Partial Least Squares Regression (part I)

Predictive Modeling & Statistical Learning

Gaston Sanchez

PLS Regression

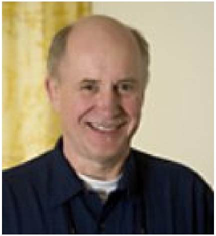
Introduction

Many real-life data sets contain (highly) correlated predictors (multicollinearity), and/or more predictors than observations ($p > n$). In these situations, regression by OLS falls short.

Partial Least Squares (PLS) Regression is one approach that provides an interesting alternative to OLS.

Introduction

PLS Regression was mainly developed in the early 1980s by Scandinavian chemometricians Svante Wold and Harald Martens.



Svante Wold

Harald Martens

The theoretical background was based on the *PLS Modeling* framework of Herman Wold (Svante's father).

THE MULTIVARIATE CALIBRATION PROBLEM IN CHEMISTRY SOLVED BY THE PLS METHOD

S. Wold, H. Martens and H. Wold

Institute of Chemistry, Umea University, Umea, Sweden,
Norwegian Food Research Institute, As, Norway, and
Department of Statistics, Uppsala University, Sweden and
Departement d'Econometrie, Université de Genève, Switzerland

Introduction.

In chemistry, slow and specialized "wet chemistry" methods are rapidly substituted by fast and general instrumentalised methods (Kowalski, 1975). One common problem is how to use spectroscopy to determine the concentrations of various constituents in complicated samples. When the constituents don't absorb light in separated frequency regions, one must utilize a combination of many spectral frequencies to estimate the concentrations. The problem of how to optimally combine the absorbtions at several frequencies (or other chemical "sensors") in order to approximate a measured set of concentrations is called the multivariate calibration problem.

In this problem we have found the PLS method with two blocks in mode A (H.Wold, 1982), to be very useful. PLS=Partial Least Squares models in latent variables. In a PLS analysis, the data matrix is divided variable-wise into a

PLSR Algorithm Pseudocode (1984)

4. Details of the estimation. The PLS algorithm, in its general, mode A formulation, deals with variables blocked in q blocks, and forms a sequence of rank one approximations to the combined data matrix. In this paper, we consider only the case with two blocks, one of them furthermore restricted to consisting of only one variable. Let the data matrices for the two blocks be X and y, and denote the combined matrix by

$$Z = [X|y].$$

We then successively form a sequence of residual matrices Z_s, using the following algorithm:

ALGORITHM PLS.

1. Start $Z_1 = [X|y]$, $b_0 = 0$.
2. For $s = 1, 2, \cdots$, until $\|Z_s\|$ is small
    1. $u_s = X_s X'_s y_s / \|X'_s y_s\|$.
    2. $c_s = Z'_s u_s / u'_s u_s$, $c_s = (a'_s, \rho_s)'$.
    3. $Z_{s+1} = Z_s - u_s c'_s$.
    4. Solve $A'_s b_s = r_s$, $r_s = (\rho_1, \cdots, \rho_s)'$ for b_s.

Super simple notation and very compact, but also cryptic. Good for programming purposes, Bad for users understanding.

About PLS Regression

- PLS Regression was developed as an algorithmic solution.
- No optimization criterion was defined.
- It was introduced in the fields of chemometrics
  where almost immediately became a hit
- It slowly attracted the attention of curious applied statisticians.
- It took a couple of years (late 1980s - early 1990s) for applied mathematicians to discover its properties.
- Nowadays there are several versions (flavors) of the algorithm to compute a PLS regression.

PLS Regression Stages

I will show you what I consider the “standard” algorithm, thoroughly describing its steps.

Then I will give you some remarks on what PLSR does, describe its properties, some variations, and review some examples.

PLS Regression Methodology

About PLS Regression

- PLS regression is somewhat close to Principal Components regression (PCR).
- Like PCR, PLSR involves projecting the response onto uncorrelated components (i.e. linear combinations of predictors).
- Unlike PCR, the way PLS components are extracted is by taking into account the response variable.

PLS Regression Idea

We seek for components $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_H]$ such that:

- they approximate the predictors: $\mathbf{X} = \mathbf{Z}\mathbf{P}^\mathsf{T} + \mathbf{E}$
- they also predict the response: $\mathbf{y} = \mathbf{Z}\mathbf{d} + \mathbf{e}$

where:

- $\mathbf{P}$ is a matrix of loadings
- $\mathbf{E}$ is a matrix of X-residuals
- $\mathbf{d}$ is a vector of regression coefficients
- $\mathbf{e}$ is a vector of y-residuals

PLS Regression Idea

As we will see, it turns out the the PLS components can be expressed as linear combinations of the predictors:

$$\mathbf{Z} = \mathbf{X}\mathbf{W}^*$$

Moreover, the PLS regression equation can be written in terms of the predictors:

$$\hat{\mathbf{y}} = \mathbf{Z}\mathbf{d} = \mathbf{X}\mathbf{W}^*\mathbf{d}$$

PLS Regression Stages

Here's an oversimplified description of PLSR recipe:

- Stage 1: Look for m orthogonal components $\mathbf{z_h}$ that explain well the predictors, and are well correlated with the response.
- m is determined by cross-validation
- Stage 2: Regression of $\mathbf{y}$ on the PLS components $\mathbf{z_h}$
- Stage 3: Express the regression in terms of $\mathbf{X}$

This obviously does not tell us how to compute the PLS components. It's just a *game plan* describing the main stages.

PLS Regression Idea

There is an implicit assumption in the PLS regression model: both $\mathbf{X}$ and $\mathbf{y}$ are assumed to be functions of a reduced number of components $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_m]$

The model for the X-space is:

$$\hat{\mathbf{X}} = \mathbf{Z}\mathbf{P}^{\mathsf{T}}$$

In turn, the prediction model for $\mathbf{y}$ is given by:

$$\hat{\mathbf{y}} = \mathbf{Z}\mathbf{d}$$

Data set `cars2004`

Data cars2004

```
                                price engine cyl  hp city_mpg
Acura 3.5 RL 4dr                43755    3.5   6 225       18
Acura 3.5 RL w/Navigation 4dr 46100    3.5   6 225       18
Acura MDX                       36945    3.5   6 265       17
Acura NSX coupe 2dr manual S    89765    3.2   6 290       17
Acura RSX Type S 2dr            23820    2.0   4 200       24
Acura TL 4dr                    33195    3.2   6 270       20
                                hwy_mpg weight wheel length width
Acura 3.5 RL 4dr                     24   3880   115    197    72
Acura 3.5 RL w/Navigation 4dr        24   3893   115    197    72
Acura MDX                            23   4451   106    189    77
Acura NSX coupe 2dr manual S         24   3153   100    174    71
Acura RSX Type S 2dr                 31   2778   101    172    68
Acura TL 4dr                         28   3575   108    186    72
```

Correlation Matrix

	engine	cyl	hp	city_mpg	hwy_mpg	weight	wheel	length	width
price	0.6	0.654	0.836	-0.485	-0.469	0.476	0.204	0.210	0.314
engine		0.912	0.778	-0.706	-0.708	0.812	0.631	0.624	0.727
cyl			0.792	-0.670	-0.664	0.731	0.553	0.547	0.621
hp				-0.672	-0.652	0.631	0.396	0.381	0.500
city_mpg					0.941	-0.736	-0.481	-0.468	-0.590
hwy_mpg						-0.789	-0.455	-0.390	-0.585
weight							0.751	0.653	0.808
wheel								0.867	0.760
length									0.752

OLS Regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	32536.025	17777.488	1.8302	6.802e-02
engine	-3273.053	1542.595	-2.1218	3.451e-02
cyl	2520.927	896.202	2.8129	5.168e-03
hp	246.595	13.201	18.6797	1.621e-55
city_mpg	-229.987	332.824	-0.6910	4.900e-01
hwy_mpg	979.967	345.558	2.8359	4.817e-03
weight	9.937	2.045	4.8584	1.741e-06
wheel	-695.392	172.896	-4.0220	6.980e-05
length	33.690	89.660	0.3758	7.073e-01
width	-635.382	306.344	-2.0741	3.875e-02

```
Call:
lm(formula = price ~ ., data = dat)

Residuals:
   Min     1Q Median     3Q    Max 
-21534  -5411   -352   4054  92763 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 32536.025  17777.488   1.830  0.06802 .  
engine      -3273.053   1542.595  -2.122  0.03451 *  
cyl          2520.927    896.202   2.813  0.00517 ** 
hp            246.595     13.201  18.680  < 2e-16 ***
city_mpg     -229.987    332.824  -0.691  0.48998    
hwy_mpg       979.967    345.558   2.836  0.00482 ** 
weight          9.937      2.045   4.858 1.74e-06 ***
wheel        -695.392    172.896  -4.022 6.98e-05 ***
length         33.690     89.660   0.376  0.70731    
width        -635.382    306.344  -2.074  0.03875 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10110 on 375 degrees of freedom
Multiple R-squared:  0.745,Adjusted R-squared:  0.7389 
F-statistic: 121.7 on 9 and 375 DF,  p-value: < 2.2e-16
```

Correlations and OLS Coefficients

```
            correlation  coefficient
engine        0.5997873 -3273.05304
cyl           0.6544123  2520.92691
hp            0.8360930   246.59496
city_mpg     -0.4854130  -229.98735
hwy_mpg      -0.4694315   979.96656
weight        0.4760867     9.93652
wheel         0.2035464  -695.39157
length        0.2096682    33.69009
width         0.3135383  -635.38224
```

Some correlation signs don't match coefficient signs

First PLS Component

First PLS Component

- We assume mean-centered variables $\mathbf{X}$ and $\mathbf{y}$.
- Possibly they can also be standardized (var $= 1$) or scaled (e.g. norm 1)
- We start by setting $\mathbf{X_0} = \mathbf{X}$.
- And setting $\mathbf{y_0} = \mathbf{y}$.

First PLS Component

- We seek a first component $\mathbf{z_1} = \mathbf{X_0 w_1}$
- The vector of weights $\mathbf{w_1}$ is obtained by computing the covariance between predictors and response:

$$\mathbf{w_1} = \mathbf{X_0^{\mathsf{T}} y} \propto cov(\mathbf{X_0}, \mathbf{y})$$

- For convenience, we normalize $\mathbf{w_1}$ to unit norm.

First PLS Component

Vector of weights $\mathbf{w_1}$

```
                   [,1]
engine    0.001782118
cyl       0.002857956
hp        0.171985612
city_mpg -0.007484109
hwy_mpg  -0.007752089
weight    0.984987298
wheel     0.004225081
length    0.008131684
width     0.003089621
```

First PLS Component

Here's another view. We get a first component $\mathbf{z_1}$ such that:

$$\mathbf{z_1} = \mathbf{X_0 w_1} = w_{11}\mathbf{x_1} + w_{12}\mathbf{x_2} + \cdots + w_{1p}\mathbf{x_p}$$

where the weights w_{1j} are obtained as:

$$w_{1j} = \frac{cov(\mathbf{x_j}, \mathbf{y})}{\sqrt{\sum_{j=1}^{p} cov^2(\mathbf{x_j}, \mathbf{y})}}$$

(these weights have the normality constraint: $\|\mathbf{w_1}\| = 1$)

First PLS Component

First pls component (first 10 elements ...)

```
                                        [,1]
Acura 3.5 RL 4dr                   344.24572
Acura 3.5 RL w/Navigation 4dr  357.05055
Acura MDX                          913.48050
Acura NSX coupe 2dr manual S  -360.90753
Acura RSX Type S 2dr          -745.89228
Acura TL 4dr                        51.39841
Acura TSX 4dr                     -300.53740
Audi A4 1.8T 4dr                  -284.07748
Audi A4 3.0 4dr                    -68.58479
Audi A4 3.0 convertible 2dr     278.15085
```

First PLS Component

We use the first PLS component $\mathbf{z_1}$ to regress each $\mathbf{x_j}$ onto $\mathbf{z_1}$ in order to obtain a vector $\mathbf{p_1}$ of loadings:

$$\mathbf{p_1} = \mathbf{X_0}^{\mathsf{T}}\mathbf{z_1}/\mathbf{z_1}^{\mathsf{T}}\mathbf{z_1}$$

```
                 [,1]
engine    0.001176718
cyl       0.001561745
hp        0.064016991
city_mpg -0.005536001
hwy_mpg  -0.006343509
weight    1.003819205
wheel     0.007551534
length    0.012276141
width     0.003862309
```

First PLS Component

We also use the first PLS component to regress $\mathbf{y_0}$ on $\mathbf{z_1}$.

The regression coefficient d_1 is given by:

$$d_1 = \mathbf{y_0}^\mathsf{T}\mathbf{z_1}/\mathbf{z_1}^\mathsf{T}\mathbf{z_1}$$

The regression equation w.r.t. $\mathbf{z_1}$ is thus:

$$\begin{aligned}\mathbf{y} &= d_1\mathbf{z_1} \\ &= 13.611\mathbf{z_1}\end{aligned}$$

First PLS Component

After obtaining the first PLS component $\mathbf{z_1}$, we **deflate** each $\mathbf{x_j}$ w.r.t. $\mathbf{z_1}$:

$$\mathbf{X_1} = \mathbf{X_0} - \mathbf{z}_1\mathbf{p}_\mathbf{1}^\mathsf{T}$$

where $\mathbf{X_1}$ is a matrix of X-residuals.

We also deflate $\mathbf{y_0}$ w.r.t. $\mathbf{z_1}$:

$$\mathbf{y_1} = \mathbf{y_0} - d_1\mathbf{z_1}$$

Second PLS Component

Second PLS Component

The 2nd PLS component $\mathbf{z_2}$ is obtained in the same way, but now operating with $\mathbf{X_1}$ and $\mathbf{y_1}$:

$$\mathbf{z_2} = \mathbf{X_1 w_2} = w_{21}\mathbf{x_{11}} + w_{22}\mathbf{x_{12}} + \cdots + w_{2p}\mathbf{x_{1p}}$$

where the weights w_{2j} are obtained as:

$$w_{2j} = \frac{cov(\mathbf{x_{1j}}, \mathbf{y})}{\sqrt{\sum_{j=1}^{p} cov^2(\mathbf{x_{1j}}, \mathbf{y})}}$$

and normality constraint $\|\mathbf{w_2}\| = 1$

Second PLS Component

Vector of weights $\mathbf{w_2}$

```
                  [,1]
engine     0.005515374
cyl        0.011808873
hp         0.983627243
city_mpg  -0.017747864
hwy_mpg   -0.012832587
weight    -0.171564444
wheel     -0.030305000
length    -0.037757269
width     -0.007039429
```

Second PLS Component

Second pls component (first 8 elements)

$$\mathbf{z_2} = \mathbf{X_1 w_2} = w_{21}\mathbf{x_{11}} + w_{22}\mathbf{x_{12}} + \cdots + w_{2p}\mathbf{x_{1p}}$$

```
                                      [,1]
Acura 3.5 RL 4dr                -12.053948
Acura 3.5 RL w/Navigation 4dr -12.878753
Acura MDX                        -7.619465
Acura NSX coupe 2dr manual S  100.553595
Acura RSX Type S 2dr            33.927682
Acura TL 4dr                    52.930801
Acura TSX 4dr                    4.785062
Audi A4 1.8T 4dr               -26.546117
```

Second PLS Component

We use the second PLS component $\mathbf{z_2}$ to regress each $\mathbf{x_{1j}}$ onto $\mathbf{z_2}$ in order to obtain loadings $\mathbf{p_2}$:

$$\mathbf{p_2} = \mathbf{X_1}^\mathsf{T}\mathbf{z_2}/\mathbf{z_2}^\mathsf{T}\mathbf{z_2}$$

```
                 [,1]
engine    0.006139666
cyl       0.011322638
hp        0.984752947
city_mpg -0.024940504
hwy_mpg  -0.019595807
weight   -0.172152033
wheel    -0.014980542
length   -0.013459092
width    -0.001586278
```

Second PLS Component

We also use the second PLS component to regress $\mathbf{y_1}$ on $\mathbf{z_2}$.

The regression coefficient d_2 is given by:

$$d_2 = \mathbf{y_1}^\mathsf{T}\mathbf{z_2}/\mathbf{z_2}^\mathsf{T}\mathbf{z_2}$$

The regression equation w.r.t. $\mathbf{z_1}$ and $\mathbf{z_2}$ is thus:

$$\begin{aligned}\hat{\mathbf{y}} &= d_1\mathbf{z_1} + d_2\mathbf{z_2} \\ &= 13.611\mathbf{z_1} + 247.077\mathbf{z_2}\end{aligned}$$

Second PLS Component

After obtaining the second PLS component $\mathbf{z_2}$, we **deflate** each $\mathbf{x_{1j}}$ w.r.t. $\mathbf{z_2}$:

$$\mathbf{X_2} = \mathbf{X_1} - \mathbf{z_2}\mathbf{p_2}^\mathsf{T}$$

where $\mathbf{X_2}$ is a matrix of $X_{(1)}$-residuals.

We also deflate $\mathbf{y_1}$ w.r.t. $\mathbf{z_2}$:

$$\mathbf{y_2} = \mathbf{y_1} - d_2\mathbf{z_2}$$

Subsequent PLS Components

Subsequent PLS Component

The procedure continues sequentially, operating with the deflated matrix $\mathbf{X_{h-1}}$ and the deflated vector $\mathbf{y_{h-1}}$

- weights: $\mathbf{w_h} \propto \mathbf{X_{h-1}^\mathsf{T} y_{h-1} / y_{h-1}^\mathsf{T} y_{h-1}}$
- component: $\mathbf{z_h} = \mathbf{X_{h-1} w_h}$
- loadings: $\mathbf{p_h} = \mathbf{X_{h-1}^\mathsf{T} z_h / z_h^\mathsf{T} z_h}$
- coeff: $d_h = \mathbf{y_{h-1}^\mathsf{T} z_h / z_h^\mathsf{T} z_h}$

In theory, up to m PLS components can be extracted, where m is the rank of $\mathbf{X}$.

In practice, however, the number of components m is determined by cross-validation.

Subsequent PLS Component

The regression equation w.r.t. $\mathbf{z_1}, \mathbf{z_2}, \ldots, \mathbf{z_9}$ is:

$$\begin{aligned}\hat{\mathrm{y}} = {} & 13.611\mathbf{z_1} + 247.077\mathbf{z_2} + 238.354\mathbf{z_3} \\ & + 565.431\mathbf{z_4} + 497.236\mathbf{z_5} + 667.214\mathbf{z_6} \\ & + 722.519\mathbf{z_7} + 1188.667\mathbf{z_8} + 1423.284\mathbf{z_9}\end{aligned}$$

PLS Regression Algorithm

PLS-Regression Algorithm

Set $\mathbf{X_0} = \mathbf{X}$ and $\mathbf{y_0} = \mathbf{y}$
for $h = 1, 2, \ldots, a$ **do**
$\mathbf{w_h} = \mathbf{X_{h-1}^\top y_{h-1}} / \mathbf{y_{h-1}^\top y_{h-1}}$
$\|\mathbf{w_h}\| = 1$
$\mathbf{z_h} = \mathbf{X_{h-1} w_h} / \mathbf{w_h^\top w_h}$
$\mathbf{p_h} = \mathbf{X_{h-1}^\top z_h} / \mathbf{z_h^\top z_h}$
$\mathbf{X_h} = \mathbf{X_{h-1}} - \mathbf{z_h p_h^\top}$
$d_h = \mathbf{y_h^\top z_h} / \mathbf{z_h^\top z_h}$
$\mathbf{y_h} = \mathbf{y_{h-1}} - d_h \mathbf{z_h}$
end for

where a is the rank of $\mathbf{X}$

PLS-Regression Algorithm

Set $\mathbf{X_0} = \mathbf{X}$ and $\mathbf{y_0} = \mathbf{y}$

for $h = 1, 2, \dots, a$ **do**

$\mathbf{w_h} = \mathbf{X}_{\mathbf{h-1}}^{\mathsf{T}} \mathbf{y_{h-1}} / \mathbf{y}_{\mathbf{h-1}}^{\mathsf{T}} \mathbf{y_{h-1}}$ (regress each $\mathrm{x_j}$ onto y)

$\|\mathbf{w_h}\| = 1$ (normalize $\mathrm{w_h}$)

$\mathbf{z_h} = \mathbf{X_{h-1}} \mathbf{w_h} / \mathbf{w}_{\mathbf{h}}^{\mathsf{T}} \mathbf{w_h}$ (regress each obs $\mathrm{x_i}$ onto $\mathrm{w_h}$)

$\mathbf{p_h} = \mathbf{X}_{\mathbf{h-1}}^{\mathsf{T}} \mathbf{z_h} / \mathbf{z}_{\mathbf{h}}^{\mathsf{T}} \mathbf{z_h}$ (regress each $\mathrm{x_j}$ onto $\mathrm{z_h}$)

$\mathbf{X_h} = \mathbf{X_{h-1}} - \mathbf{z_h} \mathbf{p}_{\mathbf{h}}^{\mathsf{T}}$ (deflation of $\mathrm{X_{h-1}}$)

$d_h = \mathbf{y}_{\mathbf{h}}^{\mathsf{T}} \mathbf{z_h} / \mathbf{z}_{\mathbf{h}}^{\mathsf{T}} \mathbf{z_h}$ (regress $\mathrm{y_h}$ onto $\mathrm{z_h}$)

$\mathbf{y_h} = \mathbf{y_{h-1}} - d_h \mathbf{z_h}$ (deflation of $\mathrm{y_{h-1}}$)

end for

where a is the rank of $\mathbf{X}$

PLS-Regression Algorithm

Helland (2006) proposes a simplified algorithm. $\mathbf{X}$ and $\mathbf{y}$ are assumed to be mean-centered. Start with $\mathbf{X}_0 = \mathbf{X}$ and $\mathbf{y}_0 = \mathbf{y}$. Then set up for $h = 1, \ldots, H$:

1. start with weights $\mathbf{w_h} = \mathbf{X}_{\mathbf{h-1}}^\top \mathbf{y_{h-1}}$
2. $\mathbf{z_h} = \mathbf{X_{h-1}} \mathbf{w_h}$
3. $\mathbf{p_h} = \mathbf{X_{h-1}}^\top \mathbf{z_h} / (\mathbf{z}_\mathbf{h}^\top \mathbf{z_h}) = \mathbf{X}^\top \mathbf{z_h} / (\mathbf{z}_\mathbf{h}^\top \mathbf{z_h})$
4. $d_h = \mathbf{y}_\mathbf{h-1}^\top \mathbf{z_h} / \mathbf{z}_\mathbf{h}^\top \mathbf{z_h} = \mathbf{y}^\top \mathbf{z_h} / \mathbf{z}_\mathbf{h}^\top \mathbf{z_h}$
5. $\mathbf{X_h} = \mathbf{X} - \mathbf{z_1} \mathbf{p}_\mathbf{1}^\top - \cdots - \mathbf{z_h} \mathbf{p}_\mathbf{h}^\top$
6. $\mathbf{y_h} = \mathbf{y} - d_1 \mathbf{z_1} - \cdots - d_h \mathbf{z_h}$

Modeling $\mathbf{y}$ as function of $\mathbf{X}$

$$
\begin{aligned}
\mathbf{z_1} &= \mathbf{X}\mathbf{w_1} = \mathbf{X}\mathbf{w}_\mathbf{1}^* \\
\mathbf{z_2} &= \mathbf{X_1}\mathbf{w_2} = \mathbf{X}(\mathbf{I} - \mathbf{w_1}\mathbf{p}_\mathbf{1}^\mathsf{T})\mathbf{w_2} = \mathbf{X}\mathbf{w}_\mathbf{2}^* \\
\mathbf{z_3} &= \mathbf{X_2}\mathbf{w_3} = \mathbf{X}(\mathbf{I} - \mathbf{w_2}\mathbf{p}_\mathbf{2}^\mathsf{T})\mathbf{w_3} = \mathbf{X}\mathbf{w}_\mathbf{3}^* \\
&\vdots \\
\mathbf{z_h} &= \mathbf{X_{h-1}}\mathbf{w_h} = \mathbf{X}(\mathbf{I} - \mathbf{w_{h-1}}\mathbf{p}_\mathbf{h-1}^\mathsf{T})\mathbf{w_h} = \mathbf{X}\mathbf{w}_\mathbf{h}^*
\end{aligned}
$$

Consequently, $\mathbf{Z_h} = [\mathbf{z_1}, \mathbf{z_2}, \ldots, \mathbf{z_h}] = \mathbf{X}\mathbf{W}_\mathbf{h}^*$

Modeling $\mathbf{y}$ as function of $\mathbf{X}$

We know that $\hat{\mathbf{y}} = d_1 \mathbf{z_1} + \cdots + d_h \mathbf{z_h} = \mathbf{Zd}$

The PLS components can be expressed in terms of the original variables:

$$\mathbf{Z} = [\mathbf{z_1}, \mathbf{z_2}, \ldots, \mathbf{z_h}] = \mathbf{XW}^*$$

Therefore:

$$\hat{\mathbf{y}} = \mathbf{Zd} = \mathbf{XW}^* \mathbf{d} = \mathbf{Xb}$$

Re-expressing PLS Coefficients

- The PLS regression equation is typically expressed in terms of the original variables $\mathbf{X}$ instead of using the deflated matrices $\mathbf{X}_{h-1}$.
- This re-arrangement is more convenient for prediction purposes.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_{\mathsf{PLS}} + \mathbf{e}$$

- Note that the PLS $\boldsymbol{\beta}$-coefficients of the regression equation are not parameters of the PLS regression model. Instead, these are post-hoc calculations for making things more maneagable.

So far

In the next slides we'll talk about the properties of the PLS regression, and how the algorithm can be simplified.

References

- **The multivariate calibration problem in chemistry solved by the PLS Method** by Wold et al (1982). In *Proc. Conf. Matrix Pencils*
- **The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses** Wold et al (1984). *Siam J. Sci. Stat. Comput. Vol. 5, No. 3. p. 735-743*.

References (French Literature)

- **La Regression PLS: Theorie et Pratique** by Michel Tenenhaus (1998). Editions, Technip.
- **Probabilites, analyse des donnees et statistique** by Gilbert Saporta (2011). *Chapter 17: La regression multiple et le modele lineaire general*. Editions Technip, Paris.