*Non-Voters, Causal Model and Path Analysis**

J. M. A. M. JANSSENS

In April 1970 the Sociological Institute of the Catholic University of Tilburg held a survey among 2,500 people entitled to vote,[1] in order to determine political attitudes and behaviour. Since attendance at the polling-booths ceased to be obligatory shortly before the elections for the Provincial States (County Councils) in March 1970, it was possible to examine for the first time in the Netherlands those factors which influence the decision whether or not to vote.

In analysing the phenomenon of 'non-voting', which is new to the Netherlands, I have employed a causal model and path analysis in preference to a table analysis on account of the inadequacies inherent in this latter method.[2]

It is not only unfamiliarity with more advanced methods, but also the nature of their data which has led many sociologists up to this time to confine themselves to table analysis. Since causal models as developed by Blalock presume data at interval-level whereas the sociologist only rarely has such data at his disposal, he is usually obliged to analyse his data with the aid of tables. Boudon[3] has shown, however, that Blalock's models can also be utilized if the researcher possesses dichotomous data or can dichotomize his variables. In my opinion the advantage of this second method, namely the possibility of analyzing the simultaneous influence of a number of variables upon the phenomenon of non-voting, outweighs the disadvantage, i.e. the coarsening of the material resulting from the dichotomization.

In analyzing the simultaneous influence of a number of variables on the withholding of votes, I have tried to integrate two approaches used in

* This article is a resumé of an article in Sociale Wetenschappen, 14e jrg., no. 4, 1971, p. 243-268: Drs. J. M. A. M. Janssens, 'Niet-stemmers; causaal model en path-analyse'.

1 The response group consisted of 1838 respondents.

2 See, for these inadequacies: T. Hirschi and H. Selvin, *Delinquency Research* (New York 1967) pp. 164-168.

3 R. Boudon, 'A new look at correlational analysis', in H. Blalock and A. Blalock, *Methodology in social research* (New York 1968), p. 216.

American political theory to explain this phenomenon; on the one hand that which attempts to explain non-voting by studying the social-economic characteristics of those entitled to vote, and on the other that which seeks an explanation in certain mental attitudes and dispositions.[4] Both approaches, however, applied separately, fall short by reason of their one-sidedness. When, for instance, in the psychological approach, one succeeds in isolating particular attitudes which may influence a person not to vote, the question inevitably arises as to the factors determining these attitudes. One then searches for social-economic variables which may precede these attitudes. If, however, one takes social factors as one's starting point and discovers a relationship between these factors and failure to vote, the question arises of how such relationships are to be interpreted and which intermediate variables must be sought in order to explain the process whereby social factors can influence a failure to vote. We must then seek intermediate variables in the form of mental dispositions and attitudes, which interpret the relationship between the antecedent social factors and non-voting. It was this last consideration which formed our chief preoccupation in constructing the model. The model must therefore show three types of variables:

1. dependent variable (voted or didn't vote).
2. intermediate variables
   a. perception of political alternatives (ability to distinguish between the various political parties and to identify with a particular political party)
   b. political involvement
   (preoccupation with politics in general)
   c. perception of the importance of elections (being interested in and realizing the importance of elections).
3. Independent variables (age, education, sex, religion, church attendance, political interest of parents, parents' political party).

Since the variables incorporated in the model had to be dichotomized, a criterion for dichotomization once they had been selected was sought. It was decided to dichotomize them in such a way as to obtain the maximum contrast with regard to non-voting in the categories thus

[4] See for the social-economic approach: R. Lane, *Political Life* (New York 1959) pp. 48-49; S. Lipset, *Political Man* (London 1960), p. 184; for the psychological approach see: A. Campbell, P. Converse, W. Miller, D. Stokes, *The American Voter* (New York 1960).

TABLE I

Variable	Dichotomy	N.	% non-voters
1. age	a. born between 1936-49	622	36.66
	b. born before 1936 + no reply	1216	20.31
2. education	a. basic level + lower level	887	29.99
	b. exparded lower level up yo + including higher (university) level + don't knows + no reply	951	21.98
3. sex	a. male	854	28.92
	b. female	984	23.17
4. religion	a. Roman Catholic, Jewish, Humanist, other religions, no religion, don't know + no reply	1264	30.32
	b. Dutch Reformed, Calvinist, Greek-Orthodox	574	16.20
5. church attendance	a. seldom, never, no religion	879	32.99
	b. always, usually, sometimes	959	19.29
6. social class	a. working class	874	30.78
	b. middle class, upper class don't know + no reply	964	21.37
7. parents' interest	a. neither parent interested in politics	782	31.07
	b. one or both parents interested in politics	1056	21.97
8. parents' political party	a. respondent doesn't know or says that neither parent has a party preference	471	35.67
	b. respondent says that one or both parents has a party preference	1367	22.46
9. perception of political alternatives	a. scores 0.1 (slight)	399	43.11
	b. scores 2, 3, 4 (high)	1439	21.06
10. political involvement	a. scores 0, 1, 2 (slight)	764	37.30
	b. scores 3, 4, 5 (high)	1074	17.69
11. perception of importance of elections	a. scores 0, 1, 2 (slight)	573	48.17
	b. scores 3, 4 (high)	1265	15.73

obtained. A yardstick for this maximum contrast is the percentage of explained variance found per variable in the first step of the contrast group analysis.[5] In order to determine in which dichotomy this per-

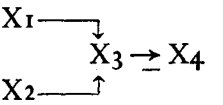
[5] For the contrast group method see: J. H. G. Segers, Ph. C. Stouthard, 'Analyse door middel van opeenvolgende contrasterende groupen' (Analysis by means of successive contrasting groups), *Sociale Wetenschappen*, vol. VI, 1963, no. 3, p. 221f.

centage is highest, this percentage is calculated per variable for all possible divisions within that variable. That dichotomy within each independent variable is then selected which has the maximum percentage of explained variance with regard to non-voting.

Table 1 gives the dichotomies obtained using this criterion. In addition the absolute number of respondents and the percentage of non-voters among them is giving for each category.

Having dichotomized the variables and calculated the correlation coefficients between the variables, Blalock's method can then be applied.[6] Blalock regards a causal model as adequate providing the (partial) correlation coefficients between variables which are not directly linked in the model by a causal arrow are equal to or approach zero. However, the strict application of this rule gives rise to two difficulties. In the first place Blalock gives no objective criterion for determining whether the value of a partial correlation coefficient diverges significantly from zero. One researcher may postulate that a partial correlation coefficient of, for example, 0.05 is too high, whereas another may regard this divergence as the result of mistakes made in measuring and base the adequacy of his model upon it. Subjective reasoning such as this, however, leaves the researcher free to choose whichever interpretation fits in best with his theoretical model. Yet there is a method of determining whether a partial correlation coefficient diverges significantly from zero.[7] To do this the value of the partial correlation coefficient is first transformed into the corresponding Z value. One then calculates $z = Z(\sqrt{N - 3 - K}$, wherein N indicates the number of respondents and K the number of variables kept constant. If then the computed z value in the normal standard distribution has a value which diverges significantly from zero on the chosen α-level, the hypothesis that the partial correlation coefficient does not diverge significantly from zero must be rejected.

A second difficulty in applying Blalock's method arises when correlations exist between the independent variables. In order to determine whether the following model is adequate,



[6] For Blalock's method see: H. Blalock, *Causal inferences in non-experimental research*, (Chapel Hill 1964).

[7] W. L. Hays, *Statistics for psychologists* (New York 1966), p. 576.

not only $r_{14.23}$ and $r_{24.13}$ must equal zero, but r_{12} as well. Assuming that r_{12} diverges significantly from zero and that the two partial correlation coefficients do not, then according to Blalock this model must be regarded as inadequate and a causal arrow will have to be drawn between X_1 and X_2. Strictly speaking, any researcher working with a number of independent variables between which correlations exist and which causally precede intermediate and dependent variables, will not only have to determine causal relationships between these latter variables but also those between his independent variables if he wishes to employ Blalock's method.

I am confronted with the same situation in my own model. I can indeed postulate the causal relationships between the intermediate and the dependent variables, but I am unable to specify the causal relationships, between the independent variables. Moreover, these relationships hold no interest for me. The only solution then is to regard a model as adequate when the partial correlation coefficients – between the independent and the intermediate variables, between the independent and the dependent, between the intermediate variables themselves and between these and the dependent variables – which correspond to arrows omitted in the model, do not diverge significantly from zero. This does not mean, however, that the correlations between the independent variables are neglected since they still have a role to play in calculating the partial correlation coefficients referred to.

In addition to these two difficulties, a third problem with regard to Blalock's method consists in the fact that Blalock assumes that the researcher is already sufficiently certain of the causal arrows to be drawn, that he is aware of the effects proceeding directly from one variable to the other and those running indirectly via other variables. This too is not the case in my model. It is, for example, not clear to me whether education influences non-voting directly or via one or more of the intermediate variables. To draw a number of arrows purely by intuition from the independent variables to the other variables seems to me an extremely haphazard solution. On the one hand one is obliged to test whether the partial correlation coefficients between variables which are not linked directly by a causal arrow diverge significantly from zero, while on the other one has no certainty at all that the arrows one has drawn are in the right place since this is not touched upon in the Blalock procedure. This problem led me to modify Blalock's method on yet a third point. I did this in such a way that the emphasis no longer lies on testing a model for adequacy but on its construction.

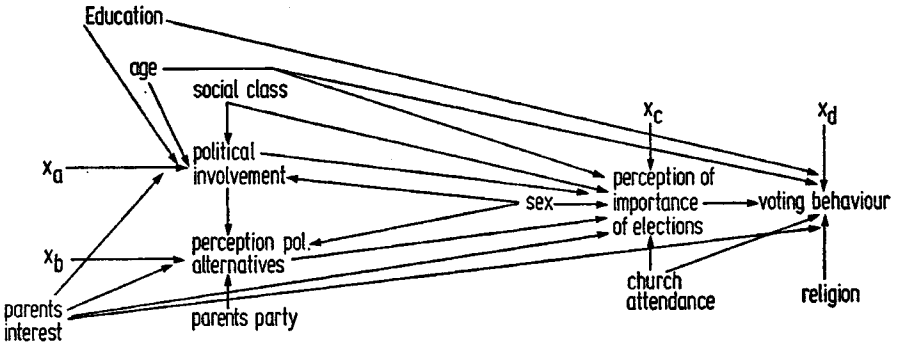
In constructing a model I start by assuming that a number of independent variables take causal precedence over a number of intermediate variables which in their turn precede the dependent. In addition the causal sequence of the intermediate variables must be postulated. This means, in relation to my model, that I assume the the eight independent variables to take causal precedence over political involvement. Causally, this political involvement precedes the perception of political alternatives which in its turn comes before the perception of the importance of elections. Finally the last variable, in time, is the dependent variable, non-voting. Once the causal priorities have been specified, one calculates for each independent variable the partial correlation coefficient with each intermediate variable, keeping constant the remaining independent variables and those intermediate variables which precede the relevant intermediate variable. Next one calculates the partial correlation-coefficients between each independent variable and the dependent variable, whereby all the remaining variables are kept constant since they precede the dependent variable in time. After this one calculates the partial correlation-coefficients between each intervening variable with each of the other intervening variables, keeping constant all independent variables and those intervening variables which take causal precedence over the relevant intermediate variable which is last in time. Finally the partial correlation coefficients are calculated between each intervening variable and the dependent variable, keeping constant all remaining variables.

When these partial correlation coefficients have been calculated, one then determines, with the aid of the procedure already indicated, whether they deviate significantly from zero. One then links by arrows those variables between which the partial correlation-coefficient diverges significantly from zero.

Having thus modified Blalock's model I am now in a position to specify the model (schematically) and to apply to it the principles of path analysis. The aim of this analysis is to measure the direct effects of a number of independent variables upon one or more dependent variables.[8]

In my model we can distinguish successively four dependent variables:

[8] For the path-analysis see: K. C. Land, 'Principles of path-analysis' and D. R. Heise, 'Problems in path-analysis and causal inference' in E. F. Borgatta, *Sociological Methodology* (San Francisco 1969); PH. C. Stouthard and H. Wassenberg, 'Analyse door middel van pijldiagrammen' (Analysis by means of arrow diagrams), *Sociale Wetenschappen*, vol. XIV, no. 1, 1971, pp. 46-64.



political involvement, perception of political alternatives, perception of the importance of elections and non-voting. It is now possible to start with each of these in turn and to measure the effects of those variables which are directly linked, in the model, with the variable in question. The results of this working method are given in table III. In addition to the value of the path-coefficient, this table also shows the values of the residual path-coefficients and the declared proportion of variation.

TABLE III. *Matrix of Path-coefficients*

	voting behaviour	perception importance of elections	perception of political alternatives	political involvement
X_2 perception importance of elections	0.4553			
X_3 perception of political alternatives		0.2501		
X_1 political involvement		0.5160	0.3311	
X_5 age	0.3075	0.0672		0.1376
X_6 church attendance	0.1383	0.0922		
X_7 religion	0.1700			
X_8 social class		0.0438		0.1205
X_9 education	0.1309			0.2102
X_{10} sex		—0.0814	0.0912	0.3131
X_{11} political interest parents	—0.0431	0.0873	0.0883	0.3338
X_{12} political party parents			0.3105	
proportion declared variation	0.4310	0.5255	0.3550	0.3658
residual path-coefficient	0.7543	0.6888	0.8031	0.7963

The following therefore appear to have a direct bearing up on failure to vote:

1. perception of importance of elections (0.4553)
2. age (0.3075)
3. religion (0.1700)
4. church attendance (0.1383)
5. education (0.1309)
6. political interest of parents (—0.0431).

In the path-analysis of non-voting, 43,10% of declared variation was accounted for, a percentage which may be termed high in view of the fact that only three intervening variables are included in the model, while others, such as citizen duty and political efficacy could not be incorporated.