KATHOLIEKE UNIVERSITEIT
LEUVEN

Faculty of Economics and Applied Economics

Testing for complementarity and substitutability in the case of multiple practices

Boris Lokshin, Martin Carree, and René Belderbos

DEPARTMENT OF MANAGERIAL ECONOMICS, STRATEGY AND INNOVATION (MSI)

MSI 0708

Testing for complementarity and substitutability in the case of multiple practices

Boris Lokshin[a], Martin Carree[a], and René Belderbos[b]

February 2007

Abstract

Recent empirical studies of firm-level performance have been concerned with establishing potential complementarity between more than two organizational practices. These papers have drawn conclusions on the basis of potentially biased estimates of pair-wise interaction effects between such practices. In this paper we develop a consistent testing framework based on multiple inequality constraints that derives from the definition of (strict) supermodularity as suggested by Athey and Stern (1998). Monte Carlo results show that the multiple restrictions test is superior for performance models with high explanatory power. If practices explain only a minor part of organizational performance no test is able to identify complementarity or substitutability in a satisfactory manner.

Keywords: Complementarity; Supermodularity; Firm performance
JEL-code: C12; D24

Corresponding Author
Prof.dr. Martin Carree
Department of Organization and Strategy
Faculty of Economics and Business Administration
Universiteit Maastricht
PO Box 616, 6200 MD Maastricht
The Netherlands
Phone: +31 43 3884981
Fax: +31 43 3254893
M.Carree@os.unimaas.nl

[a]Universiteit Maastricht
[b]Katholieke Universiteit Leuven and Eindhoven University of Technology

The authors thank Daron Acemoglu, Franz Palm, Pierre Mohnen and Scott Stern for helpful comments. The usual disclaimer applies.

1. Introduction

Researchers in the fields of industrial organization and management have long been interested in investigating complementary relations between various organizational practices of a firm. Complementarity is understood in this context to exist if the implementation of one practice increases the marginal or incremental return to other practices. Thus a joint implementation of several practices may result in economies of scope in a sense proposed by Baumol et al. (1988).[1] By the same token the implementation of one practice can decrease the marginal or incremental return to other practices. This is the case of substitutability (or subadditivity). Examples of studies of complementarity in the economics and management literature are the relationships between human resource practices and firm strategy (Ichniowski et al., 1997), firms' internal R&D and external technology sourcing (Arora and Gambardella, 1994), process and product innovation (Miravete and Pernias, 2004), labour skill and innovation strategies (Leiponen, 2005), different government innovation policies (Mohnen and Röller, 2005), information technology, workplace reorganization, and new product and service innovations (Black and Lynch, 2001; Bresnahan et al, 2002; Caroli and Van Reenen, 2001), the adoption of different information technologies in emergency health care (Athey and Stern, 2002), and between different types of labor in the determination of trade patterns (Grossman and Maggi, 2000). Siggelkow (2002) presents a theoretical model on the organizational consequences of the importance of nonsimple interactions among complementary and substitute activities.

There are two econometric approaches that can be used to test for complementarity: the "adoption' or correlation approach and the "production function" approach (e.g. Athey and Stern, 1998). The former has been popular among empirical researchers due to its simplicity (Arora, 1996). The adoption approach tests conditional correlations based on the residuals of reduced form regressions of the practices of interest on all observable exogenous variables. However, although this test can serve as supportive evidence of complementarity if practices are adopted simultaneously, it cannot serve as a definitive test. Estimated correlations between residual terms may be the result of common omitted exogenous variables or measurement errors. Even in the case of well-measured correlation between practices, there is no guarantee that decision markers were sufficiently well informed such that they indeed chose efficiency or output enhancing combinations of practices.

The 'production function' approach, in which organizational performance is related to combinations of organizational practices, does not have these drawbacks and can serve as a direct test for

[1] The related definition of supermodularity in Milgrom and Roberts (1990) is broader, as it only requires a non-negative (rather than a positive) impact of one practice on the marginal return to another practice.

complementarity or substitutability.[2] Complementarity can be investigated by examination of the cross derivative of two practices. This approach has been used in recent empirical work testing for complementarity between two practices (e.g. Veugelers and Cassiman, 2006), in which case a complementarity or substitutability test is a simple one-tailed t-test on the interaction term of the two practices. However, no robust testing procedure has been available to test for complementarity or substitutability with more than two practices, which has prevented a wide use of the production function approach in applied empirical work.[3] Studies that did adopt the production function approach have limited analysis to the estimation and examination of pair-wise interaction effects, either including all pair-wise terms (e.g. Black and Lynch, 2001; Bresnahan et al, 2002), or estimating alternating pair-wise interactions (Caroli and Van Reenen, 2001). This approach is potentially problematic. Since it ignores the impact of additional cross-terms (e.g. a triple term in case of three practices), it examines only a partial expression for the cross derivative and is prone to an omitted variable bias that affects all coefficients. As noted by Athey and Stern (1998), a proper complementarity or substitutability test requires a testing framework that considers the complete set of organizational practices. In this paper we develop such a test based on a multiple inequality restrictions framework (e.g. Kudô, 1963; Wolak, 1989) corresponding to a definition of strict supermodularity (Milgrom and Roberts, 1990). We provide Monte Carlo results comparing the power of this test with the performance of the two pair-wise tests.

The remainder of this paper is organized as follows. In the next section we review the definitions of complementarity and substitutability. Section three details the testing procedure in the case of more than two (continuous or dichotomous) practices. The Monte Carlo evidence is discussed in section four and section five concludes.

2. Complementarity and substitutability

We describe the definitions and conditions concerning complementarity and substitutability both for the case of continuously measured practices and the case of dichotomous practices. Consider an objective function[4] f of which the value is determined by the practices x_p (p=1,...,n). In case the

[2] That is, as long as the population of organizations includes a reasonable number of organizations that take non-optimal combinations of practices, e.g. because they are not well informed or face high adaptation costs. In addition, the production function approach is only reliable if there are no important performance enhancing variable omitted from the model that are positively correlated to the adoption of two or more practices.

[3] Mohnen and Roller (2005) and Leiponen (2005) do adopt a multiple inequality restrictions framework, but their testing framework does not allow for confirmation or rejection of hypotheses for the complete range of possible outcomes.

[4] The proposed framework assumes that the function $f(x,\varepsilon)$ is correctly specified and that the distributional assumptions for the error term ε are satisfied.

practices are measured continuously the following definition of complementarity holds (e.g. Baumol et al., 1988):

Definition 1 (continuous practices)

Practices x_i and x_j are considered complementary in the function f if and only if $\partial^2 f / \partial x_i \partial x_j \geq 0$ *for all values of* $(x_1, ..., x_n)$ *with the inequality holding strictly for at least one value.*

The second part of the definition, requiring the cross derivative to be positive for at least one value of the other practices, makes the definition more stringent than the definition of superrmodularity proposed in Milgrom and Roberts (1990), but is arguably the most relevant (economic) definition of complementarity.[5] The definition for substitutability is identical as definition 1 except that 'larger' is replaced by 'smaller'. We use a cross-term specification of the objective function *f* to test for complementarity or substitutability. The expressions for *n* equal to 2, 3 and 4 are:

$$f(x_1, x_2) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_{12} x_1 x_2 \quad (1)$$

$$f(x_1, x_2, x_3) = f(x_1, x_2) + \alpha_3 x_3 + \alpha_{13} x_1 x_3 + \alpha_{23} x_2 x_3 + \alpha_{123} x_1 x_2 x_3 \quad (2)$$

$$f(x_1, x_2, x_3, x_4) = f(x_1, x_2, x_3) + \alpha_4 x_4 + \alpha_{14} x_1 x_4 + \alpha_{24} x_2 x_4 + \alpha_{34} x_3 x_4 + + \alpha_{124} x_1 x_2 x_4 + \alpha_{234} x_2 x_3 x_4 + \alpha_{1234} x_1 x_2 x_3 x_4 \quad (3)$$

The cross-derivatives $\partial^2 f / \partial x_1 \partial x_2$ are equal to α_{12} for equation (1), $\alpha_{12} + \alpha_{123} x_3$ for equation (2) and $\alpha_{12} + \alpha_{123} x_3 + \alpha_{124} x_4 + \alpha_{1234} x_3 x_4$ for equation (3), respectively. This implies that there is complementarity for the case of practices 1 and 2 if $\alpha_{12} > 0$. In case of three practices, $\alpha_{12} + \alpha_{123}\, min(x_3) \geq 0$ and $\alpha_{12} + \alpha_{123}\, max(x_3) \geq 0$ with at least one of the inequalities holding. In case of four practices there are four inequalities of which at least one should hold strictly:

[5] This definition is also referred to as 'strict supermodularity' (e.g. Leiponen, 2005). Since the test is based on a supermodularity framework, it can easily be applied to determine the existence of strict super- and submodularity of the objective function in organizational design practices (e.g. Milgrom and Roberts, 1990). In case all bilateral combinations of practices satisfy complementarity, the objective function is strictly supermodular.

$$\alpha_{12} + \alpha_{123}\, min(x_3) + \alpha_{124}\, min(x_4) + \alpha_{1234}\, min(x_3) min(x_4) \geq 0,$$

$$\alpha_{12} + \alpha_{123}\, min(x_3) + \alpha_{124}\, max(x_4) + \alpha_{1234}\, min(x_3) max(x_4) \geq 0,$$

$$\alpha_{12} + \alpha_{123}\, max(x_3) + \alpha_{124}\, min(x_4) + \alpha_{1234}\, max(x_3) min(x_4) \geq 0,$$

$$\alpha_{12} + \alpha_{123}\, max(x_3) + \alpha_{124}\, max(x_4) + \alpha_{1234}\, max(x_3) max(x_4) \geq 0$$

In case the practices take on discrete values variables (step size chosen equal to one) we replace the derivative in definition 1 by a difference. If we consider the first two practices, without loss of generality, the following definition holds:

Definition 2 (discrete practices)

Practices x_1 and x_2 are considered complementary in the function f if and only if $f(x_1+1, x_2+1, x_3, ..., x_n) + f(x_1, x_2, x_3, ..., x_n) \geq f(x_1+1, x_2, x_3, ..., x_n) + f(x_1, x_2+1, x_3, ..., x_n)$ *for all values of $(x_1, ..., x_n)$ with the inequality holding strictly for at least one value.*

The case of dichotomously measured practices (practice is used or not) is a special case of this definition. In that case functions (1), (2), and (3) can also be conveniently rewritten in terms of the possible combinations of practices (cf. Mohnen and Röller, 2005). With two practices the collection of possible combinations is defined in the usual binary order as $D = \{(0,0),(0,1),(1,0),(1,1)\}$. We introduce the indicator function $I_{D=(r,s)}$, equal to one when the combination is (r,s), else zero. Similar collections of *D* with corresponding indicators functions $I_{D=(r,s,t)}$ and $I_{D=(r,s,t,u)}$ are introduced for the case of three and four practices. The functions *f* are rewritten as:

$$f(x_1, x_2) = \sum_{r=0}^{1}\sum_{s=0}^{1} \beta_{rs} I_{(x_1, x_2)=(r,s)} \tag{4}$$

$$f(x_1, x_2, x_3) = \sum_{r=0}^{1}\sum_{s=0}^{1}\sum_{t=0}^{1} \beta_{rst} I_{(x_1, x_2, x_3)=(r,s,t)} \tag{5}$$

$$f(x_1, x_2, x_3, x_4) = \sum_{r=0}^{1}\sum_{s=0}^{1}\sum_{t=0}^{1}\sum_{u=0}^{1} \beta_{rstu} I_{(x_1, x_2, x_3, x_4)=(r,s,t,u)} \tag{6}$$

The conditions of complementarity now correspond to $\alpha_{12} = \beta_{11} + \beta_{00} - \beta_{10} - \beta_{01} > 0$ for two practices, $\alpha_{12} = \beta_{110} + \beta_{000} - \beta_{100} - \beta_{010} \geq 0$ and $\alpha_{12} + \alpha_{123} = \beta_{111} + \beta_{001} - \beta_{101} - \beta_{011} \geq 0$ for three practices and the following four inequalities for four practices:

$$\alpha_{12} = \beta_{1100} + \beta_{0000} - \beta_{1000} - \beta_{0100} \geq 0$$

$$\alpha_{12} + \alpha_{123} = \beta_{1110} + \beta_{0010} - \beta_{1010} - \beta_{0110} \geq 0$$

$$\alpha_{12} + \alpha_{124} = \beta_{1101} + \beta_{0001} - \beta_{1001} - \beta_{0101} \geq 0$$

$$\alpha_{12} + \alpha_{123} + \alpha_{124} + \alpha_{1234} = \beta_{1111} + \beta_{0011} - \beta_{1011} - \beta_{0111} \geq 0 .$$

3. The testing procedure

In case of two practices the test for global complementarity is a one-sided t-test of the null hypothesis of $\alpha_{12} = 0$ in equation (1). However, in case of more than two practices, the number of inequality constraints that have to be tested simultaneously is 2^{n-2}. Statistical tests of $H_0 : R\beta = r$ versus $H_a : R\beta \geq r$ with *R* having rank *k* in the standard linear model $y = X\beta + \varepsilon$ with one of the inequalities holding strictly have been considered in Gouriéroux, Holly, and Monfort (1982). Kudô (1963, p.414) derived the theorem underlying this test. The so-called *normal orthant probability*, $P\{ \Omega \}$, being the probability that the variables with a multivariate normal distribution with mean zero and variance-covariance matrix $\Omega = R(X'X)^{-1} R'$ are all positive, plays a central role in this theorem:

Theorem 1 (the Kudô theorem):

Let $(x_1,...,x_k)$ have a multivariate normal distribution with mean zero and known variance-covariance matrix Σ and let $LR = -2 \ln \lambda$ where λ is the likelihood ratio test statistic of $H_0 : E(x_i) = 0$ for i=1,...,k versus $H_A : E(x_i) \geq 0$ for i=1,...,k where the inequality is strict for at least one value of i. Then $Pr(LR \geq c) = \sum_{M \subseteq K} Pr(\chi^2_{n(M)} \geq c) P\{ \Sigma_B^{-1} \} P\{ \Sigma_{M:B} \}$ where the summation runs over all the subsets M of $K = \{1,...,k\}$ including ø, n(M) is the number of elements in M, B is the complement of M, so that $M \cap B = ø$ and $M \cup B = K$, Σ_B is the variance-covariance matrix of x_i with $i \in B$, $\Sigma_{M:B}$ is the same for x_i with $i \in M$ but under the condition that $x_i = 0$ for $i \in B$.[6]

From this theorem it follows that in case of *p* inequality restrictions we have that the probability of *LR* exceeding *c* under the null hypothesis equals a mixed chi-square distribution of $\sum_{i=0}^{p} Pr\{ \chi^2_i \geq c \} w_{ip}$

[6] For the empty set *M*=ø we have that $\chi^2_{n(M)} = \chi^2_0$ is a constant zero and $P\{ \Sigma_M^{-1} \} = P\{ \Sigma_{M:K} \} = 1$

(see also Shapiro, 1985, p.138 and Wolak, 1989, p.214).[7] Therefore, the p-value equals $1-\sum_{i=0}^{p} Pr\{\chi_i^2 \geq LR\} w_{ip}$. The statistic can be compared to Table 1 from Kodde and Palm (1986) who provide critical values (c_l and c_u) for significance levels ranging in size from 0.25 to 0.001 and degrees of freedom from 1 to 40. In case the computed value falls in the indecision region, an exact p-value must be computed. The weights for two restrictions (n = 3) are $w_{02} = cos^{-1}(\Omega_{12} / \sqrt{\Omega_{11}\Omega_{22}})/2\pi$ where $\Omega_{ij} = R_i(X'X)^{-1}R'_j$ with R_j being the jth row of R, w_{12}= ½ and w_{22} = ½ - w_{02} (Shapiro, 1985).

For four (n = 4) or more restrictions, computation of weights requires some more work. The normal orthant probability plays a central role in this computation.[8] The weights w_{pp} and w_{0p} are equal to $P\{\Omega\}$ and $P\{\Omega^{-1}\}$, respectively, where Ω is the positive-definite covariance matrix of $(x_1,...,x_p)$. Define $P = \{1,...,p\}$ and $M(k)$ the subsets of P of exactly k elements ($\binom{p}{k}$ in number). The weights w_{kp} where $k = 1,...,p-1$ are then as follows:

$$w_{kp} = \sum_{M(k)\subseteq P} P\{\Omega_{M(k),11} - \Omega_{M(k),12}\Omega^{-1}_{M(k),22}\Omega_{M(k),21}\} P\{\Omega^{-1}_{M(k),22}\} \tag{7}$$

where $\Omega_{M(k),11}$ is the kxk-matrix obtained from Ω after only keeping the rows and columns corresponding to the elements of $M(k)$, $\Omega_{M(k),12}$ is the kx$(p$-$k)$-matrix obtained from Ω after keeping the rows corresponding to the elements of $M(k)$ and the columns corresponding to all the elements of P that are not in $M(k)$, $\Omega_{M(k),21}$ is the $(p$-$k)$xk-matrix obtained from Ω after keeping the rows corresponding to all the elements of P that are not in $M(k)$ and the columns corresponding to the elements of $M(k)$, and $\Omega_{M(k),22}$ is the $(p$-$k)$x$(p$-$k)$-matrix obtained from Ω after keeping the rows and columns corresponding to all the elements of P that are not in $M(k)$.

---

[7] Because $Pr\{\chi_0^2 \geq a\} = 0$ for all a, the summation could also run from 1 up till p. In empirical applications the variance-covariance matrix has to be estimated and the mixed chi-square distribution only holds asymptotically.

[8] Several methods are available for numerical computation of the multivariate normal integral, see e.g. Sun (1988), Genz (1993) and Hajivassiliou et al. (1996).

We illustrate (7) for the case of four practices and, hence, p equal to 4. For four practices we have that w_{24} = 1- w_{04} - w_{14} - w_{34} - w_{44} where w_{14} and w_{34} are as follows:[9]

$$w_{14} = P\{\sigma_{11}\}P\left\{\begin{pmatrix} \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{23} & \sigma_{33} & \sigma_{34} \\ \sigma_{24} & \sigma_{34} & \sigma_{44} \end{pmatrix}^{-1}\right\} + P\{\sigma_{22}\}P\left\{\begin{pmatrix} \sigma_{11} & \sigma_{13} & \sigma_{14} \\ \sigma_{31} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{43} & \sigma_{44} \end{pmatrix}^{-1}\right\}$$
$$+P\{\sigma_{33}\}P\left\{\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{24} \\ \sigma_{41} & \sigma_{42} & \sigma_{44} \end{pmatrix}^{-1}\right\} + P\{\sigma_{44}\}P\left\{\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}^{-1}\right\}$$

$$w_{34} = P\left\{\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{41} & \sigma_{32} & \sigma_{33} \end{pmatrix} - \begin{pmatrix} \sigma_{14} \\ \sigma_{24} \\ \sigma_{34} \end{pmatrix}\begin{pmatrix} \sigma_{14} \\ \sigma_{24} \\ \sigma_{34} \end{pmatrix}^{\mathrm{T}} / \sigma_{44}\right\}P\{\sigma_{44}^{-1}\} + P\left\{\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{41} & \sigma_{32} & \sigma_{33} \end{pmatrix} - \begin{pmatrix} \sigma_{14} \\ \sigma_{24} \\ \sigma_{34} \end{pmatrix}\begin{pmatrix} \sigma_{41} \\ \sigma_{42} \\ \sigma_{43} \end{pmatrix}^{\mathrm{T}} / \sigma_{33}\right\}P\{\sigma_{33}^{-1}\}$$
$$+P\left\{\begin{pmatrix} \sigma_{11} & \sigma_{13} & \sigma_{14} \\ \sigma_{31} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{43} & \sigma_{44} \end{pmatrix} - \begin{pmatrix} \sigma_{21} \\ \sigma_{32} \\ \sigma_{42} \end{pmatrix}\begin{pmatrix} \sigma_{12} \\ \sigma_{23} \\ \sigma_{24} \end{pmatrix}^{\mathrm{T}} / \sigma_{22}\right\}P\{\sigma_{22}^{-1}\} + P\left\{\begin{pmatrix} \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{42} & \sigma_{43} & \sigma_{44} \end{pmatrix} - \begin{pmatrix} \sigma_{21} \\ \sigma_{31} \\ \sigma_{41} \end{pmatrix}\begin{pmatrix} \sigma_{12} \\ \sigma_{13} \\ \sigma_{14} \end{pmatrix}^{\mathrm{T}} / \sigma_{11}\right\}P\{\sigma_{11}^{-1}\}$$

4. Monte Carlo Experiments

We compare the performance of the multiple-restrictions test with two alternative pair-wise test procedures used in recent empirical work. The alternating "single cross-term" test only incorporates the cross term of two practices at a time in the estimated equation, and infers complementarity from the estimated coefficient of the cross-term (e.g. Bresnahan et al., 2002; and Black and Lynch, 2001). The "all cross-term" test follows the same procedure but incorporates all pair-wise cross-terms $x_i x_j$ $i \neq j$ in one equation (e.g. Caroli and Van Reenen, 2001). We devise a Monte Carlo experiment to compare the power of the three test procedures. Since almost all empirical studies of complementarity in the literature examine the impact of discrete practices, we focus the experiment on the case of dichotomous variables (variables taking the values 0 or 1). We consider a performance function in the case of three practices x_1, x_2, x_3 as in equation (5); for the purpose of comparing tests we write this function in its cross-term specification:

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_{12} x_1 x_2 + \alpha_{13} x_1 x_3 + \alpha_{23} x_2 x_3 + \alpha_{123} x_1 x_2 x_3 + \varepsilon \qquad (8)$$

---

[9] In practice w_{24} is computed as $w_{24} = \sum_{i=1}^{3}\sum_{j=i+1}^{4} q_{ij}$ where $q_{ij} = P\{\Omega_{M(k),ii} - \Omega_{M(k),ij}\Omega^{-1}_{M(k),jj}\Omega_{M(k),ji}\}P\{\Omega^{-1}_{M(k),jj}\}$ and then using the summation of all weights to unity as a check of correct computation.

where $\varepsilon \sim N(0,\sigma_{\varepsilon}^{2})$. There is complementarity between practices 1 and 2 if $\alpha_{12} \geq 0$ and $\alpha_{12} + \alpha_{123} \geq 0$ with at least one of the two inequalities holding strictly. The multiple restriction specifically tests for this. The single cross term test, on the other hand, imposes $\alpha_{13} = \alpha_{23} = \alpha_{123} = 0$ and judges complementarity to exist if $\alpha_{12} > 0$. The multiple cross-term test applies the same criterion but only imposes $\alpha_{123} = 0$.

The data for our experiments are generated for a sample of 1000 observations. In the first step the coefficients α_1 through α_{123} are randomly and independently drawn from the standard normal distribution. In the second step, variables z_1, z_2, z_3 are drawn from the multivariate standard normal distribution. Variables x_1, x_2, x_3 are equal to one when z_1>0, z_2>0 and z_3>0, respectively, else zero. In order to mimic empirical research settings, the correlation structure between the practices is allowed to depend on the presence of complementarity or substitutability. If organizations possess (imperfect) information on the true state of the performance relationships between the practices and face moderate adoption costs, they are more likely to simultaneously adopt two practices if these are complementary. In case the draws of α_1 through α_{123} indicate complementarity, the correlation coefficient between x_1 and x_2 is set at 0.5 and in case of substitutability at -0.5. The correlation coefficient is set at zero if the draw indicates no complementarity or substitutability.[10] The remaining correlations are selected to make the matrix positive-definite. Their magnitudes have little effect on the test outcome. Equation (8) is used to generate data for *y*, with the relevant restrictions on parameters imposed in case of the pair-wise tests.

The outcome of the multiple-restrictions test is established by determining the log-likelihood of the unrestricted model (LL_U), the model under the complementarity constraints $\alpha_{12} \geq 0$ and $\alpha_{12} + \alpha_{123} \geq 0$ (LL_C), the model under the substitutability constraints $\alpha_{12} \leq 0$ and $\alpha_{12} + \alpha_{123} \leq 0$ (LL_S), and the model with the restriction of $\alpha_{12} = \alpha_{123} = 0$ (LL_0). The test outcomes are complementarity, substitutability, or neither. The multiple restrictions test entails first considering whether LL_C is higher than LL_S or vice versa and, depending on this comparison, testing whether LL_0 is significantly higher than LL_C or LL_S at the 5% significance level (one-sided test). The pair-wise tests consider the sign and t-statistic for $\hat{\alpha}_{12}$, with complementarity or substitutability determined to exist by a one sided test on the coefficient at the 5% significance level (t-statistic > 1.65).

[10] For comparison, we executed similar Monte Carlo simulations with correlation coefficients set at 0.8, -0.8 and 0, respectively and without systematic correlation between the practices. We found only limited changes in the accuracy of the tests. The simulations can be obtained from the authors upon request.

The above procedure has been repeated 10,000 times for models with different explanatory power. Tables 1-3 present the results of the Monte Carlo experiments for models with R^2 of approximately 10%, 50% and 90%.[11] In each of the experiments we compare the results of the three tests with the true states of complementarity and substitutability. The multiple-restrictions test generates correct predictions in 94 percent of cases and clearly outperforms both other tests for models with high explanatory power (R^2=90%). The percentage of correct predictions reduces to 83 percent for models with medium explanatory power (R^2=50%) and further to 63 percent for models with poor explanatory power (R^2=10%). The number of 'reverse' predictions is negligible in the model with high explanatory power and remains very small throughout. The results for the pair-wise tests do not show a similar increase in predictive power alongside an increase in explanatory power of the model, and perform relatively poorly in the high explanatory power model (78.9 and 73.5 percent correct predictions, respectively). In models with medium or poor explanatory power, on the other hand, the performance of the all pair-wise test is by and large equal to the multiple restrictions test. The single pair-wise test only reaches a performance comparable to the other two tests in the model with poor explanatory power, while the test also exhibits non-negligible percentage of reverse predictions in the case of models with the best fit.[12] The simulations show a changing distribution of error types if the explanatory power of the model decreases. In models with high explanatory power most errors are of type II: the tests indicate complementarity or substitutability while there is none. This occurs very often in the pair-wise tests (in more than 20 percent of all cases). For models with intermediate explanatory power type II errors are still most frequent, although the frequency of type I errors (the tests fail to confirm complementarity or substitutability) increases. For models with poor explanatory power the pattern reverses: the frequency of type II errors is relatively low but none of the tests is able to identify complementarity of substitutability in a satisfactory manner.

We conclude that the multiple restrictions test is the superior testing framework for complementarity but conditional on the presence of well specified models in which the practices of interest have a strong impact on this performance. In case of less discerning models, the simultaneous pair-wise test appears as an easily executed alternative test with similar predictive power. The single pair-wise test has the least satisfactory properties. If the practices of interest explain only a minor part of organizational performance, the results suggest that no test method is able to provide reliable predictions.

---

[11] The values of σ_{ε} = 2.4, 0.4, 0.07 are selected in order to achieve the desired R^2.

[12] We have also run experiments for continuously measured practices under identical conditions. Monte Carlo results were qualitatively similar to the dichotomous case but with smaller differences between the multiple restrictions test and the simultaneous pair-wise test. Test results as well as the Monte-Carlo programs are available from the authors upon request.

5. Conclusion

Recent empirical studies of organizational performance have been concerned with establishing potential complementarity between more than two organizational practices. These papers have drawn conclusions on the basis of potentially biased estimates of pair-wise interaction effects between such practices. This paper developed a consistent testing framework based on multiple inequality constraints that derives from the definition of (strict) supermodularity as suggested by Athey and Stern (1998), and compares the performance of this test with previously used testing methods. Monte Carlo results show that the multiple restrictions test is clearly superior for performance models with high explanatory power. A test based on estimating all pair-wise interaction terms is good alternative for empirical models with less explanatory power, while single (alternating) pair-wise complementarity tests perform less well. The accuracy of all tests for applications where the practices of interest explain only a minor part of performance is less than satisfactory as none of the tests appear to be able to identify true complementarity or substitutability well. The results may raise questions concerning the accuracy of previous empirical studies of complementarity in a multiple practices setting, and generally suggest caution in the application and interpretation of complementarity and substitutability tests in empirical research.

References

Arora, A. 1996. Testing for compementarities in reduced-form regressions: a note. *Economics Letters* 50:51-5.

Arora, A. and A. Gambardella. 1990. Complementarity and external linkages: the strategies of the large firms in biotechnology. *Journal of Industrial Economics* 38: 361-79.

Athey, S. and S. Stern. 1998. *An empirical framework for testing theories about complementarity in organizational design*. NBER working paper No. 6600. NBER, Cambridge, MA.

Athey, S. and S. Stern. 2002. The impact of information technology on emergency health care outcomes. *RAND Journal of Economics* 33(3): 399-432

Baumol, W., J. C. Panzar, and R. D. Willig. 1988. *Contestable Markets and the Theory of Industry Structure*. Harcourt Brace Jovanovich.

Black, Sandra and Lisa Lynch, 2001, How to compete: the impact of workplace practices and information technology on productivity. *The Review of Economics and Statistics*, 83(3): 434-445.

Bresnahan, T., E. Brynjolfsson, and L. M. Hit, 2002, Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence, *Quarterly Journal of Economics* 117, Feb 2002, 339-375.

Caroli, E. and J. Van Reenen, 2001, Skill-biased organizational change? Evidence from a panel of British and French establishments. *Quarterly Journal of Economics* 116, Nov. 2001: 1449-1492.

Genz, A. 1993. Comparison of methods for the computation of multivariate normal probabilities *Computing Science and Statistics* 25: 400-05.

Gouriéroux, C., A. Holly, and A. Monfort. 1982. Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* 50(1): 63-80.

Grossman, G. and G. Maggi. 2000. Diversity and Trade, *American Economic Review* 90 (5), 1255-1275.

Hajivassiliou, Vassilis, Daniel McFadden, and Paul Ruud. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. *Journal of Econometrics* 72: 85-134.

Ichniowski, C., K. Shaw, and G. Prennushi. 1997. The effects of human resource management practices on productivity. *American Economic Review* 87: 291-313.

Leiponen, Aija. 2005. 'Skills and innovation', *International Journal of Industrial Organization*, 23(5-6), 303-23.

Kodde, D. and F. Palm. 1986. Wald criteria for jointly testing equality and inequality restrictions. *Econometrica* 54(5): 1243-48.

Kudô, A. 1963. A multivariate analogue of the one-sided test. *Biometrika* 50: 403-18.

Milgrom, P. and J. Roberts. 1990. The economics of modern manufacturing: technology, strategy, and organization. *American Economic Review* 80: 511-28.

Miravete, Eugenio J; Pernias, Jose C, 2004, *Innovation Complimentarity and Scale of Production* CEPR Discussion Paper No 4483, CEPR, London.

Mohnen, P. and L-H. Röller. 2005. Complementarities in innovation policy. *European Economic Review,* 49, 1431-1450.

Shapiro, A. 1985. Asymptotic distribution of test statistics in the analysis of moment structures under inequality consraints. *Biometrika* 72(1): 133-44.

Siggelkow, N. 2002. Misperceiving interactions among complements and substitutes: Organizational consequences. *Management Science* 48(7): 900-916.

Sun, H-J. 1988. A General reduction method for n-variate normal orthant probability. *Communication in Statistical-Theoretical Methods* 17(11): 3913-3921.

Topkis, D. 1998. *Supermodularity and Complimentarity*. Princeton University Press.

Veugelers, R., and B. Cassiman. 2006. Complementarity in the innovation strategy: Internal R&D, external technology acquisition, and cooperation in R&D, *Management Science*, forthcoming.

Wolak, F. 1989. Testing inequality constraints in linear econometric models. *Journal of Econometrics* 41: 205-35.

Table 1 Monte Carlo experiment, multiple restrictions test (n=1000, draws=10000)

		Multiple restrictions test								
		R^2=90%			R^2=50%			R^2=10%		
		Compl	Nothing	Subs	Compl	Nothing	Subs	Compl	Nothing	Subs
True effect	Complements	3714	14	0	3517	209	2	2205	1504	19
	Nothing	264	1934	265	657	1177	629	341	1770	352
	Substitutes	0	19	3790	5	206	3598	28	1501	2280
	Correct, %	*94.38*			*82.92*			*62.55*		
	Reverse outcome	*0.0*			*0.07*			*0.47*		
	Type I error, %	*0.33*			*4.22*			*30.52*		
	Type II error, %	*5.29*			*12.86*			*6.93*		

Table 2 Monte Carlo experiment, all cross-term test (n=1000, draws=10000)

		All cross-term test, *all double cross terms included*								
		R^2=90%			R^2=50%			R^2=10%		
		Compl	Nothing	Subs	Compl	Nothing	Subs	Compl	Nothing	Subs
True effect	Complements	3714	14	0	3500	227	1	2251	1469	8
	Nothing	1047	389	1027	666	1181	616	224	2005	234
	Substitutes	0	19	3790	1	219	3589	13	1477	2319
	Correct, %	*78.93*			*82.70*			*65.75*		
	Reverse outcome error, %	*0.0*			*0.02*			*0.21*		
	Type I error, %	*0.33*			*4.48*			*14.90*		
	Type II error, %	*20.74*			*12.82*			*4.58*		

Table 3, Monte Carlo experiment, single cross-term test (n=1000, draws=10000)

		Single cross-term test, *x1*x2* only								
		R^2=90%			R^2=50%			R^2=10%		
		Compl	Nothing	Subs	Compl	Nothing	Subs	Compl	Nothing	Subs
True effect	Complements	3480	152	96	3320	341	67	2359	1335	34
	Nothing	1074	305	1084	847	766	850	403	1651	409
	Substitutes	117	126	3566	63	358	3388	25	1367	2417
	Correct, %	*73.51*			*74.74*			*64.27*		
	Reverse outcome error, %	*2.13*			*1.30*			*0.59*		
	Type I error, %	*4.91*			*8.29*			*27.61*		
	Type II error, %	*21.58*			*16.97*			*8.12*		