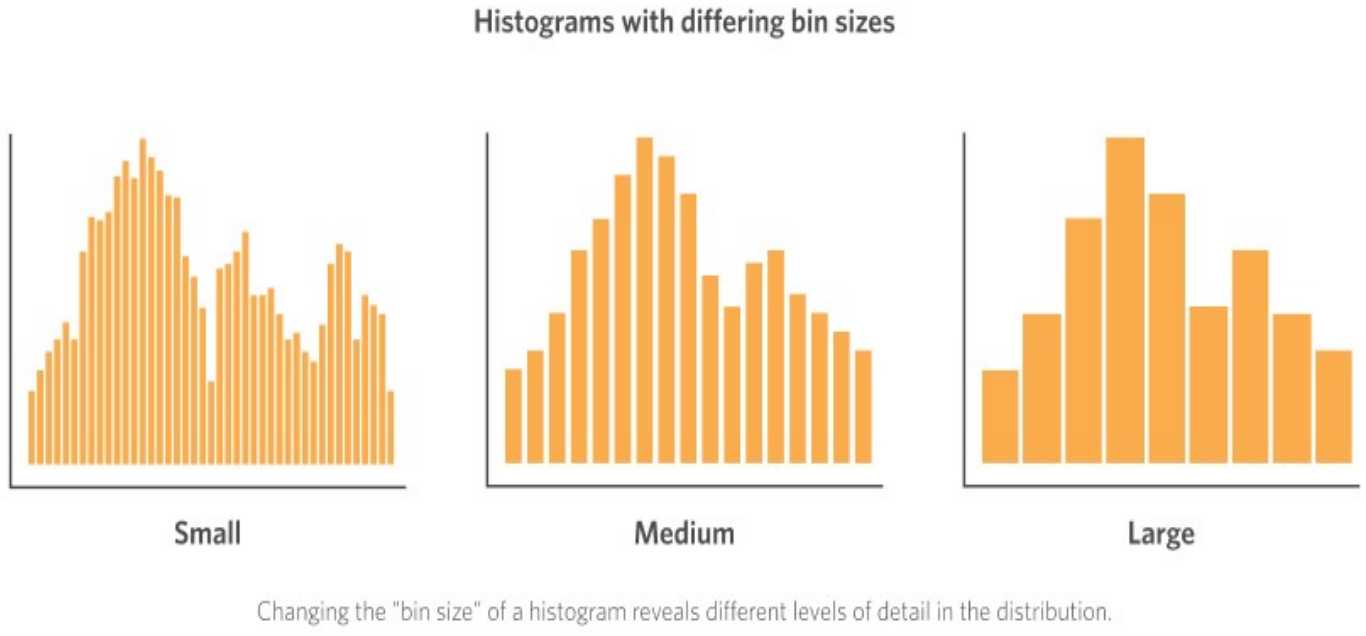


Illustration 1: "Histograms with different bin sizes", from "How to visualize your data: distribution charts" by Tom McKenzie, 2021 https://culturecounts.cc/blog/how-to-visualise-your-data-a-guide-to-distribution-charts

The reason this concept is so important as to dedicate an entire paragraph to it is that the bin width is a core variable of the next part of the estimation function; the kernel $K(\frac{x_i - x}{h})$. Here, the previously discussed variable h functions as a smoothing parameter for the future line. The variable x sans subscript represents any possible value of the data. What this all really means is that a lower h better captures multimodality *to a point*. An h that is too small will just produce a jagged line that completely mimics the bars to the point where its explanatory capability of trends is heavily diminished. On the other hand, an h that is too large will produce a line that does not capture the actual densities of the data. A general example of different bin widths is shown in the figure below:

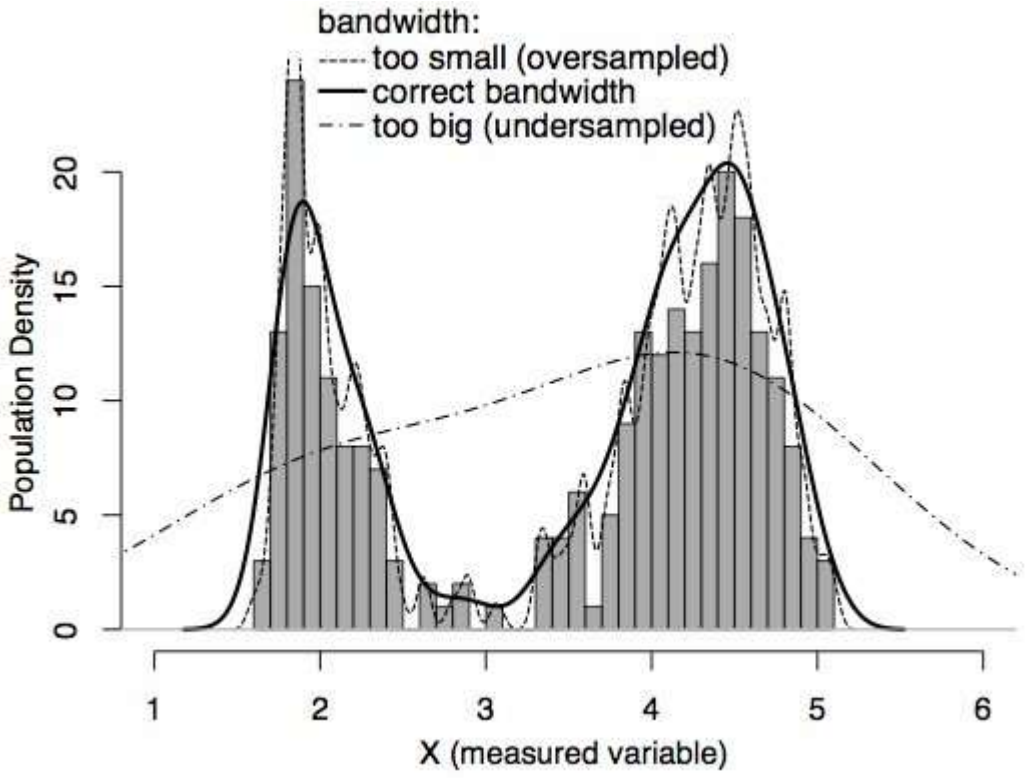


Illustration 2: "Figure 4", from "The Distribution of Olivine Compositions in Icelandic Basalts and Picrites", by Thomson, Andrew R & Maclennan, J, 2013, Journal of Petrology

4.1 Method 1: Silverman

The first method of testing for multimodality is based on the paper "Using Kernel Density Estimates to Investigate Multimodality" (Silverman, 1981). In simple terms, it tests the null hypothesis of a density *f* having *k* modes against the alternative hypothesis that the density *f* has *k+1* modes. The test statistic is the critical window width, which is defined by:

$$h_{crit}(k) = \inf [h | \hat{f} \text{ has at most } k \text{ modes}]$$

As mentioned in the previous section, *h* is the parameter that smooths out the data to obtain a kernel density estimate. If $h < h_{crit}(k)$ the density has at least *k+1* modes, meaning the distribution is multimodal. The critical bandwidth is the value of *h* where the distribution changes from unimodal to multimodal or vice versa. More formally we can say that for every

integer k, where $1 < k < n$, there exists a minimum width h_k such that the kernel density estimate has maximum k maxima (Silverman, 1981). If the null hypothesis is correct, and the distribution is unimodal, h_{crit} will be the size $n^{-\frac{1}{5}}$ (Hall & York, 2001). If the alternative hypothesis is correct, then h_{crit} may not even converge to zero with n, which means very large values of h_{crit} results in rejecting the null hypothesis.

4.2 Method 2: Hartigans Dip Test

This test discerns the maximal difference between the kernel distribution function and the unimodal distribution function that minimizes said maximum difference. Said differently, the dip is the distance between the tightest fitting unimodal distribution and the empirical kernel distribution, meaning that it is a measure of departure from unimodality (should there be one). As such, the tests null distribution is uniform, which is particularly appropriate for the purpose of this analysis.

Mathematically, the test can be expressed as follows (Henderson et al., 2002):

Define:

$$\partial(F, G) \equiv \sup_{x} |F(x) - G(x)|$$

Where both F and G are bounded functions, and the supremum (sup) is the smallest value in the upper bound. Furthermore, we have:

$$\partial(F, A) \equiv \inf_{G \in A} \partial(F, G)$$

For any arbitrary A bounded function, and the infimum (inf) is the highest value in the lower bound. Let U denote the unimodal distribution function, and D denote the dip:

$$D(F) \equiv \partial(F, U)$$

The function F is unimodal if the line is concave from $(-\infty, m]$, where m is the mode, and convex from $[m, \infty)$. Thus we have the following conditions:

$$D(F) = 0 \qquad \text{for } F \in U$$

$$D(F) > 0 \qquad \text{for } F \notin U$$

As shown by the test creators (Hartigan & Hartigan, 1985), the Dip statistic is larger for uniform distribution than any unimodal distribution with exponentially decreasing tails. As such, the Dip statistic distribution is found by sampling from the uniform distributions. As the sampling size increases, one can find if the distribution is unimodal or multimodal from the Dip-value. As shown by the conditions above, for a unimodal distribution, increasing the sample size results in the Dip approaching zero, while for a multimodal distribution it approaches some positive constant.

Both methods will be employed to investigate the modality of multiple versions of the dataset.

---

5 Results

The distribution of the dataset shows significant bimodality in both the Dip test and the Silverman test, where the two modes are located at the values 1.272 and 6.63. A secondary version of the dataset was also employed, a dataset in which none of the exclusions outlined in the theoretical section were made. This dataset also showed two modes, wherein the bimodality was also significant, where the modes were located at the values 1.25 and 9.86. A test for both the start (1972) and end (2014) of the dataset was also conducted. Tests of the 1972 dataset rejected the null hypothesis, which implied bimodality, with modes located at 1.36 and 9.77. The 2014 scores implied unimodality, with a massive shift towards higher scores, with its single mode located at 9.08.

5.1 Modality Tests

5.1.1 Modified Dataset

The "modified" dataset has excluded the categories outlined in the theoretical overview. These excluded categories include all countries coded as "Democracy", "Civil War", "Transitional Regimes", "Occupation", and "Rebel Regime". There was also a secondary filter employed, namely the "ifhpol" score, the variable being checked, which consists of the imputed average of both the Polity score and Freedom House score. The secondary filter works by removing all countries above the threshold for democracy (7), which removes any outliers that exists. (Footnote here). The modified dataset contains 3977 observations of the relevant variable.

Due to the nature of the data being examined (being ordinal variables) and the test(s) employed, most of the normal central tendencies either do not apply or does not provide worthwhile information. There is currently not an academic consensus of whether the mean provides utility when working with ordinal variables, since the exact meaning of the values in the interval between a score of 1 & 2 or 5 & 6 is not known, or even knowable in some cases (Arvidsson, 2019). Although complete consensus is not reached, a reasonable conclusion at the time of writing is that any interpretation of mean for this type of analysis is less than probable to produce usable information (Liddell & Kruschke, 2018). As such, the inclusion of any means and all its connected statistics will largely be for the sake of posterity.

n	Mean	Median	min	max	range
3977	2.91	2.42	0	7	7

The table above shows all the "generic" central tendency outputs, excluding mode. The median is found at 2.42, and the min-max (0, 7) is given by lowest score given and the cutoff-point dictated by the principles of exclusion.

Running a Dip-test of the modified dataset yields the following result:

Dip-Test	
Dip statistic	p-value
0.030425	< 2.2e-16

H_1: non-unimodal, i.e., at least bimodal

The Dip tests null hypothesis is unimodality, and the alternative hypothesis is that the distribution is multimodal (i.e., at least bimodal). The result of the Dip Test is determined by the p-value, under the common threshold of $p < 0{,}05$ for interpreting the test as reporting (significant) multimodality. Marginal significance is found when $0.05 < p < 0.10$ As such, the test rejects the null hypothesis and implies significant multimodality in the data as $p = 2.2e - 16$.

Next is the Silverman test, which has the null hypothesis of the data having *k* modes against the alternative hypothesis of having *k+1* modes. Specifying the test to $k = 1$ yields the following result:

Silverman test	
Critical Bandwidth	p-value
0.89425	<2.2e-16

H1: True number of modes is more than 1

The sufficiently low p-value combined with the high critical bandwidth reinforces the result of the Dip-test, which implies that the data is at least bimodal. In essence, the critical bandwidth is the value of the smoothing parameter h where the distribution changes from

unimodal to multimodal, so a very high value further implies multimodality. Using the same test to check for up to six modes ($k = 2,3,4,5,6$) in the modified dataset results in not rejecting the null hypothesis, further implying that the number of modes is 2.

The modes for the modified dataset are, as mentioned in the beginning of the section, located at the values 1.272705 and 6.639446, which can be seen on the following graph:

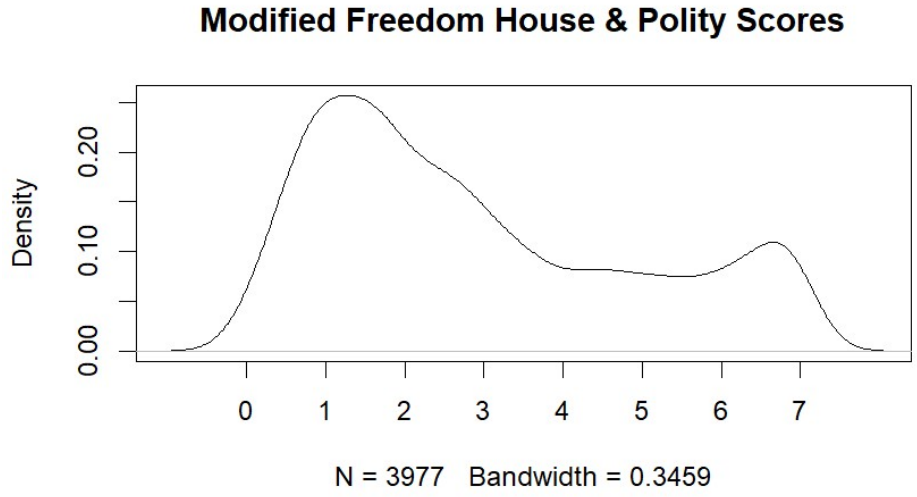


Figure 1: Density of the scores for authoritarian regimes

The bandwidth/smoothing parameter are different for each graph, calculated by using Silverman's "Rule of Thumb" method (Silverman, 2018).

5.1.2 Unmodified Dataset

The "unmodified" dataset also tested concurrent with the modified one, where the only change made to the data is the removal of NA's. The prefixes "modified" and "unmodified" will be used to qualify which version is being discussed. All the different types of tests have been run using both versions of the dataset.

For the unmodified dataset, a very similar pattern to the modified dataset occurs, where both the Dip Test and Silverman test implies bimodality. The test results are as follows:

Dip-Test	
Dip-statistic	p-value
0.064436	<2.2e-16
H_1: non-unimodal, i.e., at least bimodal	

Silverman test	
Critical Bandwidth	p-value
2.6377	<2.2e-16
H_1: True number of modes is more than 1	

The high critical bandwidth for the unmodified dataset also implies multimodality, to a stronger degree than the modified dataset, and the p-value of both tests supports rejecting the null hypothesis. The modes for the unmodified dataset are located at the values 1.24449 and 9.866495, shown on the following graph:

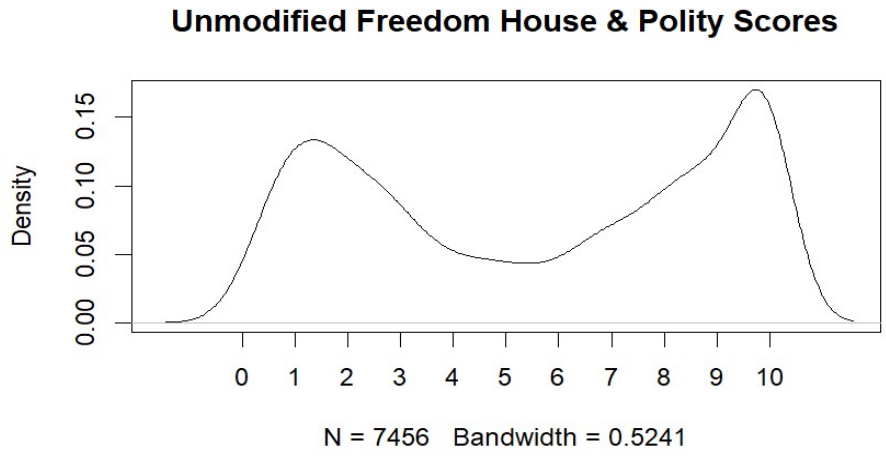


Figure 2: Density of all scores

An insight that can be gleaned from this comparison alone is that the second mode in the modified dataset is most likely a result of where the dataset is cut off. The modified dataset stops just "before" a massive influx of countries coded as democracies starts appearing.

5.1.3 Median Per Country

A potential issue with the approach thus far is that the not all observations are independent of each other, as countries within the scope can (and does) have multiple observations for different times. A country receiving the same or close to the same scores over multiple years could impact the distribution. To see if this distorts results, we have constructed two datasets with the median score for each country across the period we are looking at for both the modified and unmodified version.

5.1.3.1 Median Per Country Modified

The modes are located at the values 2.404357 & 6.213371. While the upper mode is not changed by much in terms of location, the lower mode is shifted rightward by a non-trivial amount when compared to the non-median version of the modified dataset. The Dip-test and Silverman Test of the dataset yields the following results:

Dip-Test	
Dip-statistic	p-value
0.031383	0.3853
H_1: non-unimodal, i.e., at least bimodal	

Silverman test	
Critical Bandwidth	p-value
0.97544	0.12
H1: True number of modes is more than 1	

Both tests fail to reject the null hypothesis of unimodality.

Graphically, the distribution for the modified dataset looks like:

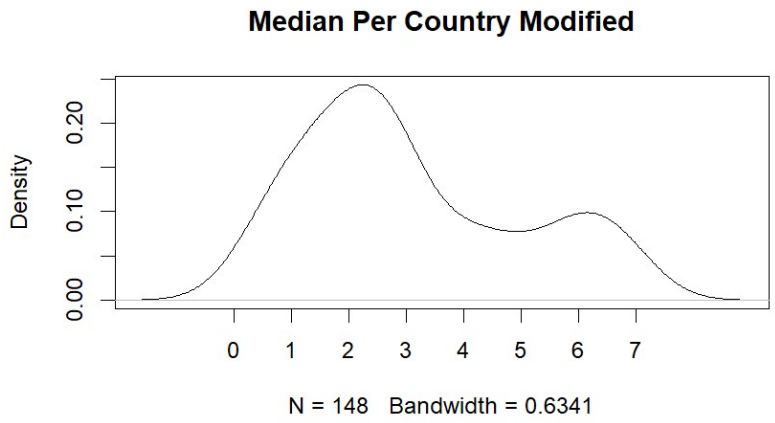


Figure 3: Density of median scores for authoritarian regimes

This result is interesting, as the relatively large dip and corresponding "mode" located towards the upper bound of the distribution makes it look multimodal. This could however be a result of the relatively low bandwidth.

5.1.3.2 Median Per Country Unmodified

In the unmodified dataset, the modes are located at 2.490975 and 9.717224. Just like for modified version the upper mode has barely changed, but the lower mode has moved by a lot. This makes sense intuitively, as the datasets are not all too different. It should be mentioned that some distortion could have occurred due to the different bandwidths. Running the Dip- and Silverman tests yields the following results:

Dip-test	
Dip-statistic	p-value
0.057549	6.40E-05

H_1: non-unimodal, i.e., at least bimodal

Silverman test	
Critical Bandwidth	p-value
2.2592	0.052

H_1: True number of modes is more than 1

Interestingly enough, the Silverman test fails to reject the null hypothesis, albeit not by much, while the Dip test does not. The Silverman test still implies bimodality, although only to a marginally significant degree. Graphically, the distribution looks like:

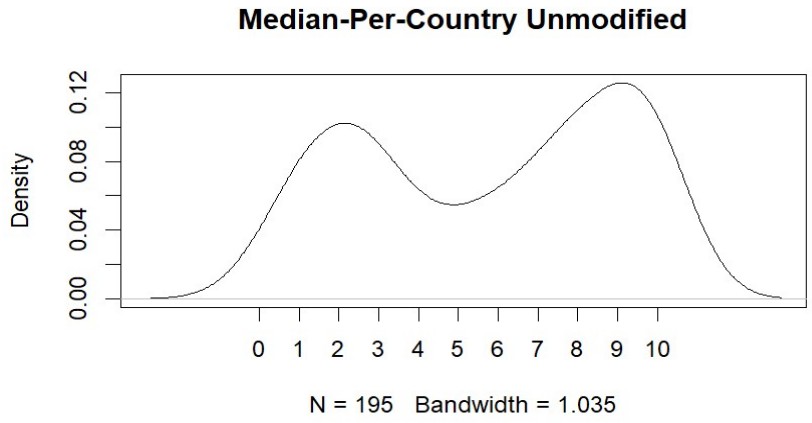


Figure 4: Density of median scores for all countries

The median version of the dataset is by and large not the most important part of this analysis, and more of an extra test. The information revealed about the distribution is less about the actual number of modes within the data, and more about seeing if the distribution behaved similarly when framed differently.

5.1.4 Start and end of period

Lastly, we look at the scores that each country received at the start and end of the period we are looking at, namely 1972 and 2014, respectively. This specific version of the dataset is based on the unmodified version of the dataset.

The Dip-test for both year's yields:

Dip-test for 1972 & 2014		
	Dip-statistic	p-value
1972	0.064466	0.0001048
2014	0.034483	0.1178
H_1: non-unimodal, i.e., at least bimodal		

The Dip test implies significant multimodality for 1972, but rejects multimodality for 2014.

Silverman-test for 1972 og 2014:

Silverman Test for 1972 & 2014		
	Critical Bandwidth	p-value
1972	2.3291	0.028
2014	1.2076	0.418
H_1: True number of modes is more than 1		

The Silverman test reaffirms the findings of the Dip-test, where the null hypothesis is rejected for 1972 but not for 2014. All tests for additional modes fails to reject the null hypothesis. Graphically, the distributions for both periods looks like:

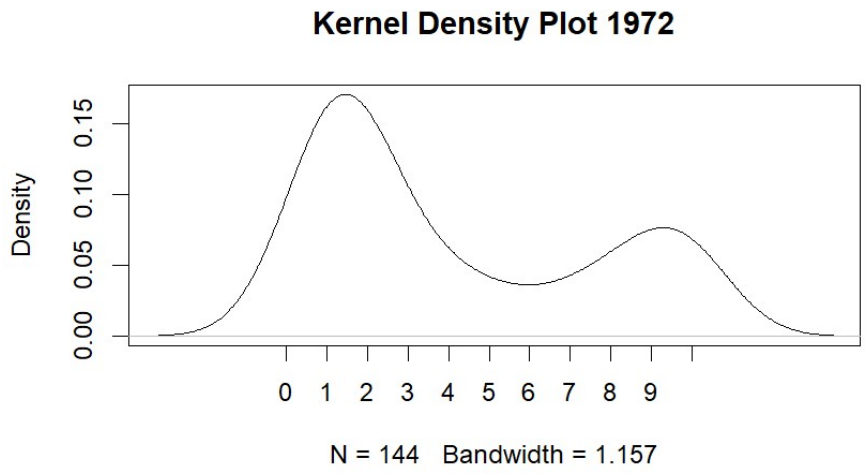


Figure 5: Density of scores in 1972

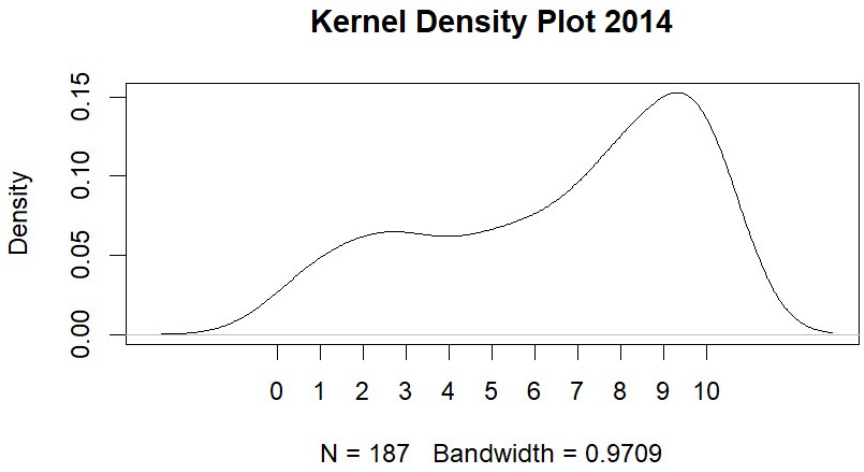


Figure 6: Density of scores in 2014

Both figures are in line with the modality tests, as the 1972 distribution are clearly bimodal and the 2014 distribution is unimodal, although with a slight dip at around 4. These two plots show a massive movement towards higher scores over the duration covers. In 1972, the modes are located at 1.36 and 9.75, with the lowest point not located in the tails (antimode) located at 5.76. The single mode in 2014 is located at 9.08. Treating the small peak before the dip as a mode and running the mode-location test, reveals that the supposed modes would be located at 2.74 and 9.83. Next, we consider the change between the central tendencies of 1972 and 2014 respectively:

	n	Mean	Median	min	max	range
1972	144	4.13	2.42	0.25	10	9.75
2014	187	6.63	7.58	0	10	10

As with the earlier central tendencies shown, most of the information is not all that relevant for this specific analysis. There are two aspects of note, firstly the massive shift in median, from 2.42 to 7.58, which further enforces the large upwards shift in scores seen. Secondly, and less overall informative, is that despite this change in median, the minimum score(s) awarded was lower in 2014 than 1972. In 2014 there were two countries "awarded" with a score of 0, Saudi Arabia and North Korea. These two countries saw a decline in their scores, where North Korea went from 0,25 to 0, and Saudi Arabia went from a 0,83 to 0.

Lastly, the difference in *n* between the periods (144 in 1972 and 187 in 2014) is a result of new countries being founded over the time period.

All in all, the modality tests affirms multimodality when all observations are inncorporated, but not when looking at the end-point of the dataset. This finding does not directly contradict the theory being tested, but it should be noted that the shift from bimodality to unimodality partially implies a strictly more efficient interval being located towards the higher end of the spectrum.

5.2 Analysis of Changes Over Time

The main point of interest in this section is seeing if there are certain values that are being avoided. What this entails is not just to check whether or not some intervals have higher degrees of movement than others, but also if certain categories retain the same countries over the period. For this section, the principles for exclusion of certain classifications have been loosened. The classifications (rebel regime, transitory regime, occupation, and civil war) were all removed due to their perceived lack of capacity to use coercion, meaning their potential lack of coercion would not be a result of self-imposed restraint. As no countries kept any of these classifications throughout the entire time period, their absence would interrupt rather than inform. The same goes for democracies, as multiple countries have moved in and out of the intervals containing democratic countries, meaning that excluding them from here on out would be counterproductive.

To illustrate the concept of movement between score intervals, consider the following figure:

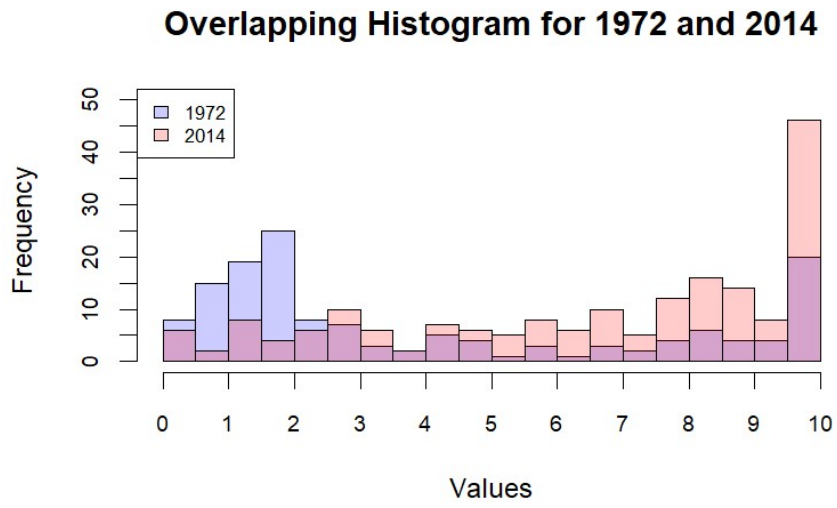
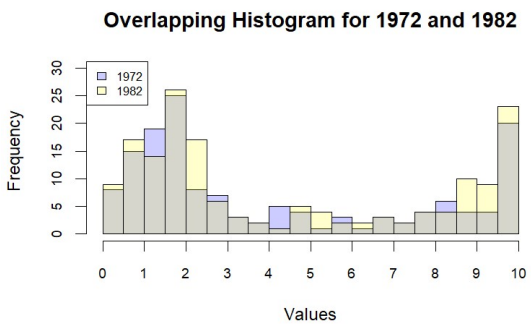
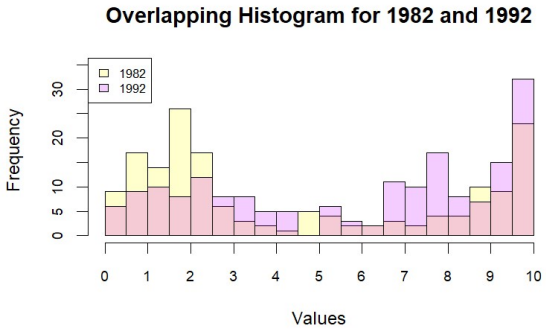


Figure 7: Number of countries in each score interval in 1972 and 2014 (overlapped)

In this histogram, the blue bars signify the different scores awarded in 1972, while the red bars represents the scores in 2014. The slightly purple areas represents the overlap. What can be seen is a substantial decrease in countries scored < 3, and nearly all scores over 5 increasing, with a massive increase seen at 10. Furthermore, it appears to have been very little change in the amount of countries situated in the interval (2.5, 5). This histogram does however not explicitly explain where the the countries in a given category moved, meaning that some or all of the countries in this interval may have been supplanted by lower scored countries that moved upwards. Said differently, just because the total number of countries that scored around 4 did not change much, it may not be the same countries, meaning that the interval may only seem efficient. Instead of looking at the entire time horizon, we can instead check on a per-decade level, to see if certain periods experienced higher levels of volatility than others:





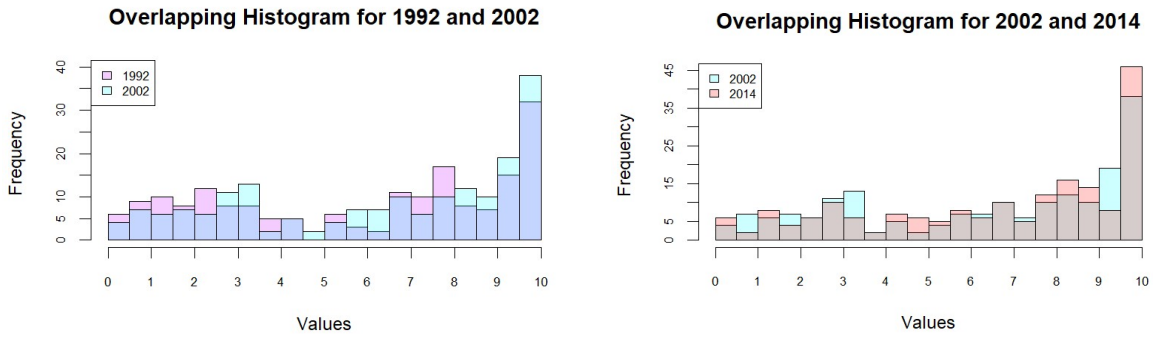


Figure 8: Overlapping histograms for each decade

What can be seen is that a majority of the biggest changes, in particular the mass exodus from the lower scores, happened in the period 1982-1992. Most of the major movements in the other periods sees the same movement patterns, but to a lesser degree with regards to the actual values.

From here the interval between each integer will be classified as a specific category. This means the interval (0,1) is classified as category 0, (1,2) is classified as category 2, and so forth. The only exception to this is category 10, which only consists of countries scored exactly 10. The following section intends to investigate the average movements of each category. Due to the exploratory nature of the subject, this section's results will be less conclusive than the investigation into the distribution and its modality.

5.2.1 Countries per Category

To gain a better insight into the change of each category its appropriate to take a closer look at the exact number of countries in each category for both periods. The following chart shows the number of countries in each category in 1972 and 2014:

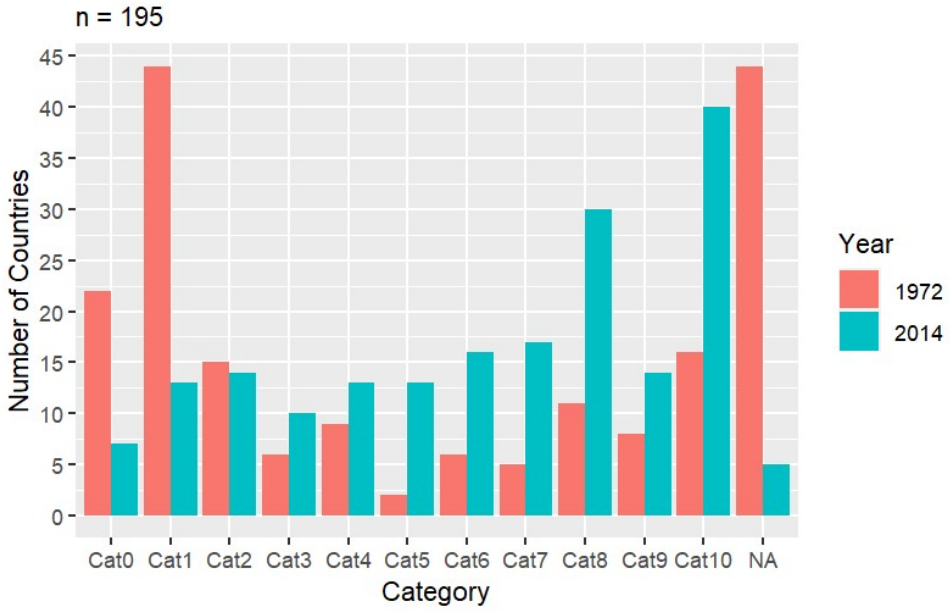


Figure 9: Number of countries in each category in 1972 and 2014 (side by side)

This figure mirrors figure (7) and (8) in many ways, as it shows a general trend towards higher scores, but it is measured by whole categories instead of the frequency of specific scores.

Pooling the scores into specific categories also shows an artefact of the datawashing not captured in the figure above: the NA's. There were 7 countries that ceased to exist (2014 NA's) and 50 countries that got founded (1972 NA's) over the period the data covers. The average score of the countries that no longer exist was 1.977 in 1972, while the average score of new countries is 7.07 in 2014. In addition, there was also a single country that both rose and fell over this duration. These countries will be included in this analysis, as the scores of both these "categories" can reveal something about how the new and dissolved countries choose/chose to position themselves.

This permutation of the data reveals that category 2 experienced the least change in total number of countries, whereas category 1 experienced the biggest decrease and category 10 experienced the biggest increase. The lower rate of change for category 2 does however not explicitly imply efficiency, as it only shows that the total number of countries in the category remained (close to) the same, not that all the countries already in category 2 remained.

One minor point of interest is the low number of countries in category 9 when compared to 10 in 2014. The fact that an interval has a little more than a third of the countries than a

neighboring single point-score is interesting. While not conclusive, it would seem like category 9 is an inefficient interval.

5.2.2 Change per Category

Grouping the dataset into categories allows for an investigation into the movement of the whole category, as opposed to disparate figures shown when looking at the individual scores. The mean absolute change and mean relative change (shown on a percent level and as a rate) per category is shown in the following table:

Table 1: Mean absolute change per category

	Category	Change
1	Cat0	3.1333333
2	Cat1	4.2317072
3	Cat2	3.6989845
4	Cat3	2.6944444
5	Cat4	1.4523809
6	Cat5	-0.9999999
7	Cat6	1.2771854
8	Cat7	-1.3569207
9	Cat8	-1.1287878
10	Cat9	0.3537046
11	Cat10	-0.0312500

The values denoted “Change” in this table describes the average movement out of each category, meaning the mean of how much the scores of the countries in each category changed over the time period. Using change as a negative indicator of efficiency for coercion usage (higher value equals lower efficiency), the following statements can be surmised:

The high rates of change found in category 0-2 is indicative of these categories higher potential for change, as well as it reflects the general trend of movement towards higher scores. Furthermore, category 10 saw the least amount of movement, which intuitively makes

sense. The highest ranked democracies being stable and efficient is less than surprising, and the little change that was seen is a result of one singular country dropping a near-trivial 0.5 points. Category 9 countries saw a small average increase.

Categories 5, 7, and 8 seeing negative change contrasts the high influx of countries to those categories. Despite growing by a large amount, the countries situated in these categories in 1972 decreased their score on average. This is particularly noticeable for category 8, which saw the second highest influx of new countries.

Graphically, the values presented in table(1) looks like:

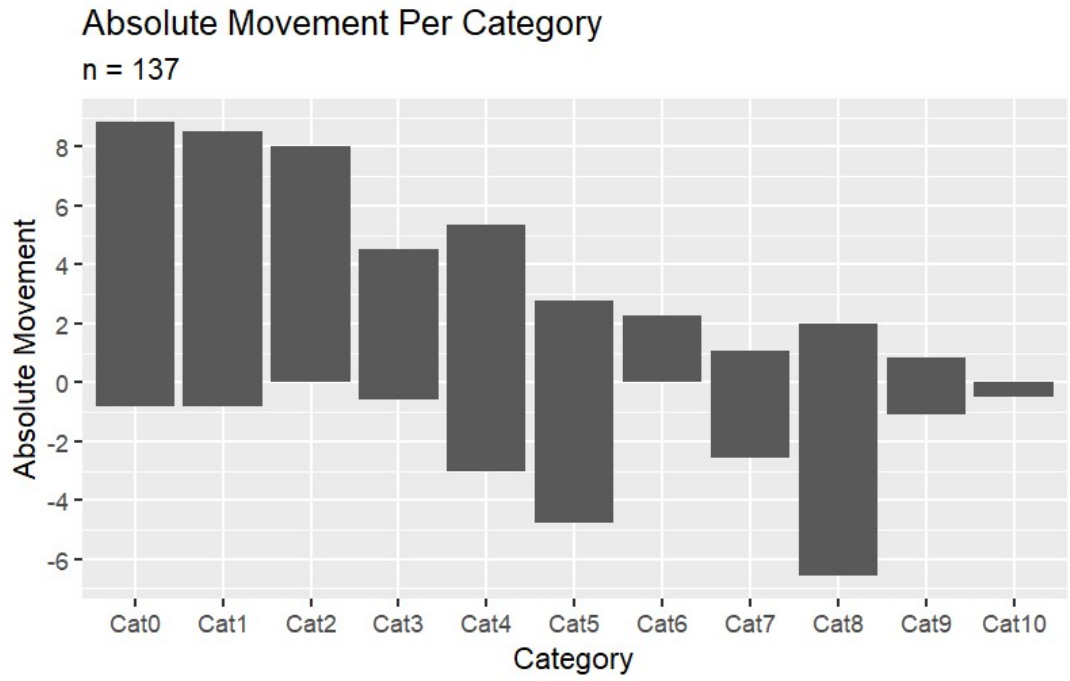


Figure 10: Absolute movement values for each category

An aspect which is revealed by figure (10), which is not present in table (1) due to the tables values being averages, is that most categories had both positive and negative movement. Barring category 10, where any positive movement is impossible, the only category which had one-sided change was category 6. Even category 0 had some internal negative movement, which means further investigation into category 6 will be necessary.

What should also be noted is that this is not a closed loop. The "experiment" is technically still ongoing, as countries still exist and change after 2014. This means that whatever category

a given country's government found themselves in in 2014 may not be their desired end goal, as a country that moved from category 0 to category 5 may only be halfway in their "growth". In contrast to this notion, movement between and across categories is not necessarily linear. A country that went through a revolution or had a significant change in government can easily "skip" categories, both on the way up and down. A very good example of this is Albania, which went from category 0 to 5 in a single year. This could be the result of some measuring issue, but it could just as easy mean that some countries experience changes in larger relative increments due to differences in how their institutions function. This non-linear change from year to year, and especially from start to end, can exacerbate the effect of extreme outliers.

The values found in this subsection only describes one part of the question, namely movement away from categories. What is ultimately interesting is seeing if there are any categories where a large proportion of countries stayed or moved, not necessarily how much they moved by. This part takes on a much more graphical approach.

5.2.3 Composition of Category Distribution

A category is composed of different countries, all of which had a category at the start of the period (NA's excluded). This next figure shows where the countries in each category came from, meaning the composition of countries based on their original category. Now let us look at the composition of the category distribution in 2014:

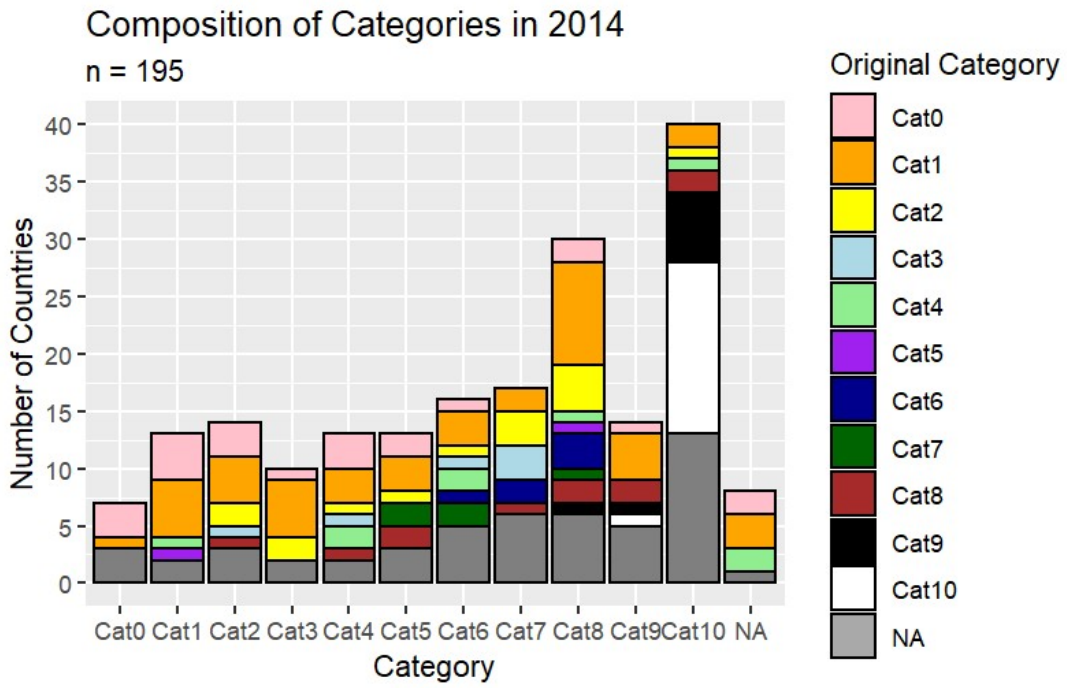


Figure 11: Number of countries in each category in 2014 based on category of origin

Figure(11) shows the number of countries in each category in 2014, and the colors signify which category these countries came from in 1972, color-coded from 0-10 top down. The NA's (grey) once again represents the countries that were founded during the time period, while the NA bar itself signifies the countries that ceased to exist. Although seemingly difficult to parse, let us use a single category as an example to explain what the figure conveys. We can see that category 0 is composed of 7 countries. Three of these countries were already category 0 in 1972, signified by the pink color. One of them came from category 1 (orange), and three are countries that were founded during the period (grey). The column denoted NA represents the countries that existed in 1972, but not in 2014. This column tells us that all countries that fell during the time period was in the interval (0, 5), since the highest rated country was in category 4. There is one exception to this, The Federal Republic of Yugoslavia, which is denoted as NA as it was both founded and dissolved within the timeframe. The country was in category 8 as of its dissolution, which breaks with the admittedly small pattern of only low scoring countries ceasing to exist.

Due to the general trend of higher scores, a secondary weak condition for efficiency may be argued for: category resilience. In this context, a category's resilience can be understood as how many countries that started in a given category stayed in said category. Resilience is not a complete condition for efficiency by any stretch of the imagination but seeing if a category retains a certain amount of countries can point to the coercion usage of these categories not creating enough coercion resentment to diminish the utility attained. There needs to be some discretionary judgement made when looking at the categories through this lens however, as it cannot explain away certain aspects.

The first characteristic of interest in the figure is that none of the previous category 0 countries reached category 10, and that the biggest influx of category 10 countries came from countries founded within the time period. We can also see that newly founded countries are relatively well spread out across the categories, and while not strictly proportional to the count in each category, there are no extreme outliers either.

It is interesting to note that all countries that ceased to exist, shown on the NA column, came from categories 0, 1, and 4. While this is a very small sample size, it seems to imply that most countries that fall are on the lower spectrum.

With a very keen eye one can also see that none of countries in category 3 and 5 in 2014 are the same as the ones in 1972, meaning all countries that resided in those categories moved out of them. This implies that they are, to some degree, inefficient intervals. The argument for category 5 being inefficient due to the resilience argument is however very weak, due to only 2 countries residing there in 1972, while 13 countries exist there in 2014. This concept condition cannot provide any explanation for the massive influx of countries seen in the category.

5.2.4 Identifying Intervals

Firstly, we can see that despite the massive decrease in total countries in category 0 and 1, a lot of their respective categories are composed of countries that remained. Category 1 in particular is composed of a (weak) majority of previous category 1 countries, which partially implies stability in spite of the large decrease in total countries. This can also be seen in the following figure, which mimics figure(10) but shows the whole category movements of each category:

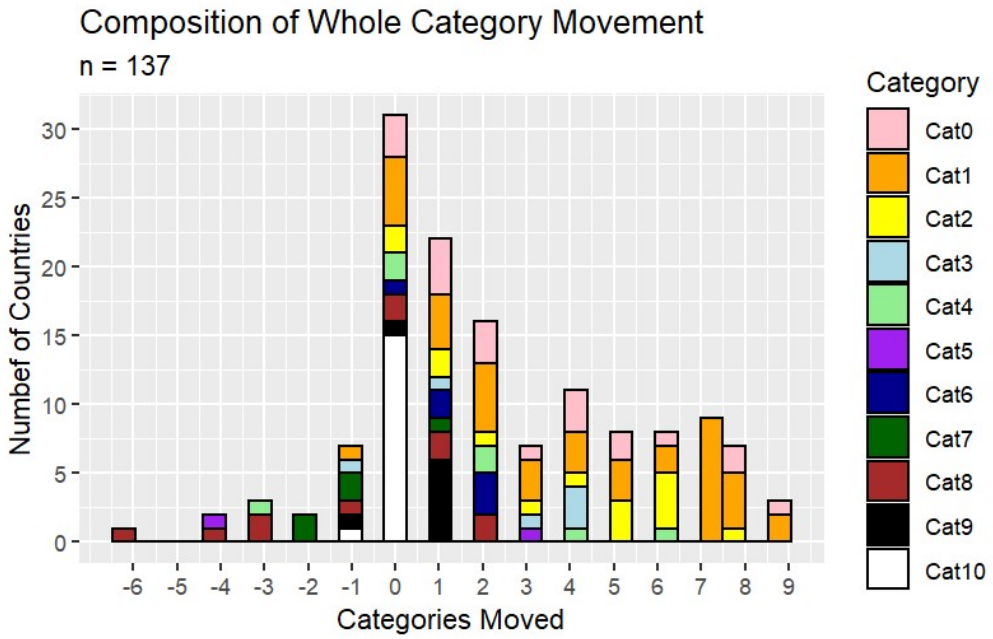


Figure 12: Movement of countries

Figure (11) shows the number of whole categories moved, where the x-axis shows how much a given country moved by, and the y-axis shows how many countries moved by a given amount. The bars themselves are composed of the original category a given country was situated in in 1972. We can see that most countries moved relatively little, with 0, 1 and 2 categories moved being the largest bars. Nearly all Category 10 countries moved by 0, while category 1 countries almost exclusively moved upwards. Additionally, we can see that only countries previously situated in categories 0:4 moved by more than four categories, which is not surprising as already high scoring countries had much less room to move upwards. Furthermore, negative movement was relatively rare overall, where a majority of countries that moved downwards came from category 7 and 8.

Revisiting the conjecture of inefficiency found at category 9, we can see that only a single country remained. This finding, in conjunction with its low amount of countries when compared to category 10, further implies that category 9 is an inefficient interval, with a "pull" towards category 10 due to its proximity.

Furthermore, there is category 7 and 8. Category 7 has the same characteristics as categories 3 and 5, where no countries remained, but as a contrast it has significant negative movement. Additionally, while the average (negative) movement of category 7 is larger than its

neighboring category 8 (17% and 12% respectively), the countries in category 7 decreased by much less and saw only one country increase their score, which was only up to category 8. Having a higher negative relative change, despite its previous countries going down by less, speaks to the sheer volume of countries employing more coercion at the end of the period.

Category 7 is also notably the first category above the democracy threshold, and the only category within this group to not have any country rise to category 10. Category 8 had very few remaining countries (2), and despite its massive increase in total countries, only one came from a higher category. This is of course partly due to there not being all too many categories above it to "take from", but combined with its average movement being negative, it seems like the countries in the category are being "pulled" downwards. Category 8 is more mixed in this regard, but still sees a higher number of countries going lower, while category 9 sees the similar trend as category 7 but the opposite direction, with only one country staying or going lower. This all heavily implies that categories 7 and 9 are part of different inefficient intervals, with category 8 being a breakpoint between positive and negative pull. This is somewhat out of the scope of this paper, as the countries above the democracy threshold are rather obviously not authoritarian regimes. However, if countries in category 7 tends to move downwards into authoritarianism, it implies that using more coercion while in this category has a positive effect on these governments.

For lower bound efficiency, there are conflicting results. Basing the concept of efficiency off movement alone alludes to there being no efficient interval on the lower end, with lower score equaling lower efficiency. Basing it off resilience implies weak efficiency at category 1, and an even weaker efficiency at category 0. This is however counterintuitive, as these were the categories who lost the most countries and whose countries moved the most. They are however the categories that retained the most amount of original countries proportional to new countries entering their intervals (excluding category 10). There not being a single clear answer is in line with the score distribution and modality tests for 2014, where a very small peak can barely be seen, but it is not pronounced enough to reject the unimodality null hypothesis. What seems relatively clear is that category 3 is explicitly not efficient, as not a single country remained, and it is completely comprised of countries from the lower categories. This last aspect alludes to category 3 being a temporary category for countries where the coercion level this category employs is dominated by the coercion resistance it

implicitly creates. This is backed up by there only being a single country which stayed within the category for more than 10 consecutive years.

The upper bound category for authoritarian states is category 6, which incidentally was the only category with no negative movements, and only a single country remained in the category. The only countries that moved downwards into this category came from category 7. As these categories exist right on the threshold of what is considered democracies, some conjecture about their relationship can be made. Countries barely over the threshold for democracies seemed to have increased their coercion levels somewhat, while countries barely below the threshold seems to have reduced it[2]. Being just below the threshold could mean coercion resentment diminishes the utility gained from coercion enough to decrease coercion. The opposite appears to be true for countries right above the threshold.

---

[2] If we apply the micro-foundation of Schøyen's theory, it seems like this two-category interval (6, 8) is where coercion and coercion resentment are in some type of inverse equilibrium. Being just above the threshold, the coercion level is not high enough to where coercion resentment increases parent socialization, which in turn incentivizes the regime to increase

6 Conclusion

6.1 Summary

The goal of this paper was to test one of the theories proposed by Schøyen. The theory states that there exist efficient and inefficient intervals of coercion as a result of an intrinsic response to coercion called coercion resistance. The existence of efficient and inefficient intervals further implies a multimodal distribution of coercion levels. The coercion resentment response diminishes the utility of coercion, but is non-linear in nature, such that only certain coercion levels diminish utility to a degree where it is rational to change coercion level. Furthermore, we assumed that the dataset, which compiles and indexes scores of certain political rights and freedoms, works as a metric for degree of authoritarianism, with the implicit assumption that this metric correlates with coercion usage.

We chose to employ two different tests using kernel density estimation, the Dip test and the Silverman test. For the entire dataset, the hypothesis was not rejected, which strengthened the hypothesis that the data was bimodal. The quality of bimodality was retained when looking exclusively at the first year of the dataset, but the data for the end of the dataset was found to be unimodal, with a general trend towards the higher scores over time. This unimodal distribution had a small dip at 3.8, but not enough of a difference to be classified as bimodal.

Furthermore, we tried to ascertain the locations of the efficient and inefficient intervals by seeing if certain sub-intervals of scores, denoted as categories, saw movement that was disproportional to its size. The results implied lower coercion levels were more efficient in general, as it produced no corresponding coercion resentment. An additional complementary way of identifying efficiency, called resilience, was devised and employed in conjunction with movement values. The very simple method checks if categories retain the same specific countries from the start to the end of the period, not just the number of countries.

The results found that the lowest interval above the threshold for democracy, specifically the interval (7,8) seemed inefficient in its coercion usage, but not to a conclusive degree. Furthermore, the only category found to be explicitly inefficient, by the metrics employed, was category 3 (3, 4), where all of the relatively few countries within came from the lower interval (0, 2.9). The inefficiency of the interval corresponds with the antimode (dip) found in the modality tests.

Category 10, comprised only of countries scored exactly 10, saw the biggest increase in number of countries and the lowest overall negative change, which is in line with the overall movement towards higher scores as a result of global democratization.

On the lower end of the spectrum, the result of this hypothesis is ambiguous. The change in values observed implies that there is strictly more efficiency the higher the score, with no efficient intervals located on the lower end. Using resilience as a metric implies some degree of efficiency in category 0 and 1, but the mass exodus of countries from these categories is too big to ignore. The relatively strong resilience does imply that employing these coercion levels creates stability for some such regimes to a degree, which does imply efficiency of sorts, although strictly less efficiency than high scoring democracies. It should be noted that countries do not exist in isolation, and in a world where a lot of economic power is concentrated within democratic countries (Acemoglu et al., 2019), more authoritarian countries are liable to be subjected to economic sanctions and interventions, which can serve as an additional incentive to democratize.

Lastly, the border between authoritarianism and democracy, categories 6 and 7, experienced growth towards each other's side of the threshold. This pattern does strengthen the hypothesis that inefficient intervals exist, but the preliminary conjecture that efficiency exists at the lowest coercion level possible for authoritarian regimes is almost outright rejected.

This paper did not reject the theory, but could only partially strengthen its assertions due to ambiguous results, but some insight was found. Certain intervals of coercion were implied to be efficient or inefficient, but anything tangible outside the general idea of "less coercion is good" is difficult to argue for. These implications are however thoroughly interesting, and can serve a function in analyzing authoritarian governmental behavior.

6.2 Concluding Remarks

The findings of this paper are far from conclusive, a fact that may seem less than optimal. However, as said multiple times throughout this paper, the point was not to provide a definitive answer to this very complicated, but important question. The main goal was to serve as an exploratory probe of sorts, to open the investigation into how the less democratically inclined countries of the world operate and use their power to coerce their

population. The theory that was tested was not rejected by the tests, which means there could be merit in further investigation and analysis.

---

7 Bibliography

Acemoglu, D., Naidu, S., Restrepo, P. & Robinson, J. A. (2019). Democracy does cause growth. *Journal of political economy*, *127*(1), 47-100.

Arvidsson, R. (2019). On the use of ordinal scoring scales in social life cycle assessment. *The International Journal of Life Cycle Assessment*, *24*(3), 604-606. https://doi.org/10.1007/s11367-018-1557-2

Byman, D. & Lind, J. (2010). Pyongyang's Survival Strategy: Tools of Authoritarian Control in North Korea. *International Security*, *35*(1), 44-74. http://www.jstor.org/stable/40784646

Froese, P. (2008). *The plot to kill God: Findings from the Soviet experiment in secularization*. Univ of California Press.

Hadenius, A. & Teorell, J. (2005). Cultural and economic prerequisites of democracy: Reassessing recent evidence. *Studies in Comparative International Development*, *39*(4), 87-106. https://doi.org/10.1007/BF02686166

Hadenius, A. & Teorell, J. (2007). Pathways from authoritarianism. *Journal of democracy*, *18*(1), 143-157.

Hall, P. & York, M. (2001). ON THE CALIBRATION OF SILVERMAN'S TEST FOR MULTIMODALITY. *Statistica Sinica*, *11*(2), 515-536. http://www.jstor.org/stable/24306875

Hartigan, J. A. & Hartigan, P. M. (1985). The dip test of unimodality. *The annals of Statistics*, 70-84.

Henderson, D. J., Attar, P. K. & Russell, R. R. (2002). Modality tests for use in applied econometrics: application to macroeconomic convergence. *University of California Riverside (Unpublished paper)*.

Liddell, T. M. & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, *79*, 328-348.

Schøyen, Ø. (2020). What limits the efficacy of coercion? *Cliometrica*, *15*(2), 267-318. https://link.springer.com/epdf/10.1007/s11698-020-00207-0?sharing_token=5rNzZ_BZ7mD-dDRUM5DfNfe4RwlQNchNByi7wbcMAY4gGGw8urNIM6WWE1Wtl0EUkbxqBs-nO7SWblFwrpzUNb8VK0zs-LGNjmKqBaSbt_CktrS1tT6TOVPTEJe0qmj_8td3pkJ5E8ptTUrZ6Glcp6wGhq4Jw1TVaCIuxPL10pA%3D

Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society: Series B (Methodological)*, *43*(1), 97-99.

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.