Fachhochschule Südwestfalen
University of Applied Sciences

Statistical Modeling

CH.9 - Working with Collinear Data

SS 2022 || Prof. Dr. Buchwitz

Wir geben Impulse

Outline

Course Contents

Session	Topic
1	Simple Linear Regression
2	Multiple Linear Regression
3	Regression Diagnostics
4	Qualitative Variables as Predictors
5	Transformation of Variables
6	Weighted Least Squares
7	Correlated Errors
8	Analysis of Collinear Data
9	Working with Collinear Data
10	Variable Selection Procedures
11	Logistic Regression
12	Further Topics

Outline

Introduction

- When multicollinearity is present, the least squares estimates of the individual regression coefficients ten to be **unstable** and can lead to erroneous inferences.
- In the last session we discussed the problem of multicollinearity and ways to diagnose this problem. We found that eliminating predictors from the analysis does not always work and in most analytical settings is not a feasible option.
- We consider two alternative approaches for dealing with multicollinearity:
  - Imposing or searching for constraints on the regression parameters.
  - Using alternative estimation techniques (e.g. principal components regression and ridge regression).

Outline

Principal Components

- The principal components method is based on the fact that any set of p predictors $X_1, X_2, \ldots, X_p$ can be **transformed** to a set of p **orthogonal** variables.
- The new orthogonal variables are known as the **principal components** and are denoted by $C_1, C_2, \ldots, C_p$.
- Each variable C_j is a linear function of the standardized variables $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_p$.

$$C_j = v_{1j}\tilde{X}_1 + v_{2j}\tilde{X}_2 + \ldots + v_{pj}\tilde{X}_p \quad \text{for} \quad j = 1, 2, \ldots, p$$

Principal Components

- The coefficients of the linear functions are chosen so that the variables $C_1, \ldots, C_p$ are orthogonal.
- The coefficients for the j-th principal components C_j are the elements of the j-th eigenvector that corresponds to the eigenvalue λ_j, the j-th largest eigenvalue of the correlation matrix of the p variables.

$$V = \begin{pmatrix} V_1 & V_2 & \cdots & V_p \end{pmatrix} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p1} & v_{p2} & \cdots & v_{pp} \end{pmatrix}$$

Example: French Econmony Data

```
P241
```

```
##    YEAR IMPORT DOPROD STOCK CONSUM
## 1    49   15.9  149.3   4.2  108.1
## 2    50   16.4  161.2   4.1  114.8
## 3    51   19.0  171.5   3.1  123.2
## 4    52   19.1  175.5   3.1  126.9
## 5    53   18.8  180.8   1.1  132.1
## 6    54   20.4  190.7   2.2  137.7
## 7    55   22.7  202.1   2.1  146.0
## 8    56   26.5  212.4   5.6  154.1
## 9    57   28.1  226.1   5.0  162.3
## 10   58   27.6  231.9   5.1  164.3
## 11   59   26.3  239.0   0.7  167.6
## 12   60   31.1  258.0   5.6  176.8
## 13   61   33.3  269.8   3.9  186.6
## 14   62   37.0  288.4   3.1  199.7
## 15   63   43.3  304.5   4.6  213.9
## 16   64   49.0  323.4   7.0  223.8
## 17   65   50.3  336.8   1.2  232.0
## 18   66   56.6  353.9   4.5  242.9
```

Data Description

`YEAR` Year of Observation.

`IMPORT` Import Volume.

`DOPROD` Domestic Production.

`STOCK` Stock Formation.

`CONSUM` Domestic Consumption.

Variables are measured in billion French francs.

Principal Components

- It can be shown that the variance of the j-th principal component is $Var(C_j) = \lambda_j$ for $j = 1, 2, \ldots, p$. Therefore the variance-covariance matrix of the principal components is

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}$$

- All the off-diagonal elements are zero because the principal components are orthogonal. The value of the j-th diagonal element λ_j is the variance of C_j, the j-th principlal component.
- The principal components are arranged so that $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_p$, which means that the first component has the largest variance.

Principal Components

```r
d <- head(P241[ ,c("DOPROD", "STOCK", "CONSUM")], 11)
d.pca <- prcomp(d, center=TRUE, scale=TRUE)
C <- d.pca$x
round(C, 4)
```

```
##        PC1     PC2     PC3
## 1  -2.1259  0.6387 -0.0207
## 2  -1.6189  0.5555 -0.0711
## 3  -1.1152 -0.0730 -0.0217
## 4  -0.8943 -0.0824  0.0108
## 5  -0.6442 -1.3067  0.0726
## 6  -0.1904 -0.6591  0.0266
## 7   0.3596 -0.7437  0.0428
## 8   0.9718  1.3541  0.0629
## 9   1.5593  0.9640  0.0236
## 10  1.7670  1.0152 -0.0450
## 11  1.9311 -1.6627 -0.0806
```

Principal Components

```r
cormat <- cor(d)
eigen(cormat)    # Eigen Decomposition of Correlation Matrix
```

```
## eigen() decomposition
## $values
## [1] 1.999155 0.998154 0.002691
##
## $vectors
##         [,1]     [,2]      [,3]
## [1,] -0.7063  0.03569  0.706982
## [2,] -0.0435 -0.99903  0.006971
## [3,] -0.7065  0.02583 -0.707197
```

```r
round(var(C),4) # Variance-Covariance Matrix of PCs
```

```
##       PC1    PC2    PC3
## PC1 1.999 0.0000 0.0000
## PC2 0.000 0.9982 0.0000
## PC3 0.000 0.0000 0.0027
```

Remember

Multicollinearity leads to heterogeneous sizes of eigenvalues so that one eigenvalue is much smaller than the others. When one eigenvalue is exactly zero a perfect linear relationship (special case of extreme multicollinearity) among the original variables exists.

The variance-covariance matrix of the new variables only has entries on the main diagonal (which correspond to the eigenvalues) and zeros in all other places (as the variables are orthogonal).

Principal Components

- The principal components lack simple interpretation as they are *a mixture* of the (standardized) original variables.
- Since λ_j is the variance of the *j*-th principal component, a value of $\lambda_j \approx 0$ shows that the respective principal componennt C_j is equal to a constant. That constant is the mean value of C_j (which is zero as the variables have been standardized).
- Inspecting the eigenvectors of the previous example shows that *only* the variables `CONSUM` and `DOPROD` play a relevant role when determining C_3.

```
##  X1_tilde  X2_tilde  X3_tilde
##  0.706982  0.006971 -0.707197
```

$$\tilde{X}_1 \approx \tilde{X}_3 \quad \text{as} \quad v_{23} \approx 0.007 \approx 0$$

Outline

Principal Component Regression

- We consider the model for the *French Economomy Dataset*

$$\text{IMPORT} = \beta_0 + \beta_1\text{DOPROD} + \beta_2\text{STOCK} + \beta_3\text{CONSUM} + \epsilon$$

- This model expressed using the standardized variables $\tilde{Y} = (y_i - \bar{y})/s_y$ and $\tilde{X}_j = (x_{ij} - \bar{x}_j)/s_{x_j}$ yields

$$\tilde{Y} = \theta_1\tilde{X}_1 + \theta_2\tilde{X}_2 + \theta_3\tilde{X}_3 + \epsilon'$$

- Utilizing the principal components of the standardized predictors the model can be written as

$$\tilde{Y} = \alpha_1 C_1 + \alpha C_2 + \alpha_3 C_3 + \epsilon'$$

Principal Component Regression

```r
# Data Preparation
d <- head(P241,11)
d_scaled <- as.data.frame(scale(d))
d_prcomp <- as.data.frame(cbind(IMPORT=d_scaled$IMPORT,
                                prcomp(d[,c("DOPROD","STOCK","CONSUM")],
                                       center=TRUE, scale=T)$x))
# Motel Estimation
mod1 <- lm(IMPORT ~  1 + DOPROD + STOCK + CONSUM, data=d)
(mod2 <- lm(IMPORT ~ -1 + DOPROD + STOCK + CONSUM, data=d_scaled))
```

```
##
## Call:
## lm(formula = IMPORT ~ -1 + DOPROD + STOCK + CONSUM, data = d_scaled)
##
## Coefficients:
## DOPROD    STOCK   CONSUM
## -0.339    0.213    1.303
```

```r
(mod3 <- lm(IMPORT ~ -1 + PC1 + PC2 + PC3, data=d_prcomp))
```

```
##
## Call:
## lm(formula = IMPORT ~ -1 + PC1 + PC2 + PC3, data = d_prcomp)
##
## Coefficients:
##   PC1    PC2    PC3
## 0.690  0.191  1.160
```

Principal Component Regression

```r
ev <- eigen(cor(d[,c("DOPROD","STOCK","CONSUM")]))$vectors

# Multiply eigenvectors with constant to match output in book
# Note: Eigenvectoren not consistent
if( all(ev[,1] < c(0,0,0))) ev[,1] <- ev[,1] * -1
if(!all(ev[,2] < c(0,1,0))) ev[,2] <- ev[,2] * -1
if(!all(ev[,3] < c(0,0,1))) ev[,3] <- ev[,3] * -1

# Eigenvectors
ev
```

```
##        [,1]     [,2]      [,3]
## [1,] 0.7063 -0.03569 -0.706982
## [2,] 0.0435  0.99903 -0.006971
## [3,] 0.7065 -0.02583  0.707197
```

Principal Component Regression

- The coefficients of the principal component regression can be calculated based on the regression coefficients from the model using the standardized values.

$$
\begin{aligned}
\alpha_1 &= 0.706\ \theta_1 + 0.044\ \theta_2 + 0.707\ \theta_3 \\
\alpha_2 &= -0.036\ \theta_1 + 0.999\ \theta_2 + -0.026\ \theta_3 \\
\alpha_3 &= -0.707\ \theta_1 + -0.007\ \theta_2 + 0.707\ \theta_3
\end{aligned}
$$

- Conversely this relationship can be turned around to obtain the coefficients from the regression with standardized variables from the principal component regression.

$$
\begin{aligned}
\theta_1 &= 0.706\ \alpha_1 + -0.036\ \alpha_2 + -0.707\ \alpha_3 \\
\theta_2 &= 0.044\ \alpha_1 + 0.999\ \alpha_2 + -0.007\ \alpha_3 \\
\theta_3 &= 0.707\ \alpha_1 + -0.026\ \alpha_2 + 0.707\ \alpha_3
\end{aligned}
$$

Principal Component Regression

```
# Calculate alpha (principal components) from theta (standardized variables)
as.vector(coef(mod2) %*% ev)
```

```
## [1] 0.6900 0.1913 1.1597
```

```
coef(mod3)
```

```
##    PC1    PC2    PC3
## 0.6900 0.1913 1.1597
```

```
# Calculate theta (standardized variables) from alpha (principal components)
as.vector(ev %*% coef(mod3))
```

```
## [1] -0.3393  0.2130  1.3027
```

```
coef(mod2)
```

```
##  DOPROD   STOCK  CONSUM
## -0.3393  0.2130  1.3027
```

Principal Component Regression

$$
\begin{aligned}
\tilde{Y} &= \theta_1 \tilde{X}_1 + \theta_2 \tilde{X}_2 + \theta_3 \tilde{X}_3 + \epsilon' \\
&= \alpha_1 C_1 + \alpha C_2 + \alpha_3 C_3 + \epsilon'
\end{aligned}
$$

- Although the above equations both hold, the C's are **orthogonal**.
- The orthogonality bypasses (but not eliminates) the multicollinearity problem, however, the resulting relationship and therefore the coefficients are **not easily interpreted**.
- The α's unlike the θ's do not have simple interpretations as marginal effects of the original (standardized) predictor variables.

The final estimation results are always restated in terms of the θ's or origninal β's for interpretation!

Principal Component Regression

Based on the coefficients obtained from regressing the standardized variables the relationship can be expressed in terms of the original β_j's using the following relationship:

$$\hat{\beta}_j = \frac{s_y}{s_j}\hat{\theta}_j \quad \text{for} \quad j = 1, 2, \ldots, p$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}_1 - \hat{\beta}_2\bar{x}_2 - \ldots - \hat{\beta}_p\bar{x}_p$$

This back-transform of the variables to the original scale is crucial for interpretation of the final results!

Reduction of Multicollinearity in the Data

- Principal component regression can be used to **reduce collinearity in the estimation data**.
- this can be achieved by using **less than the full set of principal components** to explain the variation in the response.
- When all principal components are used the OLS solution can be exactly reproduced (as seen before).

Reduction of Multicollinearity in the Data

- The C_j's have sample variances $\lambda_1, \lambda_2, \ldots, \lambda_p$ equal to their eigenvalues.

```
eigen(cor(d[,c("DOPROD","STOCK","CONSUM")]))$values
```

```
## [1] 1.999155 0.998154 0.002691
```

- Since C_3 has very small variance, the linear function defining C_3 is **approximately equal to zero** and is the source of collinearity in the data.

Reduction of Multicollinearity in the Data

- We exclude C_3 from the analysis and consider the two possible remaining regression models

$$\tilde{Y} = \alpha_1 C_1 + \epsilon$$
$$\tilde{Y} = \alpha_1 C_1 + \alpha_2 C_2 + \epsilon$$

There are two important things to note here:

1. In an regression equation where the full set of potential predictor variables under consideration are orthogonal, the estimated values of the regression **coefficients are not altered** when subsets of these variables are either introduced or deleted.
2. Both models lead to estimates for **all** three of the original standardized coefficients θ_1, θ_2 and θ_3.

Reduction of Multicollinearity in the Data

```
mod_prcomp1 <- lm(IMPORT ~ -1 + PC1            , data=d_prcomp)
mod_prcomp2 <- lm(IMPORT ~ -1 + PC1 + PC2      , data=d_prcomp)
mod_prcomp3 <- lm(IMPORT ~ -1 + PC1 + PC2 + PC3, data=d_prcomp)
```

	Model 1	Model 2	Model 3
PC1	0.69^{***}	0.69^{***}	0.69^{***}
	(0.05)	(0.03)	(0.02)
PC2		0.19^{***}	0.19^{***}
		(0.04)	(0.03)
PC3			1.16
			(0.61)
R^2	0.95	0.99	0.99
Adj. R^2	0.95	0.99	0.99
Num. obs.	11	11	11

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2: Statistical models

Reduction of Multicollinearity in the Data

```r
# Coefficients for standardized predictors when using one principal component
coefs1 <- coef(mod3)[1] * ev[,1]
names(coefs1) <- c("DOPROD","STOCK","CONSUM")
coefs1
```

```
##  DOPROD   STOCK  CONSUM
## 0.48736 0.03001 0.48750
```

```r
# Coefficients for standardized predictors when using two principal components
coefs2 <- coef(mod3)[1] * ev[,1] + coef(mod3)[2] * ev[,2]
names(coefs2) <- c("DOPROD","STOCK","CONSUM")
coefs2
```

```
## DOPROD  STOCK CONSUM
## 0.4805 0.2211 0.4826
```

Reduction of Multicollinearity in the Data

```
s <- apply(d[ ,c("IMPORT","DOPROD","STOCK","CONSUM")],2,sd)
m <- apply(d[ ,c("IMPORT","DOPROD","STOCK","CONSUM")],2,mean)
```

```
# Model with one PC for non-standardized data
coefs_org1 <- s[1]/s[2:4] * coefs1
intercept_org1 <- unname(m[1] - sum(m[2:4]*coefs_org1))
(beta_org1 <- c(Intercept=intercept_org1, coefs_org1))
```

```
## Intercept     DOPROD      STOCK     CONSUM
##  -7.74583    0.07381    0.08269    0.10735
```

```
# Model with two PCs for non-standardized data
coefs_org2 <- s[1]/s[2:4] * coefs2
intercept_org2 <- unname(m[1] - sum(m[2:4]*coefs_org2))
(beta_org2 <- c(Intercept=intercept_org2, coefs_org2))
```

```
## Intercept     DOPROD      STOCK     CONSUM
##  -9.13011    0.07278    0.60922    0.10626
```

Reduction of Multicollinearity in the Data

- It is evident that using different numbers of principal components gives substantially different results.
- I has already been argued that the **OLS estimates are unsatisfactory** (the negative coefficient of $\tilde{X}_1$ is unexpected and cannot be sensibly interpreted.
- The **third principal component** is the cause of multicollinearity as it is *almost constant*.
- Of the remaining two components the first one is associated with the effect of `DOPROD` and `CONSUM`. The second is uniquely associated with `STOCK` (as only the coefficient for `STOCK` changes when C_2 is added to the regression of `IMPORT` on C_1).

Principal component regression can be influenced by the presence of high-leverage points and outliers, which should be removed beforehand.

Reduction of Multicollinearity in the Data

- The following table shows that the coefficients are dependent on the number of incorporated principal components.
- As each component explains additional variance the R^2 increases with the number of considered principal components.

	std_PC1	org_PC1	std_PC2	org_PC2	std_PC3	org_PC3
Intercept	NA	-7.746	NA	-9.130	NA	-10.128
DOPROD	0.487	0.074	0.481	0.073	-0.339	-0.051
STOCK	0.030	0.083	0.221	0.609	0.213	0.587
CONSUM	0.488	0.107	0.483	0.106	1.303	0.287

Caution when using Principal Component Regression

- Principal component regression is not guaranteed to work with all datasets.
- The following dataset (Hald's dataset) suffers from multicollinearity issues, when calculating the principal components the following model can be estimated, where U is the standardized response and the C_j's are the principal components.

$$U = \alpha_1 C_1 + \alpha_2 C_2 + \alpha_3 C_3 + \alpha_4 C_4 + \epsilon$$

- It can be seen that in the full model only α_4 is significant, and almost the complete variability of the response ($R^2 \approx 1$) is captured. When C_4 is dropped the remaining three components account for none of the variability of ($R^2 \approx 0$).

Caution when using Principal Component Regression

```r
#TODO: Fix Dataset and show regression models with impact on R2
P278 # Dataset defect
```

```
##       Y.X1.X2.X3.X4
## 1    78.5 7 26 6 60
## 2   74.3 1 29 15 52
## 3  104.3 11 56 8 20
## 4   87.6 11 31 8 47
## 5    95.9 7 52 6 33
## 6  109.2 11 55 9 22
## 7   102.7 3 71 17 6
## 8   72.5 1 31 22 44
## 9   93.1 2 54 18 22
## 10 115.9 21 47 4 26
## 11  83.8 1 40 23 34
## 12 113.3 11 66 9 12
## 13 109.4 10 68 8 12
```

Outline

Ridge Regression

- Ridge regression provides an alternative estimation method that can be employed when the predictor variables are highly collinear.
- There are multiple computational variations of the ridge regression, the presented one is associated with the *ridge trace*.
- Ridge analysis using the ridge trace represents a unified approach to problems of detection and estimation when multicollinearity is suspected.
- The **estimators produced are biased** but tend to have a smaller mean squared error when compared to OLS estimators.

The ridge method shrinks the estimated coefficients toward zero. This class of estimators is sometimes called **Shrinkage Estimators**.

Ridge Regression

$$\tilde{Y} = \theta_1 \tilde{X}_1 + \theta_2 \tilde{X}_2 + \theta_3 \tilde{X}_3 + \epsilon'$$

- Based on the standardized form of the regression model, the estimation equations for the ridge regression are given

$$\begin{array}{ccccccccccc}
(1+k) & \theta_1 & + & r_{12} & \theta_2 & + & \ldots & + & r_{1p} & \theta_p & = & r_{1y} \\
r_{21} & \theta_1 & + & (1+k) & \theta_2 & + & \ldots & + & r_{2p} & \theta_p & = & r_{2y} \\
 & & \vdots & & & \vdots & & \vdots & & & \vdots & \\
r_{p1} & \theta_1 & + & r_{p2} & \theta_2 & + & \ldots & + & (1+k) & \theta_p & = & r_{py}
\end{array}$$

- Here r_{iy} is the correlation coefficient between the i-th predictor and and the response variable $\tilde{Y}$.

Ridge Regression

- The ridge estimates may be viewed as resulting from a set fo data that ha sbeen *slightly altered*.
- The essential parameter that distinguishes ridge regression from OLS is k. When $k = 0$ the $\hat{\theta}$'s are the OLS estimates.
- The paramterter k may be referred to as the **bias parameter**. As k increases from zero, the bias of the estimates increases.

Ridge Regression

$$\text{Total Variance}(k) = \sum_{j=1}^{p} Var(\hat{\theta}_j(k)) = \sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + k)^2}$$

$$\text{Total Variance}(0) = \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j}$$

- As k continues to increase, the regression estimates all tend toward zero.
- The idea of ridge regression os to pick a value of k for which the reduction in total variance is not exceeded by the increase in bias.
- In practice a value of k is shown by computing $\hat{\theta}_1, \ldots, \hat{\theta}_p$ for a range of k values between 0 and 1 and plotting the results against k. The resulting graph is know as the **ridge trace** and can be used to select an appropriate value for k.

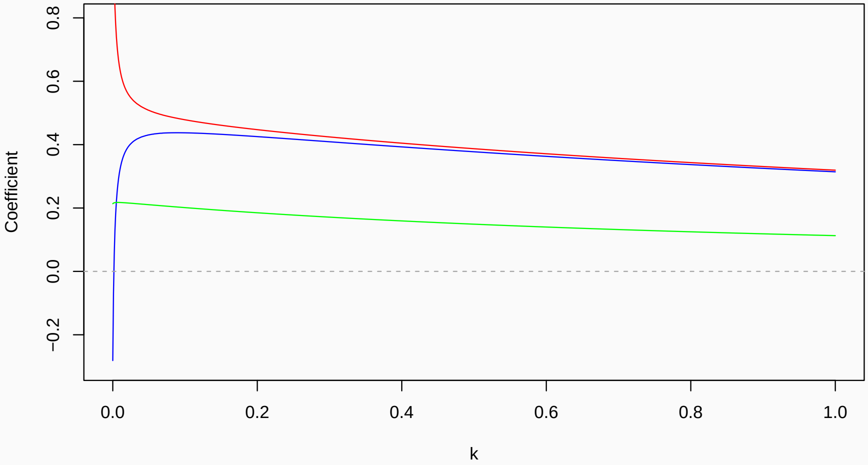
Ridge Regression Estimation

```r
y <- as.numeric(scale(head(P241,11)$IMPORT))
x <- as.matrix(scale(head(P241,11)[,c("DOPROD","STOCK","CONSUM")]))
k <- rev(seq(0,1,0.001))
mod <- glmnet::glmnet(x,y, family="gaussian", lambda = k, alpha = 0)
mod
```

```
##
## Call:  glmnet::glmnet(x = x, y = y, family = "gaussian", alpha = 0,      lambda = k)
##
##       Df %Dev Lambda
## 1      3 86.6  1.000
## 2      3 86.6  0.999
## 3      3 86.6  0.998
## 4      3 86.7  0.997
## 5      3 86.7  0.996
## 6      3 86.7  0.995
## 7      3 86.7  0.994
## 8      3 86.7  0.993
## 9      3 86.7  0.992
## 10     3 86.7  0.991
## 11     3 86.8  0.990
## 12     3 86.8  0.989
## 13     3 86.8  0.988
## 14     3 86.8  0.987
## 15     3 86.8  0.986
## 16     3 86.8  0.985
```

Ridge Trace

```
plot(k,rep(NA,length(k)),type="l",ylim=c(-0.3, 0.8), ylab="Coefficient")
lines(k, mod$beta[1,], col = "blue" ) # Coefficient for DOPROD
lines(k, mod$beta[3,], col = "red"  ) # Coefficient for CONSUM
lines(k, mod$beta[2,], col = "green") # Coefficient for STOCK
abline(h=0, lty="dashed", col="darkgrey")
```



Ridge Regression

- When working with collinear data the values for k are typically chosen at the low end of the range.
- If the estimated coefficients show *large fluctuations* for small values of k, instability has been demonstrated and collinearity is probably at work.
- The previously shown **ridge trace plot** shows that the coefficients $\hat{\theta}_{\text{CONSUM}}$ and $\hat{\theta}_{\text{DOPROD}}$ are quite unstable for small values of k. The unplausible negative coefficient disappears quickly and stabilizes around 0.4.
- The coefficient $\hat{\theta}_{\text{STOCK}}$ is unaffected by the collinearity and remains almost stable throughout the range of k.

Ridge Regression

The next step in ridge analysis is to select a value of k and to obtain the corresponding estimates of the regression coefficients. As k is a **bias parameter** it is dersirable ro select the smallest value of k for which stability occurs. Several methods have been suggested:

1. Fixed Point Method 2. Iterative Method 3. Ridge Trace

Ridge Regression: Fixed Point Method

$$k = \frac{p\hat{\sigma}^2(0)}{\sum_{j=1}^{p}[\hat{\theta}_j(0)]^2}$$

- Here $\hat{\theta}_1(0), \ldots, \hat{\theta}_p(0)$ are the parameters of the OLS estimate, when $k = 0$, and $\hat{\sigma}^2(0)$ is the corresponding mean square.

Ridge Regression: Interative Method

$$k_1 = \frac{p\hat{\sigma}^2(0)}{\sum_{j=1}^{p}[\hat{\theta}_j(k_0)]^2} \qquad k_2 = \frac{p\hat{\sigma}^2(0)}{\sum_{j=1}^{p}[\hat{\theta}_j(k_1)]^2} \qquad \dots$$

- Start with the initial estimate of k_0 which is the resulting estimate from the fixed point estimation procedure.
- Then use the previous value of k to determine the next one and repeat this process until the difference between two successive estimates of k is neglegible (until the algorithm converges).

Ridge Regression: Ridge Trace

- The behavior of $\hat{\theta}_j(k)$ as a function of k is easily observed from the ridge trace.
- The value of k selected is the smalles value for which all the coefficients $\hat{\theta}_j(k)$ are stable.
- In addition, at the selected value of k the residual sum of squares should be close to is minimum value.
- The variance inflation factors $\text{VIF}_j(k)$ should also get down to less than 10. Recall that a value of $\text{VIF}_j = 1$ is a characteristic of an orthogonal system and a value of less than 10 indicates an noncollinear or stable system.