

DOMINANCE INDEX FOR MANY-TO-MANY CORRELATION AND ITS APPLICATIONS TO SEMICONDUCTOR YIELD ANALYSIS

Amos Hong
Argon Chen

Mechanical Engineering and Industrial Engineering
National Taiwan University
1 Sec. 4 Roosevelt Rd.
Taipei, 106, TAIWAN

ABSTRACT

As more and more functionalities are packed into a single product, one-response-at-a-time correlation analysis is no longer sufficient to discover critical factors that result in poor qualities or a low yield. Though methodologies of many-to-many correlation analysis have been proposed in the literature, difficulties arise, especially when there exist multi-collinearity effects among variables, to measure the relative importance of a variable's contribution in the association between a set of responses and a set of factors. Johnson's dominance analysis (Johnson 2000) offers a general framework for determination of relative importance of independent variables in linear multiple regression models. In this article, we extend Johnson's dominance index to many-to-many correlation analysis as a measurement to summarize the association relationship between two sets of variables. Actual semiconductor yield-analysis cases are used to illustrate the method and its effectiveness in analysis of two sets of variables.

1 INTRODUCTION

Semiconductor manufacturing consists of complex fabrication steps and has a long cycle time. The semiconductor yield loss is costly and its causes are difficult to find. Since having a good control of the electrical testing (ET) parameters is an important measure for ensuring a good yield of the final integrated circuit (IC) dies, it is critical to find the potential factors causing the out-of-control ET parameters. Such a problem is a typical many-to-many correlation problem where we try to associate many in-line process factors with multiple ET parameters to identify the yield-loss causes. It is known that discovering associations between ET parameters and in-line process metrology items, such as critical dimension, thickness and concentration, etc., can help identify root causes of yield loss and, potentially, provide insight into the underlying critical process steps. The conventional one-response-at-a-time analysis approach, e.g. a simple correlation coefficient or a multiple regression analysis for a single ET parameter on multiple in-line items, often leads to a large number of ineffective comparisons. Analyzing multiple ET parameters simultaneously has the advantage of providing a much simpler and clearer picture. Many-to-many association analysis to figure out the relationship between a set of yield/quality parameters and a set of in-line engineering variables is thus needed for a more effective search of root causes. Though methodologies of many-to-many correlation analysis have been proposed in the literature, it is still difficult to measure the relative importance of a variable's contribution in the association between the responses and the factors, especially when there exist multi-collinearity effects among variables.

This paper aims to develop a measurement of variable importance to summarize the association relationship between two sets of variables. We first review the canonical correlation analysis method. Next, we extend the relative importance measure to the many-to-many correlation model. We'll demonstrate

and validate the proposed index for the many-to-many correlation analysis through hypothetical cases and a real semiconductor yield analysis case.

2 RELATIVE IMPORTANCE (RI) FOR CANONICAL CORRELATION ANALYSIS (CCA)

2.1 Canonical Correlation Analysis (CCA)

Although many-to-many association is not popular to semiconductor yield analyst, many methods have been developed in the areas of multivariate statistics analysis and process control. Two common methods to study the relationship between two set of variables are partial least square (PLS) (Wold 1975) and canonical correlation analysis (CCA) (Hotelling 1936). Since there are several advantages of CCA over PLS, we'll focus on CCA for the development of many-to-many relative importance indices. We first briefly introduce the CCA method and the notation required throughout this paper. Denote by x a p -dimensional vector in the space of variables representing the first set and similarly by y a q -dimensional vector representing the second set of variables. After collecting n observations from each set of variables, the $(n\text{-by-}p)$ matrix of zero-mean unit-norm variables X and the $(n\text{-by-}q)$ matrix of zero-mean unit-norm variables Y are expressed into the form:

$$X = \{X_{ij}\} \quad i = 1, \dots, n \quad j = 1, \dots, p, \text{ and} \quad (1)$$

$$Y = \{Y_{ik}\} \quad i = 1, \dots, n \quad k = 1, \dots, q. \quad (2)$$

Assume that X and Y are full-rank matrices ($n > p+q$). The assumption of zero-mean unit-norm variable allows simplifying the expositions although it may cause some technical problem implicitly. All such technical problems can be resolved, usually by working with the raw data.

The matrices of sample covariance among the X variables, among the Y variables, and between the two sets are

$$\Sigma_x = (1/n-1)X^t X, \quad (3)$$

$$\Sigma_y = (1/n-1)Y^t Y, \text{ and} \quad (4)$$

$$\Sigma_{yx} = \Sigma_{xy}^t = (1/n-1)Y^t X. \quad (5)$$

CCA attempts to derive a linear combination of the variables, *canonical variates*, of each data set for maximizing their correlation while PLS is to maximize the covariance. Canonical score vectors, denoted by u and v , are n observations of canonical variates of each data set. Under the consideration of scaling, u and v are expressed as follows:

$$u = X\alpha, \text{ and} \quad (6)$$

$$v = Y\beta. \quad (7)$$

where α and β are canonical coefficients, the weights of the linear combination, which are sought to maximize the square sample correlation coefficient of the two canonical score vectors, u and v :

$$\rho(v, u)^2 = \frac{(\beta^t \Sigma_{yx} \alpha)^2}{(\beta^t \Sigma_y \beta)(\alpha^t \Sigma_x \alpha)} \quad (8)$$

where $\rho(\cdot)$ denotes correlation function.

In order to simplify the optimization problem, coefficient transformation is introduced by setting

$$a = \Sigma_x^{-\frac{1}{2}} \alpha, \text{ and} \quad (9)$$

$$b = \Sigma_y^{-\frac{1}{2}} \beta. \quad (10)$$

Then the objective function becomes

$$\rho(v, u)^2 = \frac{(b^t \Sigma_y^{-\frac{1}{2}} \Sigma_{yx} \Sigma_x^{-\frac{1}{2}} a)^2}{(b^t b)(a^t a)}. \quad (11)$$

It is straight forward to show the CCA solution is converted to solve a and b in the following formulation:

$$\begin{aligned} \max : \rho(v, u)^2 &= (b^t \Sigma_y^{-\frac{1}{2}} \Sigma_{yx} \Sigma_x^{-\frac{1}{2}} a)^2 \\ \text{st. } a^t a &= 1, \quad b^t b = 1 \end{aligned} \quad (12)$$

The solution for Equation (12) is much simpler and demands lower computing power than PLS. With the half inverse of covariance matrix $\Sigma_y^{-\frac{1}{2}}$ and $\Sigma_x^{-\frac{1}{2}}$, it implies CCA deals with two transformed datasets ($Y \Sigma_y^{-\frac{1}{2}}$ and $X \Sigma_x^{-\frac{1}{2}}$) with the covariance matrix equal to an identity matrix, i.e. mutual independent datasets. Instead of iterative process of repeated computation by PLS, the whole CCA solution is thus simply given by the singular vectors of the singular value decomposition (SVD) of K :

$$K = \Sigma_y^{-\frac{1}{2}} \Sigma_{yx} \Sigma_x^{-\frac{1}{2}} = U_K S_K V_K^t. \quad (13)$$

Let ρ_r (positive real value) be the r th singular value in S_K and b_r and a_r be the corresponding r th singular vectors of U_K and V_K , respectively. The maximum number of canonical variable pairs is equal to d . The r th canonical coefficients (α_r, β_r) are then given by

$$\alpha_r = \Sigma_x^{-\frac{1}{2}} a_r, \text{ and} \quad (14)$$

$$\beta_r = \Sigma_y^{-\frac{1}{2}} b_r \quad (15)$$

and the r th canonical score pair u_r and v_r are obtained by

$$u_r = X \Sigma_x^{-\frac{1}{2}} a_r, \text{ and} \quad (16)$$

$$v_r = Y \Sigma_y^{-\frac{1}{2}} b_r \quad (17)$$

with the r th canonical correlation equal to ρ_r :

$$\rho(v_r, u_r) = b_r^t \Sigma_y^{-\frac{1}{2}} \Sigma_{yx} \Sigma_x^{-\frac{1}{2}} a_r = \rho_r. \quad (18)$$

2.2 Many-to-many Relative Importance (RI) Index

Many relative importance indices have been proposed to assess the proportion of variance in a dependent variable explained by the independent variables in a linear regression model. General dominance index (Budescu 1993) is regarded in the literature as the most plausible assessment index due to both statistical and theoretical rationales (Johnson 2000; Lebreton, Ployhart, and Ladd 2004; Chao et al. 2008). General dominance index is defined as the average increment in R -square associated with an independent variable across all possible subset models. That is, general dominance index represents the average usefulness of an independent variable when it is included with each possible combination of other variables. Since the sum of general dominance indexes over all covariates is equal to model R -square, it can be expressed as the proportion of predictable variance accounted for by an independent variable.

An extension of general dominance index to multivariate regression model is proposed by (Azen and Budescu 2006). A recent work (Huo and Budescu 2009) has also extends the dominance approach to CCA. However, dominance index has a serious shortcoming, that is, difficult and time-consuming computation (Johnson 2000; Lebreton, Ployhart, and Ladd 2004; Chao et al. 2008). This method requires the computation of R -squares for all possible sub-models. Given p covariates in the linear regression model, there are $(2^p - 1)$ sub-models. In many-to-many correlation model with p and q variables respectively, there are $(2^q - 1) \times (2^p - 1)$ sub-models. The computation demand increases in an exponential rate. For example, there will be 1023 sub-models for linear regression models with only 10 independent variables. The number of sub-models and required computing power is even staggering for many-to-many correlation models. Because the computation requirement of the general dominance index has limited its applicability, a heuristic method for estimating the general dominance index is then proposed (Johnson 2000). The heuristic dominance analysis not only requires much lower computation loading but also be shown to be highly consistent with general dominance index through real case validation (Johnson 2000), Monte Carlo experiments (Lebreton, Ployhart, and Ladd 2004) and bootstrap estimate of indices (Chao et al. 2008). We refer to this heuristic dominance index as Johnson's dominance index in this article. Johnson's dominance index considers a variable's relative importance under collinearity by using a set of uncorrelated independent variables transformed from the original variables. The process is divided into three steps. First is to transform the original variables into a new set of orthogonal variables with maximum similarity to the original variables. Second is to calculate the relative importance of the new orthogonal independent variables. Without the collinearity issue among the orthogonal variables, the final step is to distribute the relative importance of the orthogonal variables into the original ones based on their association.

The relative importance expressed by Johnson's dominance index in one-to-many correlation has been given in (Johnson 2000). Extension to many-to-many correlation starts from standardized X and Y into unit-norm and zero-mean data matrix. It is to ensure the similarity between Johnson's dominance index and general dominance index since the comparison (Johnson 2000; Lebreton, Ployhart, and Ladd 2004; Chao et al. 2008) of the two indices have been concluded with their convergence under the standardized data distribution. The first step of the extension is to transform X and Y into maximally correlated orthogonal Z_x and Z_y , respectively, by finding the SVD of X and SVD of Y , respectively, as:

$$X = U_x S_x V_x^t, \text{ and} \quad (19)$$

$$Y = U_y S_y V_y^t. \quad (20)$$

The orthogonal Z_x maximally correlated to X and the orthogonal Z_y maximally correlated to Y are obtained, respectively, as:

$$Z_x = U_x V_x^t, \text{ and} \quad (21)$$

$$Z_y = U_y V_y^t. \quad (22)$$

The matrices of sample covariance among the Z_x variables, among the Z_y variables, and between the two sets are

$$\Sigma_x^* = (n-1)^{-1} Z_x^t Z_x = (n-1)^{-1} I_p, \quad (23)$$

$$\Sigma_y^* = (n-1)^{-1} Z_y^t Z_y = (n-1)^{-1} I_q, \text{ and} \quad (24)$$

$$\Sigma_{yx}^* = \Sigma_{xy}^{*t} = (n-1)^{-1} Z_y^t Z_x. \quad (25)$$

The process is then to seek new canonical coefficient α^* and β^* to compose the new canonical scores vector u^* and v^* :

$$u^* = Z_x \alpha^*, \text{ and} \quad (26)$$

$$v^* = Z_y \beta^*, \quad (27)$$

and to maximize the objective $\rho(v^*, u^*)^2$:

$$\rho(v^*, u^*)^2 = \frac{(\beta^{*t} \Sigma_{yx}^* \alpha^*)^2}{(\beta^{*t} \Sigma_x^* \beta^*)(\alpha^{*t} \Sigma_x^* \alpha^*)} = \frac{(\beta^{*t} Z_y^t Z_x \alpha^*)^2}{(\beta^{*t} \beta^*)(\alpha^{*t} \alpha^*)}. \quad (28)$$

Similarly, the optimal sets of coefficients (α^* 's, β^* 's), are given by SVD of the matrix $Z_y^t Z_x$:

$$Z_y^t Z_x = USV^t. \quad (29)$$

Let ρ_r (positive real value) be the r th singular value in S and b_r and a_r be the corresponding r th singular vectors of U and V , respectively. It implies the new r th canonical coefficient pair (α_r^* , β_r^*) is given by the r th singular vector pair of K :

$$\alpha_r^* = a_r, \text{ and} \quad (30)$$

$$\beta_r^* = b_r. \quad (31)$$

Therefore, the r th canonical correlation coefficient is still the singular value of K (ρ_r) and not changed by the matrix transformation:

$$\rho(v_r^*, u_r^*) = \frac{b_r^t (n-1)^{-1} K a_r}{\sqrt{(b_r^t (n-1)^{-1} b_r)(a_r^t (n-1)^{-1} a_r)}} = \rho_r. \quad (32)$$

The r th pair of original standardized canonical coefficient can be obtained as follow:

$$\alpha_r = \Sigma_x^{-\frac{1}{2}} \alpha_r^*, \text{ and} \quad (33)$$

$$\beta_r = \Sigma_y^{-\frac{1}{2}} \beta_r^*. \quad (34)$$

The r th corresponding linear regression model pair with R -square equal to ρ_r^2 is re-expressed as

$$v_r^*(=Z_y\beta_r^*)=Z_xA_r+\varepsilon_{y,r}^*, \text{ and} \quad (35)$$

$$u_r^*(=Z_x\alpha_r^*)=Z_yB_r+\varepsilon_{x,r}^*, \quad (36)$$

where $\varepsilon_{y,r}^*$ and $\varepsilon_{x,r}^*$ are n -by-1 vector of errors with zero column mean, A_r , p -by-1 regression coefficient, and B_r , q -by-1 regression coefficient. A_r and B_r are easily calculated from the equations as follows:

$$A_r=\alpha_r^*\rho_r, \text{ and} \quad (37)$$

$$B_r=\beta_r^*\rho_r. \quad (38)$$

To determine the relative importance of X_s is then to determine how the proportion of the relationship (ρ_r^2) explained by $Z_{x,l}$ is associated with X_j . Since the relative importance of X contributed to the relationship should depend on the variance structure of X , it is easier to address the contribution through the orthogonal variables Z_x , the Johnson transformation of X . (39) shows how the variance in X is explained by Z_x :

$$X=Z_xA_x. \quad (39)$$

Since Z_x is a linear transformation of X , the above equation is seen as a perfect multivariate regression model with zero residuals. The regression coefficient matrix A_x is also the correlation matrix between Z_x and X (Z_x^tX):

$$A_x=\{A_{x,lj}\}=T_x^{-1}=V_xS_xV_x^t=Z_x^tX, \quad (40)$$

where $l=1, \dots, p, j=1, \dots, p$, and $A_{x,lj}$ is the element of the l th row and the j th column in A_x , representing the correlation between $Z_{x,l}$ and X_j .

Similarly, the proportionate contribution of a particular X_j to ρ_r^2 , denoted as $D_{x,rj}$, can be estimated by multiplying the proportion of variable in each X_j accounted for by $Z_{x,l}$ ($A_{x,lj}^2$) by the proportion of ρ_r^2 accounted for by $Z_{x,l}$ (A_{rl}^2) and sum the products:

$$D_{x,rj}=\sum_{l=1}^pA_{x,lj}^2A_{rl}^2 \quad \forall j \quad \forall r. \quad (41)$$

Furthermore, the correlation matrix between Z_y and Y can be expressed as follows:

$$A_y=\{A_{y,ok}\}=V_yS_yV_y^t=Z_y^tY, \quad (42)$$

where $o=1, \dots, q, k=1, \dots, q$, and $A_{y,ok}$ is the element of the o th row and k th column in A_y , representing the correlation between $Z_{y,o}$ and Y_k . Thus, the proportionate contribution of a particular Y_k to ρ_r^2 , denoted as $D_{y,rk}$, can be as:

$$D_{y,rk}=\sum_{o=1}^qA_{y,ok}^2B_{ro}^2 \quad \forall k \quad \forall r. \quad (43)$$

Proposition 1: In the r th relationship of many-to-many correlation model, the relative importance of X set ($D_{x:rj}$) and the relative importance of Y set ($D_{y:rj}$) are both summed up to be the squared r th canonical correlation coefficient:

$$\sum_{j=1}^p D_{x:rj} = \rho_r^2, \text{ and} \quad (44)$$

$$\sum_{k=1}^q D_{y:rk} = \rho_r^2. \quad (45)$$

In summary, there are at most d sets of canonical correlations. The r th square canonical correlation ρ_r^2 describes the residual relationship left over from the 1st, 2nd..., and $(r-1)$ th canonical correlation pairs. The relative importance $D_{x:rj}$ for X_j and $D_{y:rk}$ for Y_k can be then obtained to explain how the relationship is established. Furthermore, selection of significant relationships is well studied in the literature (Marriott 1952, ; Lawley 1959) and can be easily conducted.

3 APPLICATION TO SEMICONDUCTOR YIELD LEARNING

The section gives two real semiconductor yield cases of a local manufacturing foundry fab to demonstrate the extension of Johnson's dominance index to many-to-many correlation analysis. In the first variable set, there are totally eight ET parameters of interest related to a certain MOSFET structure denoted by $et_1 \sim et_8$. Two parameters, et_1 and et_2 , are measurements of the drain current in saturation I_d under specific conditions. Three parameters, et_3 , et_4 and et_5 , are to monitor the threshold voltage E_{th} . The other parameters, et_6 , et_7 and et_8 , are to monitor the charge-carrier effective mobility N .

In the second variable set, there are nine inline measurement items of interest denoted by $l_1 \sim l_9$. Their geometric relationship is represented in the sketch map of Figure 1. As can be seen, there are two correlated set. Items l_3 , l_4 , l_5 and l_6 are correlated. So are items l_7 , l_8 and l_9 .

There are different ET parameters of interest in each case while the nine inline items are all used to analyze. Besides, the false alarm rate ϕ is set to be 0.01.

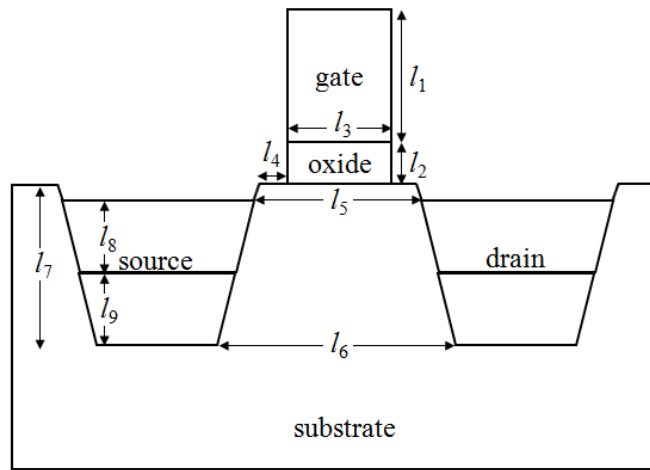


Figure 1: Geometric representation of inline metrology

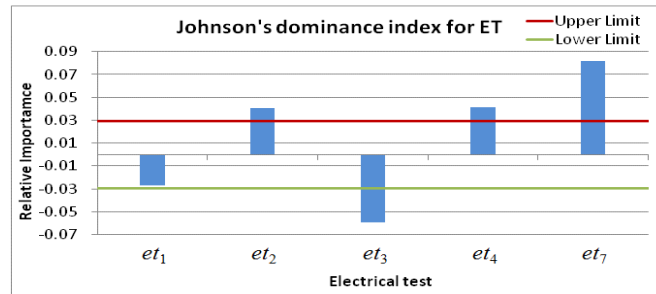
3.1 Case 1

In the case, ET Parameters of interest are et_1 , et_2 , et_3 , et_4 and et_7 . Notice that there are five ET parameters and nine inline items with sample size equal to 160. Both PLSC2A and CCA find five different relationship pairs. The strength of the relationship pairs are compared in Table 1.

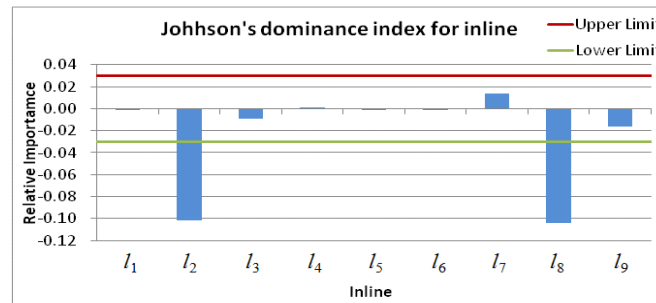
Table 1: Case 1 comparison on correlation coefficient of relationship pair by PLSC2A and CCA

pair	CCA		PLSC2A	
	correlation	<i>p</i> -value	correlation	<i>p</i> -value
1st	0.5006	0.0001	0.4163	0.0121
2nd	0.4047	0.0204	0.2669	0.6800
3rd	0.2731	0.6367	0.2582	0.7364
4th	0.2072	0.9497	0.2080	0.9480
5th	0.1392	0.9992	0.1302	0.9996

The significant relationship pairs are shown in bold font. The purpose is to find significant associated relationships explaining the yield issues and to give information to solve the issue further. The first two CCA relationship pair interest analyst and domain engineers. The result is validated by further inspection. Figure 2 shows the contribution charts, Johnson's dominance index, for the first ET-to-inline CCA pair. As shown in Figure 2(b), the critical inline items are l_2 and l_8 . Other inline items are trivial with relative importance almost equal to zero. ET parameters et_2 , et_3 , et_4 and et_7 suffer from the impact of l_2 and l_8 while et_1 is impacted mostly and then et_3 is secondary. Figure 3 shows the contribution charts for second ET-to-inline CCA pair which explains the residual relationship left by first CCA pair. As shown in Figure 3(b), the critical inline items are l_3 , l_5 and l_6 ; as in Figure 3(a), et_2 and et_1 suffer major impact. Two independent associations between ET and inline are discovered by the proposed method.

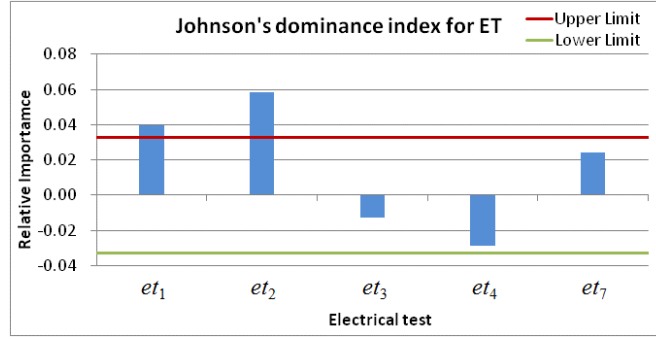


(a)

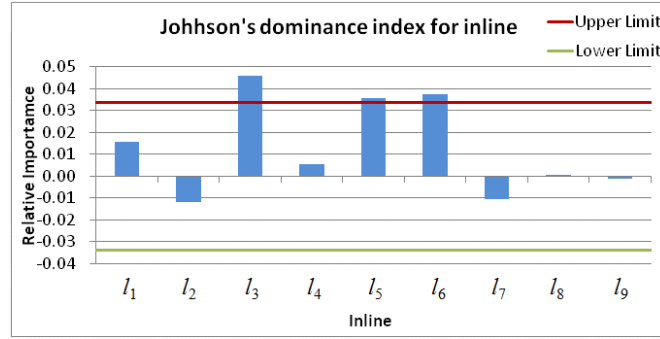


(b)

Figure 2: Case1 contribution chart for first ET-to-inline CCA pair



(a)



(b)

Figure 3: Case1 contribution chart for second ET-to-inline CCA

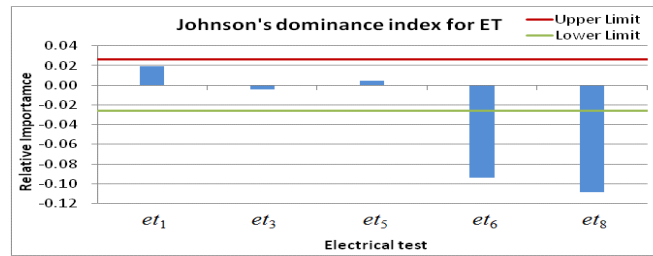
3.2 Case 2

In the case, ET Parameters of interest are et_1 , et_3 , et_5 , et_6 and et_8 . There are five ET parameters and nine inline items with sample size equal to 207. Both PLSC2A and CCA find five different relationship pairs. Their correlation coefficients are compared in Table 2.

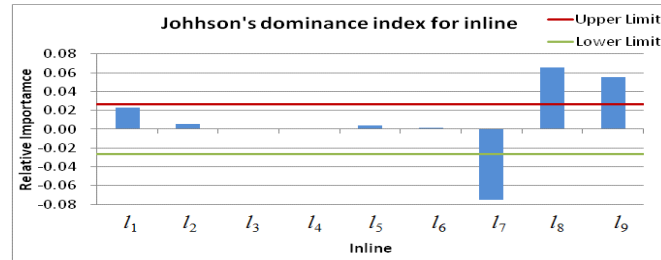
Table 2: Case2 comparison on correlation coefficient of relationship pair by PLSC2A and CCA

pair	CCA		PLSC2A	
	correlation	<i>p</i> -value	correlation	<i>p</i> -value
1st	0.4793	<0.0001	0.3248	0.0795
2nd	0.4051	0.0013	0.4047	0.0014
3rd	0.3399	0.0437	0.2854	0.2710
4th	0.2883	0.2518	0.2503	0.5470
5th	0.1532	0.9898	0.1875	0.9329

First three CCA relationship pairs are significant and provide information to solve the yield issues. Figure 4 shows the contribution charts for the dominate ET-to-inline CCA pair. As can be seen, parameters et_6 and et_8 are mostly affected by root causes l_6 , l_7 and l_8 . Figure 5 shows the contribution charts for second ET-to-inline CCA pair. As can be seen, et_3 is mostly affected and et_6 is secondarily affected by the unique root cause l_3 . In summary, the independent relationships found by the proposed method are (et_6 , et_8 vs. l_7 , l_8 , l_9), (et_3 , et_6 vs. l_3) and (et_1 , et_5 vs. l_4).

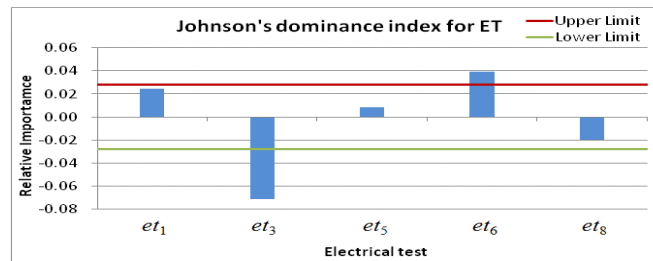


(a)

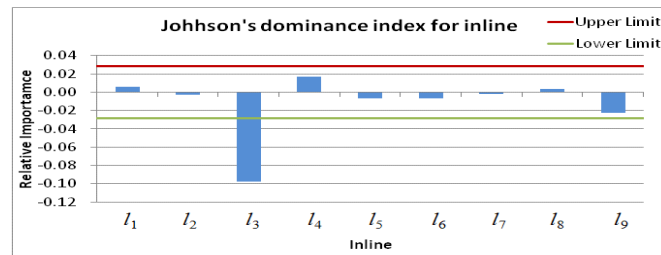


(b)

Figure 4: Case2 contribution chart for first ET-to-inline CCA pair



(a)



(b)

Figure 5: Case2 contribution chart for second ET-to-inline CCA pair

CONCLUSIONS

In this research work, we have extended the relative importance to many-to-many correlation analysis and proposed to use the dominance indices to facilitate the interpretation of the correlation results by the relative importance of each variable. Studies on two real cases have also demonstrated how the methodology is useful to pinpoint critical causes in semiconductor yield learning. As compared to conventional PLSC2A, the proposed method has a more straightforward computation and makes the explanation simpler and more interpretable.

REFERENCES

- Azen, R., and Budescu, D.V. 2006. "Comparing predictor in multivariate regression models: an extension of dominance analysis", *Journal of Education and Behavioral Statistics*, vol.31, pp.157-180, 2006.
- Budescu, D.V. 1993. "Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression", *Psychological Bulletin*, vol. 114, pp. 542-551.
- Chao, Y. E., Zhao, Y., Kupper, L. L., and Nylander-French, L.A. 2008. "Quantifying the relative importance of predictors in multiple linear regression analyses for public health studies", *Journal of Occupational and Environmental Hygiene*, vol.5, no.8, pp.519 – 529.
- Hotelling, H. 1936. "Relations between two sets of variates", *Biometrika*, vol. 28, pp. 321-377.
- Huo, Y., and Budescu, D.V. 2009. "An extension of dominance analysis to canonical correlation analysis", *Multivariate Behavioral Research*, vol. 44, pp. 688-709.
- Johnson, J.W. 2000. "A heuristic method for estimating the relative weight of predictor variables in multiple regression", *Multivariate Behavioral Research*, vol. 35, pp. 1-19.
- Lawley, D.N. 1959. "Tests of significance in canonical analysis", *Biometrika*, vol.46, no.1-2, pp.59-66.
- Lebreton, J.M., Ployhart, R.E. and Ladd, R.T. 2004. "A Monte Carlo comparison of relative importance methodologies", *Organizational Research Methods*, vol. 7, pp.258-282.
- Marriott, F.H.C. 1952. "Tests of significance in canonical analysis", *Biometrika*, vol.39, no.1-2, pp.58-64.
- Wold, H. 1975 "Path model with latent variables: the NIPALS approach", in *Quantitative Sociology: International Perspectives on Mathematical and Statistical Modeling*, H. M. Blalock et al., editors, New York: Academic Press, pp. 307-357.

AUTHOR BIOGRAPHIES

AMOS HONG received the B.S. degree in business administration and the M.S. degree in industrial engineering from National Taiwan University, Taipei, Taiwan, in 2000 and 2002, respectively. He is currently pursuing the Ph.D. degree in mechanical engineering from the Department of Mechanical Engineering, National Taiwan University, under the supervision of Dr. A. Chen. His current research interests include statistical inference, machine learning, statistical data mining, and semiconductor yield analysis. His email address is d93522002@ntu.edu.tw.

ARGON CHEN received his Ph.D. degree in industrial engineering from the Rutgers State University of New Jersey, New Brunswick, NJ. He is currently a Professor with the Graduate Institute of Industrial Engineering and the Department of Mechanical Engineering, National Taiwan University, Taipei, Taiwan. He has been working closely with the semiconductor industry in Taiwan and was a PI for research projects co-funded by Semiconductor Research Corporation, and ISMI. His current research interests include applied statistical inference, advanced statistical process control and optimization, and statistical data mining. His email address is achen@ntu.edu.tw.