Control and Cybernetics

vol. **30** (2001) No. 4

Verification of ideological classifications – a statistical approach

by

Mariusz Mazurkiewicz[a),b)], **Jacek W. Mercik**[a),b)] **and Wiesław Dobrowolski**[a)]

[a)] Institute of Production Engineering and Management
Wrocław University of Technology, Wrocław, Poland
E-mail: mazurkiewicz@ioz.pwr.wroc.pl

[b)] Wrocław College of Management and Finance

Abstract: The paper presents a statistical method of verifying ideological classifications of votes. Parliamentary votes, preclassified by an expert (on a chosen subset), are verified at an assumed significance level by seeking the most likely match with the actual vote results. Classifications that do not meet the requirements defined are rejected. The results obtained can be applied in the ideological dimensioning algorithms, enabling ideological identification of dimensions obtained.

Keywords: ideological dimension, ideological space, bootstrap.

1. Introduction

One of the major problems in any parliamentary structure modelling method is the identification of ideological dimensions. One possible solution is to indicate an ideological dimension on the basis of classifications of appropriate votes carried out by an expert. The correct ideological classification of votes is of critical importance to the quality of the constructed model.

Regardless of the approach to ideological identification of parliamentary structure (e.g. the D-Nominate model: Poole, Rosenthal, 1985,1991; Linear Probability Model: Heckman, Snyder, 1997; frequency model: Mercik *et al.*, 1997; P-nominate procedure: Mercik, Mazurkiewicz, 1997, 2001), one can consider a special case, namely the analysis of a specified set of ideologically preclassified votes. This enables identification of the obtained dimensions.

The task of preliminary definition of a set of ideological votes consists in selecting from an available set of votes those in which voting "for" is identified

The early version of this paper was presented during the 1999 Annual Meeting of the

with taking an ideological stance on a particular issue. Usually this task is delegated to an expert. This raises a number of doubts relating to the quality of such classification and has a direct effect on the results of applied procedures for the ideological identification of parliamentary structure. An additional factor that makes it difficult to obtain correct results is the small (in the statistical sense) number of the available pre-classified votes. We propose to resolve this difficulty using the bootstrap technique.

2. Ideological dimension

An ideological dimension is not an exact notion – it has not obtained an unequivocal definition. In this paper we will try to define an ideological dimension within the framework of a mathematical model of parliamentary structure. However, we must additionally assume that such a model not only generates the positions of members of parliament in a certain ideological space but also associates with those positions probabilities of certain behaviour, namely of voting "for" or "against" in each of the considered votes.

Fundamental assumption:
If some group of votes is a group of identical votes in the ideological sense (e.g. left-right, pro-anti, etc.), then there exists two opposed groups of legislators who in the indicated votes usually vote respectively "for" and "against". In other words, there are two groups, whose members have definite views on issues relating to the analysed ideological dimension, or to put it still differently, the members of the group can evaluate the conformity of their views with the subject matter put to a vote and express their position in an unequivocal way by voting "for" or "against".

The above assumption regarding the existence of such two groups of legislators may be justified indirectly. Assume that it is not so. Then searching for or defining any ideological dichotomous dimensions makes no sense. The behaviour of legislators appears to be random and there is nothing in terms of a dichotomous division that we could call a parliamentary structure.

Assume that a preliminary classification of votes in terms of their ideological character has been performed, i.e. a certain subset of the set of all votes has been identified, which contains those votes that can be classified ideologically. For instance, if we define an ideological dimension based on supporting or opposing the government (a pro attitude and an anti attitude), we will consider votes on matters relating to government initiatives. In the case of a vote of confidence in the government, voting "for" is an instance of a pro attitude and voting "against" an instance of an anti attitude; whereas in the case of a vote of no confidence in the government, voting "for" is anti and voting "against" is pro (we consider only those votes in which a member of parliament took a specific

DEFINITION 1 *We say that a member of parliament voted "for" or "yes" in the ideological sense if in the given vote the stance taken ("yes" or "no") conforms to a pro attitude. Similarly, we say that a deputy voted "against" or "no" in the ideological sense if in the given vote the stance taken ("yes" or "no") conforms to an anti attitude.*

In accordance with the above definition, before the construction of a parliamentary structure model is undertaken, ideological votes must be appropriately ordered.

Let V be the set of serial numbers of all available votes taken during one fixed legislative period. Let L_i denote the i-th legislator, $i = 1, \ldots, N$, where N – number of members of parliament. For each legislator L_i, there will be an available random sample with the cardinality $card(V)$, whose elements are random variables with an appropriately encoded voting autcome.

Now, let $V_{id} \subset V$ denote the set of serial numbers of all votes pre-classified for a particular ideological dimension, $card(V_{id}) = I$, $I < I \leq card(V)$.

Let $P_{ij}(Yes)$, $i = 1, \ldots, N$, $j = 1, \ldots, I$ denote the probability that the i-th legislator vote "for" in the ideological sense in vote j of the set V_{id}. Define $P_{ij}(No)$ analogously.

Denote by G^p_{Yes} the group of legislators voting "yes" in votes of the set V_{id}, and define analogously, where $p \in (0, 1]$ is a fixed number describing the level of support for voting "yes" for the members of G^p_{Yes} and "no" for the members of G^p_{No}.

Let X_{ij} denote a random variable that is the result of the vote of the i-th legislator in the j-th vote of the set V_{id}. Let

$$X_{ij} = \begin{cases} 1 \text{ if legislator } L_i \text{ voted "for" in the ideological sense in the } j\text{-th vote} \\ 0 \text{ if legislator } L_i \text{ did not vote "for" in the ideological sense in} \\ \hfill \text{the } j\text{-th vote.} \end{cases}$$

Then, for each legislator L_i, and, in particular, for $L_i \in G^p_{Yes}$, we have at most $card(V_{id})$ vote results from V_{id}.

3. Rough estimation of the sets G^p_{Yes} and G^p_{No}

The fundamental assumption implies the existence of the sets G^p_{Yes} and G^p_{No} connected with a certain ideological dimension. However, the idea of ideological dimension itself is not precisely defined – the only existing information is carried by the set of pre-classified votes V_{id}. The set V_{id} is identified by an expert, it may contain misclassified votes, as well as votes inappropriate for a particular ideological dimension that they should represent. The definition below describes the elements of sets G^p_{Yes} and G^p_{No} taking into account the errors connected with description of an ideological dimension.

DEFINITION 2 *A legislator $L_i \in G^p_{Yes}$ iff $\exists V_0 \subset V_{id} : \forall i \in V_0 P_{ij}(Yes) \geq p;$*

The set V_{id} may include incorrectly classified votes. The number of votes the of set V_0 expresses the number of correctly classified votes. The set V_{id} may also include votes in which legislators did not act "reasonably" (e.g. in accordance with their ideological preferences) – the number of such votes is characterised by an arbitrary value of the parameter p (to be exact, by the probability $1-p$).

Vote	Stance taken	Expert opinion	Stance taken	Expert opinion
1	"Yes"	"For" in the ideological sense	"No"	"Against" in the ideological sense
2	"Yes"	Unclassified	"No"	Unclassified
3	"Yes"	"Against" in the ideological sense	"No"	"For" in the ideological sense

Table 1. Examples of preliminary classification of hypothetical votes in a given ideological sense

In practice, the value of the parameter p should be set in such a way as to satisfy the condition $G^p_{Yes} \neq \emptyset$ and $G^p_{No} \neq \emptyset$ for $V_0 \neq \emptyset$ (under constraint: $G^p_{Yes} \cap G^p_{No} = \emptyset$) and make p as big as possible. Given the weakness of the initial assumptions (we only have an unverified preliminary classification), it is not possible to get a better estimation of p.

In reality, there are only observations of random variables X_{ij} – the results of preclassified votes. Definition 2 gives us a probabilistic sense of the sets G^p_{Yes} and G^p_{No}. However, their existence is guaranteed by fundamental assumption, and establishment of G^p_{Yes} and G^p_{No} on the basis of this definition is not possible. One can estimate only the values of the characteristic functions of the sets G^p_{Yes} and G^p_{No} for each legislator L_i by statistical approaches. Then, the condition of belonging to G^p_{Yes} or to G^p_{No}, introduced in Definition 2, may be replaced by the following statistical procedure of estimation of the characteristic function:

We say that $L_i \in G^p_{Yes}$, *if there is no basis for rejecting the hypothesis* $H_0 : P(X_{ij} = 1) \geq p$ *at a chosen significance level* α.

We say that $L_i \in G^p_{No}$, *if there is no basis for rejecting the hypothesis* $H_0 : P(X_{ij} = 0) \geq p$ *at a chosen significance level* α.

Of course, such a choice of the sets G^p_{Yes} and G^p_{No} makes them dependent on the parameter α. As the preliminary classification of the set V_{id} has been carried out by an expert, it may contain misclassified votes – their existence determines the way in which legislators are assigned to G^p_{Yes} or to G^p_{No}. However, thanks to the method defined in the procedure above, we avoid an explicit determination of the set V_0.

For each legislator L_i, $i = 1, \ldots, N$, and, in particular, for $L_i \in G^p_{Yes}$, we probability $P(X_{ij} = 1)$ may be obtained upon a temporary assumption that in each vote in V_{id} the probabilities are identical, i.e. for a given legislator L_i,

$$\forall j \in V_{id} P(X_{ij} = 1) = P_i(Yes),$$

where $P_i(Yes)$ denotes the probability of voting "yes" for legislator L_i in each vote from V_{id}. Under this assumption the sum of random variables X_{ij}, denoted by $X_i = \sum_{j \in V_{id}} X_{ij}$, has a Bernoulli distribution with the parameters $P_i(Yes)$ and n_i for a given legislator L_i, where n_i is the number of votes in V_{id}, in which the legislator L_i voted "for" or "against" in the ideological sense. Using the most powerful test (a family with a monotone likelihood quotient), we can thus verify (Lehman, 1986) the hypothesis $H_0 : P_i(Yes) \geq p$ against the alternative hypothesis $H_i : P_i(Yes) < p$ at any significance level α. In a similar way, we may consider testing $H_0 : P_i(No) \geq p$ against $H_1 : P_i(No) < p$. A legislator L_i is qualified as a member of the G^p_{Yes} group if there is no basis for rejection hypothesis $H_0 : P_i(Yes) \geq p$. A legislator L_k is qualified as a member of G^p_{No} group if there is no basis for the rejection of hypothesis $H_0 : P_k(No) \geq p$. In the methodological sense we have estimated the composition of sets G^p_{Yes}, G^p_{No}.

4. The model of parliamentary structure

The model that should be used if one attempts to identify ideological dimensions is a model that satisfies two conditions. First, it is a model that defines the positions of legislators in a certain space (for example, in ideological space). Second, for each vote the model determines the probabilities of specific behaviour (result of vote) related to the position of points corresponding to each legislator in a certain space.

It means that for each legislator L_i, for $i = 1, \ldots, N$ and for each vote $v_0 \in V$ separately, the probability of taking stance "yes" is given by the equation:

$$P(\text{stance taken by } L_i = \text{"}yes\text{"}) = f(l_i, yes_{v_0}),$$

where l_i is a position in a certain space, connected with the legislator L_i, yes_{v_0} is a position of the result "yes" in the same certain space, connected to vote v_0, f is a given function (from the model). Analogously, the probability of result of vote "no" is given by the equation:

$$P(\text{stance taken by } L_i = \text{"}no\text{"}) = f(l_i, no_{v_0})$$

where l_i is a position associated with the legislator L_i, no_{v_0} is a position of result "no" connected to vote v_0, f is a given function (from the model). The estimation of the parameters of such a model should be based on the results of each vote. Then, if we use only votes in V_{id}, the result will be a certain ideological space.

If the space generated by a model is one-dimensional, the problem of ideoon a certain axis are at the same time their positions along an ideological dimension. In contrast, a situation where the generated space is mathematically not a one-dimensional does not lead to a clear identification of the space (e.g. Poole, Rosenthal, 1985; Hołubiec et al., 1997). To achieve that, it is necessary to carry out an ideological classification of each vote in a manner that does not raise doubts.

5. Verification of V_{id}

As the set V_{id} is identified by an expert, it may contain misclassified votes, as well as votes inappropriate for a particular ideological dimension that they should represent. Thus, it is necessary to determine when a vote in the set corresponds to the analysed ideological dimension. Classification of a given vote v_0 from the set V_{id} as classified rigorously according to a definite ideological dimension is based on the results of votes of legislators belonging to the groups denoted as G^p_{Yes} and G^p_{No} in this vote. Hence, during the first step one has to establish these sets using the schema described in the section, namely, the Rough estimation of the sets: G^p_{Yes} and G^p_{No}, where the level of p is arbitrarily chosen.

Using the estimates of the sets G^p_{Yes}, G^p_{No} we can evaluate each vote from the set V_{id} in the sense of fit to a given ideological dimension. The evaluation provides us with the basis for qualification of votes to the set of verified votes $\overline{V}_{id}$ according to the following algorithm (see Chart 1):

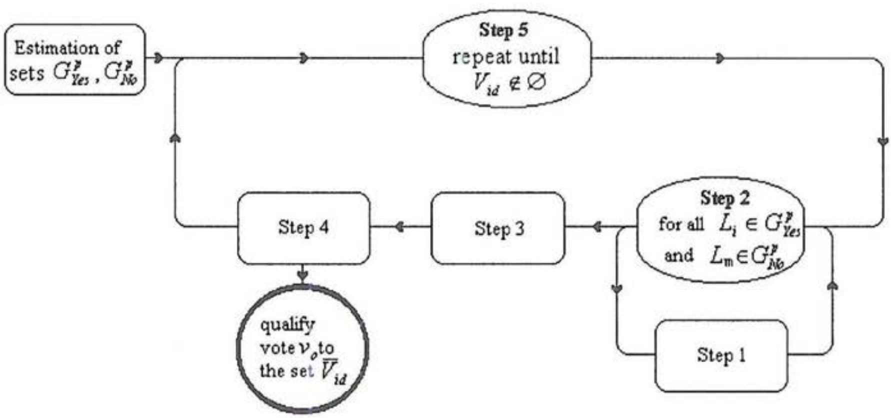

Chart 1: Scheme of the algorithm.

1. Let v_0 be a element of V_{id}. Then, using the estimated position $\widehat{l_i}$ of legislator L_i from the set G^p_{Yes}, it is possible to estimate $P_{ij}(Yes) =$

number of the legislator, j is the number of the vote v_0, and $\widehat{yes}_{v_0}$ is estimated position of result of vote v_0. Using the probability $P_{ij}(Yes)$, one can generate bootstrap vote results (Efron, 1979; Mazurkiewicz et al., 1998) in the following way: generate bootstrap observations of random variables $X_{i1}^{Yes}, X_{i2}^{Yes}, \ldots, X_{in}^{Yes}$ with a distribution defined by the estimates of probability function: $P(X_{ik}^{Yes} = 1) = P_{ij}(Yes)$, $P(X_{ik}^{Yes} = 0) = 1 - P_{ij}(Yes)$, $k = 1, 2, \ldots, n^*$, for vote v_0 number j and legislator $L_i \in G_{Yes}^p$, n^*-number of artificially generated observations. Using the estimated position $\widehat{l}_m$ of legislator L_m from the set G_{No}^p, it is possible to estimate $P_{ij}(No)$, where m is the number of the legislator, and j is the number of the vote v_0. Generate bootstrap observations of random variables $X_{m1}^{No}, X_{m2}^{No}, \ldots, X_{mn}^{No}$ with a distribution defined by the estimates of the probability function $P(X_{ms}^{No} = 1) = P_{mj}(No)$, $P(X_{ms}^{No} = 0) = 1 - P_{mj}(No)$, $s = 1, 2, \ldots, n^*$, for the vote number j and the legislator number m.

2. Reiterate Step 1 for each i such that $L_i \in G_{Yes}^p$ and for each m such that $L_m \in G_{No}^p$.
3. Combine the artificial observations (based on one vote results for vote v_0) obtained in Step 2 into one set denoted $S_{v_0}^*$.
4. Remove vote v_0 from V_{id}. Denote observations in $S_{v_0}^*$ by x_r^*, where $r = 1, \ldots, N^*$, $N^* = n^*(card(G_{Yes}^p) + card(G_{No}^p))$. Assuming that the elements of $S_{v_0}^*$ are observations of random variables $X_1^*, X_2^*, \ldots, X_N^*$ with an identical distribution defined by the estimates of the probability function $P(X_r^* = 1) = p_{v_0}$, $P(X_r^* = 0) = 1 - p_{v_0}$ for each ($p_{v_0} \in (0, 1]$), verify the hypothesis $H_0 : p_{v_0} \geq p$ against $H_1 : p_{v_0} < p$ at a chosen significance level α. If there is no basis for the rejection of H_0, qualify the vote v_0 to the set of verified votes $\overline{V}_{id}$.
5. Reiterate Steps 1–4 for the remaining votes in the set V_{id}.

The effect of the application of the algorithm is the set $\overline{V}_{id}$, which, when used again to build a model of parliamentary structure, will generate the proper ideological dimension.

6. Application of the V_{id} verification algorithm

A practical application of the algorithm described above will be presented using the example of a model of the structure of the Polish parliament. The model will be built using the P-nominate procedure (Mercik, Mazurkiewicz, 1997, 2001). The P-nominate procedure is based on the classification of votes in the ideological sense, usually carried out by experts. Correct classification makes it possible to obtain proper identifications of ideological dimensions – it is this classification that the quality of results obtained using the P-nominate method depends

The analysed classification of votes concerns three types of dimension: left-right (60 votes selected), pro-government-anti-government (36 votes), pro-European integration-anti-European integration (12 votes). The classification applies to the second term of the *Sejm* (lower house of Polish parliament), 1993–1997, which voted altogether about 8700 times. The classification (Hołubiec et al., 1997), carried out by an expert, involved selecting votes corresponding to each dimension and defining vote results in the ideological sense.

The classification verification algorithm is dependent on two additional parameters – the significance level α for the statistical tests used and the parameter p. A crucial role in the verification of a set of votes is played by the parameter p and its influence on the choice of elements, and especially the cardinality, of G^p_{Yes} and G^p_{No}. The exact effect of p on the cardinality of G^p_{Yes} and G^p_{No} in the analysed case of the Polish parliament is presented in Appendix A. It seems that the only possible technique for the selection of the value of p is to leave it to the user of a particular parliamentary structure model. Of course, one could try to develop heuristic techniques based on the cardinalities of G^p_{Yes} and G^p_{No} as functions of the argument p aiding the selection of that parameter – however, this possibility will not be investigated in this paper.

Appendix B contains charts showing the effect of p on the number of votes passing the verification procedure.

The effect of the parameter α- the significance level of the statistical test used in the estimation procedure of characteristic function for sets G^p_{Yes} and G^p_{No}, is presented in Appendix C. The charts show that the effect of the parameter α on the verification of the set of pre-classified votes is slight, even when significance levels vary considerably.

The obtained results of analysis of data from the Polish parliament of the second term confirm the correctness of the proposed algorithm. First, sets G^p_{Yes} and G^p_{No}, estimated roughly, are not empty for some $p > \frac{1}{2}$ and their cardinality for every analysed ideological dimension exceeds 20% of the number of legislators, which confirms also the basic assumption. The respective figures presented in appendix A suggest the continuity of percentage of numbers of legislators in sets G^p_{Yes} and G^p_{No} as a function of the parameter p, with only one exception of the dimension connected with the integration into the European Union. However, this case is very exceptional because pre-classification was done by an expert using only 12 votes, which is definitely statistically too little. The continuity of percentage of numbers of legislators in the sets G^p_{Yes} and G^p_{No} allows for a rational choice of the parameter p of probability of voting "yes" or "against" in the ideological sense.

In the process of verification itself one can see for every ideological dimension (Figs. 4, 5 and 6) the "jump" which describes the critical value of the parameter p, when the level of percentage of verified votes varies between 60% and 70% of pre-classified votes.

In the case of the left-right ideological dimension (Fig. 5) the verified votes for means that 30% (exactly 18 votes) were classified wrongly with the criteria used. In the case of the "pro-anti government" dimension for $p = 0.64$ the number of correctly classified votes exceeds slightly 97% (exactly only 3 votes are not conform with criteria). Instability may be observed again in the case of the "European Union" dimension – it is the result of small cardinality of the sample of pre-classified votes at disposal.

Finally, all the results allow us to say that the proposed algorithm may verify experts' opinions by objective statistical methods without another expert's help. This is very important for any ideological dimension modelling. Additionally, the use of verified expert's opinion significantly improves the credibility of any expert based model.

References

EFRON, B. (1979) Bootstrap Methods: Another Look at the Jacknife. *The Annals of Statistics* **7**, 1, 1–26.

HECKMAN, J.J. and SNYDER, J.M. JR. (1997) Linear Probability Models of the Demand for Attributes with Empirical Application to Estimating Preferences of Legislators. *Rand Journal of Economics*, **28**, 142–189.

HOŁUBIEC, J., MAŁKIEWICZ, A., MAZURKIEWICZ, M., MERCIK, J.W. and WAGNER, D. (1997) Identification of Ideological Dimensions under Fuzziness: the Case of Poland. In: J. Kacprzyk, H. Nurmi, M. Fedrizzi, eds., *Consensus under Fuzziness.* Kluwer Academic Publishers, Boston, 267–284.

LEHMANN E.L. (1986) *Testing Statistical Hypotheses.* Wiley, 2nd edition. New York.

MAZURKIEWICZ, M., MERCIK, J. and DOBROWOLSKI, W. (1998) *Bootstrap Technique in Expert Evaluation of Voting Dimensionality.* Paper presented at the European Public Choice Annual Meeting, Göteborg.

MERCIK, J.W., DOBROWOLSKI, W. and MAŁKIEWICZ, A. (1997) Program SeJM: Computer Aided Analysis of Group Decision Making. *Raporty Instytutu Organizacji i Zarzdzania Politechniki Wrocławskiej*, PRE 4.

MERCIK, J.W. and MAZURKIEWICZ, M. (1997) *Dimensional Analysis of Parliamentary Decisions.* Paper presented at MCDM Conference, Cape Town, January 6–10.

MERCIK, J.W. and MAZURKIEWICZ, M. (2001) Modelling the Behaviour of Polish Members of Parliament. *Operations Research and Decisions*, **2**, 77–90.

POOLE, K.T. and ROSENTHAL, H. (1985) A Spatial Model for Legislative Roll Call Analysis. *American Journal of Political Science*, **29**, 357–384.

POOLE, K.T. and ROSENTHAL, H. (1991) Patterns of Congressional Voting

Appendix A. Cardinalities of G^p_{Yes} and G^p_{No}

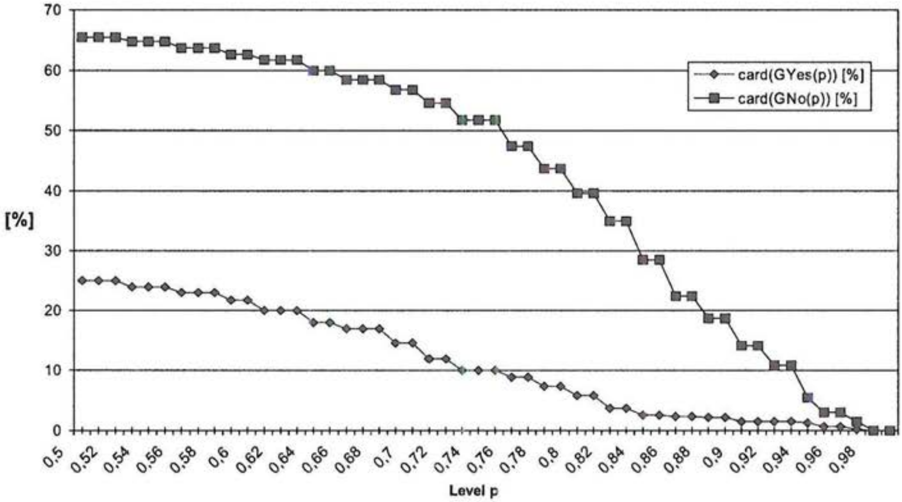

Figure 1. The pro-government-anti-government dimension; cardinalities of G^p_{Yes} and G^p_{No} as percentages of all parliamentarians, depending on the probability p.

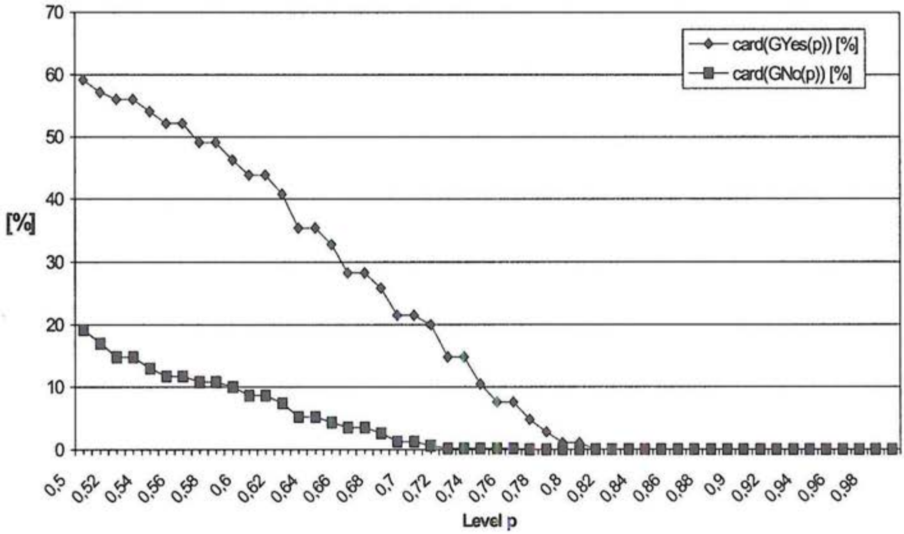

Figure 2. The left-right dimension; cardinalities of G^p_{Yes} and G^p_{No} as percentages of all

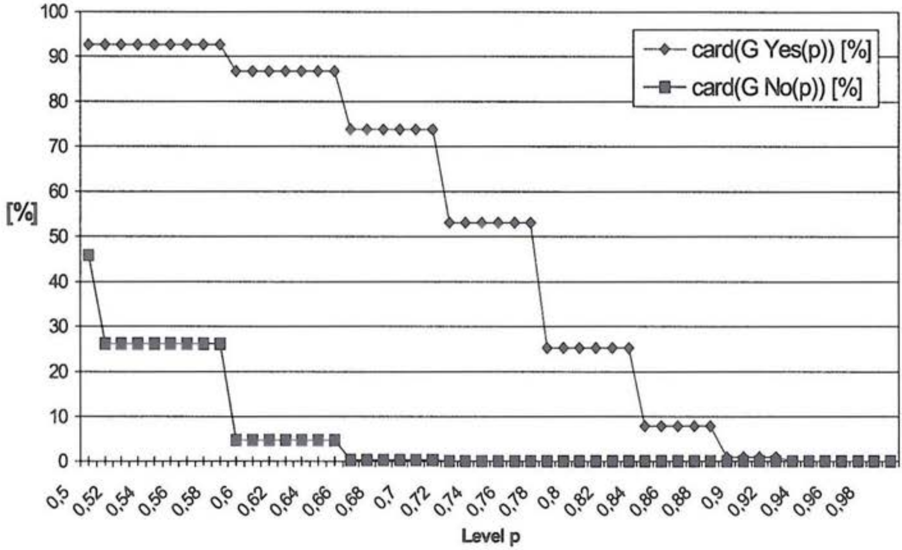

Figure 3. Attitude towards European integration; cardinalities of G^p_{Yes} and G^p_{No} as percentages of all parliamentarians, depending on the probability p.

Appendix B. Cardinalities of verified sets of votes

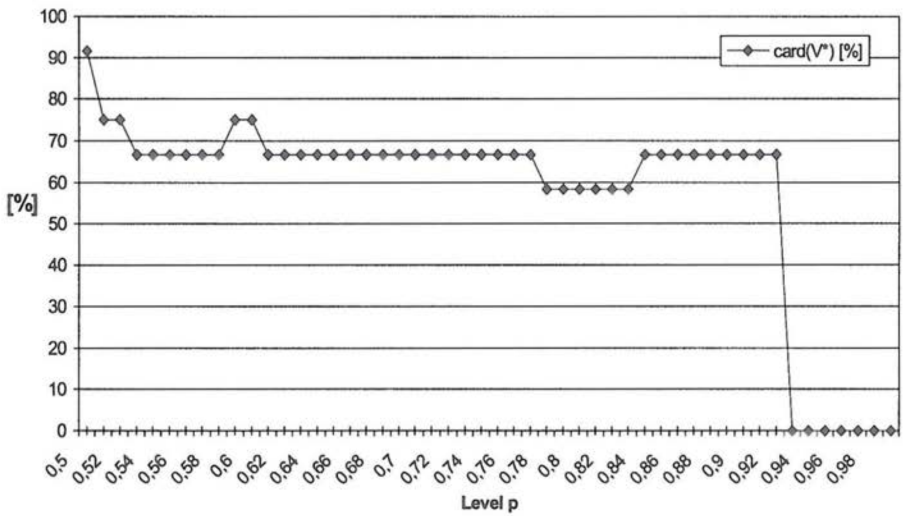

Figure 4. The European integration dimension; cardinality of the verified set of votes

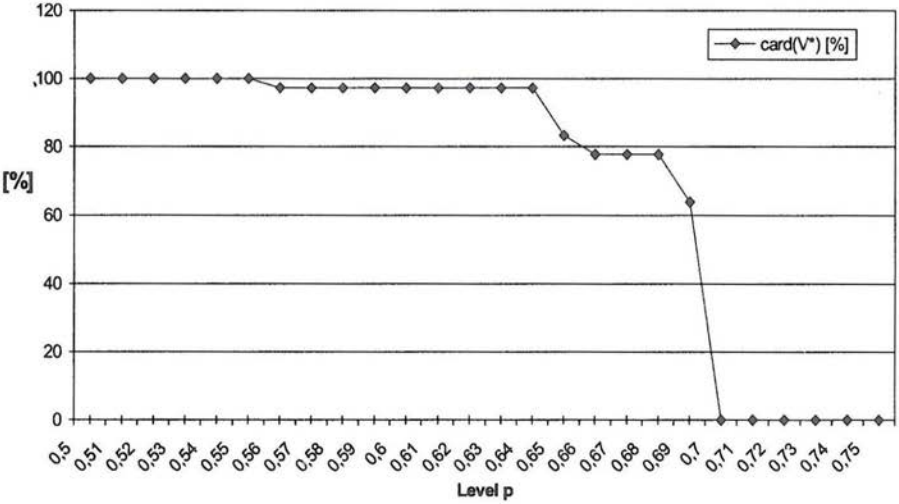

Figure 5. The pro-government dimension; cardinality of the verified set of votes as a percentage of the cadinality of V_{id}, $card(V_{id}) = 36$.

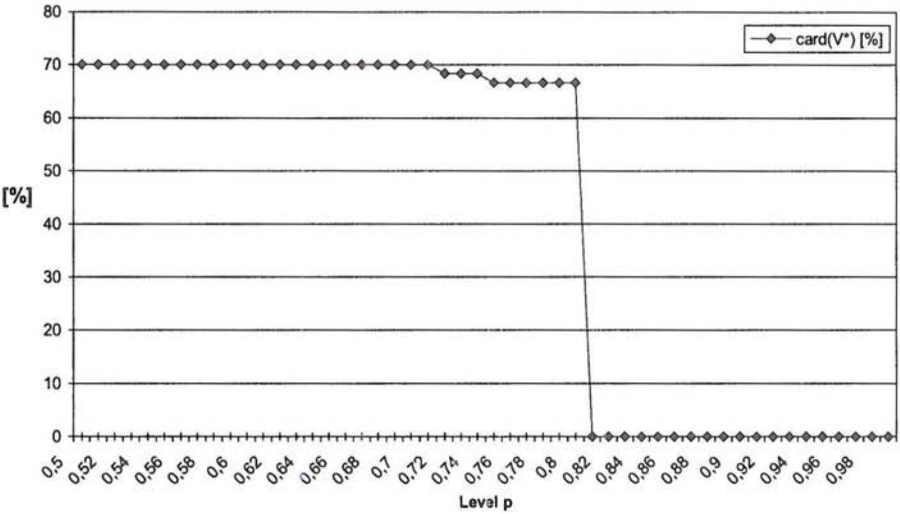

Figure 6. The left-right dimension; cardinality of the verified set of votes as a percent-

Appendix C.

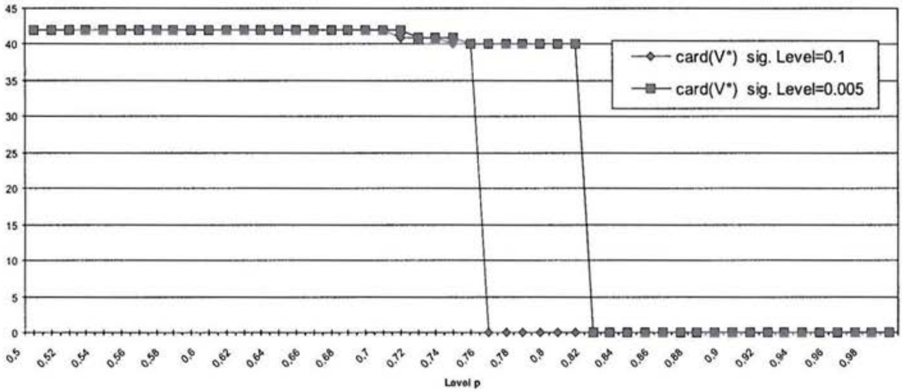

Figure 7. Cardinalities of verified sets of votes depending on the significance level – the left-right dimension.

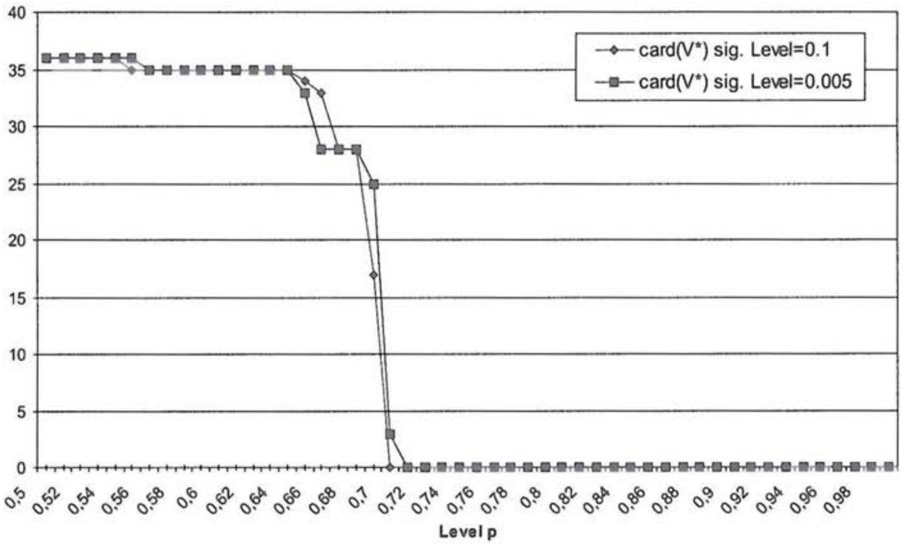

Figure 8. Cardinalities of verified sets of votes depending on the significance level – the pro-government-anti-government dimension.