

2006

# An Overview of Partial Least Squares

Dante M. Pirouz  
*Western University*

Follow this and additional works at: <https://ir.lib.uwo.ca/iveypub>



Part of the [Business Commons](#)

---

## Citation of this paper:

Pirouz, Dante M., "An Overview of Partial Least Squares" (2006). *Business Publications*. 24.  
<https://ir.lib.uwo.ca/iveypub/24>

## **An Overview of Partial Least Squares**

Dante M. Pirouz  
Doctoral Student

The Paul Merage School of Business  
University of California, Irvine  
[dpirouz04@merage.uci.edu](mailto:dpirouz04@merage.uci.edu)  
<http://www.merage.uci.edu/~dpirouz04/>

Draft  
October 10, 2006

## **Introduction**

Partial least squares analysis is a multivariate statistical technique that allows comparison between multiple response variables and multiple explanatory variables. Partial least squares is one of a number of covariance-based statistical methods which are often referred to as structural equation modeling or SEM. It was designed to deal with multiple regression when data has small sample, missing values, or multicollinearity. Partial least squares regression has been demonstrated on both real data and in simulations (Garthwaite, 1994, Tennenhaus, 1998). It has been very popular in hard science, especially chemistry and chemometrics, where there is a big problem with a high number of correlated variables and a limited number of observations. Its use in marketing has been more limited although data has similar problems (Ryan, Rayner, & Morrison, 1999).

## **History**

Path analysis and causal modeling was introduced by Wright in the 1920s (Falk & Miller, 1992, Wright, 1921). Developed in the late 1960's by Herman O.A. Wold, partial least squares regression was originally developed for use in the field of econometrics but was embraced by the chemistry field for use in analytical, physical, clinical chemistry studies (Geladi & Kowalski, 1986). Wold created partial least squares because he wanted to address weak theory and weak data (Wold, 1982, Wold, Geladi, Esbensen, & Öhman, 1987).

## **What is Partial Least Squares?**

Partial least squares designed to cope with problems in data specifically, small datasets, missing values and multicollinearity. In contrast, ordinary least squares (OLS) regression yields unstable results when data has small sample size, missing values and multicollinearity between predictor in OLS regression increases standard error of their estimated coefficients (Field, 2000). High multicollinearity increases risk of theoretically sound predictor to be rejected from regression model as non-significant variable.

The goal of partial least squares is to predict Y from X and to describe the common structure underlying the two variables (Abdi, 2003). Partial least squares is a regression method allows for the identification of underlying factors, which are a linear combination of the explanatory variables or X (also known as latent variables) which best model the response or Y variables (Talbot, 1997).

Although similar to principal components analysis (PCA) regression and canonical analysis and alternating least squares, it is considered to be a better alternative to multiple linear regression and PCA regression methods since it provides for more robust model parameters that do not change with new calibration samples from the population (Falk & Miller, 1992, Geladi & Kowalski, 1986). Furthermore, partial least squares is an improvement on PCA since the solution derived from partial least squares is constrained to the part of the covariance matrix that is directly related to the experimental manipulation or that relates to behavior (McIntosh, Chau, & Protzner, 2004).

The term partial least squares specifically means the computation of the optimal least squares fit to part of a correlation or covariance matrix (McIntosh, Chau, & Protzner, 2004, Wold, 1982). The part of the correlation or covariance matrix that the

least squares are fit to is the “cross-block” correlation between the exogenous or X variables and the dependent measures or Y variables. Partial least squares measures covariation between two or more block of variables and creates a new set of variables that is optimized for maximum covariance (not maximal correlation) using the fewest dimensions (McIntosh, Bookstein, Haxby, & Grady, 1996).

Partial least squares is sometimes called soft modeling because while OLS regression makes hard assumptions such as no multicollinearity in the independent variable, soft modeling refers to softening of these assumptions. Partial least squares is a linear technique. Partial least squares is preferred as predictive technique and not as an interpretive technique except for exploratory analysis before using interpretive techniques such as multiple linear regression or SEM.

Partial least squares optimal linear relationships are computed between latent variables and can be interpreted as the best set of predictions available for a study given all the limitations (Falk & Miller, 1992). Soft modeling is a way of estimating the likelihood of an event given information about other events.

### **Mathematics Underlying Partial Least Squares**

The basic mathematics underlying partial least squares in matrix notation are as follows:

$$N = W'Y$$

$$Y = PN + E$$

$$\text{Substituting } Y = PW'Y + E = PW'Y + (I - PW')Y$$

Where: N = a principal component

Y = a set of observed scores

W = a set of composing weights

P = a set of principal component loadings

E = residual variance

The matrix solutions are based on singular value decomposition, which does not require matrix inversions:

$$R = W'DP$$

Where W = an orthonormal matrix of left singular vectors, with

$$WW' = I, \text{ the identity matrix}$$

P = another orthonormal matrix of right singular values

D = a diagonal matrix of singular values

If D is a diagonal of eigenvalues, then the rows of P will be principal component loadings and the columns of W'D the principal component scores.

### **Assumptions and Significance Testing**

As an extension of multiple linear regression, partial least squares regression has many of the same assumptions. For example, one should be concerned with outliers and nonlinear data relationships when using partial least squares. Because the distribution of partial least squares is unknown, there is no conventional significance test. However, significance can be tested through bootstrap methods such as jackknife which is a resampling method. The problem with using a resampling method to determine significance is that although there are no specific sample size requirements, the smaller

the sample, the more likely that fitted confidence limits will be fitted to noise in data instead of true distribution.

### **Key Definitions in Partial Least Squares**

- Partial Least Squares Factors:
  - These are the latent variables extracted as linear combinations of the manifest IVs
  - Usually 3-7 factors will account for 99% of the variation accounted for
- Partial Least Squares Responses
  - These are latent variables extracted as linear combinations of manifest response variables
  - Usually 3-7 factors will account of 99% of variation
- Cross Validation
  - Partial least squares model is developed for all cases except 1, then tested on the hold out
  - Repeated N times with each case used as validation case in turn

### **Statistical Methodology**

- Partial least squares regression aims to produce a model that transforms a set of correlated explanatory variables into a new set of uncorrelated variables

- The parameter coefficients in Partial least squares regression are derived for direct correlation between predictor variables and the criterion variable
  - In OLS multicollinearity is tested using:
    - Bivariate correlations – if high
    - Variance inflation factors (VIF) >1
    - Uncentered cross-products matrix eigenvalues
      - If 1 is very large, regression is unstable
    - Survey data are very sensitive to multicollinearity!
  - Partial least squares regression is an alternative to OLS regression, canonical correlation or SEM
    - For analysis of systems of independent and response variables
  - First developed for econometrics but used in chemometric and industrial applications
  - Partial least squares is a predictive technique that can handle many IVs even with multicollinearity
  - Like canonical correlation, it can also relate set of IV to a set of multiple dependent (response) variables
  - However, partial least squares is not good as an explanatory technique since it has low power to filter out variables for minor causal importance (Tobias, 1997)
- Process:
  - 1 set of latent variables is extracted for set of manifest independents



- 1 set of latent variables is extracted simultaneously of set of manifest response (dependent) variables
  - Extraction is based on decomposition of crossproduct matrix involving both independent and response variables
- The x-scores of independent latents are used to predict y-scores or response latent(s)
- Predicted y-scores are used to predict manifest response variables
- The x and y scores are selected by partial least squares so the relationship of successive pairs of x and y scores is as strong as possible

### **Advantages of Partial Least Squares**

The following are the key advantages of partial least squares:

- Able to model multiple dependent as well as multiple independence variables
- Can handle multicollinearity in IVs
- Robust despite data noise and missing data
- Creates independent latents directly on the basis of cross products involving response variable(s) = stronger predictions
- Allows for reflective and formative latents
- Applied to small sample
- Distributional free
- Handle range of variables: nominal, ordinal, continuous

## **Disadvantages of Partial Least Squares**

Some of the disadvantages of partial least squares are as follows:

- Difficulty in interpreting loadings of independent latent variables (based on crossproduct relations with response variables not, as in conventional factor analysis, on correlations among manifest independents)
- Distributional properties of estimates not known
  - Can't get significance unless run bootstrap
- Lack of model test statistics

## **What is structural equation modeling and LISREL?**

Structural equation modeling (SEM) is a family of statistical techniques which incorporates and integrates path analysis and factor analysis. SEM is similar to multiple regression but considered more powerful, illustrative and robust since it takes into account the modeling of interactions, nonlinearities, correlated independents, measurement error, correlated error terms, multiple latent independents each measured by multiple indicators and one or more latent dependents also each with multiple indicators. It is also considered a powerful alternative to path analysis, time series analysis, and analysis of covariance. SEM is an extension of the general linear model (GLM) of which multiple regression is a part. SEM is a confirmatory rather than an exploratory procedure.

The SEM modeling process is based on two steps: validating the measurement model and fitting the structural model. SEM software is often used to create a hybrid model with both latent variables or factors and paths specified by the connected latent variables. But SEM can also be used to model in which each variable has only one

indicator which is a type of path analysis or it can be used where each variable has multiple indicators but there are no direct effects (arrows) connective the variables as a type of factor analysis. SEM is synonymous with covariance structure analysis, covariance structure modeling and analysis of covariance structure. SEM methods usually require a strong theoretical basis in order to determine the causal model relationships.

LISREL is one of a number of statistical packages for doing SEM and has been used in sociology and social science research. It estimates the coefficients in a set of structural equations. LISREL is able to analyze models with discrete latent variables (categorical or ordinal) instead of continuous variables (Kline, 2005). Also multilevel structural equation models for hierarchical data can be estimated using LISREL.

### **Differences between Partial Least Squares and SEM/LISREL**

Sample Size:

- Partial least squares can be applied to small samples unlike LISREL
- LISREL estimation requires large samples

Theoretical Basis:

- Need to set a starting point for SEM/LISREL should be theoretically based
- If no theory or theoretical basis is weak, partial least squares is better

Distribution Assumptions:

- Partial least squares is distribution free but LISREL must have strong distribution assumptions
- LISREL: assumes parameters = strong distributional assumptions

- Partial least squares: Series of interdependent OLS regression = no distribution assumptions

#### Indicators:

- Partial least squares allows for both reflective and formative latent variables but LISREL only allows reflective latent variables
- Partial least squares: forms indices = indicators of perform
- LISREL: factor = indicators of constructs

#### Significance Testing:

- LISREL = Models can be tested and falsified
- Partial least squares can not be tested and falsified
- Can estimate parameters in partial least squares and testing fit in LISREL

#### Model:

- LISREL: Causal
- Partial least squares: predictive

#### Maximizing Components:

- LISREL: factor or covariance based explain covariance
- Partial least squares: components based maximize variance
- LISREL: residual covariance are minimized – max correlation
- Partial least squares = residual variance minimized – min difference

#### Point Estimation:

- LISREL: MLE = Probability of observed data max given hypothetical model
- Partial least squares = fixed point estimation = least square

#### Optimization:

- LISREL = Poses and solves global optimization problem
- Partial least squares = limits global optimization computation thru ordinarily multiple regression

#### Software Errors:

- LISREL often has problems of inadmissible and factor indeterminacy
- Partial least squares has relatively no problems or crashing when running models

#### **Formative vs. Reflexive Variables**

An important consideration in designing a partial least squares model is which mode to be used in relating unobservables to the empirical indicators (Fornell & Bookstein, 1982). Partial least squares allows the observed indicators to be treated as either reflective or formative indicators. Reflective indicators are similar to classical test theory and factor analysis models. They attempt to account for observed variances or covariances. Formative indicators are not designed for observed variables and are instead are best used to minimize residuals in the structural relationship. The choice of whether to use reflective or formative indicators greatly affects estimation procedures. If the study's goal is to account for observed variables then reflective indicators are best. If the goal is to explain abstract or unobserved variance, then formative indicators are more explanatory. Both formative and reflexive indicators can be used within a single model.

## Steps for Doing a Partial Least Squares Study

### 1. Draw the path diagram

#### a. Falk & Miller recommend using reticular action modeling (RAM)

nomogram procedure

#### i. Based on 1) theoretical constructs = latent variables = circles, 2)

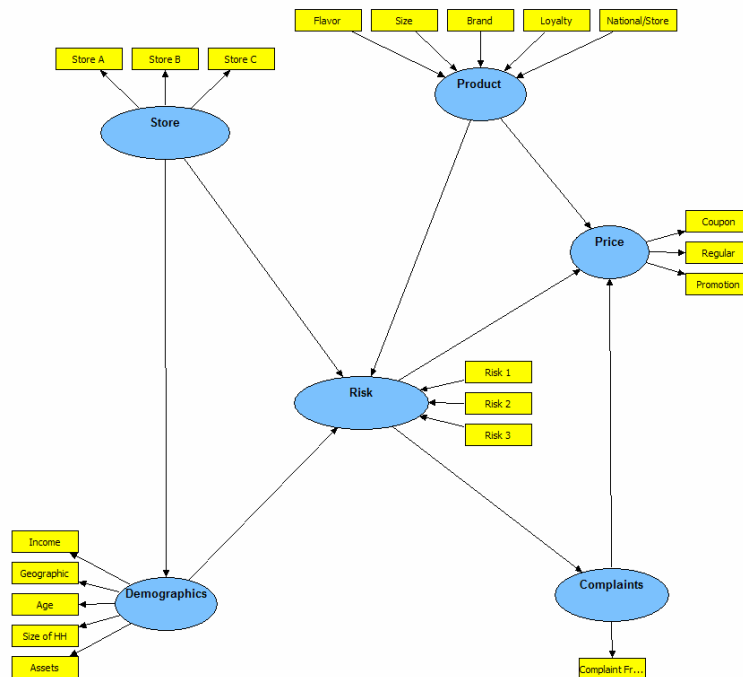
measured variables = squares, 3) asymmetrical relationships

=single headed arrow, 4) symmetrical relationships = double

headed arrow

## Nomographic Representations in Partial Least Squares

Like SEM, partial least squares has the added benefit of providing a graphical representation of the relationships between the variables. An example of this type of output is as follows:



4 features of nomograms:

- Ordering of theoretical constructs
- Specifying of arrows
- Specifying of inner model
- Blocking the manifest and theoretical variables and establishing their directions

### **Rules of Thumbs for Interpretation**

- 1) There should be at least 3 manifest variables for a latent variable
- 2) Loadings on the paths between components and manifest variables should be  $\geq .55$
- 3) The  $R^2$  or variance explained for endogenous variables should be  $\geq .10$
- 4) A predictor variable should account for at least 1.5% of the variance in a predicted variable

### **Conclusion**

Overall, partial least squares is a suitable method choice for datasets that do not fit the traditional expectations demanded by ordinary regression. When datasets are small, suffer from multicollinearity, missing values or where the distribution is unknown, partial least squares allows the user to minimize the adverse effects of these conditions. In addition, partial least squares has the added benefit of offering a nomographic representation of the relationships between variables which can offer an opportunity for additional insight into the research question. Given the often unique nature of

management and marketing data, partial least squares is an underutilized but a well suited statistical method of analysis that a researcher may consider utilizing.



## References

- Abdi, H. 2003. Partial Least Squares (PLS) Regression. In Lewis-Beck, M., A. Bryman, & T. Futing, editors, *Encyclopedia of Social Sciences Research Methods*. Thousand Oaks: Sage.
- Falk, R. Frank & Nancy B. Miller. 1992. *A Primer for Soft Modeling*. Akron: University of Akron.
- Field, A. 2000. *Discovering Statistics Using SPSS for Windows*. London: Sage
- Fornell, Claes & F.L. Bookstein. 1982. Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory. *Journal of Marketing Research*, XIX: 440-52.
- Garthwaite, Paul H. 1994. An Interpretation of Partial Least Squares. *Journal of the American Statistical Association*, 89(425): 122.
- Geladi, Paul & Bruce R. Kowalski. 1986. Partial Least Squares Regression: A Tutorial. *Analytica Chimica Acta*, 185: 1-17.
- Kline, Rex B. 2005. *Structural Equation Modeling*. New York, NY: Guilford Press.
- McIntosh, A.R., F.L. Bookstein, James V. Haxby, & C.L. Grady. 1996. Spatial Pattern Analysis of Functional Brain Images Using Partial Least Squares. *NeuroImage*, 3: 143-57.
- McIntosh, A.R., W.K. Chau, & A.B. Protzner. 2004. Spatiotemporal Analysis of Event-Related fMRI Data Using Partial Least Squares. *NeuroImage*, 23: 764-75.
- Ryan, M.J., R. Rayner, & A. Morrison. 1999. Diagnosing customer loyalty drivers: Partial least squares vs. regression. *Marketing Research* 19-26.
- Talbot, M. 1997. *Partial Least Squares Regression*.
- Tennenhaus, M. 1998. *La Regression PLS Theorie et Pratique*. Paris: Editions Technip.
- Tobias, R.D. 1997. *An Introduction to Partial Least Squares Regression*. Cary, NC: SAS Institute.
- Wold, H. 1982. Soft Modeling: The Basic Design and Some Extensions. In Wold, H. & K.G. Joreskog, editors, *Systems Under Indirect Observations: Causality, Structure, Prediction*. Amsterdam: Elsevier.
- Wold, Svante, Paul Geladi, Kim Esbensen, & Jerker Öhman. 1987. Multi-way principal components-and PLS-analysis. *Journal of Chemometrics*, 1(1): 41-56.
- Wright, Sewell. 1921. Correlation and Causation. *Journal of Agricultural Research*.