

Comments and Reviews

Søren Risbjerg Thomsen, *Danish Elections 1920–79. A Logit Approach to Ecological Analysis and Inference*

Århus: Politica, 1987, 210 pp.

Ecological inference has a long-standing tradition in political science. It addresses itself to the same substantive questions that provide the rationale for one election survey after another: estimating the floating vote and the party preferences of the social classes without the benefit of individual level data. The transition propensities and the party preferences are estimated or inferred from the aggregate-level ecological data provided by the official statistical production.

In a sense, the ecological inference techniques must overcome the obvious limitations of data matrix 1a and somehow generate the complete information which is postulated in Figure 1b.

Fig. 1a
Only ecological data

		R ₁
		R ₂
S ₁	S ₂	1,0

Fig. 1b
Complete data

P ₁₁	P ₁₂	R ₁
P ₂₁	P ₂₂	R ₂
S ₁	S ₂	1,0

The small group of social scientists who were not entirely discouraged by Robinson's warnings against the indiscriminate use of ecological data (Robinson 1950) generally advocated regression analysis as a bridge between the two levels of analysis. The models at hand were generally cast in terms of the internal of the two-by-two contingency table. It has the desirable property that one estimate is enough to fill out all the empty

Comments and Reviews

Søren Risbjerg Thomsen, *Danish Elections 1920–79. A Logit Approach to Ecological Analysis and Inference*

Århus: Politica, 1987, 210 pp.

Ecological inference has a long-standing tradition in political science. It addresses itself to the same substantive questions that provide the rationale for one election survey after another: estimating the floating vote and the party preferences of the social classes without the benefit of individual level data. The transition propensities and the party preferences are estimated or inferred from the aggregate-level ecological data provided by the official statistical production.

In a sense, the ecological inference techniques must overcome the obvious limitations of data matrix 1a and somehow generate the complete information which is postulated in Figure 1b.

Fig. 1a
Only ecological data

		R ₁
		R ₂
S ₁	S ₂	1,0

Fig. 1b
Complete data

P ₁₁	P ₁₂	R ₁
P ₂₁	P ₂₂	R ₂
S ₁	S ₂	1,0

The small group of social scientists who were not entirely discouraged by Robinson's warnings against the indiscriminate use of ecological data (Robinson 1950) generally advocated regression analysis as a bridge between the two levels of analysis. The models at hand were generally cast in terms of the internal of the two-by-two contingency table. It has the desirable property that one estimate is enough to fill out all the empty

spaces in the matrix; and, as suggested by Stokes (1969), the initial model could look like this:

$$S_1 = P_{21} + (P_{11} - P_{21})R_1$$

or

$$y = a + bx$$

S_1 , which serves as the dependent variable (y) in the regression analysis, may be thought of as the proportion of socialist votes at the regional or local level; and R_1 , which serves as the independent variable (x) in the regression, may stand for the proportion of workers at the same level of analysis. The two remaining terms in the regression equation, the constant a (P_{21}) and the coefficient b ($P_{11} - P_{21}$) yield the estimates we need in order to fill out the missing values in the matrix for the entire nation.

This is a neat and straightforward method of inference. Unfortunately, however, it works only under very special circumstances. The problem is that the model accounts only for individual effects. The 'true' cell entries in the regional or local data matrices must not deviate from one another by more than a random factor. To the extent that they do, there are regional variations in the data, which is another way of saying that the original, individual effects only, model is improperly specified. As a result it will produce unreasonable estimates (read: percentages smaller than 0 or greater than 100) which won't go away unless the model is redefined so as to include various groups and/or interaction effects. And that simply cannot be done unless we have access to other data than ecological data only.

The problems posed by regional variations have been dealt with in many ways (Stokes 1969, Shiveley 1969), sometimes with less successful results (Lewin 1972). Risbjerg Thomsen solves the problem by making his estimates of the national averages on the basis of a simple aggregation of a series of region specific estimates within *politically homogeneous regions* in Denmark. His approach, however, is not that of ecological regression but of ecological correlation. Unlike Robinson, though, Risbjerg Thomsen does not rely on the product-moment correlation (R_{xy}) but on the tetrachoric correlation which he sees as a superior alternative for the discrete, binary variables at the individual level. On the theoretical level, this neutralizes Robinson's warning against ecological correlations, and when put to an empirical test the argument proves devastating for Robinson's argument (Risbjerg Thomsen 1987, ch. 3, sect. 2.3).

Rasch's (1968) latent structure theory serves as Risbjerg Thomsen's theoretical starting point. As a consequence, the electoral behavior is seen as a function of the voters' position on a number of latent dimensions on both levels of analysis and not as a product of their previous voting behavior

or their socioeconomic status as in the case of ecological regression with respect to transition propensities or party preferences. The algorithm, which yields the region specific estimates of the individual voting behavior (the floating propensities and the political preferences of the social classes), requires a simultaneous solution of a series of binary choices where each party 'runs' against a reference party (read: the non-voting party) and each class against a reference class (read: the old-age pensioners). Like all statistical techniques, it pivots on a number of philosophical assumptions. The most important of them is the assumption of isomorphy which requires the ratio between individual and ecological variances to be constant for all dimensions of the latent variables. This constant is also assumed to be very large.

The method is validated by comparing its results with the estimates obtained with the aid of survey data. It would have been preferable if the comparison had been performed of all those election years for which survey data are available, but there is also a point to be made for Risbjerg Thomsen's decision to focus on the one election year (1973) when two national surveys were carried out. When related to them, the ecological estimates are found to be well within the error margin of survey data, and it is hard not to be impressed by the power of this method of ecological inference.

In my opinion, Søren Risbjerg Thomsen's method for ecological inference is the most promising development in ecological data research over the last thirty years. The compliance of the data with the underlying assumptions of the model could have been explored at greater length. The impact of region and of the reference parties or classes might have called for greater attention; the same can be said about the biases attributed to the rural/urban dimensions. But these are questions to which there will be ample opportunity to return in the future, and in the light of the methodological breakthrough which the dissertation represents, they are objections of minor importance.

Sten Berglund, University of Helsinki

REFERENCES

- Lewin, L. 1972. *The Swedish Electorate 1887-1968*. Stockholm: Almqvist & Wiksell.
- Rasch, G. 1968. 'A Mathematical Theory of Objectivity and its Consequences for Model Construction', Paper, European Meeting on Statistics, Econometrics and Management Science, Amsterdam.
- Risbjerg Thomsen, S. 1987. *Danish Elections 1920-79, A Logit Approach to Ecological Analysis and Inference*. Århus: Politica.
- Robinson, W. S. 1950. 'Ecological Correlations and the Behavior of Individuals', *The American Sociological Review* 15, 351-357.

- Shively, W. P. 1969. "'Ecological" Inference: The Use of Aggregate Data to Study Individuals', *The American Political Science Review* 63, 1183–1196.
- Stokes, D. E. 1969. 'Cross-Level Inference as a Game against Nature', pp. 62–83 in Bernd, J. L., ed., *Mathematical Applications in Political Science, IV*. Charlottesville: University Press of Virginia.