On comparisons of simulation experiments

Silvelyn Zwanzig

Abstract

Resampling and simulation methods like bootstrap or Simex are based on new generated observations. The name simulation experiment is introduced for the family of probability measures of these simulated samples. The family of measures of the original sample is called original experiment. Both types of experiments are defined on different probability spaces - nevertheless the wanted parameter is the same. In that situation the concept of Le Cam is applied for comparing experiments by their deficiency and distance.

Contents

1 Introduction 2
2 Nonparametric bootstrap 4
3 Parametric bootstrap in the linear model 9
4 Simex in a simple errors-in-variables model 14
5 Simex versus Bootstrap 19
  5.1 The original experiment .......................... 19
  5.2 The parametric bootstrap experiment .......... 20
  5.3 The Simex experiment .......................... 23
  5.4 Hybrid method ................................ 27
  5.5 The comparisons .............................. 29
6 Appendix 32
1 Introduction

The main intention of this paper is to show a way how to bring the Le Cam theory and simulation methods together. Le Cam stands for an abstract high theoretical statistic on the one side and simulation methods for a more engineer approach on the other side. Simulation methods today are based on good intuition and a good heuristic. Often simulations deliver the only way to handle a statistical estimation problem. Here we want to offer a chance for a more theoretical approach. In this paper different types of simulation experiments are considered: nonparametric bootstrap, parametric bootstrap in a linear model with normal distributed errors, Simex method in a simple errors in variables model. Simex and parametric bootstrap are compared in a general normal location model.

Two questions are of main interest.

• What is the effect of the additional randomization by using simulation methods?

• How we can simulate in a best way?

The main message is that additional randomization implies a loss of information. Otherwise we have upper bounds for the loss. Asymptotically, for increasing simulation size the experiments are equivalent.

The second question is only considered for parametric bootstrap experiments in the linear model. It turns out, the best way is to base the simulation on the optimal estimator otherwise the loss of information will be increase.

Simulation methods make sense because the deliver estimation methods. Have in mind, by the deficiency concept of Le Cam we compare the chances for good estimation in the related experiments - the optimal estimators are not given.

The original experiment is given by

\[ \mathcal{E} = \{ P_{\theta} , \, \theta \in \Theta \} , \]

where \( P_{\theta} = P_{\theta}^{X} = (P_{\theta}^{X})^{\otimes n} \) is a probability measure on \((\mathcal{X}^{n}, \mathcal{A}^{(n)})\).

The simulation of ”new data”

\[ \mathbf{Y} = (Y_{1},...,Y_{N^{*}})^{T} \]
use the knowledge of former observations
\[ X = (X_1, ..., X_N)^T. \]

But the generating procedure \( P^{Y/X=x} \) is independent of the parameter \( \theta \). Let be the joint probability distribution of the simulated observations \( Q_\theta \), where \( Q_\theta \) is a probability measure on \( (Y^{N^*}, B^{N^*}) \). Then the simulation experiment is
\[ \mathcal{F} = \{ Q_\theta \mid \theta \in \Theta \}. \]

The deficiency \( \delta (\mathcal{E}, \mathcal{F}) \) describes the loss of information. We say \( \delta (\mathcal{E}, \mathcal{F}) = 0 \) iff \( \mathcal{F} \) is less informative than \( \mathcal{E} \). \( \mathcal{F} \) is equivalent to \( \mathcal{E} \) iff \( \delta (\mathcal{E}, \mathcal{F}) = 0 \) and \( \delta (\mathcal{F}, \mathcal{E}) = 0 \).

**Theorem 1.** Le Cam (1986)

The deficiency \( \delta (\mathcal{E}, \mathcal{F}) \) is exactly
\[
\delta (\mathcal{E}, \mathcal{F}) = \inf_K \sup_{\theta} \frac{1}{2} \| Q_\theta - KP_\theta \|, \tag{1.0.1}
\]
where the infimum is taken over the sets of all transitions \( K \).

\( \| . \| \) denotes the \( L_1 \) norm,
\[
\| \mu \| = \sup_{f} \left\{ \int f d\mu : f \text{ measurable and } |f| \leq 1 \right\}. \tag{1.0.2}
\]

It can be easily seen that simulation experiments are less informative.

**Theorem 2.** For all experiments \( \mathcal{E}, \mathcal{F} \) it holds
\[
\delta (\mathcal{E}, \mathcal{F}) = 0.
\]

**Proof.** There exists a transition \( K_{sim} \) independent of \( \theta \)
\[
K_{sim} (B, X) = P^{Y/X=x} (B)
\]
such that for all \( B \in B^{(K)} \)
\[
Q_\theta (B) = K_{sim} P_\theta (B) = \int K_{sim} (B, x) P_\theta (dx). \tag{1.0.3}
\]

From (1.0.1) follows, that \( \delta (\mathcal{E}, \mathcal{F}) = 0. \]

Thus the more interesting question is the other direction. Can we recover the original experiment from the simulations?
2 Nonparametric bootstrap

Suppose we generate the new data independently by $P_{Y_j/X=x}$, $j = 1, ..., N^*$ which is a discrete distribution on $X = (x_1, ..., x_n)$ with $P(Y_j = x_i/X = x) = a_i$, $\sum a_i = 1$,

$$P_{Y/X=x} = \prod_{j=1}^{N} \prod_{i=1}^{n} a_i^{I_{\{x_i\}}(y_j)}, \quad I_{\{x_i\}}(y_j) = \begin{cases} 1 & \text{for } y_j = x_i \\ 0 & \text{for } y_j \neq x_i \end{cases}. \quad (2.0.4)$$

Denote by

$$n_i = \sum_{j=1}^{N^*} I_{\{x_i\}}(y_j)$$

the number of times where the value $x_i$ was resampled. We have for the random variable

$$(n_1, ..., n_n) / X = x \sim \text{Multinomial}(K, a_1...a_n), \quad \text{with } \sum_{i=1}^{n} n_i = N^* \quad (2.0.5)$$

and

$$P(n_1, ..., n_n/X = x) = \frac{N^*!}{n_1!...n_n!} a_1^{n_1} ... a_n^{n_n}. \quad (2.0.6)$$

For $a_i = \frac{1}{n}$, nonparametric bootstrap is included. Note, permutation procedures are not included.

We can show that the experiments are asymptotically equivalent. Moreover we derive an exponential bound for the deficiency.

**Theorem 3.** Under (2.0.6)

$$\delta (\mathcal{F}, \mathcal{E}) \leq \sum_{i=1}^{n} (1 - a_i)^{N^*}. \quad \text{Proof.} \quad \text{We have}$$

$$\delta (\mathcal{F}, \mathcal{E}) = \inf_{K} \sup_{\theta} \frac{1}{2} \|P_\theta - KQ_\theta\|$$

and define

$$\mathcal{X}_\geq = \left\{ x : \frac{d}{d\mu} P_\theta > \frac{d}{d\mu} KQ_\theta \right\}.$$
Then
\[
\frac{1}{2} \| P_\theta - KQ_\theta \| = P_\theta(\mathcal{X}_>) - KQ_\theta(\mathcal{X}_>) \leq \sup_{A \in A(n)} | P_\theta(A) - KQ_\theta(A) |.
\]

Define the intersections

\[ A_i(x) = \{ x_i : x \in A \} \]

such that

\[ I_A(x) = \prod_{i=1}^n I_{A_i(x)}(x_i). \]

Take the kernel

\[ K(A, y) = \prod_{j=1}^{N^*} K_A(y_j) : y_j \in \{ x_1, \ldots, x_n \} \]

with

\[ K_A(y) = I_{A_i(x)}(x_i) \text{ for } y = x_i. \]

Calculate now \( KQ_\theta(A) \) by using (2.0.4) and

\[ KQ_\theta(A) = \int \int K(A, y) K_{sim}(dy, x) P_\theta(dx) \]

with

\[
\int K(A, y) K_{sim}(dy, x) = \sum_{i_1=1}^{n} \ldots \sum_{i_K=1}^{n} a_{i_1} \ldots a_{i_K} K(A, x_{i_1}, \ldots, x_{i_K}) \tag{2.0.7}
\]

\[
= \sum_{i_1=1}^{n} \ldots \sum_{i_K=1}^{n} a_{i_1} \ldots a_{i_K} I_{A_i(x)}(x_{i_1}) \ldots I_{A_i(x)}(x_{i_K}).
\]

Applying the relation

\[
\left( \sum_{i=1}^{n} b_i \right)^{N^*} = \sum_{i_1=1}^{n} \ldots \sum_{i_K=1}^{n} b_{i_1} \ldots b_{i_K}
\]

\[
= \sum_{(n_1, \ldots, n_n) \in I(N^*)} \frac{N^*!}{n_1! \ldots n_n!} b_{n_1}^{n_1} \ldots b_{n_n}^{n_n}
\]

\[ 5 \]
with
\[ I(N^*) = \left\{ (n_1, \ldots, n_n) : \sum_{i=1}^{n} n_i = N^* \right\} \subset \{0, 1, \ldots, N^*\}^n. \]

We get
\[ \int K(A, y) K_{\text{sim}}(dy, x) = \sum_{(n_1, \ldots, n_n) \in I(N^*)} \frac{N^*!}{n_1! \ldots n_n!} a_1^{n_1} \cdots a_n^{n_n} I_{A_1(x)}(x_1)^{n_1} \cdots I_{A_n(x)}(x_n)^{n_n}. \]

We obtain
\[ KQ_\theta(A) = \sum_{(n_1, \ldots, n_n) \in I(N^*)} \frac{N^*!}{n_1! \ldots n_n!} a_1^{n_1} \cdots a_n^{n_n} E_\theta \left( I_{A_1(x)}(x_1)^{n_1} \cdots I_{A_n(x)}(x_n)^{n_n} \right). \]

Now we consider two different cases. Split up the sum in (2.0.8) by
\[ I(N^*) = I_{>0}(N^*) \cup \overline{I_{>0}(N^*)} \]

with
\[ I_{>0}(N^*) = \left\{ (n_1, \ldots, n_n) : (n_1, \ldots, n_n) \in I(N^*) : \min_i n_i > 0 \right\}. \]

For \( n_i > 0 \) it holds
\[ I_{A_i(x)}(x_i)^{n_i} = I_{A_i(x)}(x_i), \]
thus for \( (n_1, \ldots, n_n) \in I_{>0}(N^*) \)
\[ E_\theta \left( I_{A_1(x)}(x_1)^{n_1} \cdots I_{A_n(x)}(x_n)^{n_n} \right) = E_\theta \left( I_{A_1(x)}(x_1) \cdots I_{A_n(x)}(x_n) \right) = E_\theta I_A(x) = P_\theta(A). \]

We get
\[ KQ_\theta(A) = p(N^*) P_\theta(A) + \text{rest} \]
with
\[ p(N^*) = \sum_{(n_1, \ldots, n_n) \in I_{>0}(N^*)} \frac{N^*!}{n_1! \ldots n_n!} a_1^{n_1} \cdots a_n^{n_n} \]

and
\[ \text{rest} = \sum_{(n_1, \ldots, n_n) \in \overline{I_{>0}(N^*)}} \frac{N^*!}{n_1! \ldots n_n!} a_1^{n_1} \cdots a_n^{n_n} E_\theta \left( I_{A_1(x)}(x_1)^{n_1} \cdots I_{A_n(x)}(x_n)^{n_n} \right). \]
We can estimate

\[ p(N^*) = P(I_{>0}(N^*)) = 1 - P(\exists i, n_i = 0) \geq 1 - \sum_{i=1}^{n} P(n_i = 0). \]

The marginal distribution of \( n_i \) in (2.0.6) is Bin\( (N^*, a_i) \), such that

\[ P(n_i = 0) = (1 - a_i)^{N^*} \]

and

\[ 0 \leq 1 - p(N^*) \leq \sum_{i=1}^{n} (1 - a_i)^{N^*}. \]

Because \( E_\theta (I_{A_i(x_i)}(x_1)^{n_1} \ldots I_{A_n(x_i)}(x_n)^{n_n}) \leq 1 \) we get

\[ 0 \leq \text{rest} \leq P(I_{>0}(N^*)) = 1 - p(N^*). \]

Summarizing we have

\[ KQ_\theta (A) - P_\theta (A) = p(N^*) P_\theta (A) - P_\theta (A) + \text{rest} = (p(N^*) - 1) P_\theta (A) + \text{rest} \leq 0 + \text{rest} \leq 1 - p(N^*) \]

and

\[ KQ_\theta (A) - P_\theta (A) = (p(N^*) - 1) P_\theta (A) + \text{rest} \geq (p(N^*) - 1) P_\theta (A) \geq p(N^*) - 1. \]

Hence for all \( A \in A^{(n)} \)

\[ |KQ_\theta (A) - P_\theta (A)| \leq 1 - p(N^*) \leq \sum_{i=1}^{n} (1 - a_i)^{N^*}. \]

The bound \( \sum_{i=1}^{n} (1 - a_i)^{N^*} \) is minimal for the nonparametric bootstrap experiment, where \( a_i = \frac{1}{n} \).

Corollary 4. Assume \( a_i = \frac{1}{n} \)

1. For fixed \( n \)

\[ \lim_{N^* \to \infty} \delta (\mathcal{F}, \mathcal{E}) = 0 \]
2. \( N^* = N^*(n) > 2n \ln(n) \) then

\[
\lim_{n \to \infty} \delta(\mathcal{F}, \mathcal{E}) = 0
\]

Proof. Under \( a_i = \frac{1}{n} \),

\[
\sum_{i=1}^{n} (1 - a_i)^{N^*} = n(1 - \frac{1}{n})^{N^*}.
\]

The first statement follows immediately. Consider \( N^* = N^*(n) > 2n \ln(n) \)

\[
n(1 - \frac{1}{n})^{N^*} \leq \exp(\ln n + 2n \ln n \ln(1 - \frac{1}{n})) = \exp(\ln n(1 + n \ln(1 - \frac{1}{n})) \exp(\ln n \ln(1 - \frac{1}{n})) \to 0
\]

Because:

\[
\lim_{n \to \infty} (\ln n(1 + n \ln(1 - \frac{1}{n})) = \ln(2)
\]

and

\[
\lim_{n \to \infty} (n \ln(1 - \frac{1}{n})) = -1
\]

But there does not exist a series of transitions \( \{K(m)\} \), such that \( P_\theta(A) = \lim_{m \to \infty} K(m)Q_\theta(A) \) for all \( A \in \mathcal{A}^{(n)} \).

Theorem 5. Assume (2.0.4) and that the experiment \( \mathcal{E} \) is complete. Then

\[
\delta(\mathcal{F}, \mathcal{E}) > 0.
\]

Proof. The proof is indirect and is not constructive. Assume there exists a series of transitions \( \{K(m)\} \), such that for all \( A \in \mathcal{A}^{(n)} \)

\[
P_\theta(A) = \lim_{m \to \infty} \int K(m)(A, y)dQ_\theta.
\] (2.0.9)

Using (1.0.3) we get

\[
P_\theta(A) = \lim_{m \to \infty} \int \int K(m)(A, y)K_{sim}(dy, x)dP_\theta.
\]
From (2.0.4) we obtain

\[ E_m(A, x) = \int K_m(A, y)K_{sim}(dy, x) = \sum_{i_1=1}^{n} \ldots \sum_{i_N=1}^{n} a_{i_1} \ldots a_{i_N} K_m(A, x_{i_1}, ..., x_{i_N}). \]  

(2.0.10)

The assumption (2.0.9) is

\[ \int (I_A(x) - \lim_{m \to \infty} E_m(A, x)) dP_\theta = 0 \text{ for all } \theta \in \Theta \text{ and all } A \in \mathcal{A}(n). \]

From the completeness of \( \mathcal{E} \) follows

\[ I_A(x) - \lim_{m \to \infty} E_m(A, x) = 0 \text{ for all a. s. } x \text{ and all } A \in \mathcal{A}(n). \]  

(2.0.11)

Consider now a set \( A_1 \subset \mathcal{X}, A_1 \in \mathcal{A} \) with \( 0 < P_\theta(A_1) < 1 \), and \( \overline{A_1} = \mathcal{X} \setminus A_1 \) take

\[ C_1 = A_1 \times \overline{A_1} \times \mathcal{X} \times \ldots \times \mathcal{X}, \quad C_2 = \overline{A_1} \times A_1 \times \mathcal{X} \times \ldots \times \mathcal{X} \]

then for all \( x \in C_1, I_{C_1}(x) = 1 \) and \( I_{C_2}(x) = 0 \). But because of (2.0.10) \( E_m(A, x) \) is symmetric in \( x \) for all \( m \), thus \( E_m(C_1, x) = E_m(C_2, x) \). That gives the contradiction to (2.0.11).

\section{3 Parametric bootstrap in the linear model}

Consider a linear regression model with a design matrix, which has full rank

\[ \mathcal{E} = \left\{ N_n \left( X\beta, \sigma^2 I_n \right), \beta \in \mathbb{R}^p \right\}. \]

Note, we use now the denotation of linear models, where \( Y \) is the vector of the observations and \( X \) is the design matrix. Thus the original experiment is the linear model is

\[ Y = X\beta + \varepsilon, \varepsilon \sim N_n(0, \sigma^2 I_n). \]

Consider a linear unbiased estimator for \( \beta \), thus \( Y \) is fitted by \( LY \), with \( ELY = X\beta \). The least squares estimator is included as the special case, where \( L \) is the projection matrix.

\[ \hat{\beta} = (X^TX)^{-1}X^TY, \hat{Y} = X\hat{\beta} = X(X^TX)^{-1}X^TY = PY, P = X(X^TX)^{-1}. \]
The bootstrap samples are
\[ Y_j^* = LY + \varepsilon_j^* \quad \text{where} \quad \varepsilon_j^* \sim N_n(0, \sigma^2 I_n), \quad j = 1, \ldots, B^*. \] (3.0.12)

The generated \( \varepsilon_j^* \) are independent of \( Y \) and independent of each other. We use the star for the denotation in the bootstrap "world".

Derive the bootstrap experiment. Because \( P_\beta \) and \( P_{Y^*_j/Y} \) are normal distributions and the marginal distribution of \( Y^*=(Y_1^*, \ldots, Y_{B^*}^*)^T \) is normal too. It is enough to calculate the expectation and the covariance matrix. We have
\[
EY_j^* = E(LY + \varepsilon_j^*) = ELY = X\beta \quad \text{for all} \quad j = 1, \ldots, B^*
\]
and
\[
\text{Cov}(Y_j^*) = \text{Cov}(LY + \varepsilon_j^*) = \text{Cov}(\varepsilon_j^*) + \text{Cov}(LY) = \sigma^2 I_n + \sigma^2 LL^T = \sigma^2 (I + LL^T).
\]

For \( k \neq j \) it holds
\[
\text{Cov}(Y_k^*, Y_j^*) = \text{Cov}((LY + \varepsilon_k^*), (LY + \varepsilon_j^*)) = \text{Cov}(LY) = \sigma^2 LL^T.
\]

We get the bootstrap experiment with
\[ Y^* \sim N_{nB^*} \left( (1_{B^*} \otimes X)\beta, \sigma^2 \left( I_{nB^*} + 1_{B^*}1_{B^*}^T \otimes LL^T \right) \right), \]
where \( 1_{B^*} = (1, \ldots, 1)^T \). Thus
\[
\mathcal{F}^* = \left\{ N_{nB^*} \left( X^*\beta, \sigma^2 \Sigma^* \right) \mid \beta \in \mathbb{R}^p \right\},
\]
with
\[ X^* = 1_{B^*} \otimes X, \quad \Sigma^* = I_{nB^*} + (1_{B^*} \otimes L)(1_{B^*} \otimes L)^T. \]

The special case, where the bootstrap simulations are based on the least squares estimator are denoted by
\[
\mathcal{F}^*_\text{lse} = \left\{ N_{nB^*} \left( X^*\beta, \sigma^2 \Sigma^*_p \right) \mid \beta \in \mathbb{R}^p \right\},
\]
where
\[ I_{nB^*} + 1_{B^*}1_{B^*}^T \otimes P \quad \text{with} \quad P = X(X^TX)^{-1}X^T, \]
by using \((A \otimes B)(C \otimes D) = AC \otimes BD\) and \((A \otimes B)^T = A^T \otimes B^T\).
\[1_{B^*}1_{B^*}^T \otimes X(X^TX)^{-1}X^T\]
\[= (1_{B^*} \otimes X)(1 \otimes (X^TX)^{-1})(1_{B^*} \otimes X)^T\]
\[= (1_{B^*} \otimes X)(X^TX)^{-1}(1_{B^*} \otimes X)^T\]

Thus
\[\Sigma^*_P = I_{nB^*} + X^*(X^TX)^{-1}(X^*)^T.\]

For comparing the bootstrap experiment with the original experiment we calculate the equivalent sufficient experiments.

**Theorem 6.** The sufficient experiment to \(F^*\) is
\[\mathcal{F}^*_{suff} = \{ N_p(\beta, \sigma^2\Sigma_L) \ , \ \beta \in \mathbb{R}^p \},\]
with
\[\Sigma_L = (X^T(I_n + B^*LL^T)^{-1}X)^{-1}.\] (3.0.13)

Especially for \(L = P\) holds
\[\Sigma_P = (1 + \frac{1}{B^*})(X^TX)^{-1}.\] (3.0.14)

**Proof.** We apply Theorem 23 and have to calculate
\[\Sigma_L = (X^*(\Sigma^*)^{-1}X^*)^{-1}\]

Using
\[(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}\] (3.0.15)
and
\[(1_{B^*} \otimes L)^T(1_{B^*} \otimes L) = 1_{B^*}^T1_{B^*} \otimes L^TL = B^*L^TL\]
we have
\[(\Sigma^*)^{-1} = \left(I_{nB^*} + (1_{B^*} \otimes L)(1_{B^*} \otimes L)^T\right)^{-1}\]
\[= I_{nB^*} - (1_{B^*} \otimes L)(I_n + (1_{B^*} \otimes L)^T(1_{B^*} \otimes L))^{-1}(1_{B^*} \otimes L)^T\]
\[= I_{nB^*} - (1_{B^*} \otimes L)(I_n + B^*L^TL)^{-1}(1_{B^*} \otimes L)^T.\]
Now
\[ X^* X^* = (1_{B^*} \otimes X)^T (1_{B^*} \otimes X) = 1_{B^*}^T 1_{B^*} \otimes X^T X \]
\[ = B^* \otimes X^T X = B^* X^T X \]
and
\[ X^* (1_{B^*} \otimes L) = (1_{B^*} \otimes X)^T (1_{B^*} \otimes L) = B^* X^T L \]
we get
\[ X^* (\Sigma^*)^{-1} X^* = B^* X^T X - B^* X^T L (I_n + B^* L^T L)^{-1} B^* L^T X \]
\[ = B^* X^T (I_n - L (I_n + B^* L^T L)^{-1} B^* L^T ) X. \]
Using again (3.0.15) we obtain
\[ X^* (\Sigma^*)^{-1} X^* = B^* X^T (I_n + B^* LL^T)^{-1} X. \]
Now we will show (3.0.14). Set \( L = P. \) The projection is idempotent \( LL^T = P. \) Consider
\[ (I_n + B^* LL^T)^{-1} = (I_n + BP)^{-1} = (I_n + B^* X (X^T X)^{-1} X^T)^{-1}. \]
Applying (3.0.15)
\[ = I_n - B^* X (X^T X + B^* X^T X)^{-1} X^T = I_n - \frac{B^*}{1 + B^*} P. \]
Remind \( PX = X \) and we get (3.0.14) by
\[ B^* X^T (I_n + B^* LL^T)^{-1} X = B^* X^T (I_n - \frac{B^*}{1 + B^*} P) X \]
\[ = B^* X^T X (1 - \frac{B^*}{1 + B^*}) = \frac{B^*}{1 + B^*} X^T X. \]
Applying the result of Torgersen and Hansen, [3], we get that the most informative bootstrap experiment is based on the best estimator. Note the least squares estimator is the best unbiased estimator. We have that
\[ Cov\hat{\beta} \preceq Cov\bar{\beta} \] for all unbiased estimators of \( \bar{\beta}. \)
Theorem 7. It holds for all \( L \) with \( LX = X \)

\[ \delta (\mathcal{F}^*, \mathcal{F}_{lse}^*) > 0. \]

**Proof.** Using the result of Torgersen Hansen it remains to shown that \( \Sigma_L \succeq \Sigma_P \). For \( L \) with \( LX = X \) the fitted values \( LY \) are unbiased estimators of \( X\beta \). Thus \( \text{Cov}(LY) = \sigma^2 LL^T \succeq \text{Cov}(PY) = \sigma^2 P \). This implies

\[ (I_n + B^*LL^T) \succeq (I_n + B^*P) \]

and so

\[ (I_n + B^*P)^{-1} \succeq (I_n + B^*LL^T)^{-1}. \]

This gives

\[ X^T (I_n + B^*P)^{-1} X \succeq X^T (I_n + B^*LL^T)^{-1} X \]

and

\[ \Sigma_L = (X^T (I_n + B^*LL^T)^{-1} X)^{-1} \succeq (X^T (I_n + B^*P)^{-1} X)^{-1} = \Sigma_P. \]

For increasing \( B^* \) the loss of information can be neglected. Especially we have the result.

Theorem 8. It holds

\[ \delta (\mathcal{F}_{lse}^*, \mathcal{E}) = P \left( \frac{\ln \left( 1 + \frac{1}{B^*} \right)}{B^*} \right) \leq \frac{\chi^2}{p} \leq (1 + \frac{1}{B^*}) \frac{\ln \left( 1 + \frac{1}{B^*} \right)}{B^*}, \]

where \( \chi^2 \) is chi squared distributed with \( p \) degrees of freedom.

**Proof.** The result is a consequence of (3.0.14), Theorem (20) and (29).

Using the Corollary 30 we get the following result.

Corollary 9. It holds

\[ 0 < \delta (\mathcal{F}_{lse}^*, \mathcal{E}) \leq p \frac{1}{B^*}. \]
4 Simex in a simple errors-in-variables model

In [1] Cook and Stefanski introduce a new type of simulation type estimator Simex, which becomes more and more popular. The simple underlying truth of the Simex procedure is, that it is impossible to simulate new samples with less measurement errors than in the original sample - but we can generate new samples with arbitrary larger variances by adding simulated errors to the original observed regressor variables. Thus we get the chance to study the influence of their variance to the estimators. The Simex estimator is defined by fitting a parametric model to this relationship between variances and the estimators calculated from the simulated samples and by extrapolating backwards to the case of zero variance.

Suppose the simple error in variables model (EIV), which is defined as follows.

\[ y_i = \beta \xi_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \]
\[ x_i = \xi_i + \delta_i, \delta_i \sim N(0, \sigma^2), \]

where \( \varepsilon_i \) and \( \delta_i \) independently, \( i = 1, \ldots, n \) distributed. Wanted \( \beta \in \mathbb{R} \). The points \((\xi_1, \ldots, \xi_n)\) are the variables, which are observed with error only. \((\xi_1, \ldots, \xi_n)\) are nuisance parameters. Define

\[
Z = \begin{pmatrix} y_1 \\ \vdots \\ y_n \\ x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \text{and } \mu(\theta) = \begin{pmatrix} \xi_1 \beta \\ \vdots \\ \xi_n \beta \\ \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} = \left( \begin{pmatrix} \beta \\ 1 \end{pmatrix} \otimes I_n \right) \xi(n) \tag{4.0.16}
\]

with \( \theta = (\beta, \xi_1, \ldots, \xi_n) \in \Theta, \xi(n) = (\xi_1, \ldots, \xi_n)^T \) and

\[ Z \sim N_{2n}(\mu(\theta), \sigma^2 I_{2n}). \]

Summarizing, the original experiment of a simple errors in variables model is

\[
\mathcal{E}_{EIV} = \left\{ N_{2n}(\mu(\theta), \sigma^2 I_{2n}) : \theta = (\beta, \xi_1, \ldots, \xi_n) \in \Theta \right\}
\]

\[
= \left\{ N_n(\beta \xi(n), \sigma^2 I_n) \otimes N_n(\xi(n), \sigma^2 I_n) : \theta = (\beta, \xi_1, \ldots, \xi_n) \in \Theta \right\}
\]

A Simex procedure consists of the following main steps:
1. **SIMulation of new observations:**

\[ x_{ik}^* = x_i + \sqrt{\lambda_k} \delta_{ik}^*, \quad \delta_{ik}^* \sim N(0, \sigma^2) \text{ i.i.d. } i = 1, \ldots, n, \quad k = 1, \ldots, B^* \]

(4.0.17)

2. Calculate \( \hat{\beta}_k^* \), \( k = 1, \ldots, B^* \) from

\[ (y_1, \ldots, y_n, x_{1k}^*, x_{2k}^*, \ldots, x_{nk}^*) \]

3. Fit a parametric curve \( b(\lambda) \) to \( (\hat{\beta}_1^*, \ldots, \hat{\beta}_{B^*}^*) \).

4. **Extrapolation:** Define the new estimator\[ \hat{\beta}_{SIMEX} := b(-\sigma^2). \]

Here we are not interested in the special Simex estimation procedure, we just want to compare the original experiment with the simulation experiment. First we have to find out the statistical model concerning to the simulated observations in (4.0.17).

**Theorem 10.** The Simex experiment related to (4.0.17) is

\[ \mathcal{F}_{SIMEX} = \{ N_{n+nB^*} \left( X_{sim} \xi_{(n)}, \Sigma_{sim} \right), \theta = (\beta, \xi_{(n)}) \in \mathbb{R} \times \mathbb{R}^n \}, \]

where

\[ X_{sim} = I_n \otimes \begin{pmatrix} \beta \\ 1_{B^*} \end{pmatrix} \]

and

\[ \Sigma_{sim} = \begin{pmatrix} \sigma^2 I_n & 0 \\ 0 & \sigma^2 \left( (1_{B^*}1_{B^*}^T + D) \otimes I_n \right) \end{pmatrix} \]

with \( 1_{B^*} = (1, \ldots, 1)^T \) and \( D = diag(\lambda_1, \ldots, \lambda_{B^*}). \)

**Proof.** Calculate the distribution of \( \{(y_1, \ldots, y_n, x_{1k}^*, x_{2k}^*, \ldots, x_{nk}^*) : k = 1, \ldots, B^* \} \).

Because the original data and the pseudo errors are normal distributed, the simulated observations are normal too. Thus it is enough to calculate the expectation and the covariance matrix.

We have

\[ Ey_i = \beta \xi_i, \quad Ex_{ik}^* = Ex_i + \sqrt{\lambda_k} E\delta_{ik}^* = Ex_i = \xi_i. \]
and
\[ \operatorname{Var}(y_i) = \sigma^2; \quad \operatorname{Var}(x_{ik}^*) = \operatorname{Var}(x_i) + \lambda_k \operatorname{Var}(\delta_{ik}^*) = (1 + \lambda_k) \sigma^2 \]

Further
\[ \operatorname{Cov}(y_i, x_{jk}^*) = 0 \text{ for all } i = 1, \ldots, n; \quad j = 1, \ldots, n; \quad k = 1, \ldots, B^* \]

and
\[ \operatorname{Cov}(y_i, y_j) = 0 \text{ for } i \neq j, \]

and
\[ \operatorname{Cov}(x_{jk}^*, x_{il}^*) = 0 \text{ for all } i \neq j, k = 1, \ldots, B^*, l = 1, \ldots, B^* \]

but for all \( k = 1, \ldots, B^*, \quad l = 1, \ldots, B^* \)
\[ \operatorname{Cov}(x_{ik}^*, x_{il}^*) = \operatorname{Cov}(x_i + \sqrt{\lambda_k \delta_{ik}^*}, x_i + \sqrt{\lambda_l \delta_{il}^*}) = \sigma^2. \]

Define \( x_{(k)}^* = (x_{1k}^*, \ldots, x_{nk}^*)^T \) then
\[ E x_{(k)}^* = \xi(n), \quad \operatorname{Cov}(x_{(k)}^*) = (1 + \lambda_k) \sigma^2 I_n \]

and
\[ \operatorname{Cov}(x_{(k)}^*, x_{(l)}^*) = \sigma^2 I_n. \]

Rewrite it in matrix form:
\[
Z^* = \begin{pmatrix} y_1 \\ \vdots \\ y_n \\ x_{(1)}^* \\ \vdots \\ x_{(B^*)}^* \end{pmatrix}, \quad EZ^* = \begin{pmatrix} \beta \xi(n) \\ \xi(n) \\ \vdots \\ \xi(n) \end{pmatrix} = \left( \begin{pmatrix} \beta \\ 1_{B^*} \end{pmatrix} \otimes I_n \right) \xi(n)
\]

and
\[
\operatorname{Cov}(Z^*) = \begin{pmatrix} \sigma^2 I_n & 0 & \cdots & \cdots & 0 \\ 0 & (1 + \lambda_1) \sigma^2 I_n & \sigma^2 I_n & \cdots & \sigma^2 I_n \\ \vdots & \sigma^2 I_n & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & (1 + \lambda_{B^* - 1}) \sigma^2 I_n & \sigma^2 I_n \\ 0 & \sigma^2 I_n & \cdots & \sigma^2 I_n & (1 + \lambda_{B^*}) \sigma^2 I_n \end{pmatrix}
\]
Note
\[
\begin{pmatrix}
(1 + \lambda_1) \sigma_2^2 I_n & \sigma_2^2 I_n & \cdots & \sigma_2^2 I_n \\
\sigma_2^2 I_n & \ddots & \ddots & \vdots \\
\vdots & \ddots & (1 + \lambda_{B^*} - 1) \sigma_2^2 I_n & \sigma_2^2 I_n \\
\sigma_2^2 I_n & \cdots & \sigma_2^2 I_n & (1 + \lambda_{B^*}) \sigma_2^2 I_n \\
\end{pmatrix}
= \sigma_2^2 ((1_{B^*}^T 1_{B^*} + D) \otimes I_n)
\]

with \( D = \text{diag}(\lambda_1, \ldots, \lambda_{B^*}) \). Thus
\[
\text{Cov}(Z^*) = \begin{pmatrix}
\sigma_2^2 I_n & 0 \\
0 & \sigma_2^2 ((1_{B^*}^T 1_{B^*} + D) \otimes I_n)
\end{pmatrix}
\]

\[\text{Theorem 11.}\] An equivalent experiment to \( F_{SIMEX} \) is
\[
F_{\text{suff}} = \{ N_n(\beta \xi(n), \sigma^2 I_n) \otimes N_n(\xi(n), \sigma^2 (1 + \Delta) I_n), \theta = (\beta, \xi_1, \ldots, \xi_n) \in \Theta \}.
\]

with \( \Delta = \frac{1}{\sum_{k=1}^{B^*} \lambda_k^{-1}}. \) \( (4.0.18) \)

\[\text{Proof.}\] We apply Theorem 23 to \( \{ N_{nB^*}((1_{B^*} \otimes I_n) \xi(n), \sigma^2 (\Sigma \otimes I_n)) : \xi(n) \in \mathbb{R}^n \} \). We have to calculate
\[
(1_{B^*} \otimes I_n)^T (\Sigma \otimes I_n)^{-1} (1_{B^*} \otimes I_n)^{-1}.
\]

It holds \((1_{B^*} \otimes I_n)^T = 1_{B^*}^T \otimes I_n \) and \((\Sigma \otimes I_n)^{-1} = \Sigma^{-1} \otimes I_n \), thus
\[
(1_{B^*} \otimes I_n)^T (\Sigma \otimes I_n)^{-1} (1_{B^*} \otimes I_n) = (1_{B^*}^T \otimes I_n) (\Sigma^{-1} \otimes I_n) (1_{B^*} \otimes I_n)
= 1_{B^*}^T \Sigma^{-1} 1_{B^*} \otimes I_n
= 1_{B^*}^T \Sigma^{-1} 1_{B^*} \otimes I_n
\]

Using \((3.0.15)\)
\[
(1_{B^*}^T (1_{B^*} 1_{B^*}^T + D)^{-1} 1_{B^*})^{-1} = 1 + \frac{1}{1_{B^*}^T D^{-1} 1_{B^*}} = 1 + \Delta
\]

17
because $1_{B^*}^T D^{-1} 1_{B^*}$ is a scalar. Especially

$$1_{B^*}^T D^{-1} 1_{B^*} = tr(D^{-1}) = \sum_{k=1}^{B^*} \frac{1}{\lambda_k} = \Delta^{-1}.$$  

\[\square\]

**Theorem 12.** It holds with

$$\Delta = \frac{1}{\sum_{k=1}^{B^*} \lambda_k^{-1}}$$

$$\delta(\mathcal{F}_{SIMEX}, \mathcal{E}) \leq P\left( \frac{\ln(1 + \Delta)}{\Delta} \leq \frac{\chi^2}{n} \leq (1 + \Delta) \frac{\ln(1 + \Delta)}{\Delta} \right) \leq \frac{n}{\sum_{k=1}^{B^*} \lambda_k^{-1}}$$

where $\chi^2$ is chi squared distributed with $n$ degrees of freedom.

**Proof.** We compare

$$\delta(\mathcal{F}_{SIMEX}, \mathcal{E}) = \delta(\mathcal{F}_{suff}, \mathcal{E})$$

We embed $\mathcal{F}_{suff} \subset \mathcal{F}_{big}$ $\mathcal{E} \subset \mathcal{E}_{big}$ with

$$\mathcal{F}_{big} = \left\{ N_{2n}(\mu, \Sigma_{simex}) , \mu \in \mathbb{R}^{2n} \right\},$$

$$\mathcal{E}_{big} = \left\{ N_{2n}(\mu, \Sigma_{org}) , \mu \in \mathbb{R}^{2n} \right\},$$

with

$$\Sigma_{simex} = \begin{pmatrix} \sigma^2 I_n & 0 \\ 0 & \sigma^2 (1 + \Delta) I_n \end{pmatrix}, \Sigma_{org} = \sigma^2 I_{2n}.$$ 

and apply Lemma 22, that

$$\delta(\mathcal{F}_{SIMEX}, \mathcal{E}) = \delta(\mathcal{F}_{suff}, \mathcal{E}) \leq \delta(\mathcal{F}_{big}, \mathcal{E}_{big}).$$

For the upper bound we apply Theorem 28 with

$$\Sigma_A = \begin{pmatrix} \sigma^2 I_n & 0 \\ 0 & \sigma^2 (1 + \Delta) I_n \end{pmatrix}, \Sigma_B = \begin{pmatrix} \sigma^2 I_n & 0 \\ 0 & \sigma^2 I_n \end{pmatrix}$$

It holds

$$\Sigma_B^{-1} - \Sigma_A^{-1} = \frac{\Delta}{1 + \Delta} \begin{pmatrix} 0 & 0 \\ 0 & \sigma^{-2} I_n \end{pmatrix}$$
and $Z = \begin{pmatrix} Z_{(1)} \\ Z_{(2)} \end{pmatrix} \sim N_{2n}(0, \begin{pmatrix} \sigma^2 I_n & 0 \\ 0 & \sigma^2(1 + \Delta)I_n \end{pmatrix})$

$$Z^T (\Sigma_B^{-1} - \Sigma_A^{-1}) Z = \sigma^2 \frac{\Delta}{1 + \Delta} \frac{Z_{(2)}^T Z_{(2)}}{\sigma^2} \chi^2_n$$

$$\delta(\mathcal{F}_{big}, \mathcal{E}_{big}) = \frac{1}{2} E \left| \frac{1}{1 + \Delta} \exp\left(\frac{\Delta}{\sigma^2 \chi} - 1\right) \right|.$$ 

Lemma 27 delivers the explicit formulary. 

## 5 Simex versus Bootstrap

In this section we consider a more general regression model with normal errors. The model assumptions include nonlinear models, where the application of bootstrap and Simex makes sense.

### 5.1 The original experiment

Consider the following normal location model

$$Y = \xi + \varepsilon, \varepsilon \sim N_n(0, \Sigma_\varepsilon), \xi \in \Theta \subset \mathbb{R}^n, \Sigma_\varepsilon \text{ known.} \quad (5.1.1)$$

These model includes for instance nonlinear time series, with

$$y_t = f(x_t, t, \theta) + \varepsilon_t, \varepsilon_t = \rho \varepsilon_{t-1} + u_t, \ u_t \sim N(0, \sigma^2) \text{ i.i.d.}.$$ 

In this case

$$\xi = \begin{pmatrix} f(x_1, 1, \theta) \\ \vdots \\ f(x_n, n, \theta) \end{pmatrix}$$

and

$$\Sigma_\varepsilon = \sigma^2 \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \cdots & \rho^2 & \rho & 1 \end{pmatrix}.$$
Another example for (5.1.1) is the following time series

\[ y_t = f(x_t, t, \theta) + z_t + \beta z_{t-1}, \ z_t \sim N(0, \sigma^2) \ i.i.d. \]

There we have

\[ \xi = \begin{pmatrix} f(x_1, 1, \theta) \\ \vdots \\ f(x_n, n, \theta) \end{pmatrix} \]

and

\[
\sigma^2 \Sigma_{\epsilon} = \sigma^2 \begin{pmatrix}
1 & \frac{\beta}{1+\beta^2} & \cdots & 0 \\
\frac{\beta}{1+\beta^2} & 1 & \frac{\beta}{1+\beta^2} & \vdots \\
\vdots & \ddots & \ddots & \frac{\beta}{1+\beta^2} \\
0 & \cdots & \frac{\beta}{1+\beta^2} & 1
\end{pmatrix}.
\]

As third example we will mention a nonparametric regression model:

\[ y_t = f(x_t) + z_t, \ z_t \sim N(0, \sigma^2) \ i.i.d. \]

Here

\[ \xi = \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix}, \ f \in \mathcal{F} \]

and \( \Sigma_{\epsilon} = \sigma^2 I \).

Let us formulate the model (5.1.1) as experiment:

\[ \mathcal{E} = \{ N_n(\xi, \Sigma_{\epsilon}) , \ \xi \in \Theta \subset \mathbb{R}^n \} \]

5.2 The parametric bootstrap experiment

Suppose that there is reasonable linear estimator \( LY \) for \( \xi \).

This can be a moving average estimator in the time series examples

\[
\hat{y}_i = \sum_{j=i-k}^{i+k} w_j y_j, \ \sum w_i = 1.
\]
Especially for $\hat{y}_i = \frac{1}{2} w_2 y_{i-1} + w_1 y_i + \frac{1}{2} w_2 y_{i+1}$ we get

$$L = \begin{pmatrix}
    w_1 & w_2 & 0 & 0 & 0 \\
    \frac{1}{2} w_2 & w_1 & \frac{1}{2} w_2 & 0 & \ldots & 0 \\
    0 & \frac{1}{2} w_2 & w_1 & \frac{1}{2} w_2 & 0 \\
    0 & 0 & \frac{1}{2} w_2 & w_1 & \frac{1}{2} w_2 \\
    0 & \ldots & 0 & \frac{1}{2} w_2 & w_1 \\
    0 & 0 & 0 & 0 & w_2 & w_1
\end{pmatrix}.$$ 

Another examples for $LY$ are kernel estimators

$$\hat{y}_i = \frac{1}{nh} \sum_{j=1}^{n} K \left( \frac{x_i - x_j}{h} \right) y_j.$$ 

The matrix $L$ has for normal kernels $K(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2)$ the form

$$L = \frac{1}{nh\sqrt{2\pi}} \left( \exp\left( -\frac{(x_i - x_j)^2}{2h^2} \right) \right)_{i,j}.$$ 

In difference to Section 3 we don’t require, that $LY$ is an unbiased estimator. But the bootstrap samples are generated analogously to (3.0.12) by

$$Y^*_j = LY^* + \Sigma_e^{\frac{1}{2}} \epsilon^*_j, \quad \epsilon^*_j \sim N_n(0, I_n), \quad j = 1, \ldots, B^*.$$ 

The generated $\epsilon^*_j$ are independent of $Y$ and independent of each other. We use the star for the denotation in the bootstrap ”world”.

**Theorem 13.** The Bootstrap experiment related to (5.2.1) is

$$\mathcal{F}_{\text{boot}} = \{N_n B^* \cap ((1_B \otimes L)\xi, \Sigma_e^*), \xi \in \Theta \subset \mathbb{R}^n \},$$

where

$$\Sigma_e^* = I_{B^*} \otimes \Sigma_e + (1_B \otimes L)\Sigma_e(1_B \otimes L)^T.$$ 

**Proof.** Because $P^Y$ and $P^{Y_j^*|Y}$ are normal distributions and the marginal distribution of $Y^*=(Y^*_1, \ldots, Y^*_B)^T$ is normal too. It is enough to calculate the expectation and the covariance matrix. It holds $EY^*_j = LYE = L\xi$ and

$$Cov(Y^*_j) = Cov(LY) + Cov(\Sigma_e^{\frac{1}{2}} \epsilon^*_j)$$

$$= L\Sigma_e L^T + \Sigma_e^{\frac{1}{2}} I_n \Sigma_e^{\frac{1}{2}}$$

$$= L\Sigma_e L^T + \Sigma_e.$$ 

21
Further the covariance between two different bootstrap samples are calculated for \( j \neq k \):

\[
\text{Cov}(Y^*_j, Y^*_k) = \text{Cov}(LY, LY) + \text{Cov}(\Sigma^1_\varepsilon \varepsilon^*_j, \Sigma^1_\varepsilon \varepsilon^*_k)
\]

\[
= L\Sigma_\varepsilon L^T + \Sigma^1_\varepsilon \text{Cov}(\varepsilon^*_j, \varepsilon^*_k)\Sigma^1_\varepsilon
\]

Then the whole bootstrap data are

\[
Y^* \sim \mathcal{N}_{nB^*}((1_{B^*} \otimes L)\xi, \Sigma^*)
\]

where \( 1_{B^*} = (1, ..., 1)^T \) and

\[
\Sigma^* = \begin{pmatrix}
L\Sigma_\varepsilon L^T + \Sigma_\varepsilon & L\Sigma_\varepsilon L^T & \cdots & L\Sigma_\varepsilon L^T \\
L\Sigma_\varepsilon L^T & L\Sigma_\varepsilon L^T + \Sigma_\varepsilon & \cdots & L\Sigma_\varepsilon L^T \\
\vdots & \vdots & \ddots & \vdots \\
L\Sigma_\varepsilon L^T & \cdots & L\Sigma_\varepsilon L^T + \Sigma_\varepsilon & L\Sigma_\varepsilon L^T + \Sigma_\varepsilon
\end{pmatrix}
\]

That can be rewritten as

\[
\Sigma^* = I_{B^*} \otimes \Sigma_\varepsilon + 1_{B^*}1_{B^*}^T \otimes L\Sigma_\varepsilon L^T,
\]

where \( 1_{B^*}1_{B^*}^T \) is the \( B^* \times B^* \) matrix, where all elements are 1. Using twice the rule for Kronecker symbol multiplication \((A \otimes B)(F \otimes G) = (AF \otimes BG)\) we get

\[
1_{B^*}1_{B^*}^T \otimes L\Sigma_\varepsilon L^T = (1_{B^*} \otimes L)\Sigma_\varepsilon(1_{B^*} \otimes L)^T
\]

Thus

\[
\Sigma^* = I_{B^*} \otimes \Sigma_\varepsilon + (1_{B^*} \otimes L)\Sigma_\varepsilon(1_{B^*} \otimes L)^T.
\]

For comparing the bootstrap experiment with the other experiments we calculate an equivalent sufficient experiments. But in difference to the chapter above the sufficient experiment is not the minimal sufficient experiment! We reduce only the bootstrap replications, there is no reduction inside the underlying experiment. This trick make it possible to handle nonlinear relationships in the location parameter.
Theorem 14. An equivalent experiment to $F_L^*$ is

$$F_{suff}^* = \{N_n(\xi, \Sigma_{suff}) , \xi \in \Theta \subset \mathbb{R}^n\},$$

where

$$\Sigma_{suff} = \Sigma_e + \frac{1}{B^*} \left( L^T \Sigma_e^{-1} L \right)^{-1}.$$

Proof. From Theorem 23 it follows

$$F_{suff}^* = \{N_n(\xi, \Sigma_{suff}) , \xi \in \Theta \subset \mathbb{R}^n\},$$

where

$$\Sigma_{suff} = \left( (1_{B^*} \otimes L)^T \Sigma_e^{-1} (1_{B^*} \otimes L) \right)^{-1}.$$

Rewrite the covariance matrix by using Lemma 24

$$\Sigma_{suff}^{-1} = (1_{B^*} \otimes L)^T \left( I_{B^*} \otimes \Sigma_e + (1_{B^*} \otimes L) \Sigma_e (1_{B^*} \otimes L)^T \right)^{-1} (1_{B^*} \otimes L)$$

$$= (\Sigma_e + \left( (1_{B^*} \otimes L)^T (1_{B^*} \otimes \Sigma_e)^{-1} (1_{B^*} \otimes L) \right)^{-1})^{-1}$$

Using $(A \otimes B)(F \otimes G) = (AF \otimes BG)$ we get

$$(1_{B^*} \otimes L)^T (I_{B^*} \otimes \Sigma_e)^{-1} (1_{B^*} \otimes L) = 1_{B^*}^T 1_{B^*} \otimes L^T \Sigma_e^{-1} L = B^* L^T \Sigma_e^{-1} L$$

Summarizing

$$\Sigma_{suff} = \Sigma_e + \frac{1}{B^*} \left( L^T \Sigma_e^{-1} L \right)^{-1}.$$  

5.3 The Simex experiment

Let us now consider the Simex analogon to the bootstrap samples in (5.2.1). We add pseudo errors to the original observations and variate the variance of the pseudo errors by the additional parameter $\lambda_k > 0$:

$$Y^*_{\#} = Y + \sqrt{\lambda_k \Sigma^2} \varepsilon^*_{\#}, \varepsilon^*_{\#} \sim N_n(0, I_n), \ k = 1, ..., B^*,$$  \hspace{1cm} (5.3.1)

These experiment belongs to a Simex procedure for location models with the following set up

$$Y = \xi(\theta) + \varepsilon, \varepsilon \sim N_n(0, \sigma^2 \Sigma), \theta \in \Theta \subset \mathbb{R}^n, \sigma^2 \Sigma \text{ known}.$$
Wanted to improve
\[ \hat{\theta} = \arg\min_{\theta \in \Theta} Q(Y, \theta), \]
with
\[ E\hat{\theta} = \theta + \sigma^2 \text{BIAS} + O(\sigma^3). \]

Let us give the main steps for a Simex method with known \( \sigma^2 \).

1. **Simulate** \( k = 1, ..., B^* \)
   \[
   Y_k^\# = Y + \sqrt{\lambda_k \Sigma_k^2} \varepsilon_k^*, \quad \varepsilon_k^* \sim N_n(0, I_n), \quad k = 1, ..., B^*. \]

2. **Calculate** the estimators \( \hat{\theta}_k \) from \( Y_k^\# \)
   \[ \hat{\theta}_k = \arg\min_{\theta \in \Theta} Q\left(Y_k^\#, \theta\right). \]

3. **Fit** a model to \( \hat{\theta}_{(1)}, ..., \hat{\theta}_{(B)} \)
   \[ \hat{h} = \arg\min_{h \in F} \sum_{k=1}^{B^*} \left\| \hat{\theta}_k - h(\lambda_k) \right\|^2. \]

4. **Extrapolate** backwards
   \[ \hat{\theta}_{SIMEX} = \hat{h}(\sigma^2). \]

For unknown \( \sigma^2 \) the following adaptive Simex is useful.

1. **Simulate** \( k = 1, ..., B^* \)
   \[
   Y_k^\# = Y + \sqrt{\lambda_k \Sigma_k^2} \varepsilon_k^*, \quad \varepsilon_j^* \sim N_n(0, I_n), \quad k = 1, ..., B^*. \]

2. **Calculate** the estimators \( \hat{\theta}_k \) from \( Y_k^\# \)
   \[ \hat{\theta}_k = \arg\min_{\theta \in \Theta} Q\left(Y_k^\#, \theta\right) \text{ and } Q_{(k)} = \min_{\theta \in \Theta} Q\left(Y_k^\#, \theta\right). \]
3. Fit a model to $\hat{\theta}_1, \ldots, \hat{\theta}_B$

$$\hat{h} = \arg\min_{h \in \mathcal{F}} \sum_{k=1}^{B^*} \left\| \hat{\theta}(k) - h(\lambda_k) \right\|^2.$$

4. Fit a model to $Q_1, \ldots, Q_B$

$$\hat{H} = \arg\min_{H \in \mathcal{F}} \sum_{k=1}^{B^*} \left\| Q(k) - H(\lambda_k) \right\|^2.$$

5. Calculate $\hat{\lambda} = \min_\lambda \hat{H}(\lambda)$.

6. Extrapolate backwards

$$\hat{\theta}_{\text{ASIMEX}} = \hat{h}(-\hat{\lambda}).$$

Determine now the experiment of the Simex samples.

**Theorem 15.** The Simex experiment related to (5.3.1) is

$$\mathcal{F}_{\text{simex}} = \left\{ N_{nB^*} \left( (1_{B^*} \otimes I_n) \xi, \Sigma^# \right), \xi \in \Theta \subset \mathbb{R}^n \right\},$$

where

$$\Sigma^# = D \otimes \Sigma_e + (1_{B^*} \otimes I_n) \Sigma_e (1_{B^*} \otimes I_n)^T$$

and $D = \text{diag}(\lambda_1, \ldots, \lambda_{B^*})$.

**Proof.** Because all underlying distributions are normal, the marginal distribution of $Y^# = (Y^#_1, \ldots, Y^#_{B^*})^T$ is normal too. It is enough to calculate the expectation and the covariance matrix. It holds $EY^# = EY = \xi$ and

$$\text{Cov}(Y^#_k) = \text{Cov}(Y) + \text{Cov}(\sqrt{\lambda_k} \Sigma^\frac{1}{2} \varepsilon^*_j)$$

$$= \Sigma_e + \lambda_k \Sigma^\frac{1}{2} I_n \Sigma^\frac{1}{2} = (1 + \lambda_k) \Sigma_e.$$

Further the covariance between two different Simex samples are calculated for $j \neq k$, remind $\text{Cov}(\varepsilon^*_j, \varepsilon^*_k) = 0$

$$\text{Cov}(Y^#_j, Y^#_k) = \text{Cov}(Y, Y) + \text{Cov}(\sqrt{\lambda_j} \Sigma^\frac{1}{2} \varepsilon^*_j, \sqrt{\lambda_k} \Sigma^\frac{1}{2} \varepsilon^*_k)$$

$$= \Sigma_e + \sqrt{\lambda_j} \sqrt{\lambda_k} \Sigma^\frac{1}{2} \text{Cov}(\varepsilon^*_j, \varepsilon^*_k) \Sigma^\frac{1}{2} = \Sigma_e.$$
Then the whole Simex data set is

\[ \mathbf{Y}^\# \sim N_{nB^*} \left( (1_B^* \otimes I_n)\xi, \Sigma^\# \right), \]

where

\[
\Sigma^\# = \begin{pmatrix}
(1 + \lambda_1)\Sigma_e & \Sigma_e & \cdots & \Sigma_e \\
\Sigma_e & (1 + \lambda_2)\Sigma_e & \cdots & \Sigma_e \\
\Sigma_e & \cdots & \cdots & \Sigma_e \\
\Sigma_e & \cdots & \cdots & (1 + \lambda_{B^*})\Sigma_e
\end{pmatrix}
\]

\[
\Sigma^\# = D \otimes \Sigma_e + (1_B^* \otimes I_n)\Sigma_e(1_B^* \otimes I_n)^T.
\]

Rewrite the covariance matrix by using the matrix rules we get

\[
\Sigma^\# = D \otimes \Sigma_e + (1_B^* \otimes I_n)\Sigma_e(1_B^* \otimes I_n)^T.
\]

\[\blacksquare\]

For comparing the experiments we reduce \( \mathcal{F}_{\text{simex}} \) to a sufficient experiment with an \( n \) dimensional normal distribution.

**Theorem 16.** An equivalent experiment to \( \mathcal{F}_{\text{simex}} \) is

\[
\mathcal{F}_{\text{suff}}^\# = \left\{ N_n \left( \xi, \Sigma_{\text{suff}}^\# \right), \xi \in \Theta \subset \mathbb{R}^n \right\},
\]

where

\[
\Sigma_{\text{suff}}^\# = (1 + \Delta)\Sigma_e, \quad \Delta = \frac{1}{\sum_{k=1}^{B^*} \lambda_k^{-1}}.
\]

**Proof.** From Theorem 23 it follows

\[
\mathcal{F}_{\text{suff}}^\# = \left\{ N_n \left( \xi, \Sigma_{\text{suff}}^\# \right), \xi \in \Theta \subset \mathbb{R}^n \right\},
\]

where

\[
\Sigma_{\text{suff}}^\# = (1_B^* \otimes I_n)^T \Sigma_{\text{suff}}^\# (1_B^* \otimes I_n)^{-1}
\]

Using Lemma 24 we rewrite

\[
(1_B^* \otimes I_n)^T \Sigma_{\text{suff}}^\# (1_B^* \otimes I_n)
\]

\[
= (1_B^* \otimes I_n)^T (D \otimes \Sigma_e + (1_B^* \otimes I_n)\Sigma_e(1_B^* \otimes I_n)^T)^{-1} (1_B^* \otimes I_n)
\]

\[
= (\Sigma_e + ((1_B^* \otimes I_n)^T (D \otimes \Sigma_e)^{-1} (1_B^* \otimes I_n)^T)^{-1})^{-1}
\]

26
Applying \((A \otimes B)(F \otimes G) = (AF \otimes BG)\) we get

\[
(1_{B^*} \otimes I_n)^T (D \otimes \Sigma_e)^{-1} (1_{B^*} \otimes I_n)^T = 1_{B^*}^T D^{-1} 1_{B^*} \otimes \Sigma_e^{-1} = \Delta^{-1} \Sigma_e^{-1}.
\]

Remind \(1_{B^*}^T D^{-1} 1_{B^*} = \Delta^{-1}\) is a scalar. \(\blacksquare\)

5.4 Hybrid method

Let us now combine both aspects of bootstrap sampling and Simex sampling in such a way that we smooth the observations like in the parametric bootstrap procedure and add an pseudo error with varying variances like in the Simex approach.

Thus the new generated samples are a combination of both principles:

\[
Y_k^{*#} = L Y + \sqrt{\lambda_k} \Sigma_e^{1/2} \epsilon_k^*, \quad \epsilon_k^* \sim N_n(0, I_n), \quad k = 1, ..., B^*
\]  

(5.4.1)

For \(L = I_n\) we are back to Simex. For \(\lambda_k = 1\) we obtain the parametric bootstrap simulation.

Theorem 17. The hybrid experiment related to (5.4.1) is

\[
\mathcal{F}_{\text{hybrid}} = \left\{ N_{nB^*} \left( (1_{B^*} \otimes L) \xi, \Sigma^{*#} \right) , \xi \in \Theta \subset \mathbb{R}^n \right\},
\]

where

\[
\Sigma^{*#} = D \otimes \Sigma_e + (1_{B^*} \otimes L) \Sigma_e (1_{B^*} \otimes L)^T.
\]

Proof. It is enough to calculate expectation and covariance. We have

\[
EY_k^{*#} = EL Y = L \xi
\]

and

\[
\text{Cov}(Y_k^{*#}) = \text{Cov}(LY) + \text{Cov}(\sqrt{\lambda_k} \Sigma_e^{1/2} \epsilon_j^*)
\]

\[
= L \Sigma_e L^T + \lambda_k \Sigma_e^{1/2} I_n \Sigma_e^{1/2} = L \Sigma_e L^T + \lambda_k \Sigma_e
\]

and for \(j \neq k\)

\[
\text{Cov}(Y_j^{*#}, Y_k^{*#}) = \text{Cov}(LY, LY) + \text{Cov}(\sqrt{\lambda_j} \Sigma_e^{1/2} \epsilon_j^*, \sqrt{\lambda_k} \Sigma_e^{1/2} \epsilon_k^*)
\]

\[
= L \Sigma_e L^T.
\]
Then the whole hybrid data set is

$$Y^{**} \sim N_{nB^*} \left((1_{B^*} \otimes L)\xi, \Sigma^{**}\right),$$

where

$$\Sigma^{**} = \begin{pmatrix}
L\Sigma \xi L^T + \lambda_1 \Sigma \xi & L\Sigma \xi L^T & \cdots & L\Sigma \xi L^T \\
L\Sigma \xi L^T & L\Sigma \xi L^T + \lambda_2 \Sigma \xi & \cdots & L\Sigma \xi L^T \\
L\Sigma \xi L^T & \cdots & L\Sigma \xi L^T + \lambda_{B^*} \Sigma \xi \\
\end{pmatrix}$$

which can be rewritten as

$$\Sigma^{**} = D \otimes \Sigma_\varepsilon + (1_{B^*} \otimes L) \sum_e (1_{B^*} \otimes L)^T.$$

For comparing the experiments we reduce to a sufficient experiment with an $n$ dimensional normal distribution.

**Theorem 18.** An equivalent experiment to $\mathcal{F}_{\text{hybrid}}$ is

$$\mathcal{F}_{\text{suff}}^{**} = \left\{ N_n \left(\xi, \Sigma_{\text{suff}}^{**}\right), \xi \in \Theta \subset \mathbb{R}^n \right\},$$

where

$$\Sigma_{\text{suff}}^{**} = \Sigma_\varepsilon + \Delta \left(L\Sigma_\varepsilon^{-1} L^T\right)^{-1}.$$

**Proof.** From Theorem 23 it follows

$$\mathcal{F}_{\text{suff}}^{**} = \left\{ N_n \left(\xi, \Sigma_{\text{suff}}^{**}\right), \xi \in \Theta \subset \mathbb{R}^n \right\},$$

where

$$\Sigma_{\text{suff}}^{**} = \left((1_{B^*} \otimes L)^T \Sigma_{\text{suff}}^{** -1} (1_{B^*} \otimes L)\right)^{-1}$$

Using Lemma 24 we rewrite

$$\Sigma_{\text{suff}}^{** -1} = (1_{B^*} \otimes L)^T \left(D \otimes \Sigma_\varepsilon + (1_{B^*} \otimes L)\Sigma_e (1_{B^*} \otimes L)^T\right)^{-1} (1_{B^*} \otimes L)$$

$$= (\Sigma_\varepsilon + ((1_{B^*} \otimes L)^T (D \otimes \Sigma_\varepsilon)^{-1} (1_{B^*} \otimes L)^T)^{-1})^{-1}$$
Using \((A \otimes B)(F \otimes G) = (AF \otimes BG)\) we get

\[(1_{B^*} \otimes L)^T (D \otimes \Sigma_e)^{-1} (1_{B^*} \otimes L)^T = 1_{B^*}^T, D^{-1} 1_{B^*} \otimes L \Sigma_e^{-1} L^T = \Delta^{-1} L \Sigma_e^{-1} L^T,\]

because \(\Delta^{-1} = 1_{B^*}^T, D^{-1} 1_{B^*}\) is a scalar. Summarizing

\[\Sigma_{suff}^\# = \Sigma + \Delta \left( L \Sigma_e^{-1} L^T \right)^{-1}.\]

5.5 The comparisons

Introduce a notation for the Hansen Torgersen result in Theorem 28:

For \(\Sigma_A \succ \Sigma_B\)

\[Bound(\Sigma_A, \Sigma_B) = \frac{1}{2} E \left( \left( \frac{\det \Sigma_A}{\det \Sigma_B} \right)^{-\frac{1}{2}} \exp\left( -\frac{1}{2} Z^T \left( \Sigma_B^{-1} - \Sigma_A^{-1} \right) Z \right) - 1 \right),\]

where \(Z \sim N(0, \Sigma_A)\).

Theorem 19. Under the model assumptions above

1. \(\delta(\mathcal{F}_{hybrid}, \mathcal{E}) \leq Bound \left( \Sigma_e + \Delta \left( L \Sigma_e^{-1} L^T \right)^{-1}, \Sigma_e \right)\)

2. \(\delta(\mathcal{F}_{simex}, \mathcal{E}) \leq P \left( \frac{\ln(1 + \Delta)}{\Delta} \leq \frac{\lambda^2}{n} \leq (1 + \Delta) \frac{\ln(1 + \Delta)}{\Delta} \right) \leq n\Delta \leq \max_k \lambda_k \frac{n}{B^*}\)

3. \(\Delta(\mathcal{F}_{boot}, \mathcal{F}_{hybrid}) = 0, \text{ for } \Delta^{-1} = B^*\)

and

\(\delta(\mathcal{F}_{boot}, \mathcal{F}_{hybrid}) = 0 \text{ if } B^* \geq \Delta^{-1}\)
\(\delta(\mathcal{F}_{hybrid}, \mathcal{F}_{boot}) = 0 \text{ if } B^* \leq \Delta^{-1}.\)
4. For $B^* \leq \Delta^{-1}$

$$\delta (F_{\text{boot}}, F_{\text{hybrid}}) \leq \text{Bound} \left( \Sigma_e + \frac{1}{B^*} (L \Sigma^{-1}_e L^T)^{-1}, \Sigma_e + \Delta (L \Sigma^{-1}_e L^T)^{-1} \right).$$

5. 

$$\Delta (F_{\text{boot}}, F_{\text{simex}}) = 0, \text{ for } \Delta^{-1}_e \Sigma^{-1}_e = B^* L \Sigma^{-1}_e L^T$$

and

$$\delta (F_{\text{boot}}, F_{\text{simex}}) = 0 \text{ iff } \Delta^{-1}_e \Sigma^{-1}_e \preceq B^* L \Sigma^{-1}_e L^T.$$  

$$\delta (F_{\text{simex}}, F_{\text{boot}}) = 0 \text{ iff } \Delta^{-1}_e \Sigma^{-1}_e \succeq B^* L \Sigma^{-1}_e L^T.$$  

6. For $\Delta^{-1}_e \Sigma^{-1}_e \succ B^* L \Sigma^{-1}_e L^T$

$$\delta (F_{\text{boot}}, F_{\text{simex}}) \leq \text{Bound} \left( \Sigma_e + \frac{1}{B^*} (L \Sigma^{-1}_e L^T)^{-1}, (1 + \Delta) \Sigma_e \right)$$

and for $\Delta^{-1}_e \Sigma^{-1}_e \prec B^* L \Sigma^{-1}_e L^T$

$$\delta (F_{\text{simex}}, F_{\text{bound}}) \leq \text{Bound} \left( (1 + \Delta) \Sigma_e, \Sigma_e + \frac{1}{B^*} (L \Sigma^{-1}_e L^T)^{-1} \right).$$

**Proof.**

1. Because of Corollary 21

$$\delta (F_{\text{hybrid}}, \mathcal{E}) = \delta (F_{\text{suff}}^#, \mathcal{E}).$$

From Lemma 22 follows

$$\delta (F_{\text{suff}}^#, \mathcal{E}) \leq \delta (F_{\text{suff, big}}^#, \mathcal{E}_{\text{big}})$$

with

$$F_{\text{suff, big}}^# = \left\{ N_n (\mu, \Sigma_{\text{suff}}^#), \mu \in \mathbb{R}^n \right\}, \mathcal{E}_{\text{big}} = \left\{ N_n (\mu, \Sigma_e), \mu \in \mathbb{R}^n \right\}.$$  

$$\Sigma_{\text{suff}}^# = \Sigma_e + \Delta (L \Sigma^{-1}_e L^T)^{-1} \succ \Sigma_e.$$  

Applying the result of Hansen and Torgersen in the form of Theorem 28 gives the bound.
2. From Corollary 21 and Lemma 22 we get
\[ \delta\left( \mathcal{F}_{\text{simex}}, \mathcal{E} \right) = \delta\left( \mathcal{F}_{\text{suff}}, \mathcal{E} \right) \leq \delta\left( \mathcal{F}_{\text{suff,big}}, \mathcal{E}_{\text{big}} \right) \]
where
\[ \mathcal{F}_{\text{suff,big}} = \{ N_n(\mu, (1 + \Delta)\Sigma_{\varepsilon}) : \mu \in \mathbb{R}^n \} \]
Then the result follows from Theorem 29 and Corollary 30.

3. Because of Corollary 21
\[ \Delta\left( \mathcal{F}_{\text{boot}}, \mathcal{F}_{\text{hybrid}} \right) = \Delta\left( \mathcal{F}^*_\text{suff}, \mathcal{F}^\#_{\text{suff}} \right) \]
and under \( \Delta^{-1} = B^* \) it holds \( \mathcal{F}^*_{\text{suff}} = \mathcal{F}^\#_{\text{suff}} \).
Applying Corollary 21 and Lemma 22 we obtain
\[ \delta\left( \mathcal{F}_{\text{boot}}, \mathcal{F}_{\text{hybrid}} \right) = \delta\left( \mathcal{F}^*_{\text{suff}}, \mathcal{F}^\#_{\text{suff}} \right) \leq \delta\left( \mathcal{F}^*_{\text{suff,big}}, \mathcal{F}^\#_{\text{suff,big}} \right) \]
with
\[ \mathcal{F}^*_{\text{suff,big}} = \{ N_n(\mu, \Sigma^*_\text{suff}) : \mu \in \mathbb{R}^n \} \]
\[ \Sigma^*_\text{suff} = \Sigma_{\varepsilon} + B^{*-1}(L\Sigma_{\varepsilon}^{-1}L^T)^{-1} \]
\[ \mathcal{F}^\#_{\text{suff,big}} = \{ N_n(\mu, \Sigma^\#_{\text{suff}}) : \mu \in \mathbb{R}^n \} \]
\[ \Sigma^\#_{\text{suff}} = \Sigma_{\varepsilon} + \Delta(L\Sigma_{\varepsilon}^{-1}L^T)^{-1} \]
From Theorem 28 we get
\[ \delta\left( \mathcal{F}^*_{\text{suff,big}}, \mathcal{F}^\#_{\text{suff,big}} \right) \geq 0 \iff B^{*-1}(L\Sigma_{\varepsilon}^{-1}L^T)^{-1} \succeq \Delta(L\Sigma_{\varepsilon}^{-1}L^T)^{-1} \]
Thus
\[ \delta\left( \mathcal{F}^*_{\text{suff,big}}, \mathcal{F}^\#_{\text{suff,big}} \right) \geq 0 \iff B^{*-1} \geq \Delta \iff B^* \leq \Delta^{-1} \]
in other words
\[ \delta\left( \mathcal{F}^*_{\text{suff,big}}, \mathcal{F}^\#_{\text{suff,big}} \right) = 0 \iff B^* \geq \Delta^{-1} \]
hence
\[ \delta\left( \mathcal{F}_{\text{boot}}, \mathcal{F}_{\text{hybrid}} \right) = 0 \iff B^* \geq \Delta^{-1}. \]
The results follows from Theorem 29.

4. -6. This follows by the same argumentation as above.
6 Appendix

Theorem 20. Suppose $T(X)$ is a sufficient statistic for $\theta \in \Theta$. Then the experiments

$$E = \{P_{\theta}^X, \theta \in \Theta\} \text{ and } E_{\text{suff}} = \left\{P_{\theta}^{T(X)}, \theta \in \Theta\right\}$$

are equivalent:

$$\Delta (E, E_{\text{suff}}) = 0$$

Proof. Apply Theorem 1 with

$$P_{\theta}^X (A) = \int K (A, t) dP_{\theta}^{T(X)} ,$$

where

$$K (A, t) = P_{X/T(X)=t} (A).$$

(6.0.1)

Because of the sufficiency of $T(X)$, $K (A, t)$ is independent of $\theta$. Hence

$$\delta (E_{\text{suff}}, E) = 0 .$$

The other direction works for arbitrary statistics. Apply Theorem 1 with

$$P_{\theta}^{T(X)} (B) = \int K (B, x) dP_{\theta}^X ,$$

where

$$K (B, x) = I_{T^{-1}(A)} (x) ; \quad T^{-1} (A) = \{x : T(x) \in B\} .$$

Hence

$$\delta (E, E_{\text{suff}}) = 0 .$$

\[ \blacksquare \]

Corollary 21. Consider two experiments $E$, $F$ and two equivalent experiments of them $E_{\text{suff}}$, $F_{\text{suff}}$ with $\Delta (E, E_{\text{suff}}) = 0$, $\Delta (F, F_{\text{suff}}) = 0$. Then:

$$\Delta (E, F) = \Delta (E_{\text{suff}}, F_{\text{suff}})$$
Proof. $\Delta (\mathcal{E}, \mathcal{F})$ is a distance and the triangle inequality is valid. We have

$$
\Delta (\mathcal{E}, \mathcal{F}) \leq \Delta (\mathcal{E}, \mathcal{E}_{\text{suff}}) + \Delta (\mathcal{E}_{\text{suff}}, \mathcal{F}) \\
\leq \Delta (\mathcal{E}_{\text{suff}}, \mathcal{F}) \\
\leq \Delta (\mathcal{E}_{\text{suff}}, \mathcal{F}_{\text{suff}}) + \Delta (\mathcal{F}_{\text{suff}}, \mathcal{F}) \\
\leq \Delta (\mathcal{E}_{\text{suff}}, \mathcal{F}_{\text{suff}}).
$$

Otherwise

$$
\Delta (\mathcal{E}_{\text{suff}}, \mathcal{F}_{\text{suff}}) \leq \Delta (\mathcal{E}_{\text{suff}}, \mathcal{E}) + \Delta (\mathcal{E}, \mathcal{F}_{\text{suff}}) \\
\leq \Delta (\mathcal{E}, \mathcal{F}_{\text{suff}}) \\
\leq \Delta (\mathcal{E}, \mathcal{F}) + \Delta (\mathcal{F}, \mathcal{F}_{\text{suff}}) \\
\leq \Delta (\mathcal{E}, \mathcal{F}).
$$

One more small result with respect to deficiencies:

**Lemma 22.** Consider $\Theta' \subseteq \Theta$

$$
\mathcal{E}' = \{P_{\theta}, \theta \in \Theta'\} \subseteq \mathcal{E} = \{P_{\theta}, \theta \in \Theta\}
$$

and

$$
\mathcal{F}' = \{Q_{\theta}, \theta \in \Theta'\} \subseteq \mathcal{F} = \{Q_{\theta}, \theta \in \Theta\}
$$

Then

$$
\delta (\mathcal{E}', \mathcal{F}') \leq \delta (\mathcal{E}, \mathcal{F})
$$

**Proof.** Apply Theorem 1. For all $K$

$$
\sup_{\theta \in \Theta'} \| Q_{\theta} - KP_{\theta} \| \leq \sup_{\theta \in \Theta} \| Q_{\theta} - KP_{\theta} \|,
$$

Thus

$$
\delta (\mathcal{E}', \mathcal{F}') = \inf_{K} \sup_{\theta \in \Theta'} \frac{1}{2} \| Q_{\theta} - KP_{\theta} \| \leq \inf_{K} \sup_{\theta \in \Theta} \frac{1}{2} \| Q_{\theta} - KP_{\theta} \| = \delta (\mathcal{E}, \mathcal{F}) .
$$

Consider a normal experiment $Y \sim N_n (X\beta, \Sigma)$, $\theta = \beta$, $\Theta = \mathbb{R}^p$ with arbitrary known covariance matrix $\Sigma$ and known "design" matrix $X$. 
Theorem 23. Let
\[ E = \{ N_n (X\beta, \Sigma), \beta \in \mathbb{R}^p \} \]
and 
\[ E_{suff} = \left\{ N_p \left( \beta, \left( X^T \Sigma^{-1} X \right)^{-} \right), \beta \in \mathbb{R}^p \right\} \]

1. Then
\[ T(Y) = \left( X^T \Sigma^{-1} X \right)^{-} X^T \Sigma^{-1} Y \]
is a sufficient statistic in \( E \).

2. \( E \) and \( E_{suff} \) are equivalent.

Proof. The generalized least squares estimator is given by
\[ \hat{\beta}_{GLSE} = \arg \min \left( Y - X\beta \right)^T \Sigma^{-1} \left( Y - X\beta \right) = \left( X^T \Sigma^{-1} X \right)^{-} X^T \Sigma^{-1} Y \]
and \( X\hat{\beta}_{GLSE} = PY \), where \( P = X \left( X^T \Sigma^{-1} X \right)^{-} X^T \Sigma^{-1} \) is the projection with respect to the weighted norm \( \| Y \|_{\Sigma^{-1}}^2 = Y^T \Sigma^{-1} Y \).

We have the decomposition:
\[ \| Y - X\beta \|_{\Sigma^{-1}}^2 = \| Y - PY \|_{\Sigma^{-1}}^2 + \| PY - X\beta \|_{\Sigma^{-1}}^2 \]
and the factorization of the likelihood function
\[ L(\beta) = \left( \frac{1}{2\pi} \right)^{\frac{n}{2}} \left( \frac{1}{\det(\Sigma)} \right)^{\frac{1}{2}} \exp \left( -\frac{1}{2} \| Y - X\beta \|_{\Sigma^{-1}}^2 \right) \]
where
\[ h(Y) = \left( \frac{1}{2\pi} \right)^{\frac{n}{2}} \left( \frac{1}{\det(\Sigma)} \right)^{\frac{1}{2}} \exp \left( -\frac{1}{2} \| Y - PY \|_{\Sigma^{-1}}^2 \right) \]
and
\[ g(T(Y), \beta) = \exp \left( -\frac{1}{2} \| X\hat{\beta}_{GLSE} - X\beta \|_{\Sigma^{-1}}^2 \right) \].

From the Neyman factorization criterion follows that
\[ T(Y) = \hat{\beta}_{GLSE} = \left( X^T \Sigma^{-1} X \right)^{-} X^T \Sigma^{-1} Y \]
is a sufficient statistic. \( T(Y) \) is a linear function of a normal distributed vector and thus \( T(Y) \) is normal distributed with
\[ ET(Y) = \left( X^T \Sigma^{-1} X \right)^{-} X^T \Sigma^{-1} (EY) = \left( X^T \Sigma^{-1} X \right)^{-} X^T \Sigma^{-1} X\beta = \beta \]
and
\[
Cov(T(Y)) = (X^T \Sigma^{-1} X)^- \cdot X^T \Sigma^{-1} Cov(Y) \Sigma^{-1} X (X^T \Sigma^{-1} X)^-.
\]

Using the property of the generalized inverse, that \(A^{-}AA^{-} = A^{-}\) we get
\[
Cov(T(Y)) = (X^T \Sigma^{-1} X)^-.
\]

Thus
\[
T(Y) \sim N_p \left( \beta, (X^T \Sigma^{-1} X)^- \right)
\]

and the related experiment is \(E_{suff}\). Theorem 20 delivers the second statement.

A useful tool for matrix calculation gives the following lemma.

**Lemma 24.** If all necessary inverses exist and if \(A,B,C\) are matrices of the respected dimensions with \(B^T A^{-1} B = D\) then
\[
B^T(A + BC B^T)^{-1} B = (D + C)^{-1}.
\]

**Proof:**
It holds, compare for instance [2],
\[
(A + BC B^T)^{-1} = A^{-1} - A^{-1} B (C^{-1} + B^T A^{-1} B)^{-1} B^T A^{-1}.
\]

Thus
\[
B^T(A + BC B^T)^{-1} B = D - D(C^{-1} + D)^{-1} D.
\]

Applying (6.0.2) once more gives
\[
D - D(C^{-1} + D)^{-1} D = D - D(D^{-1} - D^{-1} (C + D)^{-1} D^{-1}) D
\]
\[
= D - D + (C + D)^{-1} = (C + D)^{-1}.
\]

Let us quote an result of Torgersen and Hansen from Le Cams book, p. 130 Corollary 1, using the denotations from the book.

**Theorem 25.** Let \(BB^T - AA^T\) be positive semi definite and
\[
E_A = \{ N_{n_A} (A^T \theta, I) , \theta \in \mathbb{R}^k \} , \quad E_B = \{ N_{n_B} (B^T \theta, I) , \theta \in \mathbb{R}^k \}
\]
Then

\[ \Delta (\mathcal{E}_A, \mathcal{E}_B) = \delta (\mathcal{E}_A, \mathcal{E}_B) = \frac{1}{2} E \left| \frac{\det(BB^T)}{\det(AA^T)} \right|^{\frac{1}{2}} \exp\left(-\frac{1}{2} Z^T(BB^T - AA^T)Z\right) - 1 \]

where \( Z \sim N(0, (AA^T)^{-1}) \).

**Corollary 26.** Especially for \( BB^T = cAA^T, c > 1 \) it holds

\[ \Delta (\mathcal{E}_A, \mathcal{E}_B) = \delta (\mathcal{E}_A, \mathcal{E}_B) = P\left( \frac{\ln c}{c-1} \leq \frac{\chi^2}{k} \leq c \frac{\ln c}{c-1} \right) \]

where \( \chi^2 \) chi squared distributed with \( k \) degrees of freedom.

**Proof.** The result is given in the Le Cam book, [4] - nevertheless let us make the calculations.

Under \( BB^T = cAA^T \) we have

\[ \left( \frac{\det(BB^T)}{\det(AA^T)} \right)^{\frac{1}{2}} = c^{\frac{k}{2}} \left( \frac{\det(AA^T)}{\det(AA^T)} \right)^{\frac{1}{2}} = c^{\frac{k}{2}} \]

and

\[ \exp\left(-\frac{1}{2} Z^T(BB^T - AA^T)Z\right) = \exp\left(-\frac{(c-1)}{2} Z^T(AA^T)Z\right) = \exp\left(-\frac{(c-1)}{2} \chi^2\right), \]

where \( \chi^2 \) chi squared distributed with \( k \) degrees of freedom. Thus we have to calculate

\[ \frac{1}{2} E \left| c^{\frac{k}{2}} \exp\left(-\frac{(c-1)}{2} \chi^2\right) - 1 \right| \]

The following Lemma gives the result.

**Lemma 27.** It holds:

\[ \frac{1}{2} E \left| c^{\frac{k}{2}} \exp\left(-\frac{(c-1)}{2} \chi^2\right) - 1 \right| = P\left( \frac{\ln c}{c-1} \leq \frac{\chi^2}{k} \leq c \frac{\ln c}{c-1} \right), \]

where \( \chi^2 \) chi squared distributed with \( k \) degrees of freedom.
Proof. Note

\[
\frac{1}{2} E |h(x) - 1| =
\]

\[
= \frac{1}{2} \left( \int |h(x) - 1| f(x) \, dx \right)
\]

\[
= \frac{1}{2} \left( \int_{h(x)>1} (h(x) - 1) f(x) \, dx + \int_{h(x)\leq1} (1 - h(x)) f(x) \, dx \right)
\]

\[
= \frac{1}{2} \left( - \int_{h(x)>1} f(x) \, dx + \int_{h(x)\leq1} f(x) \, dx \right.
\]

\[
+ \int_{h(x)>1} h(x) f(x) \, dx - \int_{h(x)\leq1} h(x) f(x) \, dx
\]

\[
= \frac{1}{2} \left( -P(h(x) > 1) + P(h(x) \leq 1) \right.
\]

\[
+ \int_{h(x)>1} h(x) f(x) \, dx - \int_{h(x)\leq1} h(x) f(x) \, dx
\]

\[
= \frac{1}{2} \left( 2P(h(x) \leq 1) - 1 + \int_{h(x)>1} h(x) f(x) \, dx - \int_{h(x)\leq1} h(x) f(x) \, dx \right)
\]

For the density \( f \) of the chi squared distribution with \( k \) degrees of freedom

\[
f(x) = \frac{1}{\Gamma\left(\frac{k}{2}\right)\pi^{\frac{k}{2}}k^{-\frac{k}{2}-1}} \left(\frac{1}{2}\right)^{\frac{k}{2}} \exp\left(-\frac{x}{2}\right)
\]

and for

\[
h(x) = c^\frac{k}{2} \exp\left(-\frac{(c-1)}{2}x\right)
\]

we get

\[
f(x)h(x) = f(cx)c.
\]

Hence

\[
\int_{h(x)>1} h(x) f(x) \, dx = \int_{h(x)>1} cf(cx) \, dx = \int_{h(x/c)>1} f(z) \, dz = P\left(h(x/c) > 1\right)
\]

and

\[
\int_{h(x)\leq1} h(x) f(x) \, dx = P\left(h(x/c) \leq 1\right) = 1 - P\left(h(x/c) > 1\right).
\]

We obtain

\[
\frac{1}{2} E |h(x) - 1| = P(h(x) \leq 1) - P\left(h(x/c) > 1\right).
\]
Now

\[ h(x) \leq 1 \]

\[ \iff c^k \exp(-\frac{(c-1)}{2} x) \leq 1 \iff \exp(-\frac{(c-1)}{2} x) \leq c^{-\frac{k}{2}} \]

\[ \iff -\frac{(c-1)}{2} x \leq \ln(c^{-\frac{k}{2}}) \iff (c-1)x \geq k \ln(c) \iff \frac{x}{k} \geq \frac{\ln(c)}{c-1} \]

and

\[ h\left(\frac{x}{c}\right) > 1 \]

\[ \iff c^\frac{k}{2} \exp(-\frac{(c-1)}{2c} x) > 1 \iff \exp(-\frac{(c-1)}{2c} x) > c^{-\frac{k}{2}} \]

\[ \iff \frac{x}{k} < \frac{c \ln(c)}{c-1}. \]

We get the result from above.

\[ \frac{1}{2} E |h(x) - 1| = P\left(\frac{\ln(c)}{c-1} \leq \frac{x}{k} < \frac{c \ln(c)}{c-1}\right). \]

In the original paper of Hansen, Torgersen (1974) the result was formulated as comparison. Here we give the version where the covariance matrices are the key points. It is just an equivalent formulation with the related equivalent sufficient experiments.

**Theorem 28.** Let

\[ \mathcal{E}_A = \{N_k(\theta, \Sigma_A), \theta \in \mathbb{R}^k\}, \quad \mathcal{E}_B = \{N_k(\theta, \Sigma_B), \theta \in \mathbb{R}^k\} \]

Then

\[ \delta(\mathcal{E}_A, \mathcal{E}_B) \geq 0 \quad \text{iff} \quad \Sigma_A \succeq \Sigma_B. \]

and for \( \Sigma_A \succ \Sigma_B \)

\[ \delta(\mathcal{E}_A, \mathcal{E}_B) = \frac{1}{2} E \left( \left( \frac{\det \Sigma_A}{\det \Sigma_B} \right)^{-\frac{1}{2}} \exp\left( -\frac{1}{2} Z^T \left( \Sigma_B^{-1} - \Sigma_A^{-1} \right) Z \right) - 1 \right) \]

where \( Z \sim N(0, \Sigma_A) \).
**Theorem 29.** Let
\[
\mathcal{E} = \{ N_k (\theta, \Sigma), \theta \in \mathbb{R}^k \},
\]
\[
\mathcal{F} = \{ N_k (\theta, (1 + \Delta) \Sigma), \theta \in \mathbb{R}^k \},
\]
with \( \Delta > 0 \). Then
\[
\delta(\mathcal{F}, \mathcal{E}) = P\left( \frac{\ln (1 + \Delta)}{\Delta} \leq \frac{\chi^2}{k} \leq (1 + \Delta) \frac{\ln (1 + \Delta)}{\Delta} \right)
\]
where \( \chi^2 \) is chi squared distributed with \( k \) degrees of freedom.

**Proof.** The sufficient experiments to
\[
\mathcal{E}_A = \{ N_n (A^T \theta, I), \theta \in \mathbb{R}^k \}, \quad \mathcal{E}_B = \{ N_n (B^T \theta, I), \theta \in \mathbb{R}^k \}
\]
are
\[
\mathcal{E}_{A,suff} = \{ N_n (\theta, (AA^T)^{-1}), \theta \in \mathbb{R}^k \}, \quad \mathcal{E}_{B,suff} = \{ N_n (\theta, (BB^T)^{-1}), \theta \in \mathbb{R}^k \}
\]
such that by
\[
\Delta(\mathcal{E}_A, \mathcal{E}_B) = \Delta(\mathcal{E}_{A,suff}, \mathcal{E}_{B,suff})
\]
For \( BB^T = cAA^T \), \( c > 1 \) follows \( (BB^T)^{-1} = \frac{1}{c} (AA^T)^{-1} \) and \( (AA^T)^{-1} = c(BB^T)^{-1} = (1 + \Delta)(BB^T)^{-1} \). Thus each \( P_\theta = N_n (\theta, (AA^T)^{-1}) \in \mathcal{E}_{A,suff} \) can be presented as a convolution \( P_\theta = Q * Q_\theta \), with \( Q = N(0, \Delta^2 (BB^T)^{-1}) \) and \( Q_\theta = N_n (\theta, (BB^T)^{-1}) \in \mathcal{E}_{B,suff} \). Hence \( \delta(\mathcal{E}_{B,suff}, \mathcal{E}_{A,suff}) = 0 \) and
\[
\Delta(\mathcal{E}_A, \mathcal{E}_B) = \Delta(\mathcal{E}_{A,suff}, \mathcal{E}_{B,suff}) = \delta(\mathcal{E}_{A,suff}, \mathcal{E}_{B,suff})
\]
Set \( (BB^T)^{-1} = \Sigma, \quad c = (1 + \Delta) \) and \( \mathcal{E}_{A,suff} = \mathcal{F}, \quad \mathcal{E}_{B,suff} = \mathcal{E} \). Then from Theorem 28 it follows that
\[
\delta(\mathcal{F}, \mathcal{E}) = P\left( \frac{\ln (1 + \Delta)}{\Delta} \leq \frac{\chi^2}{k} \leq (1 + \Delta) \frac{\ln (1 + \Delta)}{\Delta} \right).
\]

**Corollary 30.** It holds for \( \Delta > 0 \)
\[
0 < P\left( \frac{\ln (1 + \Delta)}{\Delta} \leq \frac{\chi^2}{k} \leq (1 + \Delta) \frac{\ln (1 + \Delta)}{\Delta} \right) < k\Delta,
\]
where \( \chi^2 \) is chi squared distributed with \( k \) degrees of freedom.
Proof. It holds for $P = P\left(\frac{\ln(1+\Delta)}{\Delta}\leq \frac{x^2}{k} \leq (1+\Delta)\frac{\ln(1+\Delta)}{\Delta}\right)$

$$\min_{x \in I} f_k(x) \ k \ln(1 + \Delta) \leq P \leq \max_{x \in I} f_k(x) \ k \ln(1 + \Delta)$$

with $I = \left[\frac{\ln(1+\Delta)}{\Delta} k, (1 + \Delta)k\frac{\ln(1+\Delta)}{\Delta}\right]$ and

$$f_k(x) = \frac{1}{\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} \left(\frac{1}{2}\right)^{\frac{k}{2}} \exp\left(-\frac{x}{2}\right)$$

We have $\max_{x \in I} f_k(x) < \max f_k(x) = f_k(k)$ and $f_k(k)$ is decreasing in $k$. Thus

$$\max_{x \in I} f_k(x) < \max f_k(x) = f_k(k) < f_1(1) = 0.2418$$

Otherwise

$$0 < \min_{x \in I} f_k(x).$$

Furthermore we use

$$1 < \frac{\ln(1 + \Delta)}{\Delta}.$$ 

References


