Unbiased Estimation + Testing = Biased Estimation

P. J. G. Teunissen[1, 2]

1 GNSS Research Centre, Curtin University of Technology, Perth, Australia
2 Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, The Netherlands
p.teunissen@curtin.edu.au

Abstract: The principle of unbiased estimation plays a prominent role in our geodetic data analytic toolbox. We do our best to free the models from any biases that otherwise may corrupt our results and we work with unbiased estimators with the aim that the unbiasedness remains intact in the computed output as well. In this contribution we will show however that the combination of unbiased estimation and statistical testing produces biased solutions.

Keywords: Unbiased Estimation, Hypothesis Testing, Biased Estimation

1 Introduction

The principle of unbiased estimation plays a prominent role in the theory of statistical inference (Dermanis, 1976, 1980, 1990; Dermanis and Rummel, 2000; Teunissen, 2000a). Popular estimators are linear least-squares estimators, linear unbiased estimators and best linear unbiased estimators. These estimators deliver unbiased results provided their input is unbiased as well. Statistical testing is then often used to validate the data with the aim to remove any biases that may be present. The consequence of this practice is that the resulting procedure is not one of estimation only, nor one of testing only, but actually one where estimation and testing are combined. We show that the nonlinearity created by this combination of estimation and testing causes the final results to be still biased. Thus as the unbiasedness property of the applied 'unbiased estimators' is undone, one may also question whether or not other estimators exist or can be constructed that do a better job in dampening the bias propagation through the combined estimation + testing procedure.

2 Estimation and Testing

2.1 Estimation under H_o and H_a

Consider the linear model

Quod Erat Demonstrandum – In quest of the ultimate geodetic insight |
Special issue for Professor Emeritus Athanasios Dermanis |
School of Rural and Surveying Engineering, AUTh, 2018

$$H_o: \; E(y) = Ax \;, \;\; D(y) = Q_{yy} \tag{1}$$

with $E(.)$ the expectation operator, $y \in \mathbb{R}^m$ the normally distributed random vector of observables, $A \in \mathbb{R}^{m \times n}$ the given design matrix of rank n, $x \in \mathbb{R}^n$ the to-be-estimated unknown parameter vector, $D(.)$ the dispersion operator and $Q_{yy} \in \mathbb{R}^{m \times m}$ the given positive-definite variance matrix of y. The linear model (1) will be referred to as our null-hypothesis H_o.

Under H_o, the best linear unbiased estimator (BLUE) of x is given as

$$\hat{x}_o = A^+ y \tag{2}$$

with least-squares (LS) inverse $A^+ = Q_{\hat{x}_o\hat{x}_o} A^T Q_{yy}^{-1}$, in which $Q_{\hat{x}_o\hat{x}_o} = D(\hat{x}_o) = (A^T Q_{yy}^{-1} A)^{-1}$ is the dispersion or variance matrix of $\hat{x}_o$.

As the BLUE's property of x_o depends on the validity of H_o, it is important that one has sufficient confidence in the assumptions underlying the null-hypothesis. Although every part of the null-hypothesis can be wrong of course, we assume here that if a mis-specification occurred that it is confined to an underparametrization of the mean of y, in which case

$$H_a: E(y) = Ax + Cb \;, \;\; D(y) = Q_{yy} \tag{3}$$

for some vector $b_y = Cb$. Experience has shown that these type of mis-specifications are by large the most common errors that occur when formulating the model. Through $b_y = Cb$ one may model, for instance, the presence of one or more blunders (outliers) in the data, cycle-slips in phase data, satellite failures, antenna-height errors, erroneous neglectance of atmospheric delays, or any other systematic effect that one failed to take into account under H_a. In the following we assume matrix $[AC] \in \mathbb{R}^{m \times (n+q)}$ to be known of rank $n+q$ and the parameter vector $b \in \mathbb{R}^q$ to be unknown. The linear model (3) will be referred to as the alternative-hypothesis H_a.

Under H_a, the BLUE of x is given as

$$\hat{x}_a = \bar{A}^+ y \tag{4}$$

with LS-inverse

$$\bar{A}^+ = (\bar{A}^T Q_{yy}^{-1} \bar{A})^{-1} \bar{A}^T Q_{yy}^{-1} \;, \;\; \bar{A}^+ = P_C^\perp A \;, \;\; P_C^\perp = I_m - C(C^T Q_{yy}^{-1} C)^{-1} C^T Q_{yy}^{-1} \,.$$

As this BLUE is based on a model with more parameters, its precision will never be better than that of $\hat{x}_o$, i.e. $D(\hat{x}_o) \le D(\hat{x}_a)$.

2.2 Testing of H_o against H_a

The estimation of x would not pose a problem if we would know which of the two models would be true. In case of H_o, we would use $\hat{x}_o$ to estimate x, but if we would know that H_a is true then we would use $\hat{x}_a$ instead. Using the estimator $\hat{x}_o$ when knowing that H_a is true should be avoided, as this would result in a biased solution, since

$$E(\hat{x}_o | H_a) = x + A^{+}Cb \tag{5}$$

The problem in practice of course is that we do not know which of the models are true. Even if we have taken the utmost care in formulating a model which we believe to be true, mis-specifications could still be present thus invalidating the model. Methods of statistical testing have therefore been developed (Baarda, 1967, 1968; Koch, 1999; Teunissen, 2000b; Imparato, 2016) that allow us to decide with some confidence which of the models to work with. In case of the above H_o and H_a, it seems reasonable to decide in favour of H_o if the BLUE of b can be considered 'insignificant'. With the BLUE of b under H_a given as

$$\hat{b} = \bar{C}^{+} y \tag{6}$$

with LS-inverse $\bar{C}^{+} = (\bar{C}^T Q_{yy}^{-1} \bar{C})^{-1} \bar{C}^T Q_{yy}^{-1}$, $\bar{C} = P_A^{\perp} C$ and variance matrix $Q_{\hat{b}\hat{b}} = (\bar{C}^T Q_{yy}^{-1} \bar{C})^{-1}$, the decision in favour of H_o is therefore taken when $\hat{b}$ lies in the acceptance region $\mathcal{A}$,

$$\hat{b} \in \mathcal{A} = \left\{ b \in \mathbb{R}^q \middle| \| b \|_{Q_{\hat{b}\hat{b}}}^2 \leq \chi_a^2(q,0) \right\} \tag{7}$$

with $\| . \|_{Q_{\hat{b}\hat{b}}}^2 = (.)^T Q_{\hat{b}\hat{b}}^{-1} (.)$ and $\chi_a^2(q,0)$ the critical value computed from the central Chi-square distribution with q degrees of freedom and chosen level of significance α. Thus H_o is rejected in favour of H_a if

$$\| \hat{b} \|_{Q_{\hat{b}\hat{b}}}^2 = \hat{b}^T Q_{\hat{b}\hat{b}}^{-1} \hat{b} > \chi_a^2(q,0) \tag{8}$$

This test is known to be a uniformly most powerful invariant (UMPI) test for testing H_o against H_a (Arnold, 1981; Teunissen, 2000b).

If the outcome of testing is to reject H_o, then not $\hat{x}_o$, but $\hat{x}_a$ is provided as the estimator for x. The three estimators, $\hat{x}_o$ (cf. 2), $\hat{x}_a$ (cf. 4) and $\hat{b}$ (cf. 6) are related as

$$\hat{x}_a = \hat{x}_o - A^{+}C\hat{b} \tag{9}$$

Thus if H_o is rejected, then $A^{+}C\hat{b}$ is the correction, which is aimed at removing the bias $A^{+}Cb$ (cf. 5) from $\hat{x}_o$.

3 Does The Bias Get Removed?

3.1 The Estimator Revisited

Although estimation and testing are often treated separately and independently, in actual practice they are usually combined. This implies that strictly speaking one cannot simply assign the properties of $\hat{x}_o$ or $\hat{x}_a$ to the actual estimator that is computed. That is, the actual estimator that is produced is not $\hat{x}_o$ nor $\hat{x}_a$, but in fact

$$\bar{x} = \begin{cases} \hat{x}_o \; if \; \hat{b} \in \mathcal{A} \\ \hat{x}_a \; if \; \hat{b} \notin \mathcal{A} \end{cases} \tag{10}$$

Hence, it is the quality of $\bar{x}$, rather than that of $\hat{x}_o$ or $\hat{x}_a$, that determines the quality of the produced results. Since ideally the goal of testing is to be able to have the bias A^+Cb removed from $\hat{x}_o$ when H_a is true (cf. 5), it is relevant to know what the mean of the actual estimator $\bar{x}$ is. By making use of the relation $\hat{x}_a = \hat{x}_o - A^+C\hat{b}$, the expectation of $\bar{x}$ can be determined as

$$\begin{aligned} E(\bar{x} \mid H_o) &= x - A^+C\textstyle\int_{\notin \mathcal{A}} \beta p_{\hat{b}}(\beta \mid H_o) d\beta \\ E(\bar{x} \mid H_a) &= x + A^+C\textstyle\int_{\in \mathcal{A}} \beta p_{\hat{b}}(\beta \mid H_a) d\beta \end{aligned} \tag{11}$$

with $p_{\hat{b}}(\beta \mid H_o)$ and $p_{\hat{b}}(\beta \mid H_a)$ being the probability density function (PDF) of b under resp. H_o and H_a.

The result (11) shows that the estimator $\bar{x}$ is biased in general, this in contrast to $\hat{x}_o$ under H_o and $\hat{x}_a$ under H_a. The conclusion reads therefore that testing does not succeed in removing (all) the bias from the contaminated data. The cause for the presence of these remaining biases is the nonlinearity involved in the mapping of (10). Thus although $\hat{x}_o$ and $\hat{x}_a$ are both individually linear functions of y, the actually produced estimator $\bar{x}$ is not. It is this nonlinearity that prohibits the unbiasedness of $\hat{x}_o$ and $\hat{x}_a$, under resp. H_o and H_a, to be passed on to $\bar{x}$.

Table 1. *The mean of the random parameters vectors $\hat{x}_o$, $\hat{x}_a$ and $\bar{x}$, under H_o and H_a respectively.*

	H_o	H_a
x_o	$E(\hat{x}_o \mid H_o) = x$	$E(\hat{x}_o \mid H_a) = x + b_{\hat{x}_o}$
x_a	$E(\hat{x}_a \mid H_o) = x$	$E(\hat{x}_a \mid H_a) = x$
$\bar{x}$	$E(\bar{x} \mid H_o) = x$	$E(\bar{x} \mid H_a) = x + b_{\bar{x}}$

Although (11) indicates that $\bar{x}$ is generally biased under both H_o and H_a, we have in our case $\int_{\notin \mathcal{A}} \beta p_{\hat{b}}(\beta | H_o) d\beta = 0$, due to the symmetry with respect to the origin of both the acceptance region $\mathcal{A}$ and the PDF $p_{\hat{b}}(\beta | H_o)$. Hence, in our case, the estimator $\bar{x}$ is fortunately always unbiased under H_o,

$$E(\bar{x} | H_o) = x \tag{12}$$

This is not true however for $\bar{x}$ under H_a. We have

$$E(\bar{x} | H_a) = x + b_{\bar{x}} \tag{13}$$

with the bias given as

$$b_{\bar{x}} = A^{+} C b_{\mathcal{A}} \quad \text{with} \quad b_{\mathcal{A}} = \int_{\in \mathcal{A}} \beta p_{\hat{b}}(\beta | H_a) d\beta \tag{14}$$

This shows that the bias in $\bar{x}$ is driven by the vector $b_{\mathcal{A}}$ and its propagation into the parameter space. The vector $b_{\mathcal{A}}$ itself is governed by the acceptance region $\mathcal{A}$ and through the PDF $p_{\hat{b}}(\beta | H_a)$, by the actual bias b and the precision with which it can be estimated, $Q_{\hat{b}\hat{b}}$. To see the effect testing has, one can compare the testing-induced bias (14), with the bias one otherwise would have when using $\hat{x}_o$ under H_a (cf. 5),

$$b_{\hat{x}_o} = E(\hat{x}_o - x | H_a) = A^{+} C b \tag{15}$$

It follows from comparing (14) with (15), since $b = E(\hat{b} | H_a) = \int_{\mathbb{R}^q} \beta p_{\hat{b}}(\beta | H_a) d\beta$, that through testing it is the component of this integral over the acceptance region $\mathcal{A}$ that is retained. We thus have $b_{\bar{x}} = b_{\hat{x}_o}$ if $\mathcal{A} = \mathbb{R}^q$, which corresponds to the case of always accepting H_o. A summary overview of the means of the random vectors $\hat{x}_o$, $\hat{x}_a$ and $\bar{x}$ is given in Table 1.

3.2 When is the bias small and large?

As mentioned above the testing induced-bias $b_{\bar{x}}$ is driven by $b_{\mathcal{A}}$ and its propagation into the parameter space. To describe its significance, we will work with the dimensionless bias-to-noise ratio (BNR) $\| b_{\bar{x}} \|_{Q_{\hat{x}_o \hat{x}_o}}$ and study its behaviour for the one-dimensional case. If $q = 1$ then matrix C becomes a vector, $C = c$, and b becomes a scalar. For this case the BNR works out as

$$\| b_{\bar{x}} \|_{Q_{\hat{x}_o \hat{x}_o}} = \frac{|b_{\mathcal{A}}|}{\sigma_{\hat{b}}} \tan\theta \tag{16}$$

with θ being the angle that vector c makes with the range space of the orthogonal complement of A, i.e. $\tan\theta = \| P_A c \|_{Q_{yy}} / \| P_A^{\perp} c |_{Q_{yy}}$ (Teunissen, 2000b), p.111. Here $P_A = AA^{+}$ and $P_A^{\perp} = I_m - AA^{+}$. In the decomposition (16), $b_{\mathcal{A}} / \sigma_{\hat{b}}$ describes the significance of $b_{\mathcal{A}}$, while $\tan\theta$ shows how it gets propagated into the parameter space.

There are two cases for which $b_{\mathcal{A}}$ will be 'small'. It will be small when the PDF $p_{\hat{b}}(\beta | H_a)$ has only a small portion of its probability mass over $\mathcal{A}$, and it will be small when it differs only a little from the PDF under H_o. To quantify this behaviour, we make use of the one-dimensional integral

$$b_{\mathcal{A}} = \frac{1}{\sqrt{2\pi}\sigma_{\hat{b}}} \int_{-\chi_a(1,0)\sigma_{\hat{b}}}^{\chi_a(1,0)\sigma_{\hat{b}}} \beta \exp\{-\frac{1}{2}\left(\frac{\beta - b}{\sigma_{\hat{b}}}\right)^2\} d\beta \tag{17}$$

from which it can be worked out that

$$\frac{b_{\mathcal{A}}}{\sigma_{\hat{b}}} = F(\chi_a(1,0)) - F(-\chi_a(1,0)) \tag{18}$$

in which $F(x) = \phi\left(\frac{b}{\sigma_{\hat{b}}} + x\right) + \frac{b}{\sigma_{\hat{b}}} \Phi\left(\frac{b}{\sigma_{\hat{b}}} + x\right)$ with $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}x^2\}$ and $\Phi(x) = \int_{-\infty}^{x} \phi(v) d(v)$.

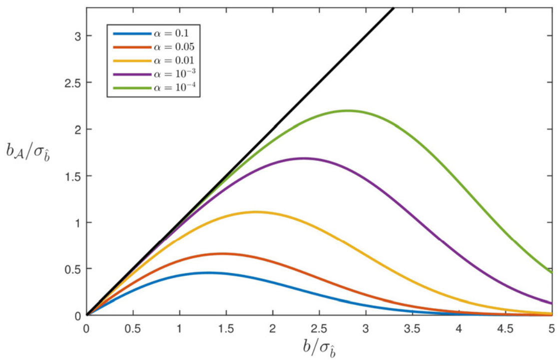

***Fig. 1**. The output bias $b_{\mathcal{A}} / \sigma_{\hat{b}}$ as function of the input bias $b / \sigma_{\hat{b}}$ for different values of α (after Teunissen et al (2016)).*

Figure 1 shows $b_{\mathcal{A}} / \sigma_{\hat{b}}$ as a function of $b / \sigma_{\hat{b}}$ for different values of α. The straight line in the figure describes the bias one would have in case no testing would be performed. As $b_{\mathcal{A}} \leq b$ for every value of b, the figure clearly shows the benefit of testing: the bias that remains after testing is always smaller than the original bias. Note that this benefit, i.e. the difference between b and $b_{\mathcal{A}}$, only kicks in after the bias b has become large enough. The difference is small, when b is small, and it gets larger for larger b, with $b_{\mathcal{A}}$ approaching zero in the limit. Also note that for smaller levels of significance α, the difference between b and $b_{\mathcal{A}}$ stays small for a larger range of b-values. This is understandable as a smaller corresponds with a larger acceptance interval $\mathcal{A}$, as a consequence of which one would have for a larger range of b-values an outcome of testing that does not differ from the no-testing scenario.

4 Summary

Although statistical testing is intended to remove biases from the data, we have shown that biases will always remain under the alternative hypothesis. The usage of estimators that are unbiased under their respective hypotheses is therefore no guarantee that the finally computed solution is unbiased as well. We have shown that the presence of such biases in the final solution can be explained by the nonlinearity created by the combination of estimation and testing. The size of these remaining biases depends on the strength of the underlying model, the chosen false alarm rate, and the size and type of the actual input bias. The size of the remaining bias will get smaller with increasing model strength and larger levels of significance. Despite the presence of these biases, the benefit of testing was demonstrated by showing that the remaining bias is always smaller than the bias one otherwise would have in the absence of testing.

Acknowledgements: This contribution derives from collaborative work of the author with Drs Davide Imparato and Christian Tiberius. The author is the recipient of an Australian Research Council (ARC) Federation Fellowship (project number FF0883188). This support is gratefully acknowledged.

References

Arnold, S. (1981): *The theory of linear models and multivariate analysis*, vol 2., Wiley, New York

Baarda, W. (1967): *Statistical concepts in geodesy.* Netherlands Geodetic Commission, Publications on Geodesy, New Series, 2(4)

Baarda, W. (1968): *A testing procedure for use in geodetic networks*. Netherlands Geodetic Commission, Publications on Geodesy, New Series, 2(5)

Dermanis, A. (1976*) Probabilistic and deterministic aspects of linear estimation in geodesy*. PhD thesis, Ohio State University.

Dermanis. A, (1980): Adjustment of geodetic observations in the presence of signals. Proceedings of the international school of advanced geodesy, *Bollettino di Geodesia e Scienze Afini* 38, pp. 419-445

Dermanis, A. (1990): A unified approach to linear estimation and prediction. *20th IUGG General Assembly*, Vienna

Dermanis A. and R. Rummel (2000): Data analysis methods in geodesy. In: *Geomatic Method for the Analysis of Data in the Earth Sciences*, Springer, pp 17-92

Imparato, D. (2016): *GNSS based Receiver Autonomous Integrity Monitoring for Aircraft Navigation*. PhD-thesis, TU Delft

Koch KR (1999): *Parameter estimation and hypothesis testing in linear models*. Springer Science and Business Media

Teunissen, P.J.G. (2000a): *Adjustment Theory: An Introduction*. Delft University Press, Series on Mathematical Geodesy and Positioning

Teunissen, P.J.G. (2000b): *Testing Theory: An Introduction*. Delft University Press, Series on Mathematical Geodesy and Positioning

Teunissen, P.J.G., D. Imparato and C.CJ.M. Tiberius (2016): Does successful RAIM produce unbiased positions? *NAVIGATION*, Journal of Institute of Navigation, pp. 1-10.