# Statistical calibration for infinite many future values in linear regression: simultaneous or pointwise tolerance intervals or what else?

April 19, 2022

**Abstract**

Statistical calibration using regression is a useful statistical tool with many applications. For confidence sets for $x$-values associated with infinitely many future $y$-values, there is a consensus in the statistical literature that the confidence sets constructed should guarantee a key *property*. While it is well known that the confidence sets based on the simultaneous tolerance intervals (STI's) guarantee this key *property* conservatively, it is desirable to construct confidence sets that satisfy this *property* exactly. Also, there is a misconception that the confidence sets based on the pointwise tolerance intervals (PTI's) also guarantee this *property*. This paper constructs the weighted simultaneous tolerance intervals (WSTI's) so that the confidence sets based on the WSTI's satisfy this *property* exactly if the future observations have the $x$-values distributed according to a known specific distribution $F(\cdot)$. Through the lens of the WSTI's, convincing counter examples are also provided to demonstrate that the confidence sets based on the PTI's do not guarantee the key *property* in general and so should not be used. The WSTI's have been applied to real data examples to show that the WSTI's can produce more accurate calibration intervals than STI's and PTI's.

*Keywords:* Confidence level; Confidence sets; Linear regression; Pointwise tolerance intervals; Simultaneous tolerance intervals; Statistical calibration.

# 1 Introduction

Statistical calibration using linear regression is a useful statistical tool, having a rich history going back to Eisenhart (1939). The problem involves a quantity of primary interest $x$ which is expensive or difficult to measure, a surrogate quantity $y$ which is cheaper or easy to measure, and the assumption that $y$ and $x$ are related by a linear regression model; we are more interested in the inference for covariate $x$ instead of response variable $y$.

For example, radon ($^{222}R_n$) is an important hazardous radioactive gas emitted from geological materials. The decay of $^{222}R_n$ is one of the most important causes of lung cancer in underground mining population (Dudney et al., 1995) and also has potential toxic effects to other human populations. The alpha particles and other radioactive solid decay products (polonium, bismuth, lead, etc.) emitted cause lung cancer by high energetic disintegration. In order to ensure the safety of environment, the alpha track detector (ATD) is often used to measure the intensity of $^{222}R_n$ with the ultimate aim of measuring the concentration of $^{222}R_n$. Suppose $x$ is the true concentration of $^{222}R_n$, while $y$ is the reading on an ATD, at a location. Since ATD has measurement errors, we should calibrate to obtain the confidence set for the true concentration of $^{222}R_n$, and so determine whether the indoor environment is safe by comparing the confidence set with the proposed standard.

A calibration problem often involves only two quantities $y$ and $x$ (or their suitable transformations), and polynomial regression model is a simple yet the most frequently used model to relate two quantities. Hence, in this paper, we focus on polynomial regression as in Han et al. (2016). Problems involving several $y$'s and so multivariate polynomial regression will be studied and reported separately in future.

As another example, cadmium ($Cd$) is a naturally occurring toxic heavy metal. Excessive intake of $Cd$ will cause $Cd$ poisoning, and so cause kidney, bone, lung and other organ

lesions (Stephens, 1920). Even with its low permissible exposure in humans, long-term exposure will cause health issues. Hence, it is necessary to measure the concentration of $Cd$ in the environment, such as in drinking water. Graphite Furnace Atomic Absorption Spectroscopy (GFAAS) is a commonly used elemental analysis technique to measure the concentration of almost all types of samples. When the sample is atomized, an absorbance signal is measured from which the sample concentration is determined. Suppose that $x$ is the cadmium ($Cd$) concentration in drinking water while $y$ is the absorbency signal measurement obtained on a stripchart recorder. While Beer's Law states that there is a linear relationship between absorbance $y$ and concentration $x$ (Ingle and Crouch, 1988), the calibrated regression line from the training data for most metals are not linear, therefore, we use polynomial regression to fit the training data, and then use statistical calibration to get the confidence set of the true concentration of $x$ given $y$, and so evaluate the quality of drinking water by comparing the confidence sets of true $Cd$ concentration with its threshold.

In order to use an observed $y$ to infer the corresponding but unobserved $x$, a calibration experiment is carried out to measure $y_{0i}$ corresponding to a known $x_{0i}$ for $i = 1, \cdots, n$. A regression model of $y$ on $x$ is then fitted by using the training data $\mathcal{E} = \{(x_{0i}, y_{0i}),\ i = 1, \cdots, n\}$ and used to infer the $x$-values corresponding to *infinitely many* $y$-values to be observed in future. The inference of the $x$-value corresponding to one single future $y$-value is considered by Eisenhart (1939), Seber (1977), Brown (1982) and Smith and Corbett (1987) among many others, and the relevant literature is reviewed in Osborne (1991), Brown (1993) and Parker et al. (2010).

This paper focuses on inference for *infinitely many* future $y$-values. Specifically, a confidence set $C(y_x)$ for the unknown $x$ corresponding to each $y_x$ observed in future is constructed in such a way that the infinite sequence of confidence sets $C(y_x)$ corresponding to an infinite

sequence of future $y_x$-values has the following property: with respect to the randomness in the training data $\mathcal{E}$, we can state with confidence $\gamma$ that the proportion of confidence sets $C(y_x)$ containing the corresponding true $x$-values is at least $\beta$, where $0 < \gamma, \beta < 1$ are pre-specified constants. This *property* can be expressed as

$$P_{\mathcal{E}}\left\{\liminf_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} I_{\left\{x_i\in C(y_{x_i})\right\}} \geq \beta\right\} = \gamma \tag{1}$$

where $y_{x_i,\ i=1,\cdots,N}$ are a sequence of future observations, $x_{i,\ i=1,\cdots,N}$ are the true but unknown $x$-values, $N$ is a very large number approaching infinity, $I_A$ denotes the indicator function of the set $A$ and hence $\frac{1}{N}\sum_{i=1}^{N} I_{\left\{x_i\in C(y_{x_i})\right\}}$ is the proportion of the $N$ confidence sets that contain the true $x$-values. It is argued by Lieberman et al. (1967), Scheffé (1973), Aitchison (1982), Mee et al. (1991), Mee and Eberhardt (1996), Mathew and Zha (1997), Mathew et al. (1998), Krishnamoorthy and Mathew (2009, Chapter 3), Han et al. (2016) and Chvosteková (2019) among others that this *property* is highly desirable in many applications. For example, a company that sells ATD's would require this *property* to be satisfied for values of $\gamma$ and $\beta$ close to one. Similar properties have been used in counting by weighing by Liu et al. (2016) and in classification by Liu et al. (2019a,b).

One standard way to construct the confidence sets $C(y_x)$ having the *property* (1) is to use the $(\beta, \gamma)$-simultaneous tolerance intervals (STI's). Construction of $(\beta, \gamma)$-STI's is considered first by Lieberman and Miller (1963) for simultaneous predictions. Most construction methods available in the literature, including Scheffé (1973), are conservative, that is, the left-hand-side probability in *property* (1) is larger than $\gamma$; see Han (2014) for a review. Stride has been made on the construction of exact $(\beta, \gamma)$-STI's by Odeh and Mee (1990) for one-sided case and Mee et al. (1991) for two-sided case. Note, however, that the $(\beta, \gamma)$-STI's of Odeh and Mee (1990) and Mee et al. (1991) are for a multiple linear regression model, where the covariates are assumed to have no functional relationships, over a special ellipsoidal

covariate region only. These STI's become conservative when applied to a polynomial regression model of order two (i.e. quadratic regression) or higher, in which the covariates are $x$, $x^2$, etc. Even for the simple linear regression (i.e. polynomial regression of order one), these STI's are only over a covariate set that is symmetric about $\bar{x}_0$, the mean of the observed covariate values $x_{0i}$ in $\mathcal{E}$. Hence these STI's are also conservative when applied to a simple linear regression model with the covariate interval $[a, b]$ not symmetric about $\bar{x}_0$. Exact one-sided $(\beta, \gamma)$-STI's for a polynomial regression model of any order and over any given interval $[a, b]$ are constructed in Han et al. (2016). Exact two-sided $(\beta, \gamma)$-STI's for the same situation are still not available in the statistical literature.

The first purpose of this paper is to introduce new tolerance intervals $[L_w(x; \mathcal{E}),\, U_w(x; \mathcal{E})]$ over $[a, b]$, called weighted simultaneous tolerance intervals (WSTI's). Under an assumption that the future $x$-values follow a specific distribution over the interval $[a, b]$, the corresponding confidence sets $C_w(y_x)$ defined as

$$C_w(y_x) = \{\, x \in [a, b] :\, L_w(x; \mathcal{E}) \le y_x \le U_w(x; \mathcal{E}) \,\} \tag{2}$$

satisfy the *property* in (1) exactly. The second purpose is to show, through the lens of the WSTI's, why the confidence sets $C_s(y_x)$ based on the STI's satisfy the *property* in (1) conservatively, and that the confidence sets $C_0(y_x)$ based on the PTI's do not guarantee the *property* in (1) in general.

The layout of this paper is as follows. In Section 2, the STI's, PTI's and the associated confidence sets $C(y_x)$ are reviewed briefly. Section 3 deals with the construction of exact $(\beta, \gamma)$-WSTI's for a polynomial regression model over a given covariate interval $[a,\, b]$. It is also shown why the exact $(\beta, \gamma)$-STI's produce conservative confidence sets $C_s(y_x)$. Section 4 provides several examples to illustrate and compare the WSTI's, STI's and PTI's. In particular, counter examples are provided to demonstrate that the confidence sets $C_0(y_x)$

based on the PTI's do not guarantee the *property* in (1). Section 5 contains some concluding remarks.

## 2 PTI's and STI's

Assume that the unknown $x$-values corresponding to all the future $y_x$'s are in a given interval $[a, b]$. For example, the true concentration of radon cannot be lower than $a = 0$ or higher than some upper threshold $b$. The $(\beta, \gamma)$-STI's $[L_s(x; \mathcal{E}),\, U_s(x; \mathcal{E})]$ over the interval $x \in [a, b]$ satisfy

$$P_{\mathcal{E}} \{\, P_{y_x} \{\, L_s(x; \mathcal{E}) < y_x < U_s(x; \mathcal{E}) | \mathcal{E} \,\} \geq \beta \;\text{ for all }\; x \in [a, b] \,\} = \gamma \tag{3}$$

where $y_x$ denotes a future $y$-value corresponding to $x$ and is independent of the training data $\mathcal{E}$, the probability $P_{y_x} \{\, \cdot \,|\mathcal{E} \,\}$ is with respect to $y_x$ and conditional on $\mathcal{E}$, and the probability $P_{\mathcal{E}} \{\, \cdot \,\}$ is with respect to $\mathcal{E}$. Then for each future $y_x$ the confidence set for the corresponding $x$, taking into consideration that $x \in [a, b]$, is defined as

$$C_s(y_x) = \{\, x \in [a, b] :\; L_s(x; \mathcal{E}) \leq y_x \leq U_s(x; \mathcal{E}) \,\}\,. \tag{4}$$

It is shown in Scheffé (1973) that these confidence sets $C_s(y_x)$ satisfy (1) conservatively, that is, the probability on the left side is at least $\gamma$. Note, however, no one has shown that confidence sets $C(y_x)$ have to been constructed via the $(\beta, \gamma)$-STI's $[L_s(x; \mathcal{E}),\, U_s(x; \mathcal{E})]$ in order to satisfy the *property* in (1) at least conservatively.

The one-sided upper $(\beta, \gamma)$-STI's have $L_s(x; \mathcal{E}) = -\infty$ in (3), and the one-sided lower $(\beta, \gamma)$-STI's have $U_s(x; \mathcal{E}) = \infty$ in (3). The confidence set $C_s(y_x)$ corresponding to the lower STI's often takes the form of an upper confidence limit, which is most relevant for the example of ATD since the company wants to monitor that the radon concentrations are

not above the safety threshold set by a government agency by using the upper confidence limits. The confidence set $C_s(y_x)$ corresponding to the upper STI's often takes the form of a lower confidence limit for the simple linear regression when covariate $x$ increases, response variable $y$ also increases.

The $(\beta, \gamma)$-pointwise tolerance intervals (PTI's) $[L_0(x; \mathcal{E}),\, U_0(x; \mathcal{E})]$ over the interval $x \in [a, b]$ satisfy

$$P_{\mathcal{E}} \{ P_{y_x} \{ L_0(x; \mathcal{E}) < y_x < U_0(x; \mathcal{E}) | \mathcal{E} \} \geq \beta \} = \gamma \quad \text{for each } x \in [a, b]. \tag{5}$$

Then for each future $y_x$ the confidence set $C_0(y_x)$ for the corresponding $x$ based on the PTI's is defined as

$$C_0(y_x) = \{ x \in [a, b] : L_0(x; \mathcal{E}) \leq y_x \leq U_0(x; \mathcal{E}) \}. \tag{6}$$

The upper $(\beta, \gamma)$-PTI's have $L_0(x; \mathcal{E}) = -\infty$ in (5), and the lower $(\beta, \gamma)$-PTI's have $U_0(x; \mathcal{E}) = \infty$ in (5). The numerical results in Mee and Eberhardt (1996) and Lee (1999) lead to the conjecture that the property in (1) is satisfied conservatively by the PTI's based confidence sets $C_0(y_x)$ in (6). It is pointed out, however, in Han et al. (2016) that this is not true in general.

# 3 WSTI's for polynomial regression

Assume that $y$ and the only covariate $x$ are related by a polynomial regression model of order $p - 1$ $(\geq 1)$:

$$y = \alpha_0 + \alpha_1 x + \cdots + \alpha_{p-1} x^{p-1} + \epsilon = \boldsymbol{x}^T \boldsymbol{\alpha} + \epsilon$$

where $\boldsymbol{x} = (1, x, \cdots, x^{p-1})^T$, $\boldsymbol{\alpha} = (\alpha_0, \cdots, \alpha_{p-1})^T$ is the vector of unknown regression coefficients, and the $\epsilon$'s are independent $N(0, \sigma^2)$ errors with unknown variance $\sigma^2 > 0$. The observed training data $\mathcal{E}$ can be represented in the usual matrix form $\boldsymbol{y} = X\boldsymbol{\alpha} + \boldsymbol{\epsilon}$, where the $n \times p$ design matrix $X$ is assumed to be of full column-rank without loss of generality. The usual estimators of $\boldsymbol{\alpha}$ and $\sigma$ are denoted by $\hat{\boldsymbol{\alpha}}$ and $\hat{\sigma}$, respectively.

Exact WSTI's of the form

$$L_w(x; \mathcal{E}) = \boldsymbol{x}^T\hat{\boldsymbol{\alpha}} - \lambda_w\hat{\sigma}\left(z_\beta + \sqrt{(p+2)\boldsymbol{x}^T(X^TX)^{-1}\boldsymbol{x}}\right) \tag{7}$$

$$U_w(x; \mathcal{E}) = \boldsymbol{x}^T\hat{\boldsymbol{\alpha}} + \lambda_w\hat{\sigma}\left(z_\beta + \sqrt{(p+2)\boldsymbol{x}^T(X^TX)^{-1}\boldsymbol{x}}\right) \tag{8}$$

will be constructed, where $\lambda_w > 0$ is a critical constant chosen so that the corresponding confidence sets $C_w(y_x)$ defined in (2) satisfy the *property* in (1), and $z_\beta$ is the $\beta$ quantile of the standard normal distribution $N(0, 1)$. Note that $L_w(x; \mathcal{E})$ and $U_w(x; \mathcal{E})$ are of the same form as the $L_s(x; \mathcal{E})$ and $U_s(x; \mathcal{E})$ in Odeh and Mee (1990), Mee et al. (1991) and Han et al. (2016) except the critical constant $\lambda_w$.

Now we assume that *the $x$-values of all the future observations actually follow a specific probability distribution on the interval $x \in [a, b]$, either discrete or continuous, with cdf $F(\cdot)$*. Under this assumption, it can be proved in a similar way as in Liu et al. (2019b, Appendix A) that, conditional on $\mathcal{E}$ (through $\hat{\boldsymbol{\alpha}}$ and $\hat{\sigma}$ only), we have

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} I_{\{x_i \in C_w(y_{x_i})\}} = \int_a^b P_{y_x}\{\, L_w(x; \mathcal{E}) \le y_x \le U_w(x; \mathcal{E}) | \mathcal{E}\,\}\, dF(x). \tag{9}$$

Hence the *property* in (1) is satisfied if the critical constant $\lambda_w$ solves the equation

$$P_{\mathcal{E}}\left\{ \int_a^b P_{y_x}\{\, L_w(x; \mathcal{E}) \le y_x \le U_w(x; \mathcal{E}) | \mathcal{E}\,\}\, dF(x) \ge \beta \right\} = \gamma. \tag{10}$$

The rudiments of this idea have appeared in Chvosteková (2019) for calibration based on the uniform distribution of $x$ on $[a, b]$, and in Liu et al. (2019b) for classification.

From the WSTI's, one can see why the confidence sets $C_s(y_x)$ based on the STI's satisfy the *property* in (1) conservatively. Note that equation (9) also holds with $C_w(y_{x_i})$ replaced by $C_s(y_{x_i})$ from equation (4) and $[L_w(x;\mathcal{E}),\, U_w(x;\mathcal{E})]$ replaced by $[L_s(x;\mathcal{E}),\, U_s(x;\mathcal{E})]$. It is also clear that, for any $F(\cdot)$,

$$\int_a^b P_{y_x}\left\{\, L_s(x;\mathcal{E}) \le y_x \le U_s(x;\mathcal{E}) | \mathcal{E}\,\right\} dF(x) \;\ge\; \min_{x\in[a,\,b]} P_{y_x}\left\{\, L_s(x;\mathcal{E}) \le y_x \le U_s(x;\mathcal{E}) | \mathcal{E}\,\right\}\,.$$

It follows therefore immediately that

$$\begin{aligned}
& P_{\mathcal{E}}\left\{ \liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} I_{\left\{x_i \in C_s(y_{x_i})\right\}} \;\ge\; \beta \right\} \\
= \;& P_{\mathcal{E}}\left\{ \int_a^b P_{y_x}\left\{\, L_s(x;\mathcal{E}) \le y_x \le U_s(x;\mathcal{E}) | \mathcal{E}\,\right\} dF(x) \;\ge \beta \right\} \\
\ge \;& P_{\mathcal{E}}\left\{ \min_{x\in[a,\,b]} P_{y_x}\left\{\, L_s(x;\mathcal{E}) \le y_x \le U_s(x;\mathcal{E}) | \mathcal{E}\,\right\} \;\ge \beta \right\} \;=\; \gamma
\end{aligned}$$

where the last equality follows directly from the construction of STI's in (3).

One can also see why the confidence sets $C_0(y_x)$ based on the PTI's may not guarantee the *property* in (1). Note that equation (9) also holds with $C_w(y_{x_i})$ replaced by $C_0(y_{x_i})$ and $[L_w(x;\mathcal{E}),\, U_w(x;\mathcal{E})]$ replaced by $[L_0(x;\mathcal{E}),\, U_0(x;\mathcal{E})]$. Hence if the interval $[L_0(x;\mathcal{E}),\, U_0(x;\mathcal{E})]$ is contained strictly inside the interval $[L_w(x;\mathcal{E}),\, U_w(x;\mathcal{E})]$ for each $x \in [a, b]$ and $\mathcal{E}$ such that

$$\int_a^b P_{y_x}\left\{\, L_0(x;\mathcal{E}) \le y_x \le U_0(x;\mathcal{E}) | \mathcal{E}\,\right\} dF(x) \;<\; \int_a^b P_{y_x}\left\{\, L_w(x;\mathcal{E}) \le y_x \le U_w(x;\mathcal{E}) | \mathcal{E}\,\right\} dF(x)$$

then

$$
\begin{aligned}
& P_{\mathcal{E}}\left\{ \liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} I_{\{x_i\in C_0(y_{x_i})\}} \ \geq \ \beta \right\} \\
= \ & P_{\mathcal{E}}\left\{ \int_a^b P_{y_x}\left\{ L_0(x;\mathcal{E}) \leq y_x \leq U_0(x;\mathcal{E}) | \mathcal{E} \right\} dF(x) \ \geq \beta \right\} \\
< \ & P_{\mathcal{E}}\left\{ \int_a^b P_{y_x}\left\{ L_w(x;\mathcal{E}) \leq y_x \leq U_w(x;\mathcal{E}) | \mathcal{E} \right\} dF(x) \ \geq \beta \right\} = \gamma
\end{aligned}
$$

where the last equality follows directly from the construction of WSTI's in (10). Examples 4.1 and 4.3 in Section 4 below are two examples in which the PTI's are strictly narrower than the WSTI's for each $x \in [a,b]$ and $\mathcal{E}$ and so the PTI's based confidence sets do not guarantee the *property* in (1).

From here on, we focus on the one-sided case since exact two-sided $(\beta,\gamma)$-STI's on any interval $[a,b]$ are still not available yet in the statistical literature as pointed out in the Introduction section. First we consider the computation of $\lambda_w$ from equation (10) for the one-sided upper WSTI's with $L_w(x;\mathcal{E}) = -\infty$, while the one-sided lower WSTI's with $U_w(x;\mathcal{E}) = \infty$ uses the same critical constant $\lambda_w$. In this case, we have

$$
\begin{aligned}
& P_{\mathcal{E}}\left\{ \int_a^b P_{y_x}\left\{ L_w(x;\mathcal{E}) \leq y_x \leq U_w(x;\mathcal{E}) | \mathcal{E} \right\} dF(x) \ \geq \beta \right\} \\
= \ & P_{\mathcal{E}}\left\{ \int_a^b P_{y_x}\left\{ y_x \leq \boldsymbol{x}^T\hat{\boldsymbol{\alpha}} + \lambda_w\hat{\sigma}\left( z_\beta + \sqrt{(p+2)\boldsymbol{x}^T(X^TX)^{-1}\boldsymbol{x}} \right) | \mathcal{E} \right\} dF(x) \ \geq \beta \right\} \\
= \ & P_{\hat{\boldsymbol{\alpha}},\hat{\sigma}}\left\{ \int_a^b \Phi\left\{ \boldsymbol{x}^T(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})/\sigma + \lambda_w(\hat{\sigma}/\sigma)\left( z_\beta + \sqrt{(p+2)\boldsymbol{x}^T(X^TX)^{-1}\boldsymbol{x}} \right) \right\} dF(x) \ \geq \beta \right\} \\
= \ & P_{\boldsymbol{Z},u}\left\{ \int_a^b \Phi\left\{ \boldsymbol{x}^T\boldsymbol{Z} + \lambda_w u\left( z_\beta + \sqrt{(p+2)\boldsymbol{x}^T(X^TX)^{-1}\boldsymbol{x}} \right) \right\} dF(x) \ \geq \beta \right\} \qquad (11)
\end{aligned}
$$

where $\boldsymbol{Z} = (\boldsymbol{\alpha}-\hat{\boldsymbol{\alpha}})/\sigma \sim \boldsymbol{N}(\boldsymbol{0},(X^TX)^{-1})$, $u = \hat{\sigma}/\sigma \sim \sqrt{\chi^2_\nu/\nu}$ with $\nu = n-p$, and $\boldsymbol{Z}$ and $u$ are independent.

It is clear that the probability in (11) has nothing to do with the unknown parameters

$\boldsymbol{\alpha}$ and $\sigma$. Furthermore, $\lambda_w$ solving equation (10) depends only on $p, \nu, (X^TX)^{-1}, \beta, \gamma$, the interval $[a, b]$ and the distribution $F(\cdot)$.

For a general $p \geq 2$ the following simple simulation-based method for finding $\lambda_w$ fast and accurately is used in our `R` code, similar to what is used in Liu et al. (2019b). From the expression in (11), in the $s$-th repeat of simulation, $s = 1, \ldots, S$, generate independent $\boldsymbol{Z}_s \sim \boldsymbol{N}(\boldsymbol{0}, (X^TX)^{-1})$ and $u_s \sim \sqrt{\chi^2_\nu/\nu}$, and find the $\lambda_w = \lambda_{w,s}$ so that

$$\int_a^b \Phi\left\{ \boldsymbol{x}^T\boldsymbol{Z}_s + \lambda_{w,s}u_s\left(z_\beta + \sqrt{(p+2)\boldsymbol{x}^T(X^TX)^{-1}\boldsymbol{x}}\right)\right\} dF(x) \;= \beta. \tag{12}$$

Repeat this $S$ times to get $\lambda_{w,1}, \ldots, \lambda_{w,S}$, order these as $\lambda_{w,[1]} \leq \ldots \leq \lambda_{w,[S]}$, and use $\lambda_{w,[\gamma S]}$ as $\lambda_w$. It is well known that $\lambda_{w,[\gamma S]}$ converges to the required critical constant $\lambda_w$ with probability one as $S \to \infty$ (cf. Serfling, 1980). This approach of using a sample quantile to approximate the population quantile has been used successfully in solving many difficult problems; see, for example, Edwards and Berry (1987) and Liu et al. (2004, 2005).

To find the $\lambda_{w,s}$ that solves the equation in (12) for each $s$, we use numerical quadrature, e.g. the `R` function `integrate`, to compute the expression on the left side of (12) for each given value of $\lambda_{w,s} > 0$, which is clearly strictly increasing in $\lambda_{w,s}$. We then find $\lambda_{w,s}$ by using a suitable search algorithm. It is noteworthy that only one-dimensional numerical quadrature is involved for a general $p(\geq 2)$.

Each $\lambda_w$ in Examples 4.1 - 4.3 of Section 4 below was computed using $S = 1,000,000$ simulations and took about 2000 seconds on an ordinary Window's PC (Core(TM2)Due CPU P8400@2.26GHz). These critical constants are expected to be accurate to at least two decimal places from our experiments with different random seeds.

For the upper $(\beta, \gamma)$-STI's $U_s(x; \mathcal{E})$ of the form

$$U_s(x; \mathcal{E}) = \boldsymbol{x}^T \hat{\boldsymbol{\alpha}} + \lambda_s \hat{\sigma} \left[ z_\beta + \sqrt{(p+2)\boldsymbol{x}^T (X^T X)^{-1} \boldsymbol{x}} \right] , \tag{13}$$

a simulation based method for computing the critical constant $\lambda_s$ that solves the equation in (3) with $L_s(x; \mathcal{E}) = -\infty$ is given in Han et al. (2016) (coded in Matlab), and re-coded in `R`. Each $\lambda_s$ in Examples 4.1 - 4.3 of Section 4 was computed using $S = 1,000,000$ simulations and took about 150 seconds on the same Window's PC. Again these critical constants are expected to be accurate to at least two decimal places from our experiments with different random seeds. For example, when $\beta = 0.95$ and $\gamma = 0.99$, the critical value $\lambda_s = 1.2557$ for Example 4.2 below given in Han et al. (2016) using Matlab is computed to be 1.2552 using our `R` code.

For the upper $(\beta, \gamma)$-PTI's $U_0(x; \mathcal{E})$ of the form

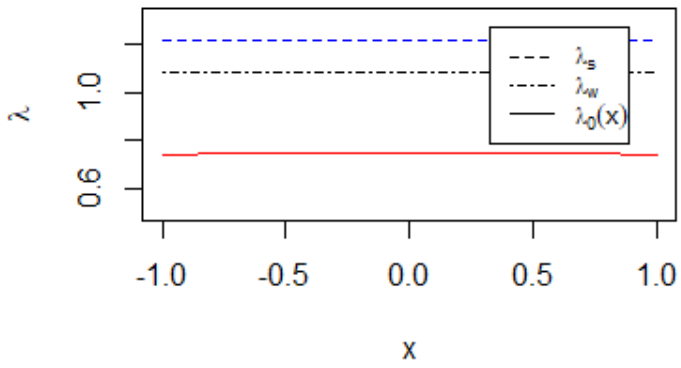
$$U_0(x; \mathcal{E}) = \boldsymbol{x}^T \hat{\boldsymbol{\alpha}} + \lambda_0 \hat{\sigma} \left[ z_\beta + \sqrt{(p+2)\boldsymbol{x}^T (X^T X)^{-1} \boldsymbol{x}} \right] , \tag{14}$$

it follows directly from the equation in (5) with $L_0(x; \mathcal{E}) = -\infty$, after a few lines of algebraic manipulation, that $\lambda_0$ is given by

$$\lambda_0 = \lambda_0(x) = \frac{\sqrt{\boldsymbol{x}^T (X^T X)^{-1} \boldsymbol{x}}}{\left( z_\beta + \sqrt{(p+2)\boldsymbol{x}^T (X^T X)^{-1} \boldsymbol{x}} \right)} t_{\gamma, \nu, z_\beta / \sqrt{\boldsymbol{x}^T (X^T X)^{-1} \boldsymbol{x}}}$$

where $t_{\gamma,\nu,\delta}$ denotes the $\gamma$-quantile of a $t$ distribution with degrees of freedom $\nu$ and non-centrality parameter $\delta$ and can be computed directly by using the `R` function `qt`. It is noteworthy that the $\lambda_0$ of the PTI's varies with $x \in [a, b]$, while the $\lambda_w$ of the WSTI's and $\lambda_s$ of the STI's are constants, and the PTI's, STI's and WSTI's differ only in their critical constants.

Figure 1: $\lambda_0(x)$ *of PTI's (solid line),* $\lambda_w = 1.085$ *of the WSTI's (dash-dot line), and* $\lambda_s = 1.214$ *of the STI's (dash line)*




# 4 Examples

## 4.1 The PTI's based confidence sets do not guarantee the property in (1)

In this example it is shown that the confidence sets $C_0(y_x)$ in (6), based on the PTI's, do not guarantee the *property* in (1).

Assume that $y$ and $x$ follow a simple linear regression model (i.e., $p = 2$), the covariate interval of interest is $[a, b] = [-1, 1]$, the training dataset contains 11 observations with the $x$-values given by equally-spaced $-1, -0.8, \cdots, 0.8, 1$ (and so $\nu = 9$), $\beta = 0.95$ and $\gamma = 0.90$. Hence $\lambda_0(x)$ in (14) can be computed and is plotted by the solid line in Figure 1 over $x \in [a, b] = [-1, 1]$. As expected, $\lambda_0(x)$ varies with $x$, but only slightly around 0.75 in this example. For STIs, the value of $\lambda_s$ in (13) is computed to be 1.215. The value of $\lambda_s$ is also computed using ten different ransom seeds, and the standard deviation of these ten $\lambda_s$'s is 0.00068. Hence $\lambda_s = 1.215$ is expected to be accurate to at least two decimal places according to the 3-$\sigma$ rule. The $\lambda_s$ is plotted by the dash line in Figure 1. As expected, $\lambda_s$

is strictly larger than $\lambda_0(x)$ for any $x \in [a, b]$.

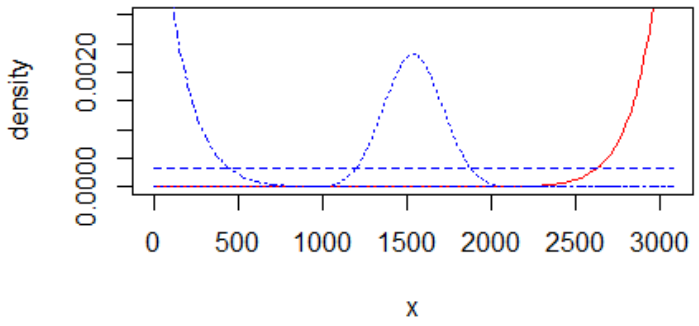
Now assume the $x$-values of all the future observations actually follow a uniform distribution on the interval $[a, b] = [-1, 1]$. The $\lambda_w$ of the WSTI's is computed to be 1.085. The standard deviation of the ten $\lambda_w$-values based on ten random seeds is 0.00047 and so $\lambda_w = 1.085$ is expected to be accurate to the two decimal place at least. Then the $\lambda_w$ is plotted by the dash-dot line in Figure 1. Note that the confidence sets $C_w(y_x)$ in (2), based on the WSTI's, satisfy the *property* in (1) exactly. Since $\lambda_0(x)$ is substantially smaller than $\lambda_w$ for each $x \in [a, b]$, the upper PTI's $U_0(x; \mathcal{E})$ are also substantially smaller than the upper WSTI's $U_w(x; \mathcal{E})$ for each $x \in [a, b]$. Consequently the PTI's based confidence sets will not guarantee the *property* in (1). This demonstrates that the confidence sets $C_0(y_x)$ based on the PTI's do not satisfy the *property* in (1) in general and so should not be used.

Finally it is noteworthy that the assumed $x$-values in the training data of this example is not at all extreme since equally spaced $x$-values on any interval $[a, b]$ can be transformed into equally spaced new $x$-values on the interval $[-1, 1]$ under a simple linear transformation of the original covariate. Also the regression coefficients $\boldsymbol{\alpha}$ and error variance $\sigma^2$ do not need to be specified.

## 4.2 Radon Example

Mee and Eberhardt (1996, Table 3) provides data from a calibration experiment of ATD's which are used to measure indoor concentrations of radon. In the experiment, the ATD's were exposed to known levels of radon $x$ in a laboratory. The response variable $y$ is an optical count of number of damage tracks, caused by alpha radioactive decays, over a specific area of the film. After the usual model diagnosis, the $n = 40$ observations are fitted by a simple linear regression model in Mee and Eberhardt (1996), with the fitted least squares line $y = 124.4 + .789x$, $\hat{\sigma} = 41.26$ and $R^2 = 0.93$.

Figure 2: *The four beta density functions on the interval* $[a, b]$*:* $B(20, 1)$ *– solid line;* $B(1, 1)$ *(i.e., uniform distribution) – dash line;* $B(40, 40)$ *– dot line;* $B(1, 20)$ *– dash-dot line.*




In order to construct upper confidence bounds on the level of radon exposures $x$ corresponding to future observed ATD readings $y$, lower STI's are required:

$$L_s(x; \mathcal{E}) = \boldsymbol{x}^T \hat{\boldsymbol{\alpha}} - \lambda_s \hat{\sigma} \left[ z_\beta + \sqrt{(p+2)\boldsymbol{x}^T (X^T X)^{-1} \boldsymbol{x}} \right] \quad \text{for all } x \in [a, b]. \tag{15}$$

Set $\beta = 0.95$, $\gamma = 0.99$, $[a, b] = [0, 3074]$ as in Mee and Eberhardt (1996), with $p = 2$ and $\nu = 38$, the critical constant $\lambda_s$ is computed to be 1.255 using our `R` code. The upper confidence bounds $C_s(y_x)$ can then be constructed by using (4), with $U_s(x; \mathcal{E}) = \infty$; see Han et al. (2016, Fig. 1) for how to get $C_s(y_x)$ from $L_s(x; \mathcal{E})$.

To what extent the STI's based upper confidence bounds $C_s(y_x)$ above could be improved if we know that the $x$-values of all the future observations actually follow a particular distribution $F(\cdot)$ on the interval $[a, b]$? For this, we have tried four beta distributions on $[a, b]$ with the two shape parameters given respectively by: $(20, 1)$, $(1, 1)$, $(40, 40)$, $(1, 20)$. The probability density functions of these four beta distributions are plotted in Figure 2. Here, $B(20, 1)$ represents the scenario that the $x$-values of most future observations are close to the upper limit $b$ of the interval $[a, b]$; for $B(1, 1)$, the $x$-values of future observations are assumed to be uniformly distributed on the interval $[a, b]$; for $B(40, 40)$, the $x$-values are

mostly scattered around the middle of the interval $[a, b]$; and for $B(1, 20)$, the $x$-values of most future observations are close to the lower limit $a$ of the interval $[a, b]$.

With each of the four beta distributions as $F(\cdot)$, the critical constant $\lambda_w$ of the lower WSTI's in (7) can be computed using our `R` code. These are given by 1.173, 1.176, 1.197, 1.203 for the beta distributions $B(20, 1)$, $B(1, 1)$, $B(40, 40)$, $B(1, 20)$, respectively. The upper confidence bounds $C_w(y_x)$ in (2) can then be constructed, with $U_w(x; \mathcal{E}) = \infty$. As expected, $\lambda_w$ is strictly smaller than $\lambda_s = 1.255$ and hence the upper confidence bounds $C_w(y_x)$ are slightly smaller than the upper confidence bounds $C_s(y_x)$.

### 4.3 Cadmium Example

As mentioned in Section 1, $Cd$ concentration in drinking water is important for environmental health, and graphite furnace atomic absorption spectroscopy (GFAAS) is an analytical technique for the determination of $Cd$ concentrations in different types of samples. In this example, the variation of (peak) absorbency $y$ (in mm) with $Cd$ concentration $x$ (parts per billion) is established by atomizing samples of known concentrations and the obtained calibration curve from the training data is used to infer the concentrations of future observed absorbances.

Lundberg and De Maré (1980, Table 1) provides data from a calibration experiment of GFAAS which contain $n = 21$ pairs of observations at four different concentrations $x = 0, 5, 15$ and 20. After the usual model diagnosis, a quadratic regression model is recommended in Lundberg and De Maré (1980) which fits the data very well with $R^2 = 0.999$.

Now suppose that one wants to construct lower confidence bounds on $x$ for future observed $y$'s. For this one can use the upper STI's in (13). For $\beta = 0.95$, $\gamma = 0.99$, $[a, b] = [0, 20]$, $p = 3$ and $\nu = 18$, the critical constant $\lambda_s$ is computed to be 1.422 using our `R` code. The lower confidence bounds $C_s(y_x)$ in (4) can then be constructed, with $L_s(x; \mathcal{E}) = -\infty$; see

Han et al. (2016, Fig. 2) for how to get $C_s(y_x)$ from $U_s(x;\mathcal{E})$.

To get some idea on the extent to which the STI's based lower confidence bounds $C_s(y_x)$ above could be improved if we know that the $x$-values of all the future observations actually follow a particular distribution $F(\cdot)$ on the interval $[a,b]$, we have also tried the four beta distributions $B(20,1)$, $B(1,1)$, $B(40,40)$, $B(1,20)$ on the interval $[a,b]=[0,20]$ in this example.

With each of the four beta distributions as $F(\cdot)$, the critical constant $\lambda_w$ of the upper WSTI's in (8) can be computed using our `R` code, and these are given by 1.265, 1.229, 1.247 and 1.258 for the beta distributions $B(20,1)$, $B(1,1)$, $B(40,40)$, $B(1,20)$, respectively. The lower confidence bounds $C_w(y_x)$ in (2) can then be constructed, with $L_w(x;\mathcal{E})=-\infty$. As expected, each $\lambda_w$ is strictly smaller than $\lambda_s=1.422$ and hence the lower confidence bounds $C_w(y_x)$ are larger than the lower confidence bounds $C_s(y_x)$.

It is noteworthy that the value of $\lambda_0(x)$ for this example is about 1.14 for any $x\in[a,b]$, which is smaller than $\lambda_w$ when $F(\cdot)$ is any one of the four beta distributions. Hence the lower confidence bounds $C_0(y_x)$ based on the PTI's do not guarantee the *property* in (1) when the $x$-values of the future observations follow any one of the four beta distributions.

# 5 Conclusion

Statistical calibration using regression is a useful statistical tool with many applications. For inference related to infinitely many future $y$-values, there is a clear consensus in the statistical literature that the confidence sets constructed should satisfy the *property* in (1) at least conservatively. While it is well-known that the confidence sets based on the STI's satisfy this *property* conservatively, it is highly desirable to construct confidence sets that satisfy this *property* exactly. Also, there is a misconception that the confidence sets based

on the PTI's also guarantee the *property* in (1).

In this paper, the WSTI's are constructed so that the confidence sets based on the WSTI's satisfy the *property* in (1) exactly if the future observations have the $x$-values distributed according to a known specific distribution $F(\cdot)$ on the covariate interval $[a, b]$. Through the WSTI's, convincing counter examples are provided to demonstrate that the confidence sets based on the PTI's do not guarantee the *property* in (1) in general and so should not be used.

If the distribution $F(\cdot)$ exists and is known then the confidence sets based on the WSTI's should clearly be used. But how to solicit the distribution $F(\cdot)$ warrants further research. It can be envisaged that in many real applications, it is difficult to solicit the distribution $F(\cdot)$. In such a case, one can fall back on the confidence sets based on the STI's, which guarantee the *property* in (1). Hence STI's still play a key role in calibration for infinitely many future observations. Exact two-sided STI's are still not available in the statistical literature even for univariate polynomial regression of order $p-1 \geq 2$ over an arbitrary interval $[a, b]$. For many interesting and challenging calibration problems involving multivariate polynomial regression (cf. Mathew et al., 1997 and 1998), only PTI's are available. All these warrant further research.

We have also compared the WSTI's proposed in this paper to the WSTI's with the form of prediction intervals, that is, $\boldsymbol{x}^T\hat{\boldsymbol{\alpha}} \pm \lambda_w^*\hat{\sigma}\sqrt{1+\boldsymbol{x}^T(X^TX)^{-1}\boldsymbol{x}}$, though the details are not reported here due to page limit. Based on our results, the average shifts of one-sided WSTI's proposed in this paper is only about 0.60% larger than the average shifts of WSTI's in the form of prediction intervals. Hence, the difference between these two forms is very small. In addition, we found that, the WSTI's proposed in our paper are narrower than the WSTI's in the form of prediction intervals when covariate $x$ is close to the mean $\bar{x}$; while the WSTI's in the form of prediction intervals are narrower when $x$ is near two ends

of the covariate range $[a, b]$.

**Acknowledgements**: We are indebted to Professor Robert Mee for a stimulating discussion which motivated this work.

## Reference

Aitchison, T.C. (1982). Discussion of Brown's paper. *JRSS(B)*, 44(3), 309-310.

Brown, P.J. (1982). Multivariate Calibration. *JRSS(B)*, 44(3), 287-321.

Brown, P.J. (1993). *Measurement, Regression and Calibration*. New York: Oxford University Press.

Chvosteková, M. (2019). Multiple use confidence intervals for a univariate statistical calibration. *Measurement Science Review*, 19(6), 264-270.

Dudney, C. S., Wilson, D. L., Bertini, H., Gammage, R. B., Otten, J. A. (1995). On the reduction of error in alpha track detector measurements of indoor 222Rn. *Health physics*, 69(4), 501-507.

Edwards, D. and Berry, J.J. (1987). The efficiency of simulation-based multiple comparisons. *Biometrics*, 43, 913-928.

Eisenhart, C. (1939). The interpretation of certain regression methods and their use in biological and industrial research. *The Annals of Mathematical Statistics*, 10, 162-186.

Han, Y. (2014). *Simultaneous Confidence Bands and Simultaneous Tolerance Bands in Linear Regression*. PhD thesis, School of Mathematical Sciences, University of Southampton, UK.

Han, Y., Liu, W., Bretz, F., Wan, F., Yang, P. (2016). Statistical calibration and exact

one-sided simultaneous tolerance intervals for polynomial regression. *Journal of Statistical Planning and Inference*, 168, 90–96.

Ingle, Jr, J. D., Crouch, S. R. (1988). *Spectrochemical analysis*. New Jersey: Prentice Hall.

Krishnamoorthy, K. and Mathew, T. (2009). *Statistical Tolerance Regions: Theory, Applications and Computation*. New York: Wiley.

Krishnamoorthy, K., Kulkarni, P.M. and Mathew, T. (2001). Multiple use one-sided hypotheses testing in univariate linear calibration. *Journal of Statistical Planning and Inference*, 93, 211-223.

Lee, Y. (1999). *Tolerance Regions in Univariate and Multivariate Linear Models*. PhD thesis, Dept of Mathematics and Statistics, University of Maryland Baltimore County, USA.

Lieberman, G.J. and Miller, R.G., Jr. (1963). Simultaneous tolerance intervals in regression. *Biometrika*, 50, 155-168.

Lieberman, G.J., Miller, R.G., Jr. and Hamilton, M.A. (1967). Simultaneous discrimination intervals in regression. *Biometrika*, 54, 133-145 (Correction: 58, p.687).

Limam, M.M.T. and Thomas, D.R. (1988). Simultaneous tolerance intervals for the linear regression model. *Journal of the American Statistical Association*, 83, 801-804.

Liu, W., Jamshidian, M. and Zhang, Y. (2004). Multiple comparison of several linear regression models. *Journal of the American Statistical Association*, 99, 395-403.

Liu, W., Jamshidian, M., Zhang, Y. and J. Donnelly (2005). Simulation-based simultaneous confidence bands in multiple linear regression with predictor variables constrained in intervals. *Journal of Computational and Graphical Statistics*, 14(2), 459-484.

Liu, W., Han, Y., Bretz, F., Wan, F., Yang, P. (2016). Counting by weighing: know your numbers with confidence. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65, 641–648.

Liu, W., Bretz, F., Srimaneekarn, N., Peng, J. and Hayter, A.J. (2019a). Confidence sets for statistical classification. *Stats*, 2(3), 332-346; https://doi.org/10.3390/stats2030024

Liu, W., Bretz, F. and Hayter, A.J. (2019b). Confidence sets for statistical classification (II): exact confidence sets. *Stats*, 2(4), 439-446; https://doi.org/10.3390/stats2040030

Lundberg, E. and De Maré, J. (1980). Interval estimates in the spectroscopy calibration problem. *Scandinavia Journal of Statistics*, 7, 40-42.

Mathew, T., Sharma, M.K. and Nordstrom, K. (1998). Tolerance regions and multiple-use confidence regions in multivariate calibration. *The Annals of Statistics*, 26, 1989-2013.

Mathew, T. and Zha, W. (1997). Multiple use confidence regions in multivariate calibration. *J Amer. Statist. Assoc.* 92, 1141-1150.

Mee, R.W. and Eberhardt, K.R. (1996). A Comparison of uncertainty criteria for calibration. *Technometrics*, 38, 221-229.

Mee, R.W., Eberhardt, K.R. and Reeve, C.P. (1991). Calibration and simultaneous tolerance intervals for regression. *Technometrics*, 33, 211-219.

Odeh, R.E. and Mee, R.W. (1990). One-sided simultaneous tolerance limit for regression. *Comm. Statist.– Simulation Comput.*, 19, 663-680.

Osborne, C. (1991). Statistical calibration: a review. *Internat. Statist. Rev.*, 59, 309-336.

Parker, P.A., Vining, G.G., Wilson, S.R., Szarka, J.L. and Johnson, N.G. (2010). The prediction properties of classical and inverse regression for the simple linear calibration

problem. *Journal of Quality Technology*, 42(4), 332-347.

Scheffé, H. (1973). A statistical theory of calibration. *Annals of Statistics*, 1, 1-37.

Seber, G.A.F. (1977). *Linear Regression Analysis*. Wiley.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley.

Smith, R.L. and Corbett, M. (1987). Measure marathon courses: an application of statistical calibration theory. *Applied Statistics*, 36, 283-295.

Stephens, G. A. (1920). Cadmium Poisoning And Plumbism. *British Medical Journal*, 2(3118), 530.