Cross-Classification Analysis in the Field of Management and Organization: Comments on the DEL-Technique

Robert A.W. Kok
Theo J.B.M. Postma
A.G.M. Steerneman

Som-theme B Innovation and interaction

Abstract

The DEL-technique, a proportionate reduction in error measure, developed by Hildebrand, Laing and Rosenthal, has been applied and portrayed as a promising prediction analysis technique to evaluate theory on the basis of cross-classification data, though it was controversial at its birth in the early 70s. According to the opponents, Goodman and Kruskal, the interpretation of DEL as a proportionate reduction in error measure of knowing a prediction rule over not knowing the prediction rule, cannot be held, because it is benchmarked against independence instead of ignorance. However, even when neglecting this criticism the DEL-measure can be easily misinterpreted as a measure of acceptance of the specified customized hypothesis as the only and best relationship between two categorical variables, when the context for the interpretation is not carefully stated in terms of the adhered research paradigm: theory-testing versus prediction logic. When taking into account this criticism, the researchers need to be acknowledged for clearly addressing some of the methodological problems in prediction research, however, an alternative proportionate reduction in error measure may generate unequivocally interpretable results and outperforms the DEL-technique.

We would like to thank José Meppelink for her assistance in conducting experiments and thoughtful suggestions in carrying out the research.

Introduction

Drazin and Kazanjian (1993) promote the use of the DEL-technique as a suitable technique for studying relationships between categorical variables in the field of management and organization. The DEL-technique has been proposed for analyzing bivariate and multivariate categorical data in which a dependent variable's state is predicted from an independent variable's state. By applying the DEL-technique and an appropriate weighting scheme the strength of a relationship between two variables X and Y is considered, taking into account the different states (categories) of the variables. We assume that X has the states $x_1,\ldots,x_c$ and Y has the states $y_1,\ldots,y_r$. In that way, patterns of association between the states of the variables can be tested (Kazanjian and Drazin, 1989; Drazin and Kazanjian, 1990). Hildebrand et al. (1977) presented the DEL-technique as a statistical procedure for analyzing cross-classification data as an alternative to classical statistical approaches. The DEL-value has been interpreted as the proportionate reduction in error of knowing the specific hypothesis (or prediction rule) over not knowing that hypothesis. Hildebrand et al. (1977) concluded that DEL "satisfies those design criteria that pertain to the measurement of prediction success attained in a population by a bivariate proposition" (p. 104). Drazin and Kazanjian (1993, p. 1380) state that "the most important strength of DEL is that it allows a researcher to develop an a priori customized prediction rule that can test a theory of the relationship between the states of two categorical variables" and "other important properties of DEL are that it is independent of sample size and is robust for small samples, and its test statistic is distributed normally". In this respect, the DEL-technique is claimed to be superior to other test methods that use test statistics like Chi-square, Lambda and Tau (Kazanjian and Drazin, 1989: p. 1496).

At a first glance, the DEL-technique offers what we need in our applications, when e.g. the Chi-square does not help us any further, once we discovered that the two categorical variables are not statistically independent. However, until now only relatively few articles have been published in which the DEL-technique has actually been used (cf. Auster, 1992; Drazin and Kazanjian, 1990 and 1993; Gankema et al., 2000; Hurry et al., 1992; Kazanjian and Drazin, 1989 and 1990; Postma and Kok, 1999). If it is true that DEL offers potential for research in the field of management and organization as argued by Drazin and Kazanjian in the 1993 Academy of Management Journal special issue on analysis techniques, this result is at least remarkable. In addition, it is conspicuous that the DEL-technique is still not available in the common statistical software packages like SPSS or SAS, given the presumed nice features of the technique. Furthermore, Postma and Kok (1999) indicated some of the practical problems in using the DEL-technique. Thus, a more critical stance towards the DEL-technique may therefore be necessary. We searched the Current Index of Statistics, yearly issued by the American Statistical Association and the Institute of Mathematical Statistics for additional references on the DEL-technique and came across the book review of Hildebrand et al. (1977) by Haberman (1977). He mentioned a fierce debate about the fundamental issues of the DEL-technique between Hildebrand, Laing and Rosenthal on the one hand and Goodman and Kruskal on the other, in volume 3 of the Journal of Mathematical Sociology (1974: pp. 163-213). Haberman also strongly criticized the DEL-technique. Moreover, it appeared that Kazanjian and Drazin (1989, 1990) and Drazin and Kazanjian (1990) only referred to Hildebrand et al. (1977). Although, Hurry et al. (1992) do refer to Hildebrand et al. (1974a), and Auster (1992) and Drazin and Kazanjian (1993) do refer to Hildebrand et al. (1974a; 1974b), they seem to be unaware of the fundamental criticism and the resulting scientific debate, for they only mentioned the claims with regard to the presumably beneficial properties of the DEL-technique that have been criticized heavily in the literature. In addition, Gankema et al. (2000) and Postma and Kok (1999) referred to Hildebrand et al. (1977) and Drazin and Kazanjian (1990; 1993). Thus, researchers still seem to use the DEL-technique on a limited scale in the management and organization field, legitimizing it by referring to studies, in which it is applied and portrayed as a promising technique. These researchers are either aware or unaware of the debate, however, it is not explicitly mentioned.

It is not the aim of this paper to restart this debate. However, potential users of the DEL-technique should not be deprived of the existence and consequences of the debate. On the one side, the proposals by Hildebrand et al. (1974a; 1974b; 1977) have presumably been correctly questioned. But, on the other side, these authors need to be acknowledged for clearly addressing some of the interrelated methodological problems that researchers in the management and organization field face when analyzing relationships between categorical variables. These problems require a technique for predicting a dependent categorical variable, given the knowledge of the value of an independent categorical variable. Similar problems exist in the field of medical diagnosing. Given the debate, our main question is how to deal with the DEL-technique? We consider some methodological issues surrounding the DEL-technique and explore plausible ways to address the problems in predicting events.

This paper is organized as follows. In the next section, we briefly describe the DEL-technique, summarize the debate about its fundamental issues and divide the main question into two research questions. In the subsequent section, we describe the research approach used to answer our research questions. This research approach consists of two experiments. By means of the first experiment, we show the consequences of applying the DEL-technique when ignoring the debate about its controversy. By a second experiment, we describe alternative ways in which to proceed in addressing methodological research problems that academic researchers and practitioners face in developing a prediction technique for analyzing bivariate categorical relationships. In the final section, we conclude that the DEL-technique has some severe limitations and summarize our findings.

Background: the debate and its consequences

Before summarizing the debate and deriving the research questions, we briefly describe the DEL-technique. The DEL-statistic is claimed to be a nonparametric measure of prediction success that can be used for bivariate and multivariate analysis of categorical variables. Categorical (or qualitative) variables are variables, which are nominally or ordinally scaled. This paper focuses on the bivariate analysis of nominally and ordinally scaled categorical variables. Although only two variables are involved, the relevant methodological issues can be addressed. The DEL is considered to be a kind of measure of the proportionate reduction in error of the prediction rule (or hypothesis) over not using that prediction rule. The basic technical features of the DEL-technique are summarized in more detail in the appendix.

In summary, the debate about the DEL-technique centers around two fundamental issues[1]. The first issue is the scientific philosophical foundation of the DEL-technique. Goodman and Kruskal (1974a) argue that Hildebrand et al. (1974a) use a simplistic approach to scientific theory development and do not go into relevant issues nor cite literature concerning the nature and role of scientific theory and the degree of confirmation of theory by bodies of observed or hypothetical data as described in the philosophical literature. The second issue in the debate concerns the methodology with respect to building an empirical evaluation model of a theory. Hildebrand et al. (1974a) do not justify their use of the independence benchmark in the DEL-procedure according to Goodman and Kruskal (1974a). Hildebrand et al. (1974a) use a mathematical term as the denominator in the formula for calculating DEL, which is considered a benchmark of statistical independence. In their reaction, Hildebrand et al. (1974b) proclaim to refute the criticism by referring to the incommensurability of research paradigms. They argue that the traditional theory-testing perspectives used by Goodman and Kruskal (measures of association and multiplicative models for probabilities in cross-classifications) are inappropriate to evaluate their measure of prediction success attained by *a priori* stated propositions in the prediction logic. Accordingly, Hildebrand et al.'s interpretation of making an *a priori* prediction is especially relevant. Replying to Hildebrand et al.'s answer, Goodman and Kruskal (1974b) strongly maintain their critique on both issues.

With reference to this debate, in our opinion there are at least two pragmatic viewpoints about what a theory could possibly be. A first interpretation is that a theory should be able to describe phenomena that researchers experience. The observations researchers have, should give sufficient support to maintaining the theory and testing statistical hypotheses can be a means to assess the validity of the theory. By means of a good theory researchers should be able to make good predictions. This works rather well in the physical sciences, but in the life sciences it is much harder to formulate theories that are strongly supported by the data and give

[1] We refer interested readers to the articles in the Journal of Mathematical Sociology, 1974, volume 3, pages 163-213.

good predictions. In the life sciences, there is a second interpretation of theory. Here, we meet theory as a model that is not necessarily very close to reality, but it is meant to be rather simple and to capture the essential features or relationships, and to approximate reality in a satisfactory way. So, the researcher assumes that the frequencies in a cross-classification table are postulated to follow a multinomial distribution. The probability that $Y=y_i$ and $X=x_j$ is denoted by P_{ij}. Now suppose that the theory states that no observations will occur in cell (1,1), then $P_{11}=0$. When, in actual practice, the observed frequency is not equal to zero, then within the framework of testing statistical hypotheses, the researcher has to reject the hypothesis that $P_{11}=0$. However, this frequency can be relatively small and, as an approximation $P_{11}=0$, it can be a good model or theory. Of course, one could also argue that the multinomial model is not completely appropriate in this case and that for instance a Poisson type of statistical model would be better. Here, we build on the notion that, given the value x_1 of the independent variable X, the researcher will predict that the outcome y_1 of the dependent variable Y will not occur. There is a calculated risk that the researcher will be wrong. So, the prediction logic will be considered a way to a priori formulate an approximate model on the basis of scientific reasoning and possibly good intuition.

Ignoring the controversy in the debate can have important consequences with respect to applying the DEL-technique and interpreting the results attained in practical evaluations of management and organization theories. Therefore, we mainly focus on the consequences resulting from the second issue in the debate, the unwarranted independence benchmark argument against the incommensurability of paradigms argument. With regard to this issue we claim that when the DEL-technique is supposed to focus on predicting events, the comparison between the DEL-technique and, for example, the classical Chi-square test is figurative. The classical Chi-square test can be used to test the independence between two categorical variables X and Y, but logically cannot be used and is not used to *predict* events. Therefore, statements that the DEL-technique is superior to the Chi-square test in this respect are questionable. We agree with Hildebrand et al. that the DEL-technique cannot be used to *test* a theory which states a specified dependence of two categorical variables X and Y. In other words, it cannot be used to accept the hypothesized relationship e.g. by testing DEL's statistical significance against the DEL=0 null-hypothesis. The DEL-value is supposed to indicate that a hypothesized relationship may help to reduce prediction errors. This is not the same as accepting the hypothesis that a theory is true. However, what happens when the researchers are unaware of the debate and apply the DEL-technique? In other words, our first question is how can the DEL-results be interpreted when the DEL-technique is applied in a theory evaluation application in the management and organization research practice?

The researchers who are aware of the debate may be hesitant to use the DEL-technique in prediction research, due to the heavy criticism. What alternatives do these researchers have? In other words, our second question is which alternative ways exist to predict events in a theory evaluation context in the field of management and organization compared to the DEL-technique? In the following section the approach is described to find answers to these two questions.

Research approach

Our research approach consists of two experiments which have been conducted under specific conditions. Experiment 1 was set up to show how DEL can be interpreted when ignoring the controversy about the DEL-technique (question one). Taking into account this controversy, experiment 2 is designed to address alternative ways in prediction research (question two). For these purposes, a particular theory is developed in a specified domain and evaluated using empirical data in each experiment. In both experiments the required knowledge (theory) about the distribution of the dependent variables plays an important role. In this respect, two kinds of research can be distinguished: a priori and ex post research. In an a priori research, theory is used to formulate hypotheses before data are gathered and analyzed. In testing theories, ideally an a priori analysis is performed. This is based on the idea that the best way to empirically proceed a (part of) theory is by testing a specific hypothesis with a certain data set in the case of a priori knowledge (theory) about the proposed association to avoid the risk of adapting the theory to the data (Armstrong, 1985).

When knowledge (theory) about the proposed association is not present or not sufficiently developed an ex post research can be executed. In an ex post research, hypotheses are formulated or adapted after the data have been collected, tabulated and analyzed. This means that a specific kind of prediction cannot be made without knowledge about the dependent variable. It is important to notice that when a method or technique (e.g. the DEL-technique), aimed at testing a customized hypothesis and making inferences about the nature and strength of the association, requires that a priori theory is used, relationships cannot be predicted a priori on the basis of ex post formulated hypotheses using the same data and standard statistical procedures. Here, we face the dangers of data mining (see e.g. Freedman, 1983; Hildebrand et al., 1977: 203; Lovell, 1983; Mayer, 1975; Steerneman, 1987: sections 1.3, 3.5, and 4.6; Steerneman and Rorijs, 1988).

The potential methodological problems with the DEL-technique and our notions about an alternative research direction will be exemplified by making use of an available database (STRATOS) in both experiments. The STRATOS database has been developed by the International STRATOS-group. The international STRATOS-research group has conducted research in the context of strategic problems in Small and Medium-sized Enterprises (SMEs). STRATOS is an acronym for Strategic Orientations of Small and Medium-sized Enterprises (see Bamberger, 1994). The STRATOS-database and different publications based on this data base (e.g. Haahti, 1989; Bamberger, 1994; Snuif and Zwart, 1994) were available for our proposed research. The STRATOS database involves data on 550 variables for 1,135 enterprises (random sample) in three different sectors (clothing: 35%, food: 39%, and electronics: 27%) in eight European countries. The data are based on structured interviews with identical questionnaires held by (owner) managers of SMEs. Two nominal variables which are part of this database are chosen, based on the following demands: their scale (nominal); the (low) number of missing cases, and the theoretical plausibility of the proposed association between these variables. The two variables are 'type of industry' (categories: clothing, food, and electronics) and 'type of R&D' (categories: basic research/ applied research/ development activities (B/A/D), applied research/ development activities (A/D), and development activities only (D)). In general, innovativeness as indicated by 'various types of R&D' varies within different industry classes (cf. Van der valk, 1998; Lejour and Nahuis, 2000).

The analysis of the specific relationship between these two variables is based on a total sample data set of 772 valid observations.

Although the two experiments have the same context of theory evaluation as a starting-point, in other words, use the same theory and the same empirical sample data set, they largely differ with respect to their purpose and methodological design. The first experiment is aimed at a priori and ex post evaluating (and improving) a theory with the DEL-technique. This theory, i.e. an hypothesized relationship between two categorical variables, is evaluated using empirical data. The second experiment is aimed at a priori and ex post developing and improving a theory which predicts events, using various alternative proportionate reduction in prediction error calculations. Predicting the events means predicting the states of the dependent categorical variable out of the states of the independent one. The two experiments are described in the following two subsections.

Ignoring the controversy: experiment 1

Ignoring the controversy means that we ignore the criticism and assume that the DEL-technique is appropriate to evaluate a theory. The research case consists of evaluating a theory which states the specified customized dependence of two categorical variables and using theory-evaluating techniques, e.g. the DEL-technique. Applying the DEL-technique also means that we measure the prediction success attained by an a priori developed and explicitly stated customized hypothesis using prediction logic (prediction logic paradigm). To this aim, we developed a research case experiment in which the DEL-technique as a presumed measure of prediction success is used. Experiment 1 consists of two parts. The two parts are described in detail below.

Part 1: an a priori theory-evaluating analysis

The first part of the case consists of an a priori analysis measuring the prediction success of a hypothesis using the total sample applying the DEL-technique and some other measures. In order to do this, a customized hypothesis (or prediction rule) was a priori developed and explicitly stated. Prediction success was measured by the DEL-value. Consequently, the null-hypothesis that the DEL-value in the population is zero was tested against the alternative hypothesis that DEL is larger than zero using the normal approximation. The hypothesis (or prediction rule) concerns the association of the type of industry with R&D-type (cf. Tidd et al. 1997). The hypothesis is contingent upon the assumption that electronics is associated with all kinds of R&D (as the most innovative industry sector), clothing is associated with only development and food is associated with medium R&D activities (applied research and development), resulting in prediction rule 1 (abbreviated as PR1). This specific model is underpinned by a recent survey of the EIM on the innovativeness of Dutch industries (Van der Valk, 1998) and by comparative research on R&D intensity for various countries at the sector level in 1990 (Lejour and Nahuis, 2000). This survey shows clearly that the electronics industry puts most effort in all three kinds of R&D; that the food industry has an intermediate position, and that the clothing industry puts less effort in R&D[2]. In addition, PR1 is formulated as follows:

- Electronics is positively associated with basic research, applied research and development (weight 0)
- Food is positively associated with applied research and development (weight 0)
- Clothing is positively associated with development (weight 0)

Table 1. Weights for Prediction Rule 1 (PR1)

Industry	R&D-type		
	B/A/D	A/D	D
Clothing	1	1	0
Food	1	0	1
Electronics	0	1	1

[2] However, we must be careful for the food and clothing industries, their position was inverted in an earlier survey of the EIM.

Table 2. Association measure statistics

Association measures	A priori total sample (n=772)
DEL	0.08 (<.0001)
Pearson Chi-square (df: 4)	42.588[a] (.000)
Lambda:[b]	0.033 (.0418)
Tau:[b]	0.38 (.000)

a: 0 % of the cells have expected count less than 5
b: 'type of R&D' is the dependent variable

Results of part 1

An a priori analysis of PR1 in the total sample was performed using different association measures to measure the prediction success of this hypothesis. Table 2 shows the results of the DEL-measure, the chi-square based measure and two other PRE-measures (Guttman Lambda; Goodman and Kruskal Tau). The Chi-square measure shows that the null-hypothesis of statistical independence is rejected. The Guttman Lambda measure reflects the reduction in prediction error, when given $X=x_j$ the model predicts the modal Y-category under this condition in comparison to always predicting the unconditional modal Y-category. Goodman and Kruskal Tau shows the reduction in prediction error of predicting y_i given $X=x_j$ with the correct conditional probability in comparison to predicting y_i with the true unconditional probability. For the Lambda and the Tau measure we deal with single state predictions. Moreover, the underlying prediction rules related to Lambda and Tau are ex post in the interpretation of Hildebrand et al. (1977). The prediction logic of Hildebrand et al. (1977) allows multi-state predictions. The DEL-value was calculated and tested on its significance against the null-hypothesis that the *DEL-value is zero* using the normal approximation. The DEL-value appears to be significant. This means that there is empirical evidence to conclude that using the prediction rule PR1 significantly reduces the prediction error with 8 % over not knowing the rule. However, a significant DEL-value does not mean that this is the only specific (hypothesized) relationship between the type of industry and the type of R&D exists, in other words, it does not mean that this relationship should be accepted. In the prediction logic paradigm, this only indicates, in terms of Hildebrand et al. (1977) that a significant proportionate reduction in error of 8 % is obtained, knowing this specific prediction rule over not knowing the prediction rule. We need to realize that there may be other prediction rules which may have a far higher reduction in prediction error. Note, that in this case it is unclear how large the exact total prediction error is, because it is a relative yardstick.

An important part of the debate is just about this yardstick. The opponents conclude that *not knowing the prediction rule* is equivalent to statistical independence. Hildebrand et al. do not agree. We are not going to restart the debate, but there is a serious question about what the a priori prediction rule would be that reflects *not knowing the prediction rule*.

Part 2: an ex post theory evaluation analysis

The second part consists of conducting an ex post analysis to enhance the measure of prediction success as performed in a number of the aforementioned studies. This analysis may logically result from the analysis in part 1, for example to improve PR1, leading to a larger reduction in prediction error; 8% is not that much. The a priori customized hypothesis is not very successful. When knowledge (theory) about the proposed association is not present or not sufficiently developed, an ex post research can be executed. For example, in case of using the DEL-technique, when theory about the specific relationship is lacking, it is tempting to adjust the prediction rule according to the analysis of the data and using the same data set and procedures to test these new prediction rules. Tests of the adjusted prediction rule will generally yield a higher DEL-value and more significant results. This is not surprising, because the researcher is fitting the prediction rule to the data, instead of using the data to support a prediction rule. However, an ex-post analysis requires a modification of the DEL-technique (formulas and distribution) or validation of the results in a new sample.

The aim here is to show that ex post measuring of prediction success of a slightly different hypothesis using a slightly different weighting scheme using the same data and the same procedure, will result in a different DEL-value, which appears to be better than the DEL-value of the a priori analysis. This is basically the same as the sensitivity analysis performed by e.g. Drazin and Kazanjian (1990) in which they formulate eleven different weighting schemes between 0 and 1 to choose the DEL-value which is most significant. For this purpose, PR1 is adjusted on the basis of an analysis of the data resulting in prediction rule 2, abbreviated as PR2:

- Electronics is positively associated with basic research, applied research and development (weight 0) and partially with basic and applied research (weight 0.5)
- Food is positively associated with development (weight 0) and partially with applied research and development (weight 0.5)
- Clothing is positively associated with development only (weight 0)

Table 3 shows PR2 which has slightly different weights compared to PR1. The DEL value for PR2 tested in the total sample (n=772) is 0.16 with a P-value less than 0.0001. The comparison of PR1 and PR2 is strictly speaking an ex post analysis. More generally, the evaluation of various choices of weights lies in the area of ex post prediction. An ex post prediction requires a modification of the DEL-technique according to Hildebrand et al. (1977: pp. 221-230). They conclude that in these situations the chi-square distribution should be applied instead of the normal distribution in testing the statistical significance. The reason is that when using the normal distribution, the error in calculating the significance levels increases, in the case the number of comparisons between weights becomes larger. The calculated significance levels for the comparisons will then be to optimistic compared to the real ones. Applying the chi-square distribution in the case of a larger number of comparisons would be better. This reasoning implies that the judgement of the statistical significance may not be correct, when using the same DEL-formula and the normal distribution (as applied in table 2 on page 323 by Drazin and Kazanjian; 1990). Something similar also holds for our P-value of PR2, although only one comparison is made.

Table 3. Weights for Prediction Rule 2 (PR2)

Industry	R&D-type		
	B/A/D	A/D	D
Clothing	1	1	0
Food	1	0.5	0
Electronics	0	0.5	1

Having the opportunity to experiment with the data, we repeated the DEL-procedure 10 times in order to illustrate the effect to study the stability of the results. In addition, the sample was randomly split in two equal parts; an estimation and a validation part. The estimation part will be used to analyze the data and to improve upon the a priori PR1 using the weights of table 1. We expect that the ex post PR2 using the weights of table 3 and the estimation data set gives larger DEL-values. Because we used the same DEL-formulas and the same data set to test PR2 against the normal approximation, PR2 was validated in a new test; in our case the validation part of the sample that had not been used yet. This validation test can be considered as an a priori analysis in which the null-hypothesis that DEL differs from zero is tested against the normal approximation.

Table 4. DEL-values in the estimation and validation samples

Test	Estimation sample		Validation sample	
	PR1	PR2	PR1	PR2
Test 1	.14 (<.0001)	.22 (<.0001)	.03 (.1335)	.10 (.0039)
Test 2	.07 (.0174)	.15 (<.0001)	.10 (.0028)	.16 (<.0001)
Test 3	.11 (.001)	.16 (<.0001)	.06 (.0367)	.16 (<.0001)
Test 4	.07 (.0301)	.12 (.0019)	.10 (.0016)	.20 (<.0001)
Test 5	.08 (.0112)	.17 (<.0001)	.09 (.0051)	.15 (<.0001)
Test 6	.10 (.0033)	.17 (<.0001)	.07 (.0162)	.16 (<.0001)
Test 7	.12 (.0003)	.19 (<.0001)	.05 (.0764)	.14 (.0003)
Test 8	.07 (.0287)	.14 (<.0001)	.10 (.0016)	.18 (<.0001)
Test 9	.09 (.0057)	.14 (<.0001)	.08 (.0104)	.18 (<.0001)
Test 10	.06 (.0559)	.16 (<.0001)	.11 (.0005)	.16 (<.0001)

(p-value between parentheses)

Results of part 2

An ex post analysis was performed using an estimation and validation sample. Comparing the estimation test results of PR1 with the test results of PR2 shows that a slightly adjusted weighting scheme on the basis of examination of the data (ex post) produces a different, far higher DEL-value in all 10 experiments (see table 4). The validation results in table 4 show that the improved hypothesis indeed seem to generate consistently higher DEL-values. Note that, as mentioned in the a priori analysis in part 1, the significance of DEL can only be interpreted as proportionate reduction in error which differs from zero and not as an acceptance of PR2 as the best prediction rule.

Conclusion experiment 1

With respect to the a priori analysis in part one it can be concluded that performing a theory evaluation procedure with the DEL-technique does not necessarily mean that it can be interpreted as a test statistic which serves as a measure of acceptance of the hypothesis according to the theory-testing paradigm. Where Chi-square shows a significant statistical dependence, however, DEL should be interpreted according to the prediction logic paradigm as a measure of predictive effectiveness, just like Lambda and Tau.

A review of the empirical studies which apply the DEL-technique shows that these studies performed a similar DEL-analysis, calculating DEL and testing it against the null-hypothesis that DEL is zero. These studies concluded that in the case the DEL-value was significant, support exists for the specified customized hypothesis compared to not knowing any hypothesis. This interpretation of DEL-value is considered to provide a measure of strength analogous the interpretation of the coefficient of determination (R^2) in regression analysis (Drazin and Kazanjian, 1990; 1993: p 1380). The square root of the coefficient of determination is R: the correlation coefficient. However, this should not be interpreted as the acceptance of this customized hypothesis as the one and only relationship between the two categorical variables. In addition, a non-significant DEL cannot be interpreted as the independence between the two categorical variables. Because, in both instances, other customized hypothesized relationships may be significant and have a higher DEL-value. Only in the case that every DEL is zero, statistical independence must hold. In this sense, testing the significance of DEL and evaluating the theory (prediction logic paradigm) is not the same as testing whether or not the null-hypothesis of independence should be rejected as in testing a theory (theory-testing paradigm). Moreover, stating that a hypothesis is empirically supported in the theory-testing paradigm is something different than stating the same in the prediction logic paradigm. When this context for interpretation of DEL is not explicitly stated, DEL can be easily misinterpreted.

From a first glance at the estimation results of ex post analysis in part two, it looks like the DEL-technique can be used to evaluate very specific customized hypotheses, because it generates different DEL-values on different hypotheses. The validation results seem to support the estimation results. In theory-testing, it is common to use the same formulas in the a priori and the ex post analysis on the condition that the results are validated. This means that the data is split in an estimation sample to improve the hypothesis and a validation to test this improved hypothesis. The reason is that in the case the theory is adjusted to the data set on the basis of analysis of the data, the risk of datamining is substantial and has to be avoided if one is strictly testing theories. Taking the prediction logic paradigm as a starting-point, Hildebrand et al. (1977: p. 203) mean something the like when they state that in hypothesis testing, the procedures as mentioned in the appendix "do not apply to testing the significance of estimated DEL" when the proposition is selected ex post facto. Moreover, these procedures "especially do not apply to testing the largest estimated DEL that can be calculated from the sample data". Taking the largest estimated DEL from the same sample data is comparable to the sensitivity analysis as executed in some of the studies which use the DEL-technique. Consequently, when little theory is available and an ex post analysis seems to be the only alternative to formulate or improve the hypothesis, the results need to be interpreted with great care. Firstly, specific adapted formulas need to be applied in an ex-post analysis (use of the chi-square distribution instead of the normal distribution, according to Hildebrand et al. (1977: pp. 221-230). Secondly, in the case of an ex-post analysis using the same formulas, data and distribution, the results can only be interpreted correctly when they are validated, for example, by randomly splitting the data set in a estimation and a validation sample in order to attain the pre-specified goal of the research.

In experiment 1, we assumed that Hildebrand et al.'s ideas are correct, for the sole reason of achieving our goal in this experiment. Given the debate about methodological issues concerning the DEL-technique, but acknowledging the importance of addressing problems in predicting events in bivariate categorical relationships, researchers may want to consider alternative ways of predicting events in a theory evaluation context. This is the subject of experiment 2.

Prediction research direction: experiment 2

Using the classical statistical techniques we can test whether two nominal variables are statistically independent. When this independence is rejected, we are left with the question what kind of relationship will exist. Given the outcome X on the independent variable, what can be said about the outcome of the dependent variable. One possible alternative way is to provide the probabilities on the possible outcomes of the dependent variable given the realization of X. However, frequently we would like to predict Y itself, e.g. by indicating which outcomes of Y will probably not occur.

Inspired by the original ideas from Hildebrand et al. (1977) and in view of the discussion in the scientific literature about the DEL-method, we will have a closer look at the research case and we develop some further ideas about predicting events. Although, we could also discuss that generalized linear models (GLM) and other PRE measures such as Guttman Lambda and Goodman and Kruskal Tau could have been used in our case, we have chosen to stay as close as possible to the ideas by Hildebrand et al. and to cope with the discussion on their theory by Goodman and Kruskal in the Journal of Mathematical Sociology in 1974. For the reader interested in GLM, we refer to e.g. Agresti (1984), McCullagh and Nelder (1989), and Santner and Duffy (1989) and with respect to Tau and Lambda we refer to Goodman and Kruskal (1954) and Guttman (1941), see also Hildebrand et al. (1977).

Table 5. Total sample frequencies

Industry	Research			Total
	B/A/D	A/D	D	
Clothing	28	51	163	242
Food	36	85	141	262
Electronics	48	116	104	268
Total	112	252	408	772

In the prediction logic of Hildebrand et al. (1977) a theory is a priori postulated according to which the relative frequencies of some cells in the population are put equal to zero. The DEL-technique appoints positive weights to these cells between 0 and 1. The other cells have weights equal to zero. Given the outcome of the explanatory variable X the dependent variable Y is predicted by giving the subset of outcomes of Y that are possible according to the theory. In fact the prediction is a subset of the possible outcomes of Y depending on X. If such a subset is equal to the set of all possible outcomes of Y, then according to the a priori formulated theory we cannot exclude any of the possible outcomes of Y. However, given the specific outcome on X, the probabilities on the various Y-outcomes may differ. In this situation there will be no prediction error. In our experiment the possible values of Y are B/A/D, A/D, and D. One can imagine that predicting by means of the whole set {B/A/D, A/D, D} may be appropriate if given X the probabilities of that B/A/D, A/D, and D occur are (approximately) equal to 1/3 each. A prediction rule can also make explicit that, for example, one possible value of Y will definitely not occur.

Now let us assume that we are forced to make a choice for one possible outcome. This is in fact the situation in table 1. Here we see prediction rule 1, presented as PR1 in table 6. If we do not have any information about the types of research and development, then complete ignorance is reflected by attaching equal probabilities (=1/3) to B/A/D, A/D, and D. This is called prediction rule 3 (PR3 in table 6). The prediction error can be calculated using the cell frequencies of the states in the sample (see table 5). According to PR 3 we expect that on the average $(112+252+408)/3 = 772/3 \approx 257$ cases will be predicted correctly. So, about 515 cases will be wrongly classified. According to PR1, $163+85+48 = 296$ cases will correctly be classified. Hence 476 cases will be wrongly predicted. The

Table 6. Some prediction rules

Prediction rule	Industry			PRE
	Clothing	Food	Electronics	
PR1	D	A/D	B/A/D	0.08
PR3	D, A/D, B/A/D with probability 1/3			---
PR4	D	D	D	0.29
PR5	D	D	A/D	0.32

proportionate reduction of error (PRE) of PR1 with respect to PR3 (complete ignorance) is equal to $1-476/515 \approx 0.08$. With respect to this particular case, one may observe that all industries have development (D), so the prediction D is the less restrictive prediction. So, another prediction rule could be: “always predict D” (PR4). This would lead to 408 successes and the PRE is 1-364/515 ≈ 0.29. If we would have known table 5, the ex post prediction rule would have been: “for Clothing and Food predict D and for Electronics predict A/D” (PR5). With this prediction rule 116+163+141 = 420 cases would be predicted correctly; the PRE is equal to 1-352/515 ≈ 0.32. Note that this is only a small improvement with regard to PR4.

It would have been of much help if we could have had a first impression of the structure of the population (by conducting an ex post analysis). Because then we would have obtained a rule with a larger proportionate reduction of error. One way of doing this is to split a sample into two parts. The first part is used to look for an appropriate prediction rule. The second part will be used to validate this rule. This is at least in line with the ideas in the work of Hildebrand et al. (1977) that additional research is necessary in case of an ex post analysis and that the DEL should not be used ex post. If we do so, then we obtain the following results, see table 7.

Table 7. The best prediction rule

Test	Best prediction rule in	
	estimation sample	validation sample
Test 1	PR5	PR5
Test 2	PR5	PR4
Test 3	PR4, PR5	PR5
Test 4	PR5	PR5
Test 5	PR5	PR4
Test 6	PR5	PR4, PR5
Test 7	PR5	PR5
Test 8	PR5	PR5
Test 9	PR5	PR5
Test 10	PR5	PR4

Table 7 shows that it is possible that using the estimation sample to find the best prediction rule, another rule is optimal according to the validation sample, see tests 3, 5, 6, and 10. However, from a practical point of view the qualities of PR4 and PR5 are comparable. We would never have chosen PR1.

Conclusion experiment 2
Firstly, the results of experiment 2 show that in a simple research case an alternative way or technique of predicting events compared to the DEL-technique can be used. This alternative technique is easy and unequivocally interpretable, for example by showing the prediction error and the reduction in prediction error in an a priori analysis and an ex post analysis. In addition, this alternative technique even seems to outperform the DEL-technique (assumed its interpretation is correct) in predicting events in a theory evaluation context.

Conclusion and discussion

When ignoring the controversy about this technique, in other words, researchers are unaware of or do not acknowledge the debate, assuming that the interpretation of DEL by Hildebrand et al. (1974a; 1974b; 1977) is correct, experiment 1 shows that the DEL-measure can easily be misinterpreted as a measure of acceptance of the specified customized hypothesis as the only and best relationship between the categorical variables at stake. This is especially the case when the context for interpretation is not carefully stated in terms of the research paradigm (prediction logic versus theory-testing). Firstly, a DEL-value should not be interpreted as a measure for independence theory-testing, i.e. falsifying a theory, which states the independence between two bivariate variables. Statistical independence only holds in the case every DEL is zero. Secondly, the DEL-value should also not be interpreted as a measure for dependence theory-testing i.e. confirming a theory as the only and best theory which states a dependence between two bivariate variables. Thirdly, in an ex post analysis, the DEL-technique requires modifications which differ from those used in an a priori analysis or, in the case no modifications are used, the results need to be validated. Actually, every ex-post analysis needs to be confirmed by follow-up studies, as also argued by Hildebrand et al. (1977: p. 230).

Taking into account the controversy about the DEL-technique, it is questionable whether the interpretation according to Hildebrand et al. (1974a; 1974b; 1977) is correct. In calculating DEL, Hildebrand et al. use an independence benchmark, which means according to Goodman and Kruskal (1974a; 1974b) that they cannot really claim a proportionate reduction in error, knowing the prediction over *not knowing the prediction rule.* Acknowledging the importance of addressing methodological problems in analyzing bivariate categorical data, e.g. predicting events, researchers may wish to explore alternative techniques. Some alternative techniques of predicting events are Generalized Linear Model and PRE-measures Lambda and Tau. These techniques are not dealt with in this research in order to stay relatively close to the original ideas of Hildebrand et al. (1974a; 1974b; 1977) and the debate in the Journal of Mathematical Sociology in 1974. Another way of predicting events is calculating a PRE-measure. Experiment 2 shows that an alternative PRE-technique generates easy, unequivocally interpretable results in an a priori analysis and a validated ex post analysis and outperforms the DEL-technique.

Summarizing, the DEL-technique can not be actually used as a theory-testing method. This paper shows the methodological intricacies connected to the application of the DEL-technique. Other approaches or techniques may be worked out and may perform better, as the alternative way in this paper shows. In future research, this line of research may be addressed.

References

Agresti, A. 1984. ***Analysis of Ordinal Categorical Data***. New York: Wiley .

Armstrong, J. Scott, 1985. ***Long-Range Forecasting: from Crystal Ball to Computer***. 2nd ed, New York: Wiley.

Auster, Ellen R. 1992. The Relationship of Industry Evolution to Patterns of Technology Linkages, Joint Ventures, and Direct Investment Between U.S. and Japan. ***Management Science***, 38: 778-792.

Bamberger, I. 1994. ***Product/Market strategies of Small and Medium-sized enterprises***, Avebury, Aldershot.

Drazin, R., Kazanjian, R.K. 1990. Research Notes and Communications; A Reanalysis of Miller and Friesen's Life Cycle Data. ***Strategic Management Journal***, 11: 319-325.

Drazin, R., Kazanjian, R.K. 1993. Applying the Del technique to the Analysis of Cross-classification Data: a Test of CEO Succession and Top Management Team Development. ***Academy of Management Journal***, 36: 1374-1399.

Freedman, D.A. 1983. A Note on Screening regressions equations. ***The American Statistician***, 37: 152-155.

Gankema, H.G.J., Snuif, H.R. and Zwart, P.S. 2000. The Internationalization Process of Small and Medium-Sized Enterprises: An Evaluation of Stage Theory. ***Journal of Small Business Management***, 38 (4): 15-27.

Goodman, L.A. and Kruskal, W. 1954. Measures of Association for Cross-Classification. ***Journal of the American Statistical Association***, 49: 732-764.

Goodman, L.A. and Kruskal, W. 1974a. Empirical Evaluation of Formal Theory. ***Journal of Mathematical Sociology***, 3: 187-196.

Goodman, L.A. and Kruskal, W. 1974b. Empirical Evaluation of Formal Theory. ***Journal of Mathematical Sociology***, 3: 211-213.

Guttman, L. 1941. An Outline of the Statistical Theory of Prediction. Supplementary Study B-1. In Paul Horst et al., Eds.. ***The Prediction of Personal Adjustment***. New York: Social Science Research Council, Bulletin 48: 213-258.

Haahti, A.J. 1989. ***Entrepreneurs' Strategic Orientation: Modeling Strategic Behavior in Small Industrial Owner-managed Firms***. PhD-dissertation, The Helsinki School of Economics and Business Administration, Helsinki.

Hildebrand, D.K., Laing, J.D., Rosenthal, H.L. 1974a. Prediction Logic: A Method for Empirical Evaluation of Formal Theory. ***Journal of Mathematical Sociology***, 3: 163-185.

Hildebrand, D.K., Laing, J.D., Rosenthal, H.L. 1974b. Prediction Logic: A Method for Empirical Evaluation of Formal Theory. ***Journal of Mathematical Sociology***, 3: 197-209.

Hildebrand, D.K., Laing, J.D., Rosenthal, H.L. 1977. ***Prediction Analysis of Cross Classifications***. New York: Wiley.

Hurry, D., Miller, A.T., and Bowman, E.H. 1992. Calls on High-Technology: Japanese Exploration of Venture Capital Investments in the United States. ***Strategic Management Journal***, 13: 85-101.

Kazanjian, R.K. and Drazin, R. 1989. An Empirical Test of a Stage of Growth Progression Model. ***Management Science***, 35: 1489-1503.

Kazanjian, R.K. and Drazin, R. 1990. A Stage-Contingent Model of Design and Growth for Technology Based New Ventures. ***Journal of Business Venturing***, 5: 137-150.

Lejour, A. and Nahuis, R. 2000. Openness, Growth and R&D Spillovers: Uncovering the Missing Link? ***Research memorandum*** no. 168, CPB, The Haque.

Lovell, M.C. 1983. Datamining. ***The Review of Economics and Statistics***, 65: 1-12.

Mayer, T. 1975. Selecting Economic Hypotheses by Goodness of Fit. ***The Economic Journal***, 85: 877-883.

Mc.Cullagh, P., and Nelder, J.A. 1989. ***Generalized Linear Models***, 2nd ed., London: Chapman & Hall.

Postma, Th.J.B.M. and Kok, R.A.W. 1999. Organizational Diagnosis in Practice: A Cross-Classification Analysis Using the DEL-Technique. ***European Management Journal***, 17 (6) : 584-597.

Santner, T.J., and Duffy, D.E. 1989. ***The Statistical Analysis of Discrete Data.*** New York: Springer-Verlag.

Snuif, H.R., and Zwart, P.S. 1994. ***Modeling new venture development as a path of configurations***, Proceedings of the 39th ICBS World Conference, Strassbourg: 263-274.

Steerneman, A.G.M. 1987. ***On the Choice of Variables in Discriminant and Regression Analysis***. PhD-thesis, University of Groningen.

Steerneman, A.G.M. and Rorijs, G. 1988. Pitfalls for Forecasters, in T. K. Dijkstra, editor 1988. On Model Uncertainty and its Statistical Implications; ***Lecture Notes in Economics and Mathematical Systems:*** 102-117 , 307, Springer-Verlag.

Tidd, J., Bessant, J., Pavitt, K. 1997. ***Managing Innovation: Integrating Technological, Market and Organizational Change***. Chichester: Wiley.

Valk, van der, W.D.M. 1998. ***De innovativiteit van de Nederlandse industrie*** (the innovativeness of Dutch industries). The Hague, EIM.

Appendix

The formulas of DEL and its variance (Hildebrand et al., 1977: p. 192) are specific for this research and refer to the sampling condition in which the researcher knows neither set of marginal proportions (probabilities in the population) and fixes neither set of sample totals. This particular sampling condition requires that values of DEL (∇) are calculated according to formula 1:

$$\nabla = 1 - \frac{\sum_i \sum_j w_{ij}\, f_{ij}}{\sum_i \sum_j w_{ij}\, f_{i.}\, f_{.j}} \qquad (1)$$

where

w_{ij} = Weight for i^{th} row and j^{th} column, 0 for predicted cells and 1 (or less) for error cells,

f_{ij} = Cell probability for the i^{th} row and j^{th} column in the sample,

$f_{i.}, f_{.j}$ = Marginal probabilities for the i^{th} row and j^{th} column in the sample, respectively.

The variance of ∇ is approximated by formula 2:

$$Var(\nabla) \cong (n-1)^{-1} \left[\sum_i \sum_j (a_{ij})^2 f_{ij} - \left(\sum_i \sum_j a_{ij}\, f_{ij}\right)^2 \right]$$

$$a_{ij} = \left(\sum_i \sum_j w_{ij}\, f_{i.}\, f_{.j}\right)^{-1} \left[w_{ij} - B(\pi_{i.} + \pi_{.j}) \right] \qquad (2)$$

$$B = \left(\sum_i \sum_j a_{ij}\, f_{ij}\right)$$

$$\pi_{i.} = \sum_j w_{ij}\, f_{.j} \qquad \pi_{.j} = \sum_i w_{ij}\, f_{i.}$$

where

n = Total frequency in the cross classification table.

The hypotheses related to the distribution of the population are: H_0: DEL= 0 vs. H_a: DEL > 0. Because the DEL of the population is unknown to the researcher, the sample estimate ∇ is used. The essence is to test whether the estimated sample ∇ cannot possibly be attributed to random sampling variation around zero. The hypotheses are tested on their significance against the standard normal distribution using the estimated Z-statistic. A rule of thumb for using the normal approximation of the sampling distribution safely with small- or moderate-sized samples is presented by formula 3:

$$\frac{5}{n} \leq \left(\sum_i \sum_j w_{ij}\, f_{ij} \right) \leq 1 - \frac{5}{n} \qquad (3)$$

According to Hildebrand et al. (1977), a continuity correction can be used to improve the quality of the normal approximation. The corrected expressions for ∇ and the estimated Z statistic are represented by the following formulas (4):

$$est.Z = \frac{\nabla^{+}}{[\, est.var\ (\nabla)]^{1/2}}$$

$$\nabla^{+} = 1 - \frac{\left(\sum_i \sum_j w_{ij}\, f_{ij} \right) + \frac{0,5}{n}}{\sum_i \sum_j w_{ij}\, f_{i.}\, f_{.j}} \qquad (4)$$