

Gründler, Klaus; Krieger, Tommy

**Working Paper**

## Using support vector machines for measuring democracy

Discussion Paper Series, No. 130

**Provided in Cooperation with:**

Chair of Economic Order and Social Policy, Julius Maximilian University of Würzburg

*Suggested Citation:* Gründler, Klaus; Krieger, Tommy (2015) : Using support vector machines for measuring democracy, Discussion Paper Series, No. 130, Julius Maximilian University of Würzburg, Chair of Economic Order and Social Policy, Würzburg

This Version is available at:

<https://hdl.handle.net/10419/110276>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

Discussion Paper No.130

## Using Support Vector Machines for Measuring Democracy

*Klaus Gründler*  
*Tommy Krieger*

April 2015

# Using Support Vector Machines for Measuring Democracy

*Klaus Gründler*

*Tommy Krieger*

Discussion Paper No.130

April 2015

Julius Maximilian University of Würzburg

Chair of Economic Order and Social Policy

Sanderring 2

D-97070 Würzburg

Phone: 0931 – 31 86177

Fax: 0931 – 82744

E-Mail:

[klaus.gruendler@uni-wuerzburg.de](mailto:klaus.gruendler@uni-wuerzburg.de)

[tommy.krieger@uni-konstanz.de](mailto:tommy.krieger@uni-konstanz.de)

# Using Support Vector Machines for Measuring Democracy

Klaus Gründler<sup>\*</sup> and Tommy Krieger<sup>\*\*</sup>

*<sup>\*</sup>University of Würzburg, Department of Economics,  
Sanderring 2, 97070 Würzburg. E-Mail:*

*klaus.gruendler@uni-wuerzburg.de*

*<sup>\*\*</sup>University of Konstanz, Department of Economics,  
Univeristätsstraße 10, 78464 Konstanz. E-Mail:*

*tommy.krieger@uni-konstanz.de*

April 17, 2015

## **Abstract**

We present a novel approach for measuring democracy, which enables a very detailed and sensitive index. This method is based on Support Vector Machines, a mathematical algorithm for pattern recognition. Our implementation evaluates 188 countries in the period between 1981 and 2011. The Support Vector Machines Democracy Index (SVMDI) is continuously on the 0-1-Interval and robust to variations in the numerical process parameters. The algorithm introduced here can be used for every concept of democracy without additional adjustments, and due to its flexibility it is also a valuable tool for comparison studies.

Keywords: Democracy, Support Vector Machines, Democracy Index  
JEL No.: C43, C65, C82, H11, P16

# 1 Introduction

The traditional way to create a democracy indicator seems easy and natural. First, it is required to choose a definition of democracy. Then a number instruments must be designed that can describe the properties of the chosen theoretical concept. Finally, it is necessary to find a suitable manner to combine the selected variables to generate a democracy index (Saward, 1994).

This, however, is often much easier said than done. One main problem concerns the definition of democracy. Neither in political science nor in political practice does a unique concept exist which is widely accepted (Bühlmann et al., 2012). The interpretations range from minimal approaches focusing primarily on the election process (Dahl, 1971, Locke, 1965, Montesquieu, 1965) to concepts which have comprehensive requirements in regard to human rights or social inequality (Habermas, 1992, Rawls, 1971). Hence, it is not surprising that various democracy indicators use different concepts and instruments. For instance, the popular Vanhanen-Index utilizes only two dimensions (*participation* and *competition* of elections) to characterize a democracy (Vanhanen, 2000). Rather minimal concepts are also used in the Polity-Index as proposed by Marshall et al. (2014), Bollen (1990) and Alvarez et al. (1996). More extensive approaches are employed in the ratings established by Freedom House (FreedomHouse, 2014a), the Democracy Barometer (Bühlmann et al., 2012) or the Democracy Index of the Economist Intelligence Unit (EIU, 2011). The Advantages and disadvantages of these indicators are extensively discussed in the literature (e.g. Munck and Verkuilen (2002), Müller and Pickel (2007), Cheibub et al. (2010)). Points of criticism often include the level of detail, an unfounded scaling or combination of the variables, and the selection of the instruments. Moreover, a common problem in all of these approaches is that they hardly allow for inclusion of alternative or additional instruments, since the aggregation of variables is limited to the original concept.

An alternative way to create a democracy indicator is to combine the information of multiple traditional indexes. For instance, Acemoglu et al. (2014) link the Freedom House Rating and the Polity Score with the indicators of Cheibub et al. (2010) and Boix et al. (2013). While these indexes may provide a more precise measure of democracy, the applied heuristic is

quite facile and only enables a binary classification. Obviously a democracy index with only two possible characteristics cannot be detailed enough to describe the real situation. In contrast, [Pemstein et al. \(2010\)](#) suggest a more complex method (UDS), which is grounded on a Bayesian latent variable approach and merges ten traditional indexes. Other combining approaches are suggested by [Lauth \(2013\)](#) or [Gugiu and Centellas \(2013\)](#). In this paper we introduce a mathematical algorithm that is able to solve the problems previous indexes are confronted with. The proposed method is very adaptive, allowing for both a traditional indicator based on certain properties of the countries and a combination of an optional number of established indexes. Our approach mainly uses Support Vector Machines (SVMs), a mathematical algorithm for pattern recognition. The increased benefits of SVMs have been shown in several applications, e.g. in medicine to categorize cancer cells ([Guyon et al., 2002](#)) or in geophysics to classify hyperspectral data ([Gualtieri, 2009](#)). SVMs use a nonlinear generalization of the Generalized Portrait algorithm developed by [Vapnik and Lerner \(1963\)](#) and [Vapnik and Chervonenkis \(1964\)](#).

A short introduction is given in Section 2. For additional information see e.g. [Steinwart and Christmann \(2008\)](#), [Vapnik \(1995\)](#), [Schölkopf et al. \(1998\)](#) or [Smola and Schölkopf \(2004\)](#).

Section 3 describes the measuring process in general and details the specific setup used in our implementation. Section 4 presents the results of our approach and compares the estimated Support Vector Machines Democracy Indexes (SVMDIs) with some established democracy indicators such as the Vanhanen-Index or the Freedom-House-Rating. Section 5 evaluates the robustness of the process. We conclude in section 6.

## 2 Support Vector Machines

Support Vector Machines (SVMs) are mathematical algorithms for pattern recognition. In our measuring process we apply this method in two different ways. The first approach is a classification tool, the second uses regressions. In this section, we provide a brief overview of the first variant, as the general ideas of the two applications are very similar. This introduction is mainly based on [Steinwart and Christmann \(2008\)](#). A detailed introduction of the regression tool is given by [Smola and Schölkopf \(2004\)](#).

The problem to be solved by the SVMs classification tool can be described as follows: Given a certain data set  $(X_1, y_1); \dots; (X_n, y_n)$ , where  $X_i \in \mathbb{R}^m$  and  $y_i \in \{-1, +1\}$ , we want to find a function  $C: \mathbb{R}^m \rightarrow \mathbb{R}$  with the property

$$C(X_i) = y_i \quad \forall i = 1, \dots, n. \quad (1)$$

The general idea of the SVMs is to find a hyperplane

$$H(a, \gamma) = \{x \in \mathbb{R}^m | a^T x = \gamma\}$$

that separates the observations according to their labels  $y_i$ . Assuming we find such a hyperplane in the original data space, the function

$$C(X_i) = \text{sign}(a^T X_i + \gamma)$$

classifies the observation according to condition (1). In most practical applications, however, it is impossible to find such a solution in  $\mathbb{R}^m$ . To circumvent this problem, SVMs do not conduct this search in the original data space but in a space with higher dimension which is called *feature space*. We use a transformation function  $\Phi(\cdot)$  to shift the information  $X_i$  into the feature space. The form of the optimal hyperplane in the feature space is evaluated by solving the optimization problem<sup>1</sup>

$$\min_{a, \gamma} \frac{1}{2} \|a\|_2^2 \quad s.t. \quad y_i \cdot (\Phi(X_i)^T a + \gamma) \geq 1 \quad \forall i = 1, \dots, n.$$

By transforming the estimated hyperplane back to original data space  $\mathbb{R}^m$ , we obtain a nonlinear classification function which can satisfy condition (1). Because the transformation function is generally unknown, we instead use kernel functions  $k(\cdot)$ . The main challenge is here to choose a suitable kernel function (Burges, 1998). We rely on the Gaussian RBF kernel, commonly used in machine learning, in both applications of the Support Vector Machines.

---

<sup>1</sup>The optimization problem of the used Support Vector regression is described briefly in Appendix B.

### 3 Method

In order to use the Support Vector Machines (SVM) to measure democracy we design a procedure consisting of ten stages which we cover below. The algorithm yields a continuous indicator, which is normalized to the  $[0; 1]$  interval and can be interpreted intuitively as the probability that a certain country-year-observation is *democratic*. We divide the description into two parts. The first part characterizes the individual steps of the algorithm in general. The second part is concerned with a more detailed description of the setup used in our implementation.

#### 3.1 Algorithm

1. We choose a set of variables that characterizes a democracy.
2. We select country-years that can be easily labeled as (*non*) *democratic*. The democratic observations receive the label 1, the non democratic the label 0.
3. We choose  $d_1$  ( $d_0$ ) of the labeled (non) democratic observations. Subsequently, we refer to this set as the *R-Set*.
4. We use the above SVMs classification tool to check the initial consistency of the *R-set*. If the SVMs confirm our selection, we continue, if not we must revise it.
5. We use a random generator to pick out  $t_1$  ( $t_0$ ) of the  $d_1$  ( $d_0$ ) country-years. These  $t = t_0 + t_1$  observations constitute the *T-Set*.
6. We apply the Support Vector regression (Appendix B) using the observations in the *T-Set* and compute a non linear function  $f(\cdot)$ .
7. We use the estimated function  $f(\cdot)$  to classify all country-years.
8. We repeat the stages 5, 6 and 7  $x$  times with  $x \in \mathbb{N}$ .
9. We conduct a democracy indicator for each country-year as the mean of the  $x$  estimations.
10. We repeat all steps from 3 to 9  $y$  times with  $y \in \mathbb{N}$ .



11. We conduct the final Support Vector Machines Democracy Index (SVMDI) for each country-year as the mean of the  $y$  estimations.

### 3.2 Setup

In our implementation we use eight variables to characterize a democracy, which are listed in Table 1. These variables enable us to measure democracy in 188 countries for the period from 1981 to 2011.

Variable	Description
<i>PRFH</i>	Rating Political Rights ( <a href="#">FreedomHouse, 2014a</a> )
<i>CRFH</i>	Rating Civil Rights ( <a href="#">FreedomHouse, 2014a</a> )
<i>FPFH</i>	Rating Freedom of the Press ( <a href="#">FreedomHouse, 2014b</a> )
<i>PART</i>	Rate of Participation ( <a href="#">Vanhanen, 2000</a> )
<i>COMP</i>	Rate of Competition ( <a href="#">Vanhanen, 2000</a> )
<i>INJUS</i>	Rating Independence of Justice ( <a href="#">Cingranelli et al., 2014</a> )
<i>PRCIRI</i>	Average of the CIRI ratings: Freedom of Foreign Movement, Freedom of Domestic Movement, Freedom of Assembly and Association, Freedom of Speech and Freedom of Religion ( <a href="#">Cingranelli et al., 2014</a> )
<i>PTS</i>	Political Terror Scale ( <a href="#">Gibney et al., 2013</a> )

**Table 1** Democracy variables.

In addition, our definition of democracy prohibits non-independent countries from being classified as democratic. The settings of the six process parameters are summarized in Table 2. Furthermore, we use the Polity-Database ([Marshall et al., 2014](#)) as a benchmark for labeling in step 2. Democratic country-year-observations are defined as having a Polity-Score of 10, whereas we classify a country year as non-democratic whenever the Polity-Score is  $-7$  or lower. This selection rule ensures a classification without mismatches (step 4).<sup>2</sup>

Parameter	$d_1$	$d_0$	$t_1$	$t_0$	$x$	$y$
<b>Setup</b>	100	100	50	50	2500	200

**Table 2** Process parameters.

<sup>2</sup>This result holds for the Gaussian RBF kernel as well as other kernel functions such as the Linear kernel or Polynomial kernels with degree  $\leq 4$ .

## 4 Results

Our implementation enables the evaluation of 188 countries in the period from 1981 to 2011. A detailed illustration of all states can be found in appendix A.<sup>3</sup> All Support Vector Machines Democracy Indexes (SVMDIs) are elements of the  $[0; 1]$  interval and can be interpreted intuitively as the probability that a certain county in a specific year can be characterized as a democracy.

This section compares our indicators with some of the established indexes. Table 3 gives a brief overview of selected correlation coefficients between the SVMDI measure and commonly used indexes in economic and political science articles.

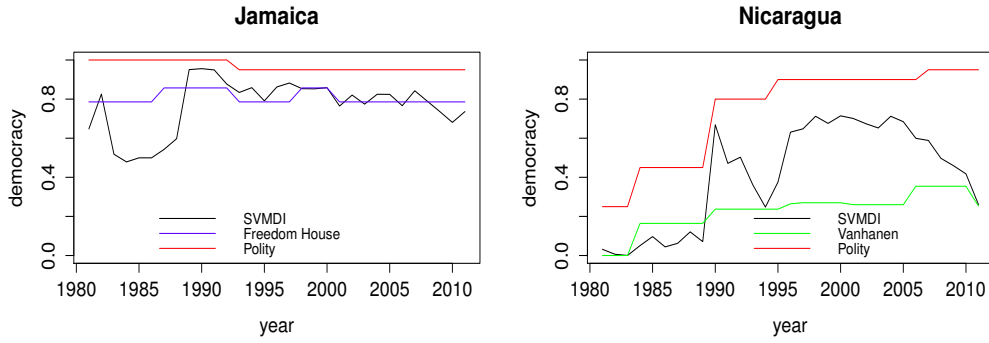
<i>year</i>	<i>FreedomHouse</i>	<i>Vanhanen-Index</i>	<i>Polity-Score</i>	<i>EIU</i>	<i>UDS</i>
1981	0.9483	0.8970	0.9433		0.9398
1986	0.9519	0.8886	0.9402		0.9403
1991	0.9534	0.8555	0.8789		0.9197
1996	0.9606	0.8024	0.8966		0.9117
2001	0.9584	0.8176	0.8831		0.9069
2006	0.9535	0.8116	0.8642	0.9036	0.9064
2011	0.9569	0.7966	0.8359	0.9063	0.9141

**Table 3** Correlations of SVMDI and commonly used democracy indexes.

The SVMDIs reveal a high correlation to each of the reported democracy measures, which is why we can conclude that our method works in general. However, the correlation does not provide information about the benefits of our approach in comparison to the established approaches. The main advantage of the method we present is that it yields more detailed and sensitive measurements, i.e. our indexes already reflect small improvements and setbacks in the process of democratization. Figure 1 serves to illustrate these properties by comparing the SVMDIs of Jamaica and Nicaragua with two established democracy indicators. In the case of Jamaica, we see that there is a huge divergence in the trend of the three indexes, especially in

<sup>3</sup>Thereby former Soviet, Yugoslavian and Czechoslovakian states are quoted in a separate manner. Their indexes from 1981 to 1990 (to 1992) have to be interpreted as the values of the USSR or SFR Yugoslavia (Czechoslovakia). The scores of Serbia and Montenegro (1991-2005) are integrated in the trends of Serbia, Montenegro and Kosovo. Moreover, Germany and Yemen have to be interpreted as West-Germany or North Yemen before reunification.

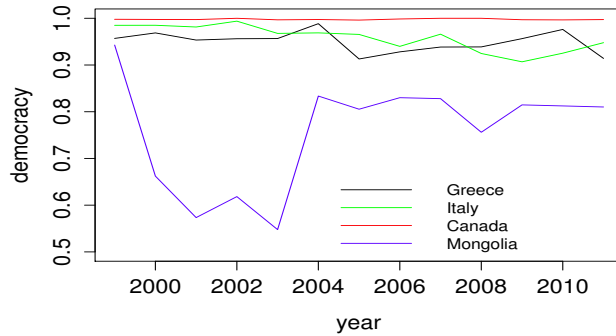
the early 1980s. While the Polity-Scores and the Freedom House-Ratings do not change significantly, our index denotes a sharp fall in the year 1983. Given the political situation in that year, our result is more plausible. The opposition 'People's National Party' boycotted the election, with the result that the ruling 'Jamaica Labor Party' won all seats in the parliament (Figueros, 1985). Therefore, there was no parliamentary opposition in the following years a situation that should be notified negatively by a democracy indicator.



**Figure 1** Democracy Indexes of Jamaica and Nicaragua.

The case of Nicaragua (Figure 1 right) highlights the fact that the Vanhanen-Index typically tends to change after elections, which in Nicaragua generally take places every five years. With the exception of the minor decline in 2011, the Vanhanen index provides no indication for a declining degree of democracy during the entire period. Likewise, the Polity-Score implies a similar phase of flourishing democracy without any indication of an interruption. In contrast, our indicators display a continuous loss of democracy since 2006. Due to the increasingly autocratic governance of president Daniel Ortega (Anderson and Dodd (2009), McConnell (2014)), who was elected in 2006, a decreasing trend is more appropriate than a constant or increasing progress.

A second useful property of our implementation is that it detects differences between established democracies. Figure 2 illustrates the trend of the SVMDIs of Canada, Mongolia, Greece, and Italy between 1999 and 2011. The Polity-Index suggests no variation in the degree of democratization in any of these countries, as it assumes a constant value of 10 during the whole period, i.e. those observations can reside in the *R-set*. In con-



**Figure 2** Support Vector Machines Democracy Indexes of Canada, Greece, Italy and Mongolia.

trast to the Polity-Index we see that our indicators reveal some differences in the trend and level of democratization of those states. The decrease in the Mongolian SVMDIs may be the result of the 2000 parliamentary election, in which the Mongolian People’s Revolutionary Party (MPRP) won 72 of 76 seats ([Severinghaus, 2001](#)). Hence, Mongolia was close to a single-party system in the period 2000-2004, which does not correspond” with the typical interpretations of democracy. The gap between Italy, Greece and Canada is also reasonable, since the level of corruption in Italy and Greece is substantially higher than in Canada ([Transparency-International, 2011](#)).

## 5 Robustness of the Process

A possible point of criticism concerns the choice of the process parameters, as there are no theoretical reasons why we choose these values. Without additional remarks it could be argued that the setup influences the indicators in a significant manner. To examine the robustness of our results, this Section conducts two different types of sensitivity analyses. The first part is concerned with the outer robustness of our approach, i.e. we illustrate that an increase in  $y$  does not affect the degree of democratization attributed to a country-year. The second part deals with the internal robustness, i.e. we show that neither a shift in  $x$ ,  $t$  nor  $t_1$  yields substantial changes in the results. Subsequently, we assume without loss of generality (w.l.o.g.) that  $y = 1$ .

As benchmarks we use the *average* of all observations, which provides

information about the shift of the democracy indicators, and the *Gini-coefficient*, which informs us of the change in their distribution. Furthermore, we examine the *Pearson correlation coefficient* and the *maximal absolute derivation* between the basic setup (Table 2) and the tested alternative setup.

## 5.1 Outer robustness

Note that we select our *democratic* observations as a subset of all country-year-combinations that have the Polity-Score 10. In the period from 1981 to 2011 there is a total of 929 elements fulfilling this condition. The number of possible *non democratic* country-year-combinations, which have a polity score of  $-7$  or less, is similarly high.

$y$	100	200	300	500	750	1000
Mean	0.4457	0.4461	0.4457	0.4456	0.4453	0.4456
Gini	0.5011	0.5005	0.5011	0.5012	0.5014	0.5011
Corr	0.9999	1	0.9999	0.9999	0.9999	0.9999
MAD	0.0112	0	0.0046	0.0071	0.0081	0.0069

**Table 4** Influence of the number of outer iterations.

The first important question is whether the number of  $y$  is sufficiently high enough to exclude a selection bias. Table 4 shows the four control variables dependent upon the number of outer iterations. We perceive that — in comparison to the basic setup — a rise in  $y$  does not cause a significant variation of the democracy indicators.

## 5.2 Internal robustness

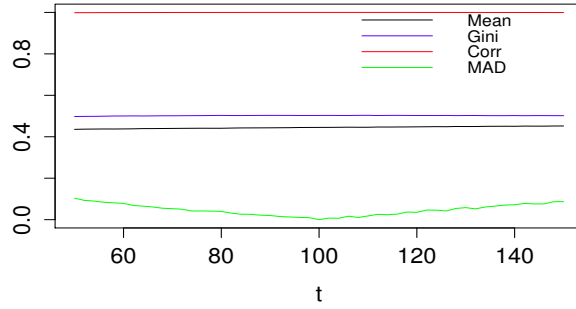
Now we investigate the number of iterations  $x$ .

$x$	500	1000	2500	5000	7500	10000
Mean	0.4462	0.4457	0.4451	0.4451	0.4454	0.4453
Gini	0.5019	0.5022	0.5028	0.5027	0.5026	0.5027
Corr	0.9999	0.9999	1	0.9999	0.9999	0.9999
MAD	0.0125	0.0053	0	0.0036	0.0045	0.0040

**Table 5** Influence of the number of internal iterations.

Table 5 shows that neither an increase nor a decrease in  $x$  influences the SVMDis in a significant way.

Furthermore, we have to examine the consequences of a change in the size  $t$  of the  $T$ -set. Similarly to the standard setup, we assume that the number of *democratic* (1) and *non democratic* (0) country-year-observations in the  $T$ -set is equal, i.e.  $t_0 = t_1 = \frac{t}{2}$ . Figure 3 displays the relation between the four control variables and the parameter  $t$  in the range from 50 to 150. On the one hand, it becomes apparent that variation in the mean, the Gini coefficient, and the correlation is very low. On the other hand, we perceive that several democratic indicators can change slightly if  $t$  rises or shrinks substantially. Yet the level of this worst case shift is lower than 0.1.



**Figure 3** Influence of the size of the  $T$ -set.

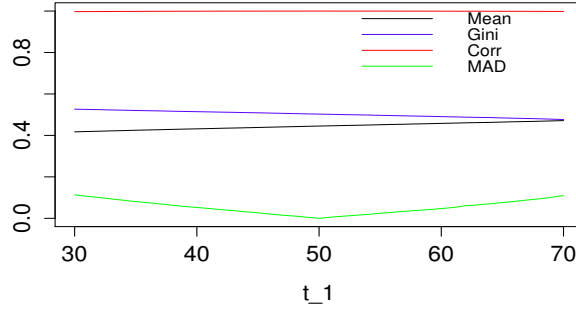
An obvious question would be how many country-year-combinations are affected by such a variation. To answer this, we exemplarily compare the basic setup and the alternative setup. Table 6 displays the distribution of the absolute derivations. We can observe that only a very small amount of the evaluated country-year-combinations is heavily affected by a huge shift in  $t$ . For instance, more than 95 percent of the observations have a deviation of less than 0.04.

<b>Derivation</b>	$[0; 0.02)$	$[0.02; 0.04)$	$[0.04; 0.06)$	$[0.06; 0.08)$	$[0.08; 0.1)$
<b>Observations</b>	4688	764	219	30	3

**Table 6** Distribution of the deviations.

The last aspect of our robustness test confirms the relation between the number of *democratic* and *non democratic* observations in the  $T$ -set. There-

fore we investigate the effects of a shift in  $t_1 \in [30; 70]$ , where  $t_0 = t - t_1$  and  $t = 100$  is fixed. Note that in our basic setup we assume that they are equal, i.e.  $t_1 = t_0 = 50$ . This assumption implies that the heterogeneity of both subsets is equal. Figure 4 provides the results of this analysis, which look similarly to the above. Indeed, the mean, the Gini-coefficient and the maximal absolute deviation vary more than in Figure 3, but this effect can only be recognized if  $t_1$  increases or decreases strongly. Furthermore, neither theoretical nor empirical reasons exist as to why we should suppose that the heterogeneity of the *democratic* country-year-combinations is much higher or lower than the heterogeneity of the *non democratic* observations. Quite the contrary, we might expect only small differences, and for these we observe very low variations in the democracy indicators. For instance, if we compare the case  $t_1 = 60$  to the basic setup, we obtain that the highest shift is less than 0.05, whereby approximately 70 percent of the observations have an absolute variation below 0.02.



**Figure 4** Influence of the distribution in the T-set.

## 6 Conclusions

While the mathematical tools of our method for measuring democracy are complex, the basic idea of the presented approach is very simple and intuitive. We select a certain set of country-year-combinations which can be easily labeled as *democratic* (1) and *non democratic* (0). Furthermore, we choose some variables which can be used to characterize a democracy. Based on these selections, the Support Vector Machines recognize the pattern in the chosen record. This information can be used to classify all

observations in the data set. We obtain continuous indicators in the range from 0 to 1, which can be interpreted intuitively as the probability that a certain country in a specific year is a democracy.

Based on the suggested implementation in which 188 countries can be evaluated in the period from 1981 to 2011, we receive very sensitive indexes.<sup>4</sup>

The Support Vector Machines Democracy Indexes (SVMDIs) can explicitly illustrate increases or decreases in the process of democratization and also reveal differences between established democracies, even if they are labeled as possible elements of the *R-set* (see Figure 2).

In comparison to other approaches there are basically three advantages of our method. First, our method can be used for every combination of variables without scaling and changes in the form of their aggregation. Therefore our method is a useful tool for use in comparing different concepts of democracy. Second, the aggregation of the variables does not rely on an arbitrary formula but is based upon an optimization model. Under the given assumptions this is the best way to combine the explanatory variables. Third, the estimated democracy indicators enable a more detailed description of the democratization process.

It is obvious that a numerical heuristic cannot work without some process assumptions and in our case we have to choose them without any theoretical foundation. Yet our sensitivity analysis implies that the method is quite robust to changes in the underlying parameters, i.e. the indicators do not vary significantly if we modify the parameter setup in a moderate manner. A second point of criticism may be the choice of the kernel function, which is not unique and can also be achieved by an empirical analysis (Borges, 1998). The Gaussian RBF kernel is standard in common literature; however, the selection of an alternative kernel function may yield different results. Another problem relates to the selection of the elements in the *R-set*, which is also not unique either. The choice is only based on the preferences of the operator, but the quality of the rules and their selection is tested during the estimation process (stage 4).

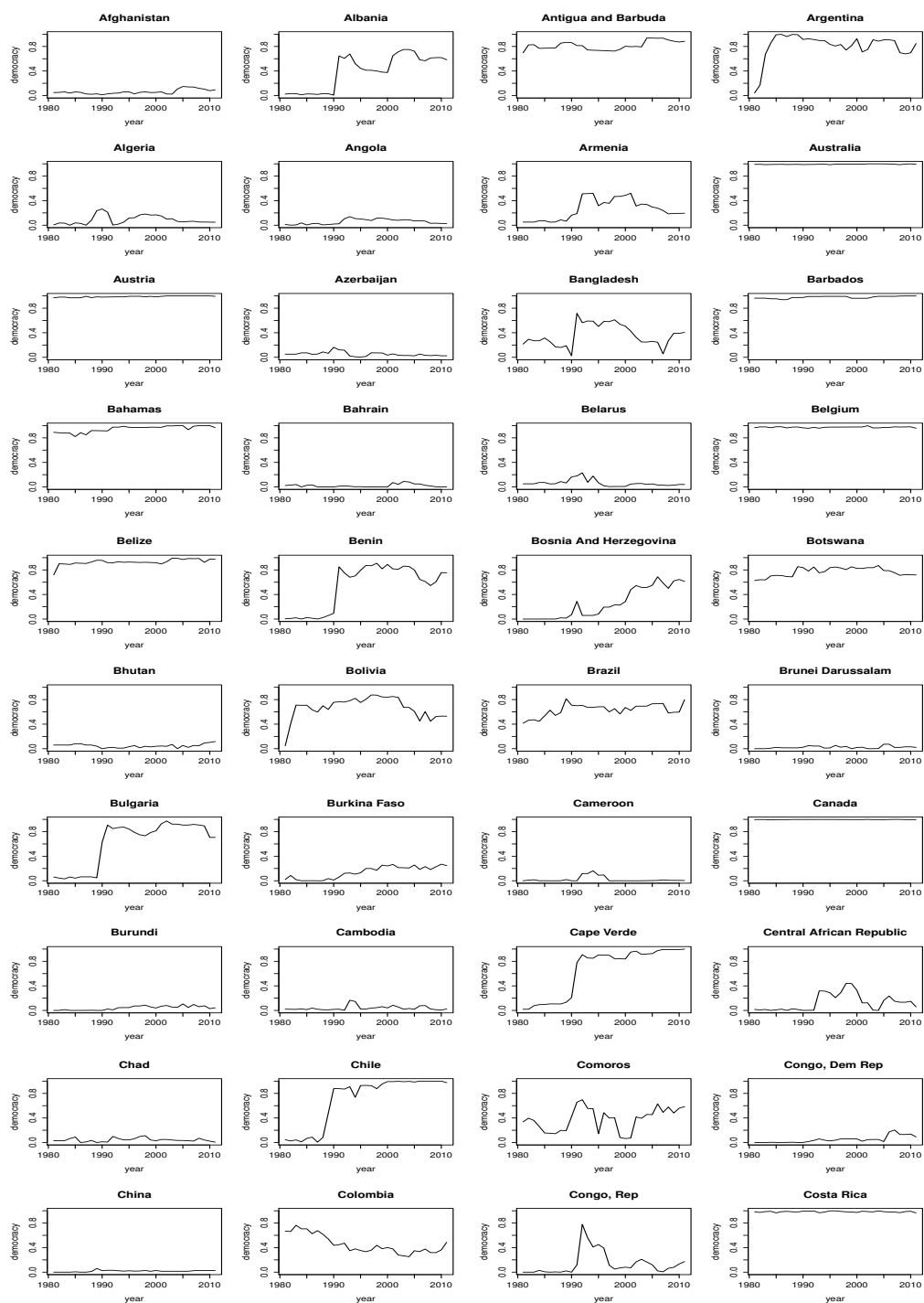
Nevertheless, the concerns are negligible in comparison to the advantages of the SVMDIs. Thus the presented method is a very useful alternative for measuring democracy.

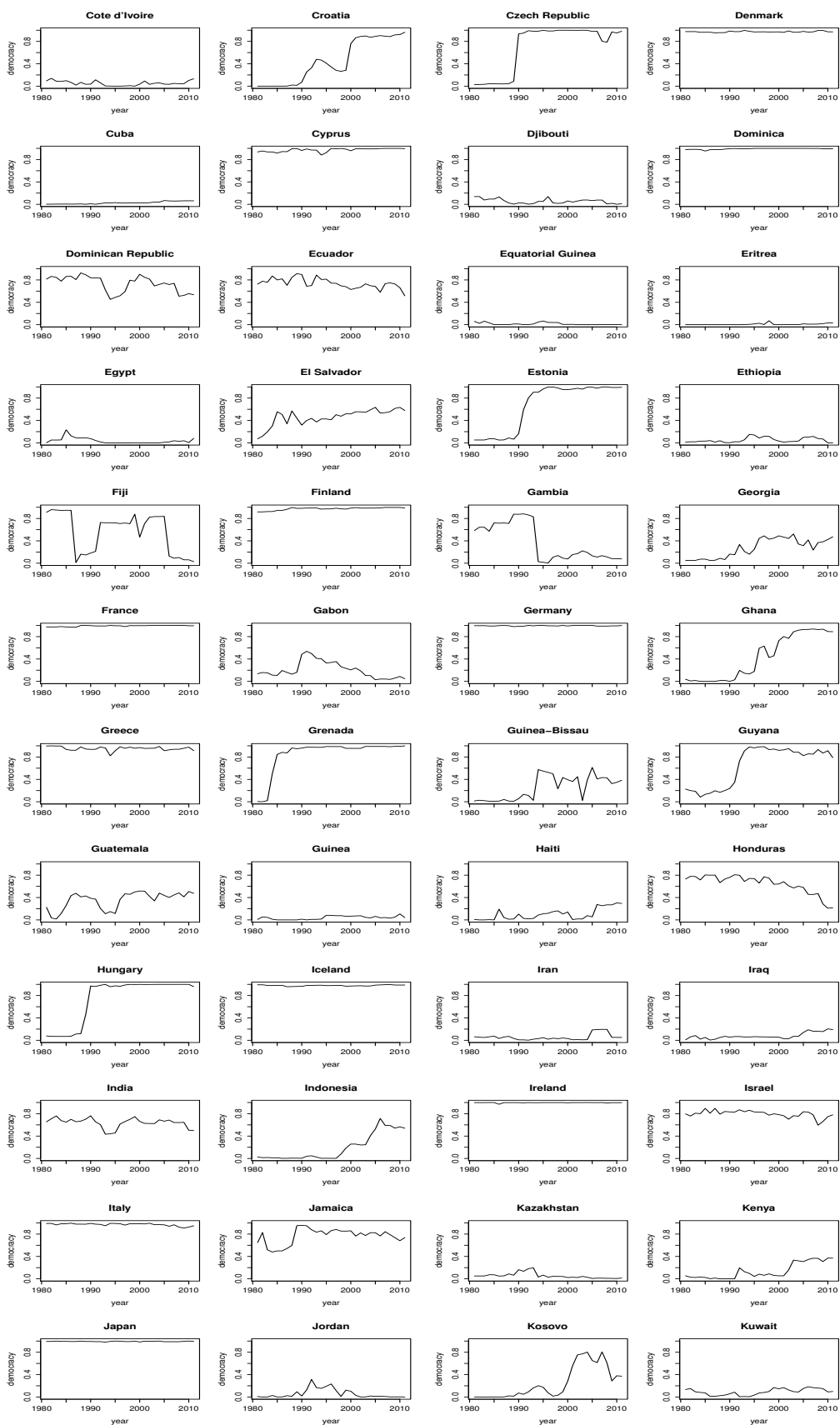
---

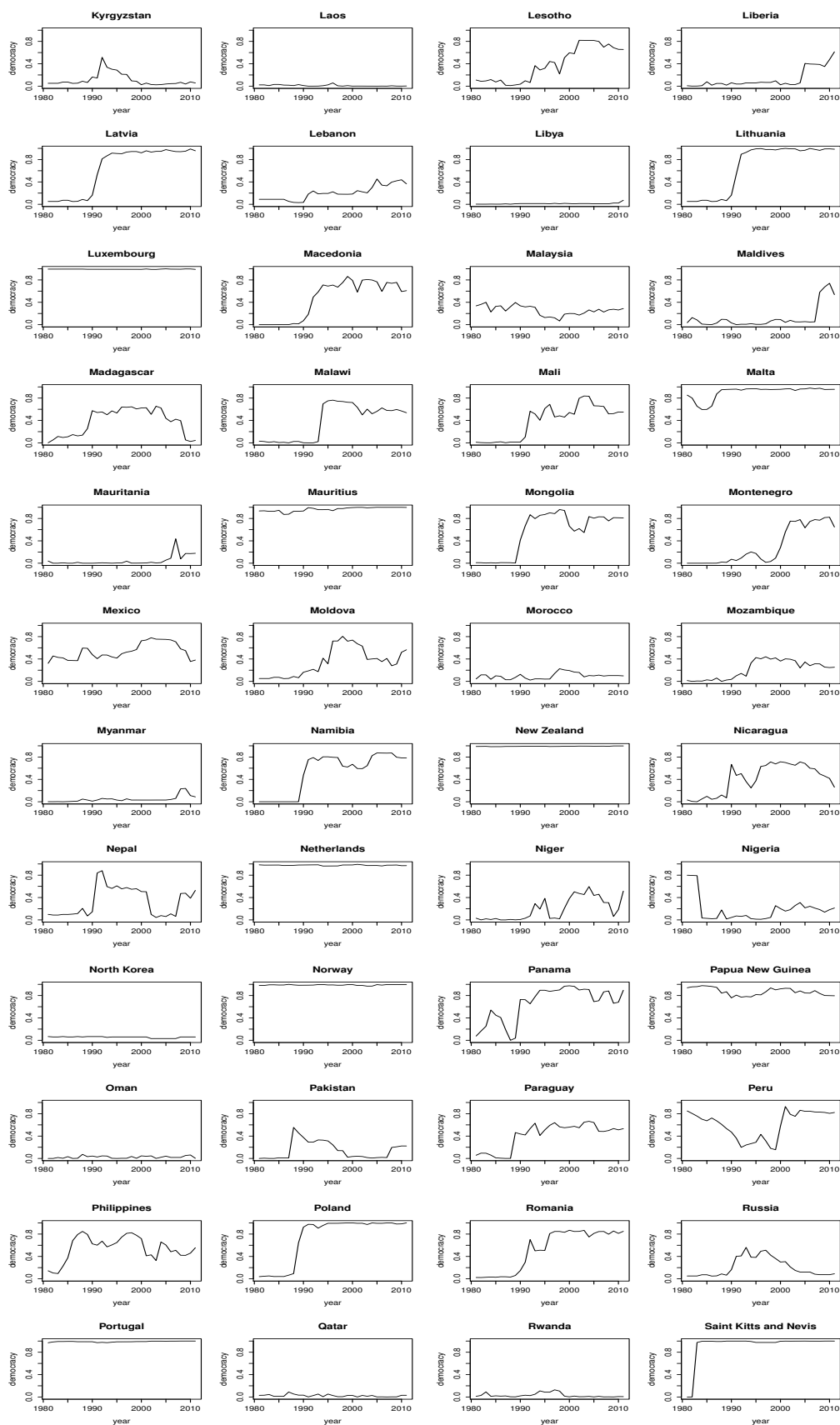
<sup>4</sup>All SVMDIs are available for download, see [http://www.wiwi.uni-wuerzburg.de/lehrstuhl/vwl4/data/svmdi\\_dataset/](http://www.wiwi.uni-wuerzburg.de/lehrstuhl/vwl4/data/svmdi_dataset/).

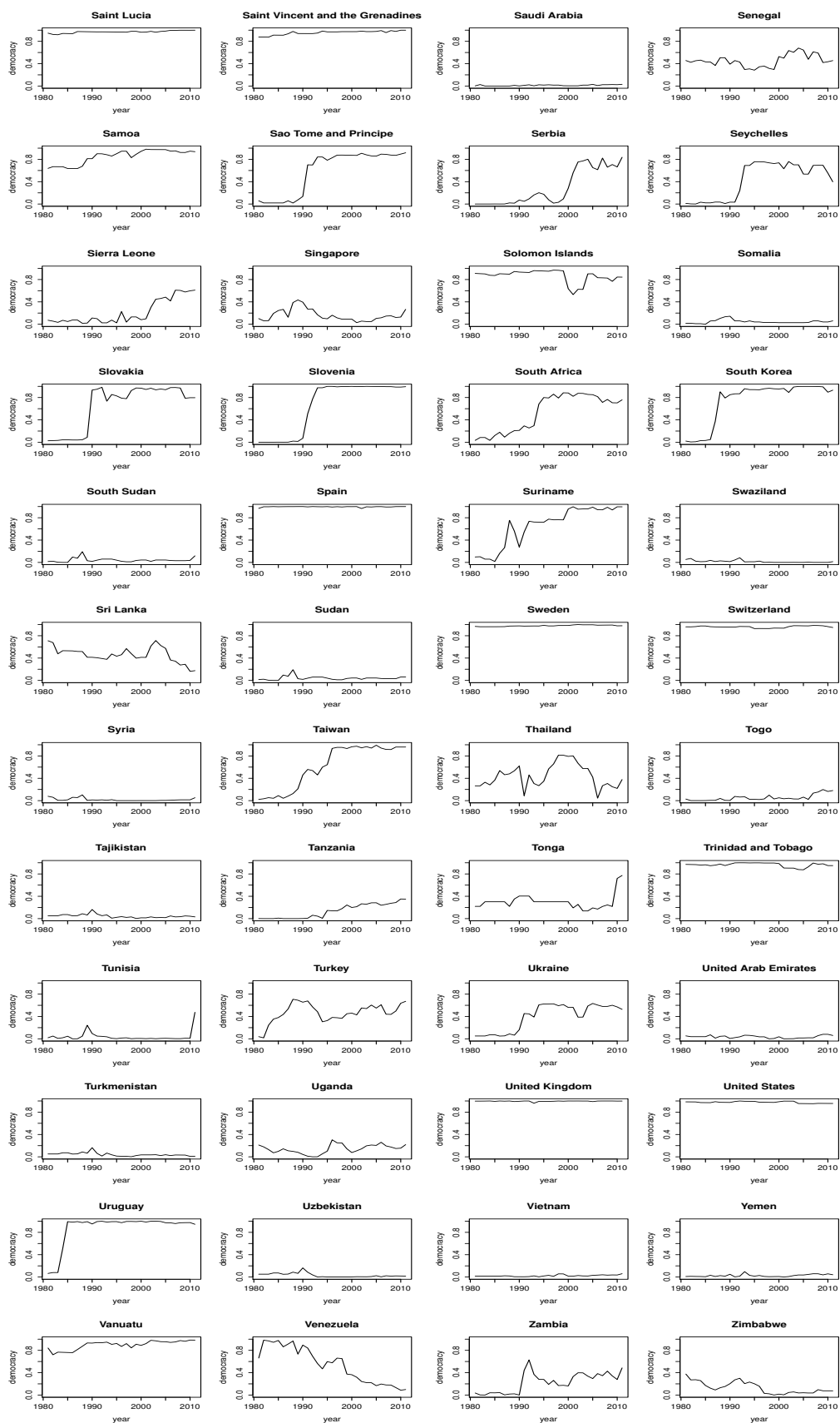


# A Individual trends









## B Support Vector Regression

The form of the optimal regression hyperplane is evaluated by solving the optimization problem

$$\begin{aligned} \max_{\alpha, \alpha^*} & -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(X_i, X_j) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i(\alpha_i - \alpha_i^*) \\ \text{s.t.} & \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \quad \text{and} \quad \alpha_i, \alpha_i^* \in [0, C], \end{aligned}$$

where  $C > 0$ ,  $\varepsilon > 0$  and  $k(\cdot, \cdot)$  is a kernel function. It is easy to show ([Smola and Schölkopf, 2004](#)) that under these assumptions the regression function has the form

$$f(X) = \sum_{i=1}^n (\alpha_i - \alpha_i^*)k(X_i, X) + b.$$

A more detailed description regarding the application of Support Vector Machines for function estimation can be found in [Smola and Schölkopf \(2004\)](#), especially with regard to the computation of the constant  $b$  and the conditions required for the kernel  $k(\cdot, \cdot)$ .

## References

- Acemoglu, D., Naidu, S., Restrepo, P., and Robinson, J. A. (2014). Democracy Does Cause Growth. NBER Working Papers 20004, National Bureau of Economic Research, Inc. 2
- Alvarez, M., Cheibub, J. A., Limongi, F., and Przeworski, A. (1996). Classifying political regimes. *Studies in Comparative International Development*, 31(2):3–36. 2
- Anderson, L. E. and Dodd, L. C. (2009). Nicaragua: progress amid regress? *Journal of Democracy*, 20(3):153–167. 8
- Boix, C., Miller, M., and Rosato, S. (2013). A complete data set of political regimes, 1800–2007. *Comparative Political Studies*, 46(12):1523–1554. 2
- Bollen, K. A. (1990). Political democracy: Conceptual and measurement traps. *Studies in Comparative International Development*, 25(1):7–24. 2
- Bühlmann, M., Merkel, W., Müller, L., Giebler, H., and Weßels, B. (2012). Demokratiebarometer: ein neues instrument zur messung von demokratiequalität. *Zeitschrift für vergleichende Politikwissenschaft*, 6(1):115–159. 2
- Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167. 4, 13
- Cheibub, J. A., Gandhi, J., and Vreeland, J. R. (2010). Democracy and dictatorship revisited. *Public Choice*, 143(1-2):67–101. 2
- Cingranelli, D., Richards, D., and Clay, K. C. (2014). *The CIRI Human Rights Dataset*. CIRI Human Rights Data Project. 6
- Dahl, R. A. (1971). Polyarchy: Participation and democracy. *New Haven*. 2
- EIU (2011). *Democracy index 2011: Democracy under stress*. Economist Intelligence Unit. 2
- Figueros, M. (1985). An assessment of overvoting in jamaica. *Social and Economic Studies*, pages 71–106. 8

- FreedomHouse (2014a). *Freedom in the World 2014: The Annual Survey of Political Rights and Civil Liberties*. Rowman & Littlefield. 2, 6
- FreedomHouse (2014b). Freedom of the press 2014. 6
- Gibney, M., Cornett, L., Wood, R., and Haschke, P. (2013). Political terror scale 1976-2012. 6
- Gualtieri, J. A. (2009). The support vector machine (svm) algorithm for supervised classification of hyperspectral remote sensing data. In Camps-Valls, G. and Bruzzone, L., editors, *Kernel Methods for Remote Sensing Data Analysis*. John Wiley, Chichester (UK). 3
- Gugiu, M. R. and Centellas, M. (2013). The democracy cluster classification index. *Political Analysis*, 21(3):334–349. 3
- Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3):389–422. 3
- Habermas, J. (1992). Faktizität und geltung: Beiträge zur diskurstheorie des rechts und des demokratischen rechtsstaates. *Frankfurt: Suhrkamp*. 2
- Lauth, H.-J. (2013). Kombiniertes index der demokratie (kid) 1996-2012. *Wuerzburg: Institut für Politikwissenschaft und Soziologie*. 3
- Locke, J. (1689 (1965)). *Two treatises of government*. Awnsham and John Churchill. 2
- Marshall, M. G., Gurr, T. R., and Jagers, K. (2014). Polity iv project — political regime characteristics and transitions, 1800-2013. *Center of Systemic Peace*. 2, 6
- McConnell, S. A. (2014). The 2011 presidential and legislative elections in nicaragua. *Electoral Studies*, 34:300–303. 8
- Montesquieu, C.-L. d. S. (1748 (1965)). Baron de la brède et de, vom geist der gesetze. *Reclam, Stuttgart*. 2

- Müller, T. and Pickel, S. (2007). Wie lässt sich demokratie am besten messen? zur konzeptqualität von demokratie-indizes. *Politische Vierteljahresschrift*, 48(3):511–539. 2
- Munck, G. L. and Verkuilen, J. (2002). Conceptualizing and measuring democracy evaluating alternative indices. *Comparative political studies*, 35(1):5–34. 2
- Pemstein, D., Meserve, S. A., and Melton, J. (2010). Democratic compromise: A latent variable analysis of ten measures of regime type. *Political Analysis*, 18(4):426–449. 3
- Rawls, J. (1971). *A Theory of Justice*. Cambridge (US). 2
- Saward, M. (1994). Democratic theory and indices of democratization. *SAGE MODERN POLITICS SERIES*, 36:6–24. 2
- Schölkopf, B., Simard, P., Smola, A. J., and Vapnik, V. (1998). Prior knowledge in support vector kernels. *Advances in neural information processing systems*, pages 640–646. 3
- Severinghaus, S. R. (2001). Mongolia in 2000 the pendulum swings again. *Asian Survey*, 41(1):61–70. 9
- Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222. 3, 18
- Steinwart, I. and Christmann, A. (2008). *Support vector machines*. 3
- Transparency-International (1995-2011). Index, corruption perceptions. URL: [http://http://www.transparency.org/cpi2014](http://www.transparency.org/cpi2014). 9
- Vanhanen, T. (2000). A new dataset for measuring democracy, 1810-1998. *Journal of Peace Research*, 37(2):251–265. 2, 6
- Vapnik, V. (1995). The nature of statical learning theory. 3
- Vapnik, V. and Chervonenkis, A. (1964). A note on one class of the algorithms of pattern recognition. *Automation and Remote Control*, 25(6):821–837. 3
- Vapnik, V. and Lerner, A. (1963). Generalized portrait method for pattern recognition. *Automation and Remote Control*, 24(6):774–780. 3