STATISTICA, anno LXIII, n. 2, 2003

AN EMPIRICAL STUDY ON NEW PRODUCT DEVELOPMENT PROCESS BY NONPARAMETRIC COMBINATION (NPC) TESTING METHODOLOGY AND POST-STRATIFICATION

L. Corain, L. Salmaso

1. INTRODUCTION

In many scientific disciplines and industrial fields researchers and practitioners are often faced with complex problems when dealing with comparisons between two or more groups using classical parametric methods although real problems rarely agree with the stringent assumptions required by such methods. The NPC methodology (Pesarin, 2001) offers an innovative but well tested approach that frees the researcher from stringent assumptions of parametric methods and allows a more flexible analysis both in terms of specification of multivariate hypotheses and in terms of the nature of the variables involved in the analysis. It does so by reducing the problem to a set of simpler sub-problems, each provided with its own suitable solution. Moreover, one of the most relevant features of NPC Test is that it does not need a modelling for dependence among variables.

The case study was based on a sample of 85 firms, working in a B2B market in two specific manufacturing industries. Applying the NonParametric Combination (NPC) of dependent rankings methodology (Pesarin and Lago, 2000), the sample was divided into two groups on the basis of external performances, i.e. market, product and financial success criteria, for New Products launched over a three year period. The Best firms were those above the median score on a global success ranking, taking into account all three of the above mentioned success criteria. The other companies were called Rest firms. In the study we tested the hypothesis that Best firms have better internal performances, i.e. Time performances (Launch on Time and Time To Market Reduction) and better Quality Capability of New Product performance, than the Rest. We also wish to test if Best firms have greater internal environment support (e.g. NP Strategic Guide, Capabilities, etc.) and use some practices and drivers more intensively than Rest firms do. Results confirmed that Best firms do have greater internal environment support.

Due to the multivariate nature of the problem, the testing procedure was properly broken down into a set of simpler sub-problems, each provided with its permutations solution; then, by nonparametric combination of them, we obtained an overall test concerned with the global null hypothesis. Moreover a confounding factor expresses by innovation level was tested as possible discriminating variable by some intermediate partial tests related to sample strata.

2. SOME REMARKS ON THE NPC METHODOLOGY

From a methodological point of view, when comparing NPC methodology to unconditional parametric testing it should be remembered that the latter suffers from the constraint that it is appropriate and applicable only when a set of conditions concerned with the likelihood model are all satisfied (Pesarin, 2002). Only if all conditions are jointly satisfied is the extension of inferential results to the population possible and appropriate. Otherwise when these conditions fail, especially if selection-bias procedures are used for data collection processes as in most real applications, most parametric inferential extensions are generally wrong or misleading.

Moreover, when all the above conditions are satisfied, in practice other assumptions regarding the validity of the parametric method, such as normality, are rarely satisfied so consequent inferences, when not improper, are necessarily approximated and their approximations are often difficult to assess. However, there are circumstances in which conditional testing procedures may be unavoidable as in the case of multivariate problems, when some variables are categorical and others are quantitative or when multivariate alternatives are subjected to order restrictions (for a detailed list of these circumstances see Pesarin, 2002).

3. NONPARAMETRIC COMBINATION (NPC) OF DEPENDENT PERMUTATION TESTS

Without loss of generality, let us refer to a one-way MANOVA layout. The data structure is defined as follows. Let us denote by $\mathbf{X}$ the data set represented as:

$$\mathbf{X}=[\mathbf{X}_1,..., \mathbf{X}_j, ..., \mathbf{X}_c]'=[\mathbf{X}_1,\ldots, \mathbf{X}_i,\ldots, \mathbf{X}_k],$$

where $\mathbf{X}_j$, j=1,...,C, (C$\geq$2) represents the j-th $n_j \times k$ group, $n_j \geq 2$ and $\Sigma_j n_j = n$, and $\mathbf{X}_i$ is the i-th univariate aspect of $\mathbf{X}$, i=1,...,k ($k \geq 1$); moreover let X_{ji} represent the i-th univariate aspect of $\mathbf{X}_j$.

In the context of NonParametric Combination (NPC) of dependent permutation tests a set of conditions should be jointly satisfied:

i) we suppose that for $\mathbf{X}=[\mathbf{X}_1,...,\mathbf{X}_c]'$ an appropriate probabilistic k-dimensional distribution structure P exists, $P_j \in F$, j=1,...,C, belonging to a (possibly non-specified) family F of non-degenerate probability distributions.
ii) the null hypothesis H_0 states the equality in distribution of the multivariate distribution of the k variables in all C groups:

$$H_0 : [P_1 = ... = P_C] = [\mathbf{X}_1 \overset{d}{=} ... \overset{d}{=} \mathbf{X}_C].$$

Null hypothesis H_0 implies the exchangeability of the individual data vector with respect to the groups. Moreover H_0 is supposed to be properly decomposed into k sub-hypotheses H_{0i}, i=1,...,k, each appropriate for partial (univariate) aspects, thus H_0 (multivariate) is true if all the H_{0i} (univariate) are jointly true:

$$H_0 : [\bigcap_{i=1}^{k} X_{1i} \overset{d}{=} ... \overset{d}{=} X_{Ci}] = [\bigcap_{i=1}^{k} H_{0i}].$$

H_0 is called the global or overall null hypothesis, and H_{0i}, i=1,...,k, are called the partial null hypotheses.

iii) The alternative hypothesis H_1 is represented by the union of partial H_{1i} sub-alternatives:

$$H_1 : [\bigcup_{i=1}^{k} H_{1i}],$$

so that H_1 is true if at least one of sub-alternatives is true.

In this context, H_1 is called the global or overall alternative, and H_{1i}, i=1,...,k, are called the partial alternatives.

iv) let $\mathbf{T}=\mathbf{T}(\mathbf{X})$ represent a k-dimensional vector of test statistics, $k \geq 1$, whose components $T_i=T_i(\mathbf{X}_i)$, i=1,...,k, represent the partial univariate and non-degenerate partial test appropriate for testing the sub-hypothesis H_{0i} against H_{1i}. Without loss of generality, all partial tests are assumed to be marginally unbiased, consistent and significant for large values (for more details see Pesarin, 2001).

At this point, in order to test the global null hypothesis H_0, the key idea comes from the partial (univariate) tests which are focused on k partial aspects, and then, combining them with an appropriate combining function, from a global (multivariate) test which is referred to as the global null hypothesis.

However, before introducing the combination methodology, we should observe that in most real problems, when the sample size is great enough, there is a clash over the problem of computational difficulties in calculating the conditional permutation space. This means it is not possible to calculate the exact p-value of observed statistic T_{i0}. This is brilliantly overcome by using the CMCP (Conditional Monte Carlo Procedure).

The CMCP on the pooled data set $\mathbf{X}$ is a random simulation of all possible permutations of the same data under H_0 (for more details refer to Pesarin, 2001). Hence, in order to obtain an estimate of the permutation distribution under H_0 of all test statistics, a CMCP can be used. Every resampling without replacement $\mathbf{X}^*$ from the pooled data set $\mathbf{X}$ actually consists of a random attribution of individual data vectors to the C samples. In every $\mathbf{X}_r^*$ resampling, r=1,...,B, the k partial tests are calculated to obtain the set of values [$T_{ir}^*=T(X_{ir}^*)$, i=1,..,k; r=1,…,B], from the B independent random resamplings.

It should be emphasized that CMCP only considers permutations of individual data vectors, so that all underlying dependence relations which are present in the component variables are preserved. From this point of view, the CMCP is essentially a multivariate procedure.

3.1. *The two-phases algorithm*

Once we have defined the hypothesis system and an appropriate set of k statistics $T_i=T_i(\mathbf{X}_i)$, i=1,...,k, the natural way to test the global null hypothesis consists of two sequential phases:

1. performing k partial tests;
2. combing them in a second-order global test.

We point out that this two-step procedure can be characterized by several intermediate combinations if there is a more complex data configuration where the most interesting cases are given by testing in presence of stratification, closed-testing, multi aspect testing and repeated measures.

Assuming that the partial tests have real values and are marginally unbiased, consistent and significant for large values, then the first phase consists in:

1.a calculating the k-vector of observed values of test statistics $\mathbf{T}_0$:

$$\mathbf{T}_0=\mathbf{T}(\mathbf{X})=[T_{i0}(X_i),\ i=1,..,k];$$

1.b considering a data permutation of $\mathbf{X}$ by a random resampling $\mathbf{X}_r^*$, in order to randomly assign every individual data vector to a proper group and then calculate the vector statistics $\mathbf{T}_r^*$:

$$\mathbf{T}_r^* = \mathbf{T}_r^*(\mathbf{X}_r^*) = [\mathrm{T}_{ir}^*(\mathrm{X}_{ir}^*),\ i=1,\ldots,k];$$

1.c carrying out B independent repetitions of step 1.b; the result is a set $\mathbf{T}^*$ of $B\times k$ CMC

$$\mathbf{T}^*=[\mathbf{T}_r^*,\ r=1,\ldots,B]=[\mathbf{T}_1^*,\ldots,\ \mathbf{T}_r^*\ ,\ldots,\mathbf{T}_B^*]'$$

is thus a random sampling from the permutation k-variate distribution of vector test statistics $\mathbf{T}$;

1.d the k-variate EDF (Empirical Distribution Function) $\hat{F}_B(\mathbf{z}\,|\,\mathbf{X})$

$$\hat{F}_B(\mathbf{z}\,|\,\mathbf{X}) = [1/2+\textstyle\sum_r \mathbf{I}(\mathbf{T}_r^* \le \mathrm{z})]/(B+1),\ \forall \mathbf{z}\in \mathbb{R}^k,$$

where $\mathbf{I}(\cdot)$ is the indicator function, and gives an estimate of the corresponding k-dimensional permutation distribution $F_B(\mathbf{z}\,|\,\mathbf{X})$ of $\mathbf{T}$. Moreover

$$\hat{L}_i(z\,|\,\mathbf{X}) = [1/2+\textstyle\sum_r \mathbf{I}(\mathrm{T}_{ir}^* \ge z)]/(B+1),\ i=1,...,k,$$

gives an estimate $\forall z \in \mathbb{R}^1$ of the marginal permutation significance level function $L_i(z \mid \mathbf{X}) = \Pr\{T_i^* \geq z \mid \mathbf{X}\}$; thus

$$\hat{L}_i(T_{i0} \mid \mathbf{X}) = \hat{\lambda}_i$$

gives an estimate of the marginal *p*-value $\lambda_i = \Pr\{T_i^* \geq T_{i0} \mid \mathbf{X}\}$ relative to test T_i, i=1,…,k. All these are unbiased and consistent estimates of corresponding true values;

1.e if $\hat{\lambda}_i < \alpha$, the null hypothesis H_{0i} relating to the *i*-th variable is rejected at the significance level α.

The second phase, based on a nonparametric combination of the dependent tests previously obtained, consists in the following steps:

2.a the combined observed value of the second-order test is evaluated through the same CMC results as the first phase, and is given by:

$$T_0'' = \psi(\hat{\lambda}_1, ..., \hat{\lambda}_k);$$

2.b the *r*-th combined value of vector statistics (step 1.d) is then calculated by:

$$T_r''^* = \psi(\hat{\lambda}_{1r}^*, ..., \hat{\lambda}_{kr}^*),$$

where $\lambda_{ir}^* = \hat{L}_i(T_{ir}^* \mid \mathbf{X})$, i = 1,…,k, r =1,…,B;

2.c hence, the *p*-value of combined test T'' is estimated as:

$$\lambda_\psi'' = \sum_r \mathbf{I}(T_r''^* \geq T'')/B;$$

2.d if $\lambda_\psi'' \leq \alpha$, the global null hypothesis H_0 is rejected at significant level a;

where ψ is an appropriate combining function.

Remember that, in order to preserve the underlying dependence relations among variables, permutations must always be carried out on individual data vectors, so that all component variables and partial tests must be jointly analyzed.

It can be seen that under the general null hypothesis the CMC procedure allows a consistent estimation of the permutation distributions, both marginal and combined, of the k partial tests. In the nonparametric combination procedure, Fisher's combination function is usually considered, principally for its good properties which are both finite and asymptotic (Pesarin, 2001). Of course, if it were considered appropriate, it would be possible to take into consideration any other combining function (Folks, 1984; Pesarin, 2001). The combined test is unbiased and consistent; it also has interesting asymptotic properties (Pesarin, 2001).

Figure 1 summarizes graphically the complete framework of NPC solution.

A general characterization of the class of combining functions is given by the following three main features for the combining function ψ:

a) it must be non-increasing in each argument:

$\psi(\ldots,\lambda_i,\ldots) \geq \psi(\ldots,\lambda_i',\ldots)$ *if* $\lambda_i < \lambda_i'$, $i \in \{1,\ldots,k\}$;

b) it must attain its supreme value $\bar{\psi}$, possibly non finite, even when only one argument reaches zero:

$\psi(\ldots,\lambda_i,\ldots) \rightarrow \bar{\psi}$ *if* $\lambda_i \rightarrow 0$, $i \in \{1,\ldots,k\}$;

c) $\forall \alpha > 0$, the critical value of every ψ is assumed to be finite and strictly smaller than the supreme value:

$T''_\alpha < \bar{\psi}$.

The above properties define the class C of combining functions. Some of the functions most often used to combine independent tests (Fisher, Lancaster, Liptak, Tippett, Mahalanobis, etc.) are included in this class. If in the overall analysis distinguishing the importance of partial tests by using appropriate weights opportunely fixed: $w_i \geq 0$, $i = 1,..,k$, with at least one strong inequality is considered more suitable, then the combined test using the Fisher combination is:

$$T'' = -\sum_i w_i \cdot \log(\lambda_i).$$

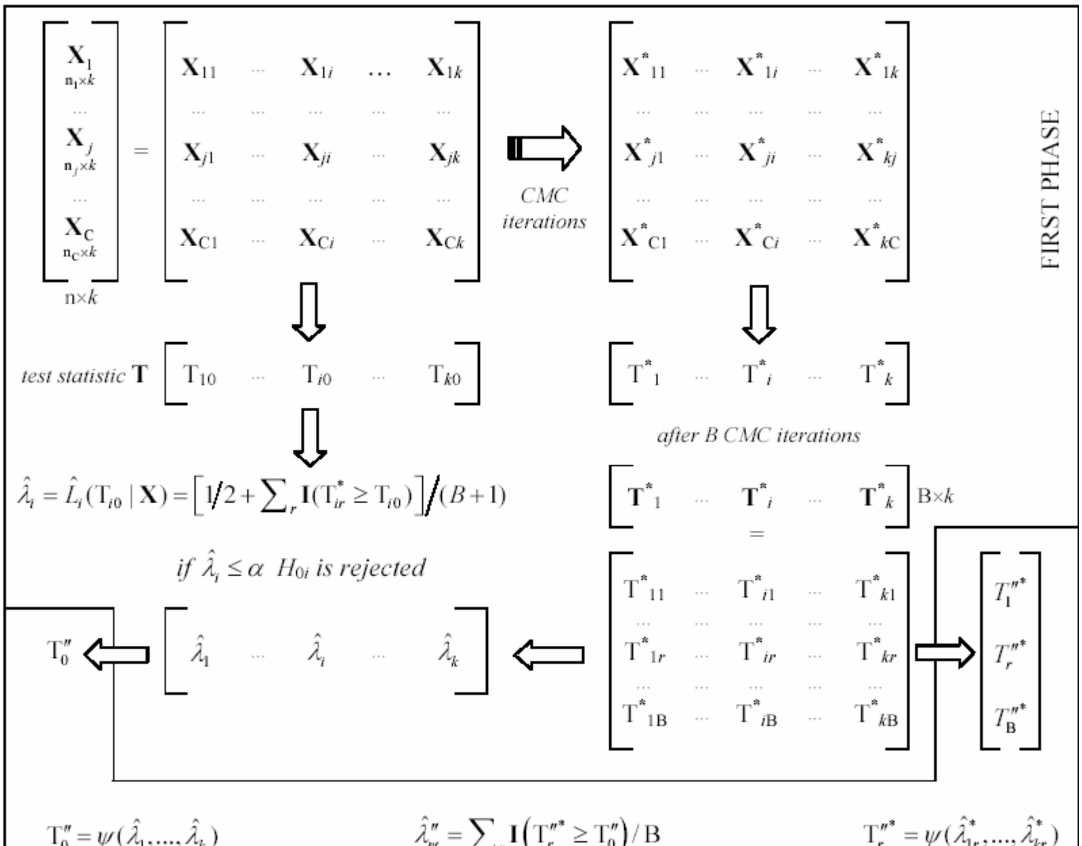

Figure 1 – Graphical description of two-phase NPC solution.

3.2. *Stratified analysis*

In many real problems a stratification factor can be considered in order to take into account some possible confounding factors such as class, sex, age and so on, or for performing a testing procedure more adherent to the complexity of the survey.

In these cases the null hypothesis (for two-group comparisons) becomes a set of S multivariate independent tests, with S equal to the number of levels of stratification variable:

$$H_0 : \bigcap_{s=1}^{S} [{}_sP_1 = {}_sP_2] = \bigcap_{s=1}^{S} [{}_s\mathbf{X}_1 \stackrel{d}{=} {}_s\mathbf{X}_2],$$

where the left-low index s=1,…,S represents the stratum.

The following table represents the data set configuration along with a graphical representation of the partial test ${}_1\mathrm{T}_1$, related to VAR_1 within the first stratum; note that the number of partial tests is equal to $S \times k$.

TABLE 1

Dataset configuration for stratified analysis

GROUP	STRATUM	VARIABLE			
		VAR_1	VAR_2	….	VAR_k
1: $\mathbf{X}_1$	1: ${}_1\mathbf{X}_1$	${}_1X_{11}$			
	2: ${}_2\mathbf{X}_1$				
	…	${}_1\mathrm{T}_1$			
	S: ${}_S\mathbf{X}_1$				
2: $\mathbf{X}_2$	1: ${}_1\mathbf{X}_2$	${}_1X_{21}$			
	2: ${}_2\mathbf{X}_2$				
	…				
	S: ${}_S\mathbf{X}_2$				

Therefore, with a stratified analysis the system of hypotheses can be rewritten in one of the following two configurations:

1. $$H_0 : \bigcap_{s=1}^{S} \left[\bigcap_{i=1}^{k} {}_sX_{1i} \stackrel{d}{=} {}_sX_{2i} \right] = \bigcap_{s=1}^{S} \left[\bigcap_{i=1}^{k} {}_sH_{0i} \right],$$

2. $$H_0 : \bigcap_{i=1}^{k} \left[\bigcap_{s=1}^{S} {}_sX_{1i} \stackrel{d}{=} {}_sX_{2i} \right] = \bigcap_{i=1}^{k} \left[\bigcap_{s=1}^{S} {}_sH_{0i} \right].$$

The first one is called within strata and stresses the importance of stratum with respect to the variables; the second one, called within variables, focuses on variables with respect to the strata.

It is clear that to obtain a global test we have to consider a new intermediate step and the testing solution becomes a three-phase algorithm. In order to display the results, Table 2 could be a particularly suitable *p*-value table.

In the table each *p*-value represents a possible significant comparison between groups, for any given variable and within a given stratum.

It is worth noting that we have to choose a priori one of the two possible second phase combinations according to the main objective of the testing problem. In fact the two analyses are different and, in general, they lead to different inferential results.

TABLE 2

p-values table for stratified analysis

STRATUM	VAR_1	VAR_2	….	VAR_k	WITHIN STRATA
1	$_1p_1$	$_1p_2$	…	$_1p_k$	$_1p_\bullet$
2	$_2p_1$	$_2p_2$	…	$_2p_k$	$_2p_\bullet$
…	…	…	…	…	…
S	$_Sp_1$	$_Sp_2$	…	$_Sp_k$	$_Sp_\bullet$
					$p_{\bullet\bullet}$
					Global Test
WITHIN VARIABLES	$_\bullet p_1$	$_\bullet p_2$	…	$_\bullet p_k$	$p_{\bullet\bullet}$

3.3. *Nested combinations*

Let us suppose that the k variables describing the testing problem can be classified into $m_1 < k$ classes according to some meaningful criteria. Moreover, the m_1 classes could themselves be put together in a further grouping, obtaining $m_2 < m_1$ classes and so on. After $T < k$ steps, this nested classification rule leads to only one final class which includes all variables. It is clear that in such a situation, before carrying out the global test by nonparametric combination of k partial tests, it is more appropriate to introduce T intermediate combination phases that reflect the meaningful classification rules. This nested procedure can be represented by a graph (Figure 2) in which, from top to bottom, each node indicates a partial test (the corresponding p-value is displayed), and each arch indicates a nonparametric combination into a higher order test.

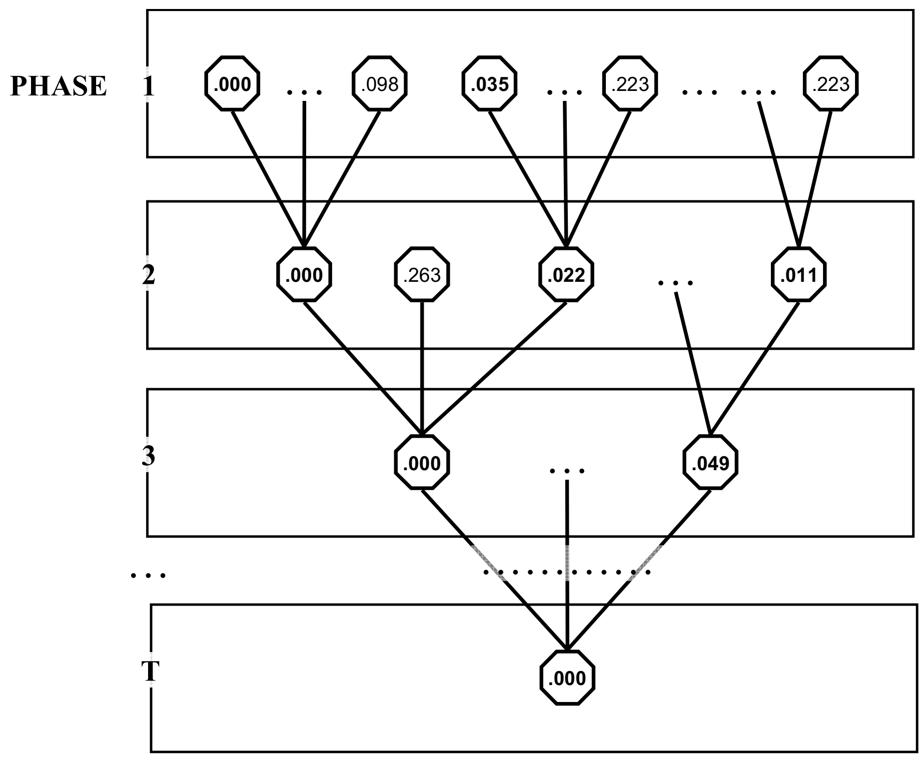

Figure 2 – Graphical representation of nested combinations.

Note that it is not necessary for all partial tests to be involved in every phase. Some could be included after a given phase.

4. NPC TEST© 2.0: STATISTICAL SOFTWARE FOR MULTIVARIATE NONPARAMETRIC PERMUTATIONS TESTS

NPC Test 2.0 (more details at www.methodologica.it) completely implements the NPC methodology offering both flexibility and a user-friendly interface. The available multivariate analyses are two or C samples with dependent variables (highlighting the dependence among responses) and two or C samples with repeated measures. Readers are reminded that in NPC Test there are no limitations to the number of observations with respect to the number of variables, i.e. there are no problems regarding a possible lack of degrees of freedom. It is possible to consider one or more stratification factors in order to solve problems with extremely complex experimental designs.

Data sets can be either created and manipulated inside the program on a normal spreadsheet or can be pasted or directly imported from most common formats. All kinds of variables are dealt with (numeric or continuous, nominal, ordered categorical or binary) and each one is provided with an appropriate set of test statistics suitable for effectively managing missing values as well. Every partial alternative hypothesis may be specified as being either one or two tailed.

Finally, all performed tests are kept in an effective report that can easily be integrated and customised by means of an efficient text editor.

5. NONPARAMETRIC COMBINATION (NPC) OF DEPENDENT RANKINGS

In many real situations we encounter the need to compare entities of a different nature (products, services, companies, behavior and so on) in order to obtain a ranking among the considered statistical units. If the comparison is based on only one feature the result is obtained in a trivial way but difficulties may arise when we are dealing with two or more informative variables jointly. We can build up as many rankings as the number of features we are dealing with. Apart from the case where units occupy the same position in every ranking, the need to summarize a set of rankings into one single global ranking arises.

The main purpose of the method (Pesarin and Lago, 2000) is to obtain a single ranking criterion for the statistical units under study, which summarizes many starting partial (univariate) criteria. This method is defined as nonparametric since it needs neither the knowledge of the underlying statistical distribution for the variables being studied, nor the dependence structure among variables, apart from the assumption that all dependences are monotonic regressions.

5.1. *Methodology*

Given a multivariate phenomenon $\mathbf{X}=[X_1, X_2,\ldots, X_k]$, observed on N statistical units, and once we have calculated the k partial rankings $R_1, R_2,\ldots, R_k$, starting from the variables X_i, $i=1,\ldots,k$, each one being informative about a partial aspect of phenomenon $\mathbf{X}$, we want to build up a global combined ranking Y:

$$Y=\psi(X_1,\ X_2,\ldots,\ X_k;\, w_1,\, w_2,\ldots,w_k),\ \psi:\mathbb{R}^{2k}\to\mathbb{R}^1,$$

where ψ is a real function allowing us to combine the partial dependent rankings and where w_1, w_2,…, w_k is a set of weights, defined on the basis of technological, functional or economic considerations, which measure the relative degree of importance among the k aspects of $\mathbf{X}$.

In order to build up $\mathbf{Y}$ we introduce a set of minimal reasonable conditions related to the variables X_i $i=1,\ldots,k$:

1. for each of the k informative variables a partial ordering criterion is well established, in the sense that "large is better"; if it is not so, it is possible to recode the variables by means of any appropriate transformation φ:
   a. if "large is worse" $\Rightarrow$ $\varphi(X)=1/X$ or $\varphi(X)=-X$;
   b. if "δ is better" (central target value) $\Rightarrow$ $\varphi(X)=|X-\delta|$;
2. regression relationships within the k informative variables are monotonic (increasing or decreasing)
3. the marginal distribution of each informative variable is non-degenerate.

Moreover, we need not make any further assumptions either on the statistical distribution of the informative variables, or on their dependence structure. Finally, notice that we do not need to assume the continuity of X_i $i=1,\ldots,k$, so that the probability of ex-equo can be different from zero.

Let us define the set of variables X_i as $\{Z_{ji}, i=1,\ldots,k, j=1,\ldots,N\}$, possibly after proper transformations. Without loss of generality, they are assumed to behave in accordance with the rule "large is better". In this setting, we consider the rank transformations R_{ji} (partial rankings):

$$\{R_{ji}= \mathbf{R}(Z_{ji}) = \#\,(Z_{ji} \geq Z_{hi}),\ i=1,\ldots,k,\ j,h=1,\ldots,N\}.$$

Associated with these ranks are the scores:

$$\left\{\lambda_{ji} = \frac{R_{ji}+0.5}{N+1},\ i=1,...,k,\ j=1,...,N\right\}.$$

Once a combining function ψ (for details of combining functions see paragraph 2.1 above) has been chosen, we compute the transformation

$$\psi: \{Y_j = \psi(\lambda_{j1},\ldots, \lambda_{jk}; w_1,\ldots, w_k), j=1,\ldots,N\},$$

and finally, applying the rank transformation, we obtain the global combined ranking Y:

$$\{Y_j= \mathbf{R}(Y_j) = \#\,(Y_j \geq Y_h),\ j,h=1,\ldots,N\}.$$

In the global ranking $\mathbf{Y}$, each statistical units is ranked in a unique way, by taking into consideration the whole set of the k informative variables.

6. CASE STUDY: WHAT DOES DISTINGUISH THE BEST FIRMS, IN THE NEW PRODUCT DEVELOPMENT (NPD) PROCESS?

Over the past decade the New Product Development (NPD) process has been analysed in a number of works, both from an academic and a practitioner's point of view (Booz *et al.* 1982; Madique and Zirger, 1985; Link, 1987; Cooper, 1990, 1993; Pittiglio *et al.*, 1995; Griffin, 1997, 1998). These works aimed at identifying NPD performance drivers, that is to say, all those practices, specific process configurations and internal business contexts which underlie the achievement of superior performances and company objectives.

However, these studies were carried out in different contexts and used both different measures of success and different methods of analysis. Griffin and Page (1993), in their literature review, identified 75 different measures previously used in papers on this topic, and classified them in the following groups: customer acceptance, financial performance, product level measures, firm based measures and program measures.

In general terms, in different industries and market types (i.e. B2C versus B2B) the relationship between drivers and performances and the appropriate set of measures of success to be considered may be different. For example, in a B2B marketplace a supplier involved in NP design, can be successful if he is able to meet the specific needs of the client at a low cost and to carry out the task within an established time (Ragatz *et al.*, 1997; Droge *et al.*, 2000). A company which produces industrial goods must consider the specific requirements of the customers and offer customised or semi-customised products. This can be done by using specific approaches and practices in NP development; for example by making an effort to develop a partnership with customers (Hartley *et al.*, 1997; Swink and Mabert, 2000; Tuten and Urban, 2001).

Recent studies have laid emphasis on the configuration of different drivers distinguishing between Best and Rest at a company level, considering the whole of the product the company developed in the last three or five years, i.e. the development program. Griffin (1997, 1998), for example, considered the NP program over a five year period and to do so, divided the sample on the basis of three sets of measures: market and financial success, relative success of the program in terms of meeting its objectives and, overall industry success. Companies were classified as best when they were in the top third of their industry for NPD success and, also, were above the mean of the entire sample regarding the relative success of the program and market–financial success.

6.1. *Context of the study, framework and key variables*

This study aims to identify the differences found in driver configurations in successful and unsuccessful firms working in a B2B marketplace in two specific industries (Machinery Manufacturing, SIC35, and Electrical, Electronic Machinery, Equipment and Supplies, SIC36). We have considered all the products developed and launched on the market by each firm in the last three years. Successful firms were those above the median position in a global ranking, taking market and financial criteria and product success criteria into account. This study has considered many different types of drivers: practices and process, strategic guide and internal environment which supports NP development.

The research considers companies which develop and produce industrial goods such as machinery, equipment and appliances to sell to other companies which use them in their production processes, or products, modules and components which will be incorporated into the client company's final products (in other words, these companies have other companies as clients, so their operations and businesses are conditioned by, for example: i) the importance of the interaction between customer and supplier, so the NP department plays an important role in designing products based on the specific needs of the customer; ii) a limited number of customers with different requirements, iii) a short distribution channel and often direct sales; iv) a different and sometimes more critical role of marketing and promotion compared to a B2C environment; v) customisation or semi-customisation of products; vi) a limited number of competitors (often companies that work in a niche or specialised market).

In this study we consider four categories of variables, referring to a three year NPD program:

- Market, Products and Financial Success;
- NPD Operational Performances (Time and Quality);
- Practices and Processes of NPD;
- NP Strategic Guide and Internal Environment.

Market, Products and Financial success

The variables belonging to this category and considered in the present study are:
- Meet Profit Goals;
- Overall Product Success;
- Meet Revenue Goals.

NPD Operational Performances

Operational Performances are those that depend on the NPD process, practices and environment support. Three types of performances are considered and are related to the time and quality dimensions of the development.
- Launch on Time;
- Time To Market Reduction;
- Product quality capability.

Practices and Processes of NPD

NPD Practices refer to a set of techniques used during the various phases of the development process. Some of them concern technological aspects, like engineering tools (CAD, QFD etc.) or the technical approach on product architecture (standardisation, modularisation and platform approach); others are concerned with organisational practices (PM, team, integration etc.).
- Engineering Tools;
- Product Architecture Approach (multi-item scale);
- Project Manager Use;
- Customer Involvement (multi-item scale);
- Integration Design – Marketing;
- Integration Design – Manufacturing;
- Supplier Involvement (multi-item scale);
- Team Use.

An NPD Process concerns the phases of the development itself and the overlapping level between these phases. The variables measure in how many cases during the development program each phase or approach has been used.
- Product Concept Development;
- Product Concept Test;
- Preliminary Design (multi-item scale);
- Late Engineering Changes (i.e. Early modifications);
- Overlapping Approach.

NP Strategic Guide and Internal Environment

NP performances and success do not only depend on best practices and well defined process but also on the internal environment which supports NP development. This support can come from the management of the company (top management support, strategic guide) and from the capabilities of the employees.

- Up – Front Capabilities (VOC) (multi-item scale);
- Top Management Support;
- NP Strategic Guide (multi-item scale);
- Company Innovation Culture;
- Technological Capabilities (multi-item scale).

Distinguishing Best and Rest firms on the basis of measures of Market, Products and Financial Success, we hypothesise that (see Figure 3, reference framework):

- H1: Best firms have higher NPD Operational Performances;
- H2: Best firms have high use of development Practices and Process drivers;
- H3: Best firms have stronger Internal Environment Support.

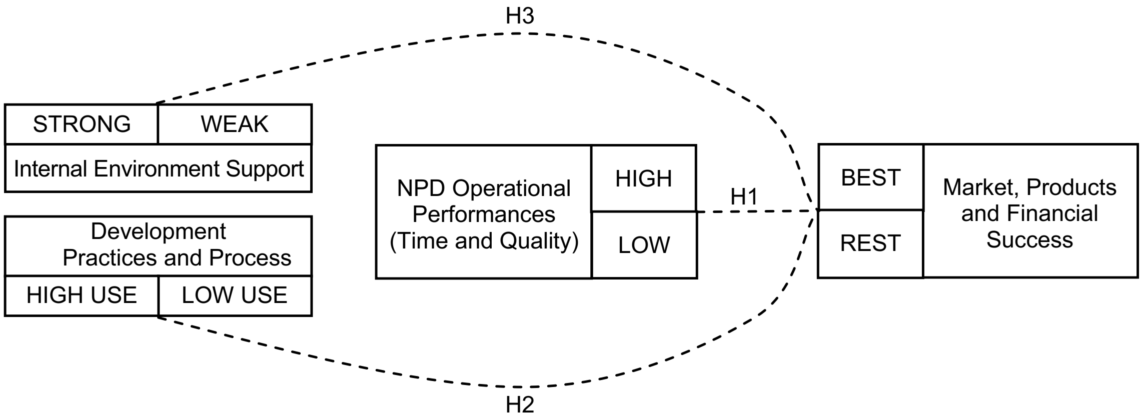

Figure 3 – Reference framework.

H1 tries to explore whether superior market, product and financial performances are associated to high time or quality performances, while H2 refers to one of the main issue in NPD studies, that is the association between success and different kinds of drivers. H3 tries to find the association with the internal environment support, which can sustain the development process.

We have considered a set of variables belonging to the above mentioned categories at NP program level. Almost all these variables have been used in the previous research mentioned in the introduction. The variables are categorical and, in general, answers are gathered considering five percentage intervals of NP that have obtained a result or performance or have used each one driver.

7. RESEARCH METHOD, SAMPLE AND DATA COLLECTION

In the empirical analysis conducted during the year 2000, we considered all NPs marketed from 1997 to 1999 by each company: this was defined as the NPD program. Market, product and financial measures of success refer to the results obtained as a result of the NPD program. For operational performances we considered the percentage of new products that have obtained high operational performances. As regards the drivers, in almost all cases we asked the company the percentage of projects which had adopted a certain driver. In other cases (i.e. capabilities and internal culture) we obtained the level of presence in the company as a whole, because it is practically impossible to discern the adoption percentage among projects for this type of variable.

Data and information were gathered through a questionnaire mailed to Italian manufacturing companies working in the B2B market in the mechanical and electronic sectors (SIC codes 35 and 36), with more than 100 and less than 1000 employees and a revenue of more than 20 billion Lire per year (approximately 10 million Euro). The addresses of the companies we mailed the questionnaire to were taken from Dun&Bradstreet's Business to Business database. The questionnaire was addressed to the new product development department manager. Phone assistance was provided to ensure that the information gathered was both complete and correct and some mangers were interviewed. The sample was made up of 85 companies. Table 3 shows the composition of the sample used for the data analysis.

TABLE 3

Sample used for the data analysis

Code	Description	N
SIC35	Machinery Manufacturing	60
SIC36	Electrical, Electronic Machinery, Equipment & Supplies	25
Tot. Sample Size		85

During the three year period considered (1997-1999), the firms launched a total of about 900 new products classified by the companies themselves as follows:

- 41% new products for new markets;
- 33% partially or totally substitute products;
- 26% products with significant improvements with respect to existing ones.

8. BEST AND REST DEFINITION

The method of nonparametric combination of dependent rankings has proved to be particularly useful for the problem of finding a meaningful classification criterion for the sample in groups, distinguishing companies which develop successful products from those which develop less successful ones. In fact, once we have applied the method to the variables measuring the Market, Products and Financial success, we obtain a global ranking of the firms, taking into account all the three success criteria. Therefore, in this global combined ranking the successful companies are those in the upper positions while the worst companies are those in the lower positions. As a discrimination rule we adopt the median positions:

those companies above the median position in the global ranking are chosen as Best companies and the remaining are labelled Rest companies.

In this way the total sample of 85 firms is divided into 40 Best companies and 45 Rest companies.

As a sensitivity analysis we perform an NPC testing procedure to verify whether the division is significant or not, that is to say whether Best companies reveal a significantly higher level of success variables. As the associated *p*-values in Table 4 show, we can verify that at a 5% significance α-level the Best companies are characterized by higher levels in all three success criteria and in the global test, taking into account the multivariate distribution of all three variables (for the interpretation of the label Innovation see the next paragraph).

TABLE 4

Sensitivity analysis on best/rest definition

Innovation	*Success* Meet Profit Goals	Overall Prod. Succ.	Meet Rev. Goals	**Combined**
Low	**.002**	**.000**	**.004**	**.000**
High	**.012**	**.048**	**.000**	**.000**
Global				**.000**

9. STRATIFICATION BY TECHNOLOGICAL INNOVATION

In order to test the hypothesis that the innovation level of the new developed products could in some way affect the results of the comparison between Best and Rest, we stratify the firms into two classes, i.e. High and Low Technological Innovators, according to the fact that all or the majority of the new products developed by the Firm are technologically innovative compared to products developed in previous years. Such firms are called High Innovators and the remaining firms are labelled Low Innovators, i.e. none, a few or about half of their newly developed products are technologically innovative compared to products developed in previous years. The sample composition resulting from the stratification procedure is shown in the following table.

TABLE 5

Sample composition after stratification

Innovation	**Best/Rest** B	R	Tot.
Low	16	24	40
High	24	21	45
Tot.	40	45	85

10. RESULTS AND DISCUSSION

The NPC Test aims to identify the significant differences of the considered variables which characterise Best firms versus Rest, i.e., high performers versus low.

A set of *p*-value tables is presented below for each one of the three tested hypotheses, reflecting the particular features of the testing problem considered, that is (i) the nested data set configuration in correspondence to multi-item scale variables (we do not use the graph representation, as in Figure 2, but a standard table), and (ii) the stratification factor given by innovation level. As a result the testing procedure is split up into the following steps:

1.1. is only for multi-item variables (they are included in H2 and H3 hypotheses), performing the partial tests in each of the two stratum (Low and High Innovation) and
1.2. combines them within strata;
2. performs the other partial tests, in each stratum, for the remaining variables and
3. combines them within strata, along with the within strata combinations of step 1.2;
4. finally, combines the two combined tests of step 3 (one for each stratum) in a global final test, which is informative on the global null hypothesis.

As far as H1 is concerned, Best firms have significantly superior levels at 5% α-level of Operational Performances from a multivariate point of view (*p*-value=.012).

TABLE 6

H1: best firms have higher NPD operational performances

Innovation	*Operational Performances*			Combined
	Launch on Time	Time to Market Red.	Quality Capability	
Low	**.045**	.128	**.031**	**.017**
High	.084	.435	.111	.092
Global				**.012**

Note that only the low innovation stratum shows a significant difference (*p*-value=.017) as a result of the contribution of two variables out of three: Launch on Time and Quality Capability of Products.

Hypothesis H2 explores the differences revealed by Development Process and Practices variables. As mentioned in the first part of the paper, many studies have already investigated the drivers of success or failure, taking a lot of variables into account and sometimes offering some "golden rules" or "best practices". Nowadays, most best practices have already been adopted by companies. In our sample, for example, almost all the firms used a lot of engineering tools (i.e. CAD, DFX, etc.), Project Manager, Teams, multistage development processes and so on. This is particularly true in the case of the B2B sample considered, where firms have to develop complex, often customised or semi-customised products. Given this it is hardly surprising to find that process and practices variables are not significantly different between Best and Rest. As shown in Table 7 and 8, the only significant difference concerns Product Architecture, within a high innovation stratum.

TABLE 7

H2: best firms do not use development practices drivers greatly

Innovation	*Development Practices*								**Combined**
	Engeneering tools	Prod. Arch. Approach	Project Manager	Multifunct. Team	Integr. Des. Marketing	Integr. Des. Manufact.	Customer Involvement	Supplier Involvement	
Low	.822	.087	.821	.238	.804	.656	.088	.968	.480
High	.707	**.019**	.321	.474	.467	.609	.926	.180	.302
Global									.425

TABLE 8

H2: best firms do not use development process drivers greatly

Innovation	*Development Process*					**Combined**
	Pr. Concept Develop.	Pr. Concept Test	Prelim. Design	Late Engin. Changes	Overlapping Approach	
Low	.635	.554	.686	.862	.493	.854
High	.181	.321	.083	.805	.121	.178
Global						.436

By analysing the multi-item variables of the H2 hypothesis in detail (Table 9), we highlight the fact that high innovator best firms prove to be significantly superior in the use of the Product Architecture Approach compared to rest firms. More specifically, two out of the three aspects of Product Architecture, i.e. standardisation and modularisation, show a difference between Best and Rest. What conclusions can we draw from this result? Some managers of the companies studied think that modular architecture and a platform logic may obtain lower costs for new products, if one considers both development costs (which in general is quite high in this type of company) and production costs. They think this is one of the main drivers for determining financial and revenue success. On the contrary, lower costs make it possible to satisfy one of the dimensions of Quality Capability, namely the price of products.

Moreover the architecture based innovation gives a competitive advantage, especially in the market segments where the technological push is more important. On the other hand this is not a differentiating driver in those segments where other aspects are more important in order to obtain a low cost, e.g. the volume of production or the cost of supplies.

TABLE 9

H2: multi-item variables of development practices and process

Innovation	*Product Architecture Approach* Standardis.	Modularisat.	Platform	**Combined**
Low	.689	**.012**	.238	.087
High	**.005**	**.026**	.246	**.019**

Innovation	*Customer Involvement* Customer Involv. 1	Customer Involv. 2	**Combined**
Low	.281	**.040**	.088
High	.847	.901	.926

Innovation	*Supplier Involvement* Supplier Involv. 1	Supplier Involv. 2	**Combined**
Low	.951	.876	.968
High	.435	.085	.180

Hypothesis H3 tests the difference in internal environment support between Best and Rest firms.

TABLE 10

H3: best firms have stronger internal environment support

Innovation	*Internal Environment Support* Top Manag. Support	NP Str. Guide	Up – Front Cap.	Tech. Capab.	Culture	**Combined**
Low	.115	.588	.436	**.000**	.056	**.007**
High	.153	**.032**	.531	.079	.060	**.049**
Global						**.003**

Results confirm the extreme importance of this kind of variable for success, showing a global *p*-value of .003 and two within strata combined *p*-values, both significant at 5% α-level, where the result is particularly strong within low innovation strata. For high innovators the most relevant variable is the Technical Capabilities while for low innovators it is the NP Strategic Guide.

Table 11 shows the *p*-values associated with multi-item variables of H3 hypothesis in detail.

TABLE 11

H3: multi-item variables of internal environment support

Innovation	*NP Strategic Guide*			Combined
	NP Str. Guide 1	NP Str. Guide 2	NP Str. Guide 3	
Low	.638	.467	.545	.588
High	**.021**	.095	.178	**.032**

Innovation	*Up – Front Capab. (VOC)*		Combined
	Up – Front Cap. 1	Up – Front Cap. 2	
Low	.606	.271	.436
High	.337	.795	.531

Innovation	*Technological Capabilities*		Combined
	Tech. Capab. 1	Tech. Capab. 2	
Low	**.000**	.099	**.000**
High	**.034**	.272	.079

11. SOME REMARKS ON THE CASE-STUDY

In short, Best firms have better NP Operational Performances and stronger internal environment support than Rest firms.

Best firms are those that perform better in Launch on Time and Quality Capability of Products. These two latter performances, in a B2B market, could be considered order winning criteria, while Time To Market Reduction seems to be an order qualifying criteria. Internal Environment Support plays a major role in Market, Products and Financial success. The results obtained are interesting because of the specific characteristics of the companies considered (B2B market, Mechanical and Electronics Industries). The contingent approach of the research appears to offer important possibilities for research in this field of and such targeting should perhaps be encouraged as a valid alternative to studies carried out on heterogeneous samples.

Furthermore the stratified analysis highlights the differences between drivers within each stratum. These results show that the successful development strategy has to consider the various contextual factors characterising the companies environment. Thus, for example, taking the different required level of technical innovation into consideration, some drivers show different levels of importance in order for the company to be successful.

This research could be improved in several directions. Firstly, the sample is an Italian sample, and thus the results cannot be generalised. Secondly, the sample is rather small and companies should be encouraged to participate in future research. Lastly, the fact that no results were obtained for some of the drivers which are, traditionally, considered to be important, such as teams, project management or engineering tools, requires further, additional study.

12. CONCLUSIONS

The nonparametric combination of dependent permutation partial tests is a method for the combination of significance levels or rejection probabilities. As we have shown in the developed example, the nonparametric combination method is suitable and effective for many multivariate testing problems which, in a parametric framework, are very difficult or even impossible to solve.

One major feature of the nonparametric combination of dependent tests, provided the permutation principle applies, is that one must pay attention to a set of partial tests, each appropriate for the related sub-hypotheses, because the underlying dependence relation structure is nonparametrically and implicitly captured by the combining procedure. In particular, the researcher is not explicitly required to specify the dependence structure of response variables. This aspect is of great importance especially for non-normal or categorical variables, in which dependence relations are generally too difficult to define and, even when well-defined, are hard to cope with (see Joe, 1997). The researcher is only required to make sure that all partial tests are marginally unbiased, a sufficient condition which is generally easy to check.

It should be emphasized that the nonparametric combination procedure may also be effective when one overall test is not directly available. In such a situation, it is usually convenient to analyse data first by examining a set of k partial aspects, each interesting in a marginal sense, and then to combine all captured information, provided that side-assumptions allow for the proper breakdown of hypotheses and the k partial tests are marginally unbiased.

In principle, it is possible to apply a proper single overall permutation procedure directly, if known, and then to avoid the combination step. But in most complex situations such a single test is not directly available or is not easy to justify. In addition, the direct analysis of the dependence relation structure is often very difficult, in the general case, due to nonlinear regression forms, monotonic functional relationships, heteroscedasticity, or other irregularities caused by categorical and/or mixed data, missing values, repeated measurements, etc.

In a way, the nonparametric combination procedure for dependent tests may be viewed as a two-phase testing procedure. The first phase considers a simulation from the permutation sample space $\mathcal{X}_{/x}$ by means of a CMC method based on B iterations, in order to estimate $F(\mathbf{z}|\mathbf{X})$. The second phase considers the combination of estimated p-values of partial tests, in order to estimate the overall p-value λ'' by using the same CMC results as the first phase. Of course, the two phases are jointly processed, so that the procedure always remains multivariate in its own right.

As a final remark, from a general point of view and in very mild conditions, the nonparametric combination method may be considered as a way of reducing the degree of complexity of most testing problems.

Dipartimento di Tecnica e Gestione dei sistemi industriali LIVIO CORAIN
Università di Padova LUIGI SALMASO

REFERENCES

F. BOOZ, A. G. ALLEN, L. S. HAMILTON (1982), *New Products Management for the 1980s*, Booz, Allen and Hamilton, New York.

R.G. COOPER (1990), *New Products: What Distinguishes the Winners*, "Research and Technology Management", November - December, 27-31.

R.G. COOPER (1993), *Winning at New Products: Accelerating the Process from Idea to Launch*, Addison-Wesley, Reading, MA.

C. DROGE, J. JAYARAM, S.K. VICKERY (2000), *The Ability to Minimize the Timing of New Product Development and Introduction: An Examination of Antecedent Factors in the North American Automobile Supplier Industry*, "Journal of Product Innovation Management", 17(1), pp. 24-40.

J.L. FOLKS (1984), *Combinations of independent tests*, in P.R. Krishnaiah and P.K. Sen (eds.), *Handbook of Statistics*, 4, 113-121, North-Holland, Amsterdam.

A. GRIFFIN (1997), *PDMA Research on New Product Development Practices: Updating Trends and Benchmarking Best Practices*, "Journal of Product Innovation Management", 14(6), pp. 429-458.

A. GRIFFIN (1998), *Drivers of new product development success: multivariate results from PDMA's best practices research*, in: "EIASM - 5th International Product Development Management Conference", Como, I, pp. 491-502.

A. GRIFFIN, A.L. PAGE (1993), *An Interim Report on Measuring Product Development Success and Factors*, "Journal of Product Innovation Management", 10(3), pp. 291-308.

J.L. HARTLEY, B.J. ZIRGER, R.R. KAMAT (1997), *Managing the buyer – supplier interface for on-time performance in product development*, "Journal of Operations Management", 15(1), pp. 57-70.

M. HOLLANDER, D.A. WOLFE (1999), *Nonparametric Statistical Methods*, 2nd Edition, Wiley, New York.

H. JOE (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.

P.L. LINK (1987), *Keys to New Product Success and Failure*, "Industrial Marketing Management", 16, pp. 109-118.

M.A. MADIQUE, B.J. ZIRGER (1985), *A Study of Success and Failure in Product Innovation: The Case of the U.S. Electronics Industry*, "IEEE Transactions on Engineering Management", 31(4), pp. 192-203.

L.S. NELSON (1992), *A randomization test for ordered alternatives*, "Journal of Quality Technology", 24, pp. 51-53.

F. PESARIN, A. LAGO (2000), *Nonparametric combination of dependent rankings with application to the quality assessment of industrial products*, "Metron", LVIII, 1-2.

F. PESARIN (2001), *Multivariate Permutation Tests with Applications in Biostatistics*, Wiley, Chichester.

F. PESARIN (2002), *Extending permutation conditional inferences to unconditional ones*, "Journal of the Italian Statistical Society", 11, 161-173.

PITTIGLIO, RABIN, TODD & MCGRATH (1995), *Product Development leadership for Technology – Based Companies: Measurement and Management – A prelude to Action*, Pittiglio Rabin Todd & McGrath, Weston, MA.

G.L. RAGATZ, R.B. HANDFIELD, T.V. SCANNELL (1997), *Success Factors for Integration Suppliers into New Product Development*, "Journal of Product Innovation Management", 14, pp. 190-202.

M.L. SWINK, V.A. MABERT (2000), *Product Development Partnership: Balancing the Needs of OEMs and Suppliers*, "Business Horizons", 43(3), pp. 59-68.

T.L. TUTEN, D.J. URBAN (2001), *An Expanded Model of Business-to-Business Partnership Formation and Success*, "Industrial Marketing Management", 30(2), pp. 149-164.

RIASSUNTO

Uno studio empirico sul processo di sviluppo del nuovo prodotto attraverso la metodologia Nonparametric Combination (NPC) testing e post-stratificazione

Questo lavoro analizza, attraverso una applicazione empirica della metodologia di verifica di ipotesi della Combinazione NonParametrica (NPC), il diverso comportamento che distingue quelle aziende che sviluppano prodotti di successo da quelle che sono di minor successo. La metodologia della Combinazione NonParametrica (NPC) di test di permutazione dipendenti, particolarmente utile per studi osservazionali ed in presenza di dati non normali e/o categoriali, consiste in un innovativo metodo di verifica di ipotesi che consente al ricercatore di superare molti limiti dei test parametrici, come la natura multivariata della maggior parte dei problemi reali e la dimensione relativamente piccola dei dataset disponibili, ove talora il numero di variabili è maggiore del numero di osservazioni disponibili.

SUMMARY

An empirical study on new product development process by Nonparametric Combination (NPC) testing methodology and post-stratification

This paper explores through an empirical application of NonParametric Combination (NPC) testing methodology, the different behaviours that distinguish those firms that develop successful products from those that are less successful. The NonParametric Combination (NPC) of dependent permutation tests methodology, particularly useful with observational studies and in presence of non-normal and/or categorical data, consists of an innovative testing method that allows the researcher to go beyond some usual parametric testing constraints, such as the multivariate nature of most real problems and the relative small size of the available datasets, when sometimes the number of variables can be greater than the number of available observations.