

# Analyzing Dependent Variables with Multiple Surrogates in Accounting Research

Enyi Enyi<sup>1</sup>

<sup>1</sup>Babcock University

December 03, 2018

## Abstract

The paper contains details a research carried out to show that the use of geometric mean to unify multivariate dependent variables in financial performance studies gives better and more practical results than the multiple abstraction analysis provided using advanced econometric tools such as TLS, PLS, MCA, Canonical correlations etc.

The study used the logistic regression analysis to compare the *a priori* expectations of 30 Ph.D research theses with their actual outcomes using econometric tools and the actual outcome using geometric means. The study used Ph.D theses in accounting and finance sourced from the libraries of four universities in Nigeria.

The study was a desktop research using publicly available literary resources and as such requires no ethical clearance.

Enyi Patrick Enyi

Babcock University

e.p.enyi@gmail.com

Department of Accounting, School of Management Sciences

Accounting and finance-based researchers often face the problem of using multiple surrogates to capture the true essence of a dependent variable in order to study its predictive relationship with explanatory variables. The major pitfall with this, is the inability to directly connect the study results with the major objective of the research forming the topic. This study explores the various methods that can be adopted to unify the engaging properties of the multiple parameters used in identifying a dependent variable. Using logistic regression, the study converted the *a priori* expectations of 30 Ph.D research theses in finance and accounting with four dependent surrogates into a probabilistic log values and compared them with the individual surrogate performance on the one hand and the surrogates geometric mean on the other hand. While the geometric mean showed no significant difference between itself and the probabilistic expectation ( $\beta = .278, t_{(30)} = .695, R^2 = .077, p > .10$ ), the individual surrogates results revealed both combined and individual significant differences with the theses *a priori* expectations (*Adj. R*<sup>2</sup> = .0291, *F*<sub>(4, 25)</sub> = 22.598, *p* < .05). This paper recommends unifying multivariate dependent variables with geometric means for successful financial performance relational studies.

**Keywords :** performance, dependent variable, surrogate, proxy, logistic regression

Researchers and scholars in the social and management sciences are often faced with the problem of identifying and operationalizing the main ingredients forming the dependent or outcome variable in performance assessment research. This is particularly a big issue when such a study is extended in scope or necessitated by the requirements to fulfil the conditions for the award of a degree. For those studying behavioral traits and similar psychological phenomena that has no accounting or financial data foundation, this will not present

a problem as there are several methods already developed to tackle the issue. Some of these methods include the use of general linear model (GLM) multivariate and repeated measures procedures using computer statistical software, while others may include the use of canonical correlations and specially constructed regression equations (Helwig, 2017; NCSS, 1989; Steiger, n.d.). The use of means as a converging point was also advocated by other researchers especially when multiple dependent variables represent different measures of the same construct. When such differing variables are measured on the same scale, researchers have the option of combining them into a single measure of that construct. The use of binary response model (BRM) was also in the picture, but as posited by Cameron (2015), the description of the use of means looks more practical and less ambiguous to comprehend (Cameron, 2015; Gruszczy, 2009; Price, Jhangiani, & Chiang, 2015). However, the imprecise and unsure nature of the various methods in current use makes most research findings and conclusions meaningless and unrelated to the major objective of the research. This is very worrisome especially to graduate level research students and their advisors and has proven to be of particular negative implications in the fields of accounting and finance in management sciences where precision in conclusions is of utmost importance in guiding investors and decision makers correctly.

## Research objective

The objective of this paper is to empirically investigate and determine whether the present method of conducting, assessing and concluding corporate performance researches using multiple surrogate findings on a dependent variable is optimally satisfactory and capable of guiding correct decision making on the basis of such study.

## Theoretical framework

As stated earlier, there are many ways to deal with multi-surrogate dependent variables (DV) when analyzing cause-and-effect relationships in research studies that bothers more on behavioral phenomena than real-life business-related problems. This is premised on the fact that business and applied economic problems demand the use of past transaction data not based on assumptions or mere conjecture in setting relational models that can be used to direct the affairs of an organization. To deal with the issues of behavioral phenomenon with more than one DV surrogate, it was suggested in a community of online discussants that one could run two separate regression equations in a case where there are two DV surrogates; one for each DV, but the general concern is that such treatment might likely not capture the interspersing relationship between all the DV surrogates. Also, fitting all surrogates' regressions separately will indeed be equivalent to formulating multivariate relationships with a matrix of dependent variables (Transaction Processing Performance Council, 2011). However, if one is interested in describing a two-block structure, this could be done using partial least square regression (PLS). Partial least square is a regression framework which relies on the idea of building successive (orthogonal) linear combinations of the variables belonging to each block such that their covariance is maximal. It is a method for relating two data matrices, X and Y, by a linear multivariate model but goes beyond traditional regression in that it models also the structure of X and Y and also derive its usefulness from its ability to analyse data with many, noisy, collinear, and even incomplete variables in both X and Y (Wold, Sjostrom, & Eriksson, 2001).

Many social scientists on the other hand, prefer the use of the GLM multivariate and repeated measures ANOVA to fit the multiple equations resulting from the use of more than one dependent variable into a single equation for the purpose of getting a unified analytical result; however, a number of others prefer the use of more exotic methods such as binary response model (BRM), multiple classification analysis (MCA), and canonical correlation among others, to obtain the same effect. Particularly, the GLM Multivariate procedure allows the analyst to model the values of multiple dependent scale variables, based on their relationships to categorical and scale predictors. In a case of ordinary GLM, there is always a single dependent variable, with a prediction mean error of zero (0) and a variance that can be computed after the GLM is fitted. But when there are multiple dependent variables, each of the dependent variables will have a prediction error (Helwig, 2017; NCSS, 1989; Steiger, n.d.).

In chemometrics analysis, the use of PLS is favoured because of the multiplicity of the inputs and outputs of

most chemical processes. Chemometrics is the use of mathematical and statistical methods to improve the understanding of chemical information and to correlate quality parameters or physical properties to analytical instrument data (Bu, 2007). Chemometrics analysis is a fascinating one because it is interdisciplinary and employs the extensive use of such tools as principal components analysis (PCA), multivariate statistics, three-pass regression, LPLS regression, latent structure regression, partial least square structural equation modeling (PLS-SEM), covariance based structural equation modeling (CB-SEM), and shrinkage structure analysis (Abdi, 2010; Afthanorhan, 2013; Helland, 1990; Kelly & Pruitt, 2015; Lingjaerde & Christophersen, 2000; Saebo, Almoya, Flatberg, Aastveit, & Martens, 2008). Chemometrics also employ the use of total least square (TLS) and Deming regression in analyzing multiple dependent variables because of the reasons earlier adduced. TLS is a method of fitting that is appropriate when there are errors in both the observation vector and in the data matrix (Golub & Van Loan, 1980). Deming regression on the other hand is a special case of TLS which allows for any number of predictors and complicated error structure to be analyzed (Jensen, 2007).

Though, most of the analytical tools enunciated above employ techniques that will eventually end in bringing out the mean values that will be used to fit the final model of the intended research relationship, such may not readily or necessarily suit the secondary nature of the data usually extracted for financial performance analysis. Besides, the average accountant or financial analyst are not expected to acquire the deep knowledge of econometric and statistical analysis necessary to undertake such intricate computations in the absence of a computer software. These are, however, the least of the problems.

In accounting and finance, ratios are used to convey performance information to stakeholders in a business atmosphere. These ratios are often relational in nature, meaning that they try to tell us what fraction, level or percentage of efficiency was achieved in the use of certain resources; in other cases, the ratios might be engaged to do comparative/differential analysis between one period's transactions and another's or even to compare the performance of different projects or activities. These are the kinds of information that investors and management need to guide them in their daily decisions and divisional performance evaluation exercises - not the abstraction thinking involved in advanced econometric measurements which have no legal substance in business and commercial transactions. In addition, the measurements used in arriving at financial and accounting ratios are ways different from the means, errors, variations and covariations produced and fitted into most econometric and statistical models of measurement by other social science researchers. Though, means and averages can be employed in financial and accounting performance measurements, the way and mode of their employment will vary significantly with those used for pure econometrics studies.

### *The use of Means*

When faced with two or more groups of data relating to surrogates with values analogous to the measure of a dependent variable, the overall mean can be computed as the average of the groups' means which then assumes the responsibility of representing the dependent variable as a single group which can now be regressed in an ordinary GLM against the underlying independent variables (Carey, n.d.; Conference & Pisa, 2007; Fritz & Berger, 2015; Fritz, Berger, Fritz, & Berger, 2015). The use of the mean as a unifying factor becomes inevitable in financial performance studies involving multiple surrogates of a dependent variable. However, the question remains - which type of mean do one employ: the arithmetic mean or the geometric mean? To answer this question, it is necessary to assess the effect of each type of mean. The *Arithmetic Mean (AM)* is a simple average derived by adding all-inclusive elements which have numerical values together and dividing by the number of elements added. Despite its extensive use to report the central tendency of a data distribution, it suffers from statistical robustness because it is greatly influenced by outliers or extreme values included in the distribution. The *Geometric Mean (GM)* on the other hand also measures the central tendency but it does so by multiplying the numerical elements involved in the set and finding then  $n$ th root of their product. A geometric mean is often used to find a single "figure of merit" for items with multiple properties when comparing different items. It is also used to analyse a set of numbers whose values are meant to be multiplied together or are exponential in nature such as investment interest rate or human population growth rate. It, however, applies only to positive numbers. Nevertheless, the geometric mean (GM) is more

respectful of the intrinsic differences across all the dimensions of the data distribution than the arithmetic mean (Transaction Processing Performance Council, 2011; UNDP, 2011). For the reasons as earlier adduced, the use of geometric mean, is, therefore, favoured and will be used to unify the multiple surrogates of the dependent variables studied in this work.

### *Accounting and financial performance indicators*

Accounting and general financial performance indicators are usually expressed as ratios derived from information contained on corporate financial statements. While ratios such as return on assets (ROA), return on equity (ROE), return on capital employed (ROCE), debt equity ratio (DER), net profit margin (NOM), gross profit margin (GPM), and inventory turnover rate (ITR) are fractional ratios that can be expressed as mere fractions or in percentages, others such as earning per share (EPS), receivable turnover ratio (RTR) or average collection period (ACP) and others may be expressed in monetary or time denominated terms as kobo or cents per share or days. The objective of whichever form the ratio takes is to convey a useful information to the recipient of such information.

The knotty point in the use of accounting ratios, however, is the fact that accounting performance data may include different elements in its composition, such as fractional values, days and monetary denominations which must be fused together to obtain the geometric mean. For instance, supposing a researcher wants to study the relationship between a company's performance as proxied with ROCE, ROA, EPS, and ACP surrogates over a 20-year period, and the personnel (PER), administrative (ADM), financial (FIN), and marketing (MKT) costs; the researcher has the option of studying the relationship between each surrogate of the performance variable and the four independent variables of PER, ADM, FIN, and MKT or combining the four surrogates into one using a geometric mean. If he decides to study them individually, he will end up with four models and four conclusions, which might conflict with one another thereby rendering the process an exercise in futility. However, if he decides to unify them with a mean, he will have to deal with the problem of bringing all the variables under a common denominator since ROCE and ROA will be in fractions while EPS will be either book value or market value denominated, and ACP will be time denominated as ACP is always expressed in days.

To resolve the issue, the EPS and the ACP must be converted to fractions in line with ROCE and ROA. To convert the EPS into a fraction, it is necessary to divide the EPS with the market value of the share of the firm, if it is market regulated or with the par/book value of the share if the firm is not listed or quoted. To convert the ACP to fraction, two steps must be followed – the first step is to divide the ACP with 365, the number of days in a year whilst the second step is to deduct the resulting fraction away from one. This two-prong approach to convert ACP to fraction is necessary because efficiency in credit administration relies much on the shortness of the debtors' collection period. The shorter the period, the more efficient the firm's credit administration is adjudged. When a relatively short ACP is converted to a fraction with the first step, it will show a small value which will be suboptimal when used in a regression analysis, but when this value is deducted from one, it will reveal the true fraction or percentage of efficiency achieved in credit administration by the firm. For instance, assuming companies A and B achieved 32- and 45-days ACP respectively, at first step fractional conversion, company A will have ACP fraction of 0.0877 (8.77%) while company B will return ACP fraction of 0.1233 (12.33%) which is indeed confusing and totally incorrect because from the face value of it, company A performed better than B in credit control, and not the other way round. In order to correct this anomaly, it will be necessary to take away the earlier computed fractions from one, such that the new ACP fractions become 0.9123 (91.23%) for company A and 0.8767 (87.67%) for company B which echoes the reality of credit control events in the two firms. Succinctly, the formula for the EPS and ACP fractional conversions are given in equations (1) and (2) below:

$$\text{EPS (in fraction)} = \frac{\text{EPS}}{\text{Share Price}} \quad (1)$$

$$\text{ACP (in fraction)} = 1 - \frac{\text{ACP}}{365} \quad (2)$$

To complete our sample analysis, it is necessary to bring all the four surrogates of the financial performance dependent variable into one using the geometric mean formula as follows:

$$FP = \sqrt[4]{ROCE} \times ROA \times EPS \times ACP \quad (3)$$

The number 4 in the formula (equation 3) was used there because the number of elements requiring a geometric mean is four. That is to say that the geometric mean of the product of the four elements is the fourth (4<sup>th</sup>) root. If the number is five, then the geometric mean will be the fifth (5<sup>th</sup>) root, and so on.

Where,

FP = Financial Performance ROCE, ROA, EPS, and ACP as previously defined.

With the variables now defined and brought under a uniform measurement platform, we can now fit the ordinary GLM as follows:

$$FP = \beta_0 + \beta_1 PER + \beta_2 ADM + \beta_3 FIN + \beta_4 MKT + \epsilon \quad (4)$$

With this overall unified dependent variable GLM regression model, it will be easy to predict the impact of each of the four cost elements on the overall fortunes or profitability of the company for the 20-year period under review.

## Methodology

The study employed comparative and empirical literature review to compare the methods adopted by 30 Ph.D researchers selected from concluded studies in finance and accounting from four universities in Nigeria. Sample selection targeted only studies on financial performance with four dependent-variable surrogates. Logistic regression was used to convert the *a priori* expectations of the 30 Ph.D research theses into a probabilistic log values' data. The probability values ranged from 0.24 to 0.99 in accordance with the expressed expectation to the actual expectation. The log values were then compared with the individual surrogate performance on the one hand and the surrogates' geometric means on the other hand using ANOVA and logistic regression analysis.

## Data Analysis, Results and Discussions

Table 1 below shows how the final results of the theses agreed with their pre-study *a priori* expectations.

Table 1: Agreement with Theses *A Priori* Expectations

Level Achieved	Number of Theses	Percentage
100%	5	16.67
75%	8	26.67
50%	11	36.66
25%	6	20.00

Source: *Multiproxy by Enyi, 2018*

Table 1 shows that only five of the thirty theses analysed representing 16.67% achieved 100% of their pre-study expected conclusions, while 11 representing 36.66% achieved only 50% of their *a priori* expectations. Eight or 27% were able to achieve 75%; 6 or 20% achieved only 25% of their pre-study conclusions while none had a zero agreement. Since the researchers of the theses under study carried out their analyses with the individual surrogates of the multivariate dependent variables, the level achieved was based on the number of the surrogates that agreed with their pre-study *a priori*.

Table 2: Summary Statistics

Item	Logit	Ln(ROA)	Ln(ROE)	Ln(ROCE)	Ln(EPS)	Ln(Unified)
Mean	0.9	2.12	1.63	1.98	2.89	2.15
Median	0.08	1.96	1.36	1.96	2.87	2.04

Item	Logit	Ln(ROA)	Ln(ROE)	Ln(ROCE)	Ln(EPS)	Ln(Unified)
Max	4.6	3.69	2.73	3.36	3.3	3.27
Min	-1.15	-0.13	-1.97	-0.21	2.17	0.14
Std Dev	1.771	0.909	1.091	0.863	0.59	0.774
Obs	30	30	30	30	30	30

Source: IBM-SPSS 25 Analysis by Enyi, 2018

The resulting models from the data analysis using GLM regression are expressed as follows:

$$\text{Logit1} = -0.678 + 0.734\text{Ln}(\text{Unified}) + \epsilon \quad (5)$$

$$\text{Logit2} = -2.99 - 2.538\text{Ln}(\text{ROA}) + 1.097\text{Ln}(\text{ROE}) + 1.709\text{Ln}(\text{ROCE}) + 1.427\text{Ln}(\text{EPS}) + \epsilon \quad (6)$$

Table 3: Comparative Analysis of Logit1 and Logit2 using IBM-SPSS 25 Output

Performance Parameter Proxies	Influence	Beta Coefficient	t - Statistics	t - Statistics	F - Statistics	F - Statistics
			Value	Sig	Value	Sig
Unified Perf. Variable	0.734	0.278	0.695	> .10	2.426	> .05
ROA	-2.538	-1.183	-1.342	< .10*	22.598	< .05**
ROE	1.097	0.653	-0.790	> .10	22.598	< .05**
ROCE	1.709	0.760	-1.183	< .10*	22.598	< .05**
EPS	1.427	0.215	-2.198	< .10*	22.598	< .05**

\* significant at 10% confidence level \*\* significant at 5% confidence level

Source: IBM-SPSS Analysis by Enyi, 2018

Table 4: ANOVA Test for Unified Performance Variable (Logit1)

Model	Sum of Sqs	Df	Mean Sqrd	F	Sig
Regression	7.196	1	7.196	2.342	.137
Residual	86.039	28	3.073		
Total	93.236	29			

$$R^2 = 0.0772$$

$$\text{Adj. } R^2 = 0.0442$$

$$\text{Durbin-Watson Statistics} = 2.383$$

Source: IBM-SPSS Analysis by Enyi, 2018

Table 5: ANOVA Test for Individual Surrogate Dependent Variables (Logit2)

Model	Sum of Sqs	Df	Mean Sqrd	F	Sig
Regression	15.202	4	3.800	1.216	.328
Residual	78.034	25	3.121		
Total	93.236	29			

$$R^2 = 0.163$$

$$\text{Adj. } R^2 = 0.0291$$

$$\text{Durbin-Watson Statistics} = 2.432$$

Source: IBM-SPSS Analysis by Enyi, 2018

## Discussions

The results of the foregoing analyses have been quite revealing. It is not surprising that the results of the empirical analysis agree largely with the findings from the literature review process. While the use of geometric mean to unify all performance parameter proxies showed no significant difference between itself and the *a priori* expectations or the theses researchers' pre-study perceptive conclusions indicated in the analysed theses, using 10% level of significance ( $\beta_{(30)} = .278$ ,  $t_{(30)} = .695$ ,  $R^2 = .0772$ ,  $p > .10$ ), all the individual surrogate variables excepting ROE returned significant figures for  $t$  at the 10% significant level. Buttrressing the  $t$  statistics findings with  $F$  statistics, the geometric mean unified variable result was also not significant both at the 5% and 10% levels of significance ( $\text{Adj. } R^2 = .0442$ ,  $F_{(1, 28)} = 2.426$ ,  $p > .05$ ), as against the test for the individual performance parameter surrogates which returned overall significant difference in relationship with the analysed theses pre-study expectations ( $\text{Adj. } R^2 = .0291$ ,  $F_{(4, 25)} = 22.598$ ,  $p < .05$ ).

In other words, what the results of the various analysis is telling us is that the theses researchers' pre-conceived conclusions were no different from the results obtained using the geometric mean unified dependent variables. Looking at the figures on table 3, it will be interesting to note that while the geometric mean unified performance variable maintained a positive relationship with the theses *a priori* expectations, one out of the four surrogate variables (ROA) showed strong negative relationship indicating that its individual influence inhibits the positive pull of the other variables, thereby distorting logic of joint conclusion. This trait was also exhibited on table 2 (summary statistics) which showed all the figures of the unified variable as positive but three of the individual surrogate variables returned negative minimum figures.

## Summary, Conclusion and Recommendations

There is no doubt that this paper has been able to adequately address the objectives it set out to achieve. It started by introducing the subject of study and identifying the problem it set out to tackle. While reviewing extant literature frameworks on the subject of multivariate dependent variables, the paper introduced the various methods and formula adaptations that can be employed to unify financial performance data for the purpose of relational research analysis. From the literature review and the earlier discussions of the results of research data analysis, it is evident that financial performance research utilizing multivariate dependent variable has very little space to maneuver with the existing econometric methods of multivariate data analysis. Financial and accounting information are made up of already processed and validated economic data and values which require careful and exact practical treatment which the abstract assumptions of most econometric measurements and analysis may not adequately handle, resulting in the need to adopt a more practical and less abstraction assumptions and methods to manipulate them. The use of geometric means to unify multivariate dependent variables into a single variable is considered the best way to effectively execute a financial performance relational research because, according to Transaction Processing Performance Council (2011), the geometric mean presents a single figure of merit which respects the intrinsic differences across all dimensions of the data distribution (UNDP, 2011).

In view of the aforementioned findings, this paper recommends that more attention be paid to the peculiar nature of financial performance research and the important role that such researches ought to play in the economy of every nation when deciding on the method and tool of analysis to adopt. It will not be out of place if we have a statistical analysis type that takes the peculiar interest of accounting and finance into consideration. If chemistry can develop a special branch of statistics called *chemometrics*, it is possible to have either *finometrics* or *accometrics*.

## References

- Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS-Regression). *Wiley Interdisciplinary Reviews: Computational Statistics* , 2 , 97–106. <https://doi.org/10.1002/wics.51>
- Afthanorhan, W. M. A. B. W. (2013). A comparison of Partial Least Square Structural Equation Modeling (PLS-SEM) and Covariance Based Structural Equation Modeling (CB-SEM) for confirmatory Factor Analysis. *International Journal of Engineering Science and Innovative Technology (IJESIT)* , 2 (5).
- Bu, D. (2007). Bridging the Gap between the State and Measurement of a Chemical System. Retrieved November 29, 2018, from <https://www.camo.com>
- Cameron, R. (2015). Mixed Methods Research. In *Mixed Methods Research Workshop* . Melbourne, Australia: Australian & New Zealand Academy of Management.
- Carey, G. (n.d.). *Chapter 13 GLM : Multiple dependent variables* .
- Conference, E., & Pisa, P. (2007). Multiple Classification Analysis ( MCA ). An , unfortunately , nearly forgotten method for doing linear regression with categorical variables. *Analysis* .
- Fritz, M., & Berger, P. D. (2015). *Improving the User Experience Through Practical Data Analytics* . *Improving the User Experience Through Practical Data Analytics* . <https://doi.org/10.1016/B978-0-12-800635-1.00010-0>
- Fritz, M., Berger, P. D., Fritz, M., & Berger, P. D. (2015). Chapter 10 – Can you relate in multiple ways? Multiple linear regression and stepwise regression. In *Improving the User Experience Through Practical Data Analytics* . <https://doi.org/10.1016/B978-0-12-800635-1.00010-0>
- Golub, G. H., & Van Loan, C. F. (1980). An Analysis of the total least squares problem. *SIAM Journal of Numerical Analysis* ,17 (6), 883–890. Retrieved from <https://epubs.siam.org>
- Gruszczy, M. (2009). Quantitative Methods in Accounting Research. *Metody Ilosciowe W Badaniach Ekonomicznych* , 10 (June 2009), 76–87. <https://doi.org/10.2139/ssrn.1803638>
- Helland, I. S. (1990). PLS regression and statistical models. *Scandinavian Journal of Statistics* , 17 (2), 97–114.
- Helwig, N. E. (2017). Canonical Correlation Analysis Copyright c 2017 by Nathaniel E . Helwig, 1–39.
- Jensen, A. C. (2007). Deming regression. *MethComp Package* . Gentofte, Denmark: Steno Diabetes Center.
- Kelly, B., & Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics. High Dimensional Problems in Econometrics* . ,186 (2), 294–316. <https://doi.org/10.1016/j.jeconom.2015.02.011>
- Lingjaerde, O.-C., & Christophersen, N. (2000). Shrinkage structure of Partial Least Squares. *Scandinavian Journal of Statistics* ,27 (3), 459–473. <https://doi.org/10.1111/1467-9469.00201>
- NCSS. (1989). Canonical Correlation. Retrieved November 30, 2018, from <http://www.ncss.com>
- Price, P. C., Jhangiani, R. S., & Chiang, I.-C. A. (2015). *Research Methods in Psychology* . Retrieved from <https://opentextbc.ca/>
- Saeboa, S., Almoya, T., Flatbergb, A., Aastveita, A. H., & Martens, H. (2008). LPLS-regression: a method for prediction and classification under the influence of background information on predictor variables. *Chemometrics and Intelligent Laboratory Systems* , 91 (2), 121–132. <https://doi.org/10.1016/j.chemolab.2007.10.006>
- Steiger, J. H. (n.d.). Canonical Correlation Analysis.



Transaction Processing Performance Council. (2011). TPC-D-Frequently Asked Questions (FAQ). Retrieved November 30, 2018, from <https://www.tpc.org>

UNDP. (2011). Frequently Asked Questions - Human Development Reports. Retrieved November 30, 2018, from <https://hdr.undp.org>

Wold, S., Sjostrom, M., & Eriksson, L. (2001). PLS-regression : a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* , 58 (2), 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)