Prediction Intervals for Regression Models

David J. Olive*

Southern Illinois University

March 6, 2006

Abstract

This paper presents simple large sample prediction intervals for a future response Y_f given a vector $\boldsymbol{x}_f$ of predictors when the regression model has the form $Y_i = m(\boldsymbol{x}_i) + e_i$ where m is a function of $\boldsymbol{x}_i$ and the errors e_i are iid. Intervals with correct asymptotic coverage and shortest asymptotic length can be made by applying the shorth estimator to the residuals. Since residuals underestimate the errors, finite sample correction factors are needed.

As an application, three prediction intervals are given for the least squares multiple linear regression model. The asymptotic coverage and length of these intervals and the classical estimator are derived. The new intervals are useful since the distribution of the errors does not need to be known, and simulations suggest

*David J. Olive is Associate Professor, Department of Mathematics, Southern Illinois University, Mailcode 4408, Carbondale, IL 62901-4408, USA. E-mail address: dolive@math.siu.edu. The author thanks the referees and editors for suggestions that improved the article.

that the large sample theory often provides good approximations for moderate sample sizes.

KEY WORDS: multiple linear regression; prediction intervals.

1 INTRODUCTION

Regression is the study of the conditional distribution $Y|\boldsymbol{x}$ of the response Y given the $p \times 1$ vector of predictors $\boldsymbol{x}$. An important regression model is

$$Y_i = m(\boldsymbol{x}_i) + e_i \tag{1.1}$$

for $i = 1, ..., n$ where m is a function of $\boldsymbol{x}_i$ and the errors e_i are continuous and iid. Many of the most important regression models have this form, including the multiple linear regression model and many time series, nonlinear, nonparametric and semiparametric models. If $\hat{m}$ is an estimator of m, then the ith residual is $r_i = Y_i - \hat{m}(\boldsymbol{x}_i) = Y_i - \hat{Y}_i$.

Notation is needed for the population and sample percentiles. Let ξ_α be the α percentile of the error e, i.e., $P(e \leq \xi_\alpha) = \alpha$. Let $\hat{\xi}_\alpha$ be the sample α percentile of the residuals, e.g., as computed by the *R/Splus* function `quantile`.

An important topic in regression analysis is predicting a future observation Y_f given a vector of predictors $\boldsymbol{x}_f$ where $(Y_f, \boldsymbol{x}_f)$ comes from the same population as the past data $(Y_i, \boldsymbol{x}_i)$ for $i = 1, ..., n$. Let $1 - \alpha_2 - \alpha_1 = 1 - \alpha$ with $0 < \alpha < 1$ and $\alpha_1 < 1 - \alpha_2$ where $0 < \alpha_i < 1$. Then

$$P[Y_f \in (m(\boldsymbol{x}_f) + \xi_{\alpha_1}, m(\boldsymbol{x}_f) + \xi_{1-\alpha_2})] = P(m(\boldsymbol{x}_f) + \xi_{\alpha_1} < m(\boldsymbol{x}_f) + e_f < m(\boldsymbol{x}_f) + \xi_{1-\alpha_2})$$

$$= P(\xi_{\alpha_1} < e_f < \xi_{1-\alpha_2}) = 1 - \alpha_2 - \alpha_1 = 1 - \alpha.$$

A large sample $100(1 - \alpha)\%$ prediction interval (PI) has the form $(\hat{L}_n, \hat{U}_n)$ where $P(\hat{L}_n < Y_f < \hat{U}_n) \xrightarrow{P} 1 - \alpha$ as the sample size $n \to \infty$. See Patel (1989) for a review. To derive a simple PI, assume that $\hat{m}$ is consistent: $\hat{m}(\boldsymbol{x}) \xrightarrow{P} m(\boldsymbol{x})$ as $n \to \infty$. Then

$$r_i = Y_i - \hat{m}(\boldsymbol{x}_i) \xrightarrow{P} Y_i - m(\boldsymbol{x}_i) = e_i \text{ and } \hat{\xi}_\alpha \xrightarrow{\mathrm{P}} \xi_\alpha.$$

Consequently,

$$P[Y_f \in (\hat{m}(\boldsymbol{x}_f) + \hat{\xi}_{\alpha_1}, \hat{m}(\boldsymbol{x}_f) + \hat{\xi}_{1-\alpha_2})] \xrightarrow{P} 1 - \alpha$$

as the sample size $n \to \infty$. Typically the squared residuals underestimate the squared errors. Hence $(\hat{m}(\boldsymbol{x}_f) + \hat{\xi}_{\alpha_1}, \hat{m}(\boldsymbol{x}_f) + \hat{\xi}_{1-\alpha_2})$ has less than $100(1-\alpha)\%$ coverage in small samples. Multiplying $\hat{\xi}_{\alpha_1}$ by a_n and $\hat{\xi}_{1-\alpha_2}$ by b_n where, for example, $a_n = b_n = 1 + 15/n$, can greatly improve the small sample performance of the PI. If $a_n \xrightarrow{P} 1$ and $b_n \xrightarrow{P} 1$ as $n \to \infty$, then

$$(\hat{L}_n, \hat{U}_n) = (\hat{m}(\boldsymbol{x}_f) + a_n\hat{\xi}_{\alpha_1}, \hat{m}(\boldsymbol{x}_f) + b_n\hat{\xi}_{1-\alpha_2}) \tag{1.2}$$

is a large sample $100(1-\alpha)\%$ PI for Y_f.

Preston (2000) suggested the PI (1.2) with $a_n = b_n \equiv 1$ for simple linear regression. The following section will give illustrations of (1.2) and show how to choose the finite sample correction factors a_n and b_n.

2 Examples

The location model is

$$Y_i = \mu + e_i \tag{2.1}$$

for $i = 1, ..., n$. Hence $\boldsymbol{x}_i = 1$ and $m(\boldsymbol{x}_i) = \mu$. If $\hat{m}(\boldsymbol{x}_i) = \hat{\mu}$ for all i, then the ith residual $r_i = Y_i - \hat{\mu}$, and the sample percentiles $\hat{\xi}_\alpha$ of the residuals are related to the sample percentiles $\hat{\xi}_\alpha(Y)$ of Y by $\hat{\xi}_\alpha = \hat{\xi}_\alpha(Y) - \hat{\mu}$. Thus $\hat{m}(\boldsymbol{x}_i) + \hat{\xi}_\alpha = \hat{\mu} + \hat{\xi}_\alpha(Y) - \hat{\mu} = \hat{\xi}_\alpha(Y)$. If $a_n = b_n \equiv 1$, then the PI (1.2) becomes the usual nonparametric PI

$$(\hat{\xi}_{\alpha_1}(Y), \hat{\xi}_{1-\alpha_2}(Y)). \tag{2.2}$$

Example 2.1. The Buxton (1920) data contains *heights* of 87 men (in mm) but 5 heights were recorded to be about 0.75 inches tall. Deleting these 5 cases and setting $\alpha_1 = \alpha_2 = 0.05$ yields a large sample 90% PI (1590,1790). It turns out that the 5 outliers were recorded under *head length* and were 1755, 1537, 1650, 1675 and 1610. Hence 4 out of 5 of these values fell within the PI.

Parametric PIs often try to find a pivotal quantity based on $Y_f - \hat{m}(\boldsymbol{x}_f)$. Assume that $(Y_f, \boldsymbol{x}_f)$ is independent of the past data and that $\text{VAR}(e) = \sigma^2$. Since $\hat{m}$ is based on the past, Y_f and $\hat{m}$ are independent and $\text{VAR}(Y_f - \hat{m}(\boldsymbol{x}_f)) = \text{VAR}(Y_f) + \text{VAR}(\hat{m}(\boldsymbol{x}_f)) = \text{VAR}(e_f) + \text{VAR}(\hat{m}(\boldsymbol{x}_f)) = \sigma^2 + \text{VAR}(\hat{m}(\boldsymbol{x}_f))$. If $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$ and $\hat{V}$ is an estimator of $\text{VAR}(\hat{m}(\boldsymbol{x}_f))$ such that $\hat{V} \xrightarrow{P} 0$, then the pivotal quantity T satisfies

$$T - \frac{e_f}{\sigma} = \frac{Y_f - \hat{m}(\boldsymbol{x}_f)}{\sqrt{\hat{\sigma}^2 + \hat{V}}} - \frac{e_f}{\sigma} \xrightarrow{P} 0.$$

Thus the percentiles of T estimate the percentiles of e/σ, asymptotically.

The most important regression model is the multiple linear regression (MLR) model

$$Y_i = x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + e_i \tag{2.3}$$

for $i = 1, \ldots, n$. In matrix notation, these n equations become

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \tag{2.4}$$

where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times p$ matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, and $\boldsymbol{e}$ is an $n \times 1$ vector of unknown errors. We will assume that $x_{i,1} \equiv 1$ and that the iid errors have 0 mean and constant variance σ^2. Note that the 0 mean assumption can be made without loss of generality since if

$Y_i = \tilde{\beta}_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + \tilde{e}_i$ where $E(\tilde{e}_i) \equiv \mu$, then $Y_i = \beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i$ where $e_i = \tilde{e}_i - \mu$ and $\beta_1 = \tilde{\beta}_1 + \mu$. Thus $E(Y_i) = m(\boldsymbol{x}_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}$.

Under regularity conditions, the least squares (OLS) estimator $\hat{\boldsymbol{\beta}}$ satisfies

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(0, \sigma^2 \ \boldsymbol{W}) \tag{2.5}$$

when

$$\frac{\boldsymbol{X}^T \boldsymbol{X}}{n} \to \boldsymbol{W}^{-1}.$$

This large sample result is analogous to the central limit theorem and is often a good approximation if $n > 5p$ and the error distribution has "light tails," i.e., the tails go to zero at an exponential rate or faster. For error distributions with heavier tails, much larger samples are needed, and the assumption that the variance σ^2 exists is crucial, e.g., Cauchy errors are not allowed. Also, outliers can cause OLS to perform arbitrarily poorly.

Under regularity conditions, much of the inference for MLR that is valid when the iid errors $e_i \sim N(0, \sigma^2)$, is approximately valid when the e_i are iid with 0 mean and constant variance if the sample size is large. For example, confidence intervals for β_i are asymptotically correct, the MSE can be used to estimate σ^2 (see Seber and Lee 2003, p. 45) and variable selection procedures perform well (see Olive and Hawkins 2005).

However, parametric prediction intervals made under the assumption that $e_i \sim N(0, \sigma^2)$ may not perform well. Following Seber and Lee (2003, p. 132), the classical parametric $100(1-\alpha)\%$ PI is

$$\hat{Y}_f \pm t_{n-p,1-\alpha/2} \sqrt{MSE} \sqrt{(1 + h_f)} \tag{2.6}$$

where $P(T \leq t_{n-p,\alpha}) = \alpha$ if T has a t distribution with $n-p$ degrees of freedom and the

"leverage"

$$h_f = \boldsymbol{x}_f^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_f.$$

Notice that if the e_i have constant variance σ^2, then $\mathrm{VAR}(Y_f - \hat{Y}_f) = \mathrm{VAR}(Y_f) + \mathrm{VAR}(\hat{Y}_f) \approx \sigma^2 + \sigma^2 h_f$ using the fact that Y_f is independent of $\hat{Y}_f$ and the fact that for large samples $\hat{\boldsymbol{\beta}} \approx N_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$ so that $\boldsymbol{x}_f^T\hat{\boldsymbol{\beta}} \approx N(\boldsymbol{x}_f^T\boldsymbol{\beta}, \sigma^2\boldsymbol{x}_f^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_f)$. Here the approximation that $\mathrm{VAR}(\hat{Y}_f) = \mathrm{VAR}(\boldsymbol{x}_f^T\hat{\boldsymbol{\beta}}) \approx \sigma^2\boldsymbol{x}_f^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_f = \sigma^2 h_f$ is being used. If the errors $e_i \sim N(0, \sigma^2)$, then the approximation is exact and the pivotal quantity

$$T = \frac{Y_f - \hat{Y}_f}{\sqrt{MSE(1 + h_f)}} \sim t_{n-p}.$$

We assume that $h_f \leq \max_{i=1,\ldots,n} h_i$ since otherwise extrapolation is occuring, i.e., $(Y_f, \boldsymbol{x}_f)$ is not from the same population as the past data $(Y_i, \boldsymbol{x}_i)$ and the PI can not be expected to be valid. Then typically $h_f \to 0$ and

$$T - \frac{e}{\sigma} \xrightarrow{P} 0.$$

Notice that the PI

$$\hat{Y}_f \pm t_{n-p,1-\alpha/2}\sqrt{MSE}\sqrt{(1 + h_f)} = \hat{Y}_f \pm z_{1-\alpha/2}\sqrt{MSE}\frac{t_{n-p,1-\alpha/2}}{z_{1-\alpha/2}}\sqrt{(1 + h_f)}.$$

Thus the quantity

$$a_n = b_n = \frac{t_{n-p,1-\alpha/2}}{z_{1-\alpha/2}}\sqrt{(1 + h_f)}$$

can be regarded as a finite sample correction factor if $e_i \sim N(0, \sigma^2)$ where $P(Z \leq z_\alpha) = \alpha$ if $Z \sim N(0, 1)$.

Let $1 - \delta$ be the asymptotic coverage of the classical nominal $(1 - \alpha)100\%$ PI (2.6).

Then

$$1-\delta = P(-\sigma z_{1-\alpha/2} < e < \sigma z_{1-\alpha/2}) \geq 1 - \frac{1}{z^2_{1-\alpha/2}} \tag{2.7}$$

where the inequality follows from Chebyshev's inequality.

Next we find $a_n = b_n$ in order to tailor the PI (1.2) for MLR. Notice that

$$E(MSE) = E\left(\sum_{i=1}^{n} \frac{r_i^2}{n-p}\right) = \sigma^2 = E\left(\sum_{i=1}^{n} \frac{e_i^2}{n}\right)$$

suggests that

$$\sqrt{\frac{n}{n-p}} r_i \approx e_i.$$

Using

$$a_n = \left(1 + \frac{15}{n}\right)\sqrt{\frac{n}{n-p}}\sqrt{(1+h_f)}, \tag{2.8}$$

a large sample semiparametric $100(1-\alpha)\%$ PI for Y_f is

$$(\hat{Y}_f + a_n\hat{\xi}_{\alpha/2}, \hat{Y}_f + a_n\hat{\xi}_{1-\alpha/2}). \tag{2.9}$$

This PI is very similar to the classical PI except that $\hat{\xi}_\alpha$ is used instead of σz_α to estimate the error percentiles ξ_α. The term $\sqrt{\frac{n}{n-p}}$ is needed since the squared residuals underestimate the squared errors, and the term $1 + 15/n$ was found to work well in the simulation study described below. Stine (1985) used the bootstrap to provide nonparametric PIs while Schmoyer (1992) gave asymptotically valid PIs based on the quantiles of a convolution of the empirical distribution of the residuals and the limiting normal distribution of the parameter estimates.

An asymptotically conservative (ac) $100(1-\alpha)\%$ PI has asymptotic coverage $1-\delta \geq 1-\alpha$. We used the (ac) $100(1-\alpha)\%$ PI

$$\hat{Y}_f \pm \sqrt{\frac{n}{n-p}} \max(|\hat{\xi}_{\alpha/2}|, |\hat{\xi}_{1-\alpha/2}|)\sqrt{(1+h_f)} \tag{2.10}$$

which has asymptotic coverage

$$1-\delta = P[-\max(|\xi_{\alpha/2}|, |\xi_{1-\alpha/2}|) < e < \max(|\xi_{\alpha/2}|, |\xi_{1-\alpha/2}|)]. \tag{2.11}$$

Notice that $1-\alpha \leq 1-\delta \leq 1-\alpha/2$ and $1-\delta = 1-\alpha$ if the error distribution is symmetric.

Example 2.2. For the Buxton (1920) data suppose that the response Y = *height* and the predictors were a constant, *head length*, *nasal height*, *bigonal breadth* and *cephalic index*. Five outliers were deleted leaving 82 cases. Figure 1 shows a fit response plot of the fitted values versus the response Y with the identity line added as a visual aid. If the model is good then the plotted points should scatter about the identity line in an evenly populated band. The triangles represent the upper and lower limits of the semiparametric 95% PI (2.9). Notice that 79 (or 96%) of the Y_i fell within their corresponding PI while 3 Y_i did not.

In the simulations below, $\hat{\xi}_\alpha$ will be the sample percentile for the PIs (2.9) and (2.10). A PI is asymptotically optimal if it has the shortest asymptotic length that gives the desired asymptotic coverage. An asymptotically optimal PI can be created by applying the shorth(c) estimator to the residuals where $c = \lceil n(1-\alpha) \rceil$ and $\lceil x \rceil$ is the smallest integer $\geq x$, e.g., $\lceil 7.7 \rceil = 8$. That is, let $r_{(1)}, ..., r_{(n)}$ be the order statistics of the residuals. Compute $r_{(c)} - r_{(1)}, r_{(c+1)} - r_{(2)}, ..., r_{(n)} - r_{(n-c+1)}$. Let $(r_{(d)}, r_{(d+c-1)}) = (\hat{\xi}_{\alpha_1}, \hat{\xi}_{1-\alpha_2})$ correspond to the interval with the smallest distance. See Grübel (1988) and Rousseeuw and Leroy (1988). Then the 100 $(1-\alpha)$% PI for Y_f is

$$(\hat{Y}_f + a_n \hat{\xi}_{\alpha_1}, \hat{Y}_f + b_n \hat{\xi}_{1-\alpha_2}). \tag{2.12}$$

In the simulations, we used $a_n = b_n$ where a_n is given by (2.8). See Di Bucchianico,

Einmahl, and Mushkudiani (2001) for related intervals for the location model.

A small simulation study compares the PI lengths and coverages for sample sizes $n = 50, 100$ and 1000 for several error distributions. The value $n = \infty$ gives the asymptotic coverages and lengths. The MLR model with $E(Y_i) = 1 + x_{i2} + \cdots + x_{i8}$ was used. The vectors $(x_2, ..., x_8)^T$ were iid $N_7(\mathbf{0}, \boldsymbol{I}_7)$. The error distributions were N(0,1), t_3, exponential(1) -1, uniform$(-1, 1)$ and $0.9N(0, 1) + 0.1N(0, 100)$. Also, a small sensitivity study to examine the effects of changing $(1 + 15/n)$ to $(1 + k/n)$ on the 99% PIs (2.9) and (2.12) was performed. For n=50 and k between 10 and 20, the coverage increased by roughly 0.001 as k increased by 1.

The simulation compared coverages and lengths of the classical (2.6), semiparametric (2.9), asymptotically conservative (2.10) and asymptotically optimal (2.12) PIs. The latter 3 intervals are asymptotically optimal for symmetric error distributions in that they have the shortest asymptotic length that gives the desired asymptotic coverage. The semiparametric PI gives the correct asymptotic coverage if the errors are not symmetric while the PI (2.10) gives higher coverage (is conservative). The simulation used 5000 runs and gave the proportion $\hat{p}$ of runs where Y_f fell within the nominal $100(1-\alpha)\%$ PI. The count $m\hat{p}$ has a binomial$(m = 5000, p = 1-\delta_n)$ distribution where $1-\delta_n$ converges to the asymptotic coverage $(1-\delta)$. The standard error for the proportion is $\sqrt{\hat{p}(1-\hat{p})/5000} =$ 0.0014, 0.0031 and 0.0042 for $p = 0.01, 0.05$ and 0.1, respectively. Hence an observed coverage $\hat{p} \in (.986, .994)$ for 99%, $\hat{p} \in (.941, .959)$ for 95% and $\hat{p} \in (.887, .913)$ for 90% PIs suggests that there is no reason to doubt that the PI has the nominal coverage.

Tables 1-5 show the results of the simulations for the 5 error distributions. The letters *c, s, a* and *o* refer to intervals (2.6), (2.9), (2.10) and (2.12) respectively. For the

normal errors, the coverages were about right and the semiparametric interval tended to be rather long for $n = 50$ and 100. The classical PI asymptotic coverage $1 - \delta$ tended to be fairly close to the nominal coverage $1 - \alpha$ for all 5 distributions and $\alpha = 0.01, 0.05,$ and 0.1. The classical PI was the most conservative for the uniform $(-1, 1)$ distribution. The classical PI had about 3% under coverage for the 99% PI when the errors were from the mixture distribution even though the length of PI was far shorter than the optimal asymptotic length of 32.9.

3 Conclusions

The large sample MLR prediction intervals presented in this paper are useful to practitioners since the normality assumption of the errors can be relaxed. For the importance of prediction intervals in data analysis, see Carroll and Ruppert (1991). The fit response plot should always be made to check the adequacy of the model (1.1) and adding the prediction limits as in Figure 1 is a valuable aid for explaining prediction intervals to students and consulting clients.

Large sample intervals similar to (2.9), (2.10) and (2.12) can be used for model (1.1) with $\hat{Y}_f = \hat{m}(\boldsymbol{x}_f)$, but may perform very poorly for realistic sample sizes if good finite sample correction factors can not be found. For semiparametric models such as spline and kernel fits, theory for the bias of r_i and the variability of $\hat{m}(\boldsymbol{x}_f)$ is needed to find useful correction factors. For example, theory for the bias of r_i and the variability of $\hat{Y}_f$ suggested Equation (2.8) as a correction factor for OLS MLR.

If there is a lot of data (e.g. ≥ 50 cases) at $\boldsymbol{x}_f$, then model free prediction intervals can

be created by applying the location model PI to the Ys at $\boldsymbol{x}_f$. If the constant variance assumption for MLR is violated, one could use the cases in a narrow vertical slice about $\boldsymbol{x}_f^T\hat{\boldsymbol{\beta}}$ in the fit response plot to make a PI. Both of these suggestions require much larger amounts of data than the simple model based PI (1.2).

A referee noted that PIs can be improved by applying optimal design techniques as in Müller and Kitsos (2004).

The *R/Splus* functions `piplot` and `pisim`, used to create Figure 1 and for the simulations, are included in the collection of functions *rpack.txt*. The Buxton data and *rpack.txt* are available from the website (http://www.math.siu.edu/olive/ol-bookp.htm).

4 References

Buxton, L.H.D., 1920. The anthropology of Cyprus. The Journal of the Royal Anthropological Institute of Great Britain and Ireland 50, 183-235.

Carroll, R.J., Ruppert, D., 1991. Prediction and tolerance intervals with transformation and/or weighting. Technometrics 33, 197-210.

Di Bucchianico, A., Einmahl, J.H.J., Mushkudiani, N.A., 2001. Smallest nonparametric tolerance regions. Ann. Statist. 29, 1320-1343.

Grübel, R., 1988. The length of the shorth. Ann. Statist. 16, 619-628.

Müller, C.H., Kitsos, C.P., 2004. Optimal design criteria based on tolerance regions. mODa 7–Advances in model-oriented design and analysis, eds. A. Di Bucchianico, H. Läuter, H.P. Wynn, Physica-Verlag, Heidelberg, 107-115.

Olive, D.J., Hawkins, D.M., 2005. Variable selection for 1D regression models. Techno-

metrics 47, 43-50.

Patel, J.K., 1989, Prediction intervals – a review. Commun. Statist. Theory and Methods 18, 2393-2465.

Preston, S., 2000. Teaching prediction intervals. J. Statist. Educat. 3, available from (http://www.amstat.org/publications/jse/secure/v8n3/preston.cfm).

Rousseeuw, P., Leroy, A., 1988. A robust scale estimator based on the shortest half. Statist. Neerlandica 42, 103-116.

Schmoyer, R.L., 1992. Asymptotically valid prediction intervals for linear models. Technometrics 34, 399-408.

Seber, G.A.F., Lee, A.J., 2003. Linear Regression Analysis, 2nd Edition. Wiley, New York.

Stine, R.A., 1985. Bootstrap prediction intervals for regression. J. Amer. Statist. Assoc. 80, 1026-1031.

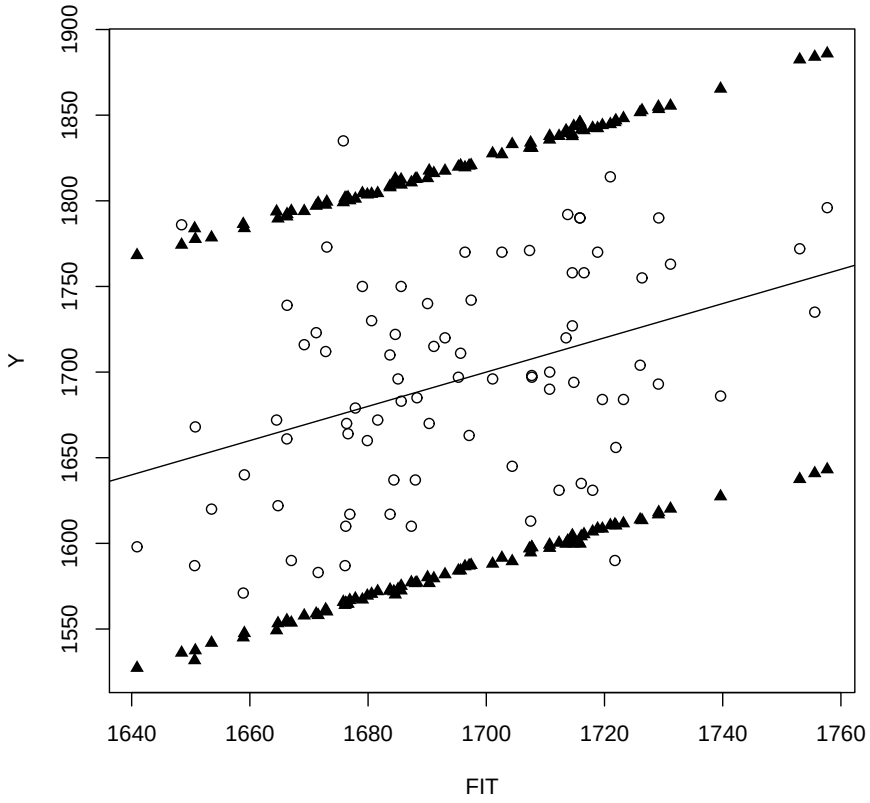

Figure 1: 95% PI Limits for Buxton Data

Table 1: N(0,1) Errors

α	n	clen	slen	alen	olen	ccov	scov	acov	ocov
0.01	50	5.860	6.172	5.191	6.448	.989	.988	.972	.990
0.01	100	5.470	5.625	5.257	5.412	.990	.988	.985	.985
0.01	1000	5.182	5.181	5.263	5.097	.992	.993	.994	.992
0.01	∞	5.152	5.152	5.152	5.152	.990	.990	.990	.990
0.05	50	4.379	5.167	4.290	5.111	.948	.974	.940	.968
0.05	100	4.136	4.531	4.172	4.359	.956	.970	.956	.958
0.05	1000	3.938	3.977	4.001	3.927	.952	.952	.954	.948
0.05	∞	3.920	3.920	3.920	3.920	.950	.950	.950	.950
0.1	50	3.642	4.445	3.658	4.193	.894	.945	.895	.929
0.1	100	3.455	3.841	3.519	3.690	.900	.930	.905	.913
0.1	1000	3.304	3.343	3.352	3.304	.901	.903	.907	.901
0.1	∞	3.290	3.290	3.290	3.290	.900	.900	.900	.900

Table 2: t_3 Errors

α	n	clen	slen	alen	olen	ccov	scov	acov	ocov
0.01	50	9.539	12.164	11.398	13.297	.972	.978	.975	.981
0.01	100	9.114	12.202	12.747	10.621	.978	.983	.985	.978
0.01	1000	8.840	11.614	12.411	11.142	.975	.990	.992	.988
0.01	∞	8.924	11.681	11.681	11.681	.979	.990	.990	.990
0.05	50	7.160	8.313	7.210	8.139	.945	.956	.943	.956
0.05	100	6.874	7.326	7.030	6.834	.950	.955	.951	.945
0.05	1000	6.732	6.452	6.599	6.317	.951	.947	.950	.945
0.05	∞	6.790	6.365	6.365	6.365	.957	.950	.950	.950
0.1	50	5.978	6.591	5.532	6.098	.915	.935	.900	.917
0.1	100	5.696	5.756	5.223	5.274	.916	.913	.901	.900
0.1	1000	5.648	4.784	4.842	4.706	.929	.901	.904	.898
0.1	∞	5.698	4.707	4.707	4.707	.935	.900	.900	.900

Table 3: Exponential(1) -1 Errors

α	n	clen	slen	alen	olen	ccov	scov	acov	ocov
0.01	50	5.795	6.432	6.821	6.817	.971	.987	.976	.988
0.01	100	5.427	5.907	7.525	5.377	.974	.987	.986	.985
0.01	1000	5.182	5.387	8.432	4.807	.972	.987	.992	.987
0.01	∞	5.152	5.293	8.597	4.605	.972	.990	.995	.990
0.05	50	4.310	5.047	5.036	4.746	.946	.971	.955	.964
0.05	100	4.100	4.381	5.189	3.840	.947	.971	.966	.955
0.05	1000	3.932	3.745	5.354	3.175	.945	.954	.972	.947
0.05	∞	3.920	3.664	5.378	2.996	.948	.950	.975	.950
0.1	50	3.601	4.183	3.960	3.629	.920	.945	.925	.916
0.1	100	3.429	3.557	3.959	3.047	.930	.943	.945	.913
0.1	1000	3.303	3.005	3.989	2.460	.931	.906	.951	.901
0.1	∞	3.290	2.944	3.991	2.303	.929	.900	.950	.900

Table 4: Uniform$(-1, 1)$ Errors

α	n	clen	slen	alen	olen	ccov	scov	acov	ocov
0.01	50	3.394	3.088	2.539	3.177	1.00	.998	.981	.999
0.01	100	3.158	2.589	2.361	2.542	1.00	.996	.985	.994
0.01	1000	2.991	2.068	2.068	2.060	1.00	.995	.993	.993
0.01	∞	2.975	1.980	1.980	1.980	1.00	.990	.990	.990
0.05	50	2.535	2.768	2.267	2.748	.979	.990	.954	.988
0.05	100	2.391	2.328	2.115	2.277	.988	.984	.956	.978
0.05	1000	2.275	1.937	1.935	1.927	1.00	.960	.955	.951
0.05	∞	2.263	1.900	1.900	1.900	1.00	.950	.950	.950
0.1	50	2.110	2.505	2.041	2.403	.919	.974	.904	.956
0.1	100	1.998	2.133	1.937	2.076	.935	.963	.916	.943
0.1	1000	1.908	1.827	1.825	1.811	.949	.910	.910	.898
0.1	∞	1.899	1.800	1.800	1.800	.950	.900	.900	.900

Table 5: 0.9 N(0,1) + 0.1 N(0,100) Errors

α	n	clen	slen	alen	olen	ccov	scov	acov	ocov
0.01	50	18.296	27.425	26.958	30.829	.964	.975	.980	.977
0.01	100	17.566	29.226	30.774	26.265	.961	.977	.982	.972
0.01	1000	17.072	32.306	34.056	31.100	.960	.988	.990	.987
0.01	∞	17.010	32.898	32.898	32.898	.960	.990	.990	.990
0.05	50	13.623	15.636	15.262	14.829	.945	.945	.943	.942
0.05	100	13.200	13.901	15.235	11.676	.949	.945	.954	.938
0.05	1000	12.971	13.257	14.656	12.354	.948	.948	.952	.945
0.05	∞	12.942	13.490	13.490	13.490	.948	.950	.950	.950
0.1	50	11.455	9.973	8.931	8.526	.937	.919	.901	.910
0.1	100	11.140	7.513	7.546	6.620	.941	.909	.907	.906
0.1	1000	10.871	4.939	5.096	4.791	.944	.904	.908	.901
0.1	∞	10.862	4.638	4.638	4.638	.934	.900	.900	.900