Implementation of Errors-in-Variables Regression and Monte Carlo Uncertainty Evaluation into Force Calibration Reporting at NIST

Thomas W. Bartel

NIST

National Institute of Standards and Technology

U.S. Department of Commerce

NIST Technical Note 1942

Implementation of Errors-in-Variables Regression and Monte Carlo Uncertainty Evaluation into Force Calibration Reporting at NIST

Thomas W. Bartel
Quantum Measurement Division
Physical Measurement Laboratory

November 2016

Abstract

The National Institute of Standards and Technology (NIST) is commencing the implementation of certain statistical procedures, namely errors-in-variables regression and Monte Carlo uncertainty evaluation, into the data analysis and reporting for the force calibration service provided by the NIST Physical Measurement Laboratory (PML). This document is intended to serve as an overview of these procedures to the users of NIST's force calibration service, by presenting computational details of the analyses and introducing the corresponding format for the force calibration report.

Key Words

Errors-in-variables; Monte Carlo method; Force calibration; Ordinary least squares; Regression computation; Uncertainty assessment; Uncertainty bands

1. Introduction

The calibration service for force measuring instruments provided by NIST follows procedures specified by documentary standards such as ASTM E 74-13a [1] and ISO 376:2011(E) [2]. The traditional analysis of force calibration data at NIST, and the associated uncertainty, have been described previously [3]. Implicit in this analysis, which employs ordinary least squares regression to fit the calibration data to a polynomial equation, is the assumption that the uncertainties in the calibrated forces applied to the transducer are small relative to other uncertainties contributed by the transducer and indicating system. This assumption has been challenged by recent refinements in transducer technology; thus, for modern transducers designed for high precision force metrology, it is no longer appropriate to ignore the uncertainties in the applied forces during the regression analysis.

A more appropriate regression method, termed errors-in-variables (EIV), together with a Monte Carlo-based uncertainty analysis that serves as a tractable companion for EIV regression, has been explored recently at NIST [4]. This exploration has motivated an effort to incorporate, in a practical manner, these new statistical methods into NIST's analysis and reporting for force calibrations. A workable procedure is now at hand for conducting the necessary calculations on a routine basis and incorporating the results into the official report for NIST force calibrations.

The statistical foundation and justification for EIV regression, accompanied by Monte Carlo uncertainty analysis, has been well described for application to force calibration data [4], and need not be repeated here. It is necessary, however, to introduce the force community to NIST's impending implementation of these procedures in the routine processing of force calibrations. Because of the adherence of NIST's traditional analysis to documentary standards, the new analysis methods are not being forwarded as replacements to the current methods; rather, the existing results will still be regarded as the "official" results of calibration and the new results will be reported in parallel in the form of a report supplement. For the near future, this supplement will emphasize a comparison between the traditional and new methods for the particular transducer calibration being reported. It is anticipated that, as the force community becomes accustomed to obtaining this comparison for ongoing calibrations, the relevant documentary standards committees will consider approval of the new methods as replacements.

2. Computational Details

The statistical algorithms, which have been described previously [4] for conducting the EIV and Monte Carlo evaluations, are implemented in the R environment for statistical computing and graphics [5]. The traditional force calibration analysis and reporting at NIST is conducted with the worksheet software operating under Microsoft Excel. Excel is also used

for recently developed computations that estimate the true forces being applied to the transducer, by accounting for the vertical gradient of the acceleration due to gravity over a set of deadweights, and by calculating the air density for each applied weight using real-time measurements of atmospheric pressure, temperature, and humidity. These Excel workbooks have been adapted to create an input worksheet for access by the R-code, which then is executed separately. When finished, the R-code leaves its results in an output worksheet for subsequent use by Excel to generate the EIV report supplement.

The Monte Carlo evaluations involve computationally intensive R-code algorithms performing several thousand uncertainty simulations, and can require twenty minutes to execute on a moderate Windows desktop computer containing a 6-core processor. Since these simulations can be conducted in parallel, one way to reduce this execution time is to make use of a virtual computing environment to define an effective processor with a much greater number of cores, enabling massively parallel processing. The R system has proved to work well with such virtual computing, such as that offered by Amazon Web Services under the name High Performance Computing[1].

3. Report Supplement for EIV and Monte Carlo Methods

This section highlights the elements of the revision to the Report of Calibration currently generated for each force transducer submitted to NIST for calibration. The revision takes the form of a supplement which is attached to the traditional, currently official report. Since the official report tabulates all of the transducer response data to the applied calibration forces, as well as the deviations of each data point from the calibration function derived by means of ordinary least-squares regression, the new supplement does not repeat this information. Instead, the supplement emphasizes a comparison between the results of the ordinary least-squares (OLS) and errors-in-variables (EIV) regressions for the calibration function. In addition, it presents a graphical comparison between the uncertainty interval obtained from the traditional uncertainty analysis accompanying the OLS regression (as detailed in [3]) and the uncertainty bands yielded by the Monte Carlo evaluation associated with the EIV regression.

The supplement first presents, in tabular form, the polynomial coefficients for the calibration function derived from OLS regression (which constitute the currently official result from the Report of Calibration) and the corresponding coefficients derived from EIV regression for a polynomial of the same order. The tabulation also lists the transducer responses predicted

---

[1] Certain commercial equipment, instruments, or materials are identified in this article in order to describe the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

from the calibration function by OLS regression with the responses predicted by means of EIV regression, for each nominal applied force (where the term "nominal" relates to the target forces that the deadweight masses were adjusted to achieve during the mass determination phase of the machine construction). The differences between the responses predicted by the OLS and EIV regressions are also included.

An example of this tabulation is shown in Figure 1, for a force transducer having a capacity of 266.89 kN (60 klbf), which exhibits exceptional control of its sensitivity to orientation within the deadweight force standard machine. The regressions are carried out for a third-order polynomial, and the standard deviation of the residuals of the measured responses from the predicted calibration functions are shown (having relative values of 8.54×10^{-6} for OLS and 9.14×10^{-6} for EIV, where these values are relative to the transducer response at the maximum applied force).

EIV - OLS Comparison Table for Compression Mode				
Nominal Force: F (N)	OLS predicted response: $R_1 = f_1(F)$	EIV predicted response: $R_2 = f_2(F)$	$R_2 - R_1$ (response units)	$R_2 - R_1$ relative to response at max. force (0.0001 %)
13344.66	0.201442	0.201444	0.000002	0.374
26689.33	0.402883	0.402885	0.000001	0.324
53378.66	0.805851	0.805852	0.000001	0.227
80067.99	1.208906	1.208906	0.000000	0.080
106757.32	1.612020	1.612019	-0.000001	-0.176
133446.65	2.015162	2.015160	-0.000002	-0.604
160135.98	2.418304	2.418299	-0.000005	-1.264
186825.31	2.821414	2.821405	-0.000009	-2.219
213514.64	3.224465	3.224450	-0.000014	-3.530
240203.97	3.627425	3.627403	-0.000021	-5.257
266893.30	4.030265	4.030235	-0.000030	-7.463

The coefficients of the following equation, determined by the OLS and EIV analyses, are tabulated below. The units for force (N) and response are the same as shown in the table above. The standard deviation is computed from the difference between the measured responses and the fitted equation.

Response = A + B(force) + C(force)^2 + D(force)^3

Coefficients	OLS	EIV
A	3.394627E-05	3.569805E-05
B	1.509145E-05	1.509143E-05
C	1.035176E-13	1.037273E-13
D	-2.603816E-19	-2.625509E-19

3.440554E-05 = OLS standard deviation of residuals in response units (8.54 x 0.0001 %)

3.683544E-05 = EIV standard deviation of residuals in response units (9.14 x 0.0001 %)

Figure 1. EIV and OLS results compared in tabular form

Whereas Figure 1 lists the differences between the transducer responses predicted by the EIV and OLS regressions for a set of discrete force points, the difference between the two polynomials is represented by a smooth curve. Figure 2 shows this curve for the example of Figure 1. Such a curve is provided in the EIV supplement to accompany the report for force transducer calibrations conducted at NIST. The scale for each axis in Figure 2 is given in relative units –the relative applied force in percent of the maximum calibration force applied to the transducer, and the relative difference between the predicted transducer responses from the EIV and OLS regressions expressed as 10^{-6} of the response at maximum applied force.

For the example shown here, the maximum response difference has a relative value of about -7.5 x 10^{-6} at the upper end of the force range. While the maximum response difference can occur elsewhere within the range, it has been seen to be within 10 x 10^{-6} for data sets currently examined.

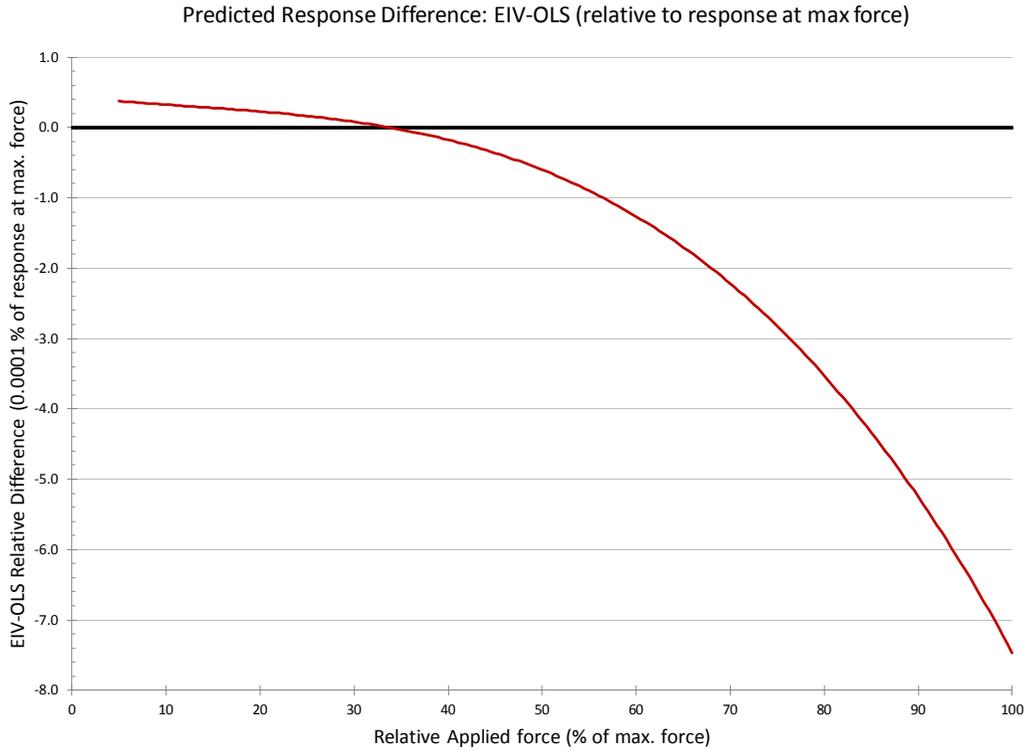Predicted Response Difference: EIV-OLS (relative to response at max force)

Figure 2. Difference between EIV and OLS calibration functions in graphical form

Figure 3 presents the deviations, or residuals, between the measured response at every applied force point and the corresponding responses predicted from the EIV and OLS regressions, for the force calibration example shown in Figure 1. Figure 3, which is included in the EIV supplement, is analogous to the deviations plot for OLS regression which appears in the Report of Calibration, and thus is familiar to the customers who receive these reports. The axis scales are given in the same relative values described for Figure 2. The OLS residuals (open circles in Figure 3) represent the same values appearing in the deviations plot of the report, but without the connecting lines joining the points for each run – where a run is one cycle through the sequence of force points (eleven in this case), with three runs seen here representing three repetitions of this sequence performed at angular orientation intervals of 120 ° about the vertical axis of the deadweight machine.
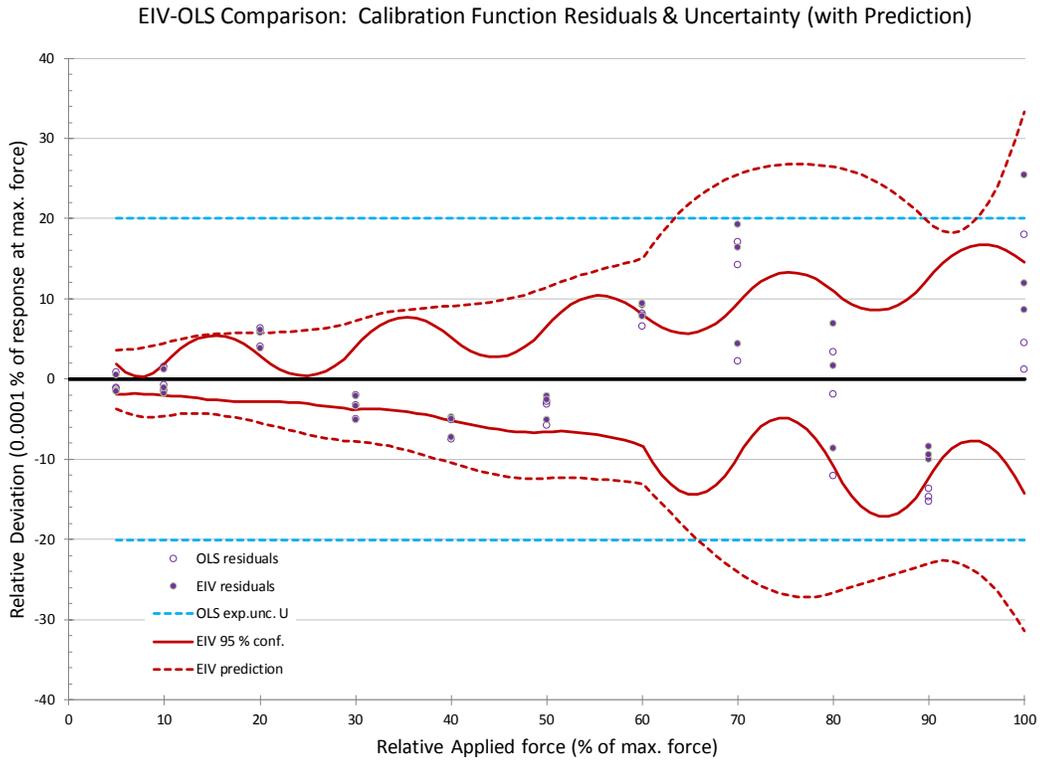
Figure 3. Calibration function residuals with uncertainty bands

In addition to providing the difference between the EIV and OLS regressions in a graphical form that is familiar to the users of NIST's force calibration service, Figure 3 compares the uncertainty interval from the traditional analysis with uncertainty bands yielded by the Monte Carlo evaluation conducted for the EIV regression. The expanded uncertainty interval, conservatively derived from a conventional uncertainty propagation as described in Ref. [3] and reported in each Report of Calibration, is shown in Figure 3 by horizontal dashed blue lines. Superimposed are the uncertainty bands, shown in red, from the Monte Carlo computation described in detail in Ref. [4]. The solid red lines depict, for the calibration function derived from EIV regression, the confidence bands for a coverage probability of 0.95. These bands are often termed the 95 % confidence bands or 95 % coverage bands; they enclose the area that is considered to contain the true calibration function with a confidence level of 95 %. Thus they indicate how well the estimated calibration function, assumed to have the form of a polynomial of chosen order and fitted to the calibration data by means of EIV regression, represents the true function relating the transducer's response to the applied force. These confidence bands correspond to the horizontal interval described earlier for the conventional uncertainty propagation.

The dashed red lines in Figure 3 depict the corresponding "prediction bands", which enclose the area that would be estimated to contain, based on the derived calibration function with 95 % confidence, the values of the transducer response that would be obtained from future measurements if conducted under the same conditions used for the calibration measurement data. The prediction bands incorporate both the uncertainty of the calibration function derived from EIV regression as well as the variability in the individual measurement points of the calibration data set over which the EIV regression was performed. Thus these bands, which are necessarily wider than the 95 % confidence bands, represent the region where new individual measurement points, if conducted, may be expected to fall.

The EIV supplement also provides the plot shown in Figure 4, depicting two additional uncertainty bands calculated by the Monte Carlo method, which are associated with the measurement function of the transducer. The identification of the 95 % confidence and prediction bands shown here are the same as described above for the EIV regression calibration function. The measurement function is calculated by the R-code from the inverse of the EIV calibration function and thus yields force as a function of transducer response. The following section provides additional guidance for determination of the measurement function by the user.
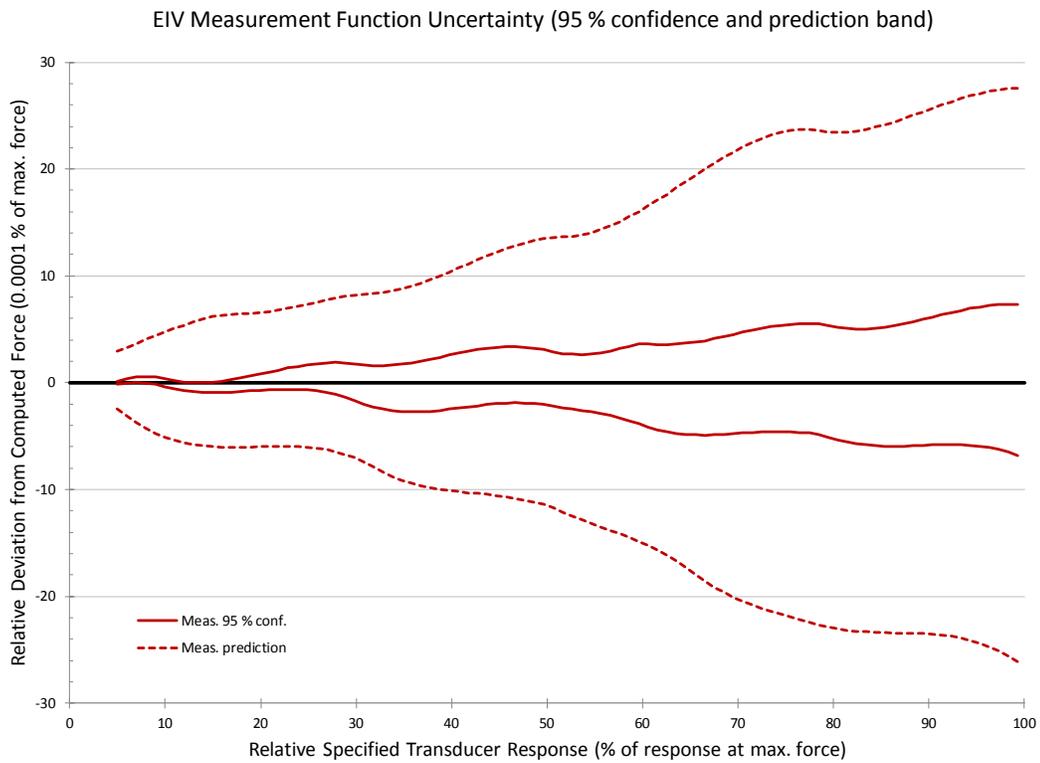

Figure 4. Measurement function uncertainty bands

4. Measurement Function Determination

The NIST force calibration service is directed toward characterizing the transducers being submitted for calibration. Thus calibrated forces are applied to a transducer, the transducer's responses to these forces are measured, and an equation, termed the calibration function, is derived from a regression method that gives the best estimate possible of the transducer response as a function of the applied force. The user of this calibration service, upon return of the calibrated force transducer and its calibration function provided by NIST, needs to perform the reverse procedure: apply an unknown force to the transducer, measure the transducer's response to this force, and make use of a measurement function to determine the value of the force that was applied. The measurement function gives the best estimate of the applied force as a function of the measured transducer response.

Historically NIST has provided only the calibration function, and not the measurement function. The user of the transducer is left with the task of obtaining the measurement function, by determining the inverse of the calibration function. In the past the calibration function was usually provided as a 2nd order polynomial, and the user could readily employ the quadratic equation to find the root, and thus the inverse, of the calibration function to obtain the force for a given measured transducer response. Increasingly complex closed-form expressions are available for the roots of 3rd and 4th order polynomial equations, but not for orders greater than four; thus polynomials of order three or higher are generally addressed by some numerical iterative root-solving procedure.

It is attractive to consider the derivation of the measurement function directly from the calibration data set, by employing a regression process to fit some function, such as a polynomial, to the data to provide the applied force as a function of the measured response. Performing this process with OLS regression would be implicitly invalid, because that would ascribe all of the measurement uncertainty to the applied forces and none to the measured responses (which incorporate the complications of transducer characteristics such as the sensitivity to orientation relative to the mechanism for applying the forces). However, EIV regression (performed with the abscissas and ordinates reversed from the process used for the calibration function) would not have this objection, since it correctly accounts for uncertainties along both axes.

Acknowledging that the inverse of a polynomial is not a polynomial, it is realized that either the calibration function or the measurement function, but not both, can be a true polynomial. Thus, the use of polynomials for both functions would have the result that a force value, used as input to the calibration function to determine response, would not exactly agree with the force yielded by the measurement function when input with that response. It may, however, be possible to approximate the measurement function as a polynomial that would yield a

8

(force, response) pair that agrees with the corresponding pair yielded by the calibration function to a desired, specified precision, for forces within the range of the calibration.

The inverse of the calculation function can be approximated to any desired precision by means of tables with interpolation or by root-solving iteration methods. The final page of the EIV Supplement to the Report of Force Calibration presents such a table, containing sixty force values uniformly spaced over the range of calibration, with the EIV predicted response that is calculated from the calibration function for each force. The user of the calibration service can, of course, readily generate a table of much finer increments.

Computer applications to solve for the roots of the calibration function are available that employ mathematical algorithms for numerical analysis. An example is given below of such a construction for MS Excel using the Newton-Raphson method that can be readily implemented for polynomials up to 5th order as encountered in force calibrations. Both the Setup Section and the Iteration Table that are shown below are incorporated into a single Excel worksheet of convenient length. The user inputs are: the coefficients A through F of the calibration function determined by the EIV regression; the specified transducer response (denoted as r_{set} in the example below) for which the user wants to know the applied force; and the desired relative tolerance in the calculated force. All other values are calculated by formulas with internal Excel functions.

The algorithm assumes the transducer response to be monotonically increasing (or decreasing) over the calibration range, so that the roots to the calibration function are single-valued. The Newton-Raphson method employs the principle that the extrapolation of the tangent to the function $y(x)$, denoted below as $y = f_{EIV}(x) - r_{set}$, to the $y=0$ axis for an iteration starting value x_a , yields a better approximation x_b to the true root of $y(x)$. The Iteration Table shown below provides for twenty iterations, giving the automatically calculated values of x_a and x_b for each iteration; this number of iterations is more than sufficient for force transducers encountered in normal practice.

The calculation illustrated here is conducted for one value at a time; the user gives one input value (response), and the worksheet returns one answer (force). The user can repeat the process for new input values, one at a time. A separate worksheet can be combined with the worksheet shown below, employing a macro to implement a short "Do...Loop", to readily enable multiple calculations to be automatically conducted from an input table that provides any arbitrary set of measured responses; by referencing the worksheet shown below, the macro yields the corresponding applied forces without additional user intervention.

Setup for Calculation of Measurement Function (Inverse of Calibration Function)

4.00000000	r_{set}, the specified response (mV/V or other response unit); this is the **INPUT VALUE**
1.1146029E+05	x_0 , the initial estimate of the force for response r_{set} , from linear extrapolation (force unit)
4.0053089E+00	r_0, the calculated response from calibration function for force x_0 (mV/V or other response unit)
1.00E-12	desired relative tolerance in calculated force (iterations are sufficient when improvement reaches tolerance)
3	number of iterations necessary to reach tolerance (calculated from iteration table)
111312.614079	x_{root} , the calculated value of the force for which the calibration function will return the response r_{set} ; this is the **ANSWER**
-1.307E-16	relative force change (ending - beginning) for the iteration reaching tolerance, from iteration table
4.00000000	$r = f_{EIV}(x_{root})$, the response calculated from the calibration function for the force x_{root}; this should be equal to the specified response r_{set}

Calibration Function, $r = f_{EIV}(x)$, from EIV computation by R-code, where x = applied force,
r = transducer response. Thus $r = A + Bx + Cx^2 + Dx^3 + Ex^4 + Fx^5$

coefficient A (intercept)	coefficient B (1st order term)	coefficient C (2nd order term)	coefficient D (3rd order term)	coefficient E (4th order term)	coefficient F (5th order term)
5.0748811E-05	3.5886764E-05	7.2177090E-13	-2.6416476E-18	0.0000000E+00	0.0000000E+00

Function from which root is calculated, y = f(x), where $f(x) = f_{EIV}(x) - r_{set}$
Thus $y = f(x) = (A - r_{set}) + Bx + Cx^2 + Dx^3 + Ex^4 + Fx^5$
The problem is to find the root, (x_{root}), of y = f(x) = 0

intercept	coefficient for 1st order term	coefficient for 2nd order term	coefficient for 3rd order term	coefficient for 4th order term	coefficient for 5th order term
-3.9999493E+00	3.5886764E-05	7.2177090E-13	-2.6416476E-18	0.0000000E+00	0.0000000E+00

Derivative Function $y' = f'(x) = B + 2Cx + 3Dx^2 + 4Ex^3 + 5Fx^4$

intercept	coefficient for 1st order term	coefficient for 2nd order term	coefficient for 3rd order term	coefficient for 4th order term	coefficient for 5th order term
3.5886764E-05	1.4435418E-12	-7.9249427E-18	0.0000000E+00	0.0000000E+00	

10

Iteration Table for Calculation of Measurement Function								
For each iteration below, x_a is the starting value, x_b is the new value : $x_b = x_a - f(x_a)/f'(x_a)$; x_b should approach x_{root}								
Iteration no.	x_a	$f(x_a)$	$f'(x_a)$	x_b	relative force change $(x_b - x_a)/x_b$	calculated response $r = f_{eiv}(x_b)$	$(r - r_{set})/r_{set}$	iterations achieving tolerance
1	1.1146029223E+05	5.3089161E-03	3.5949207E-05	1.1131261398E+05	-1.325E-03	3.9999999965E+00	-8.79E-10	
2	1.1131261398E+05	-3.5146172E-09	3.5949255E-05	1.1131261408E+05	8.783E-10	4.0000000000E+00	2.22E-16	
3	1.1131261408E+05	3.4998046E-16	3.5949255E-05	1.1131261408E+05	-1.307E-16	4.0000000000E+00	-1.11E-16	*
4	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
5	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
6	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
7	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
8	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
9	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
10	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
11	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
12	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
13	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
14	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
15	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
16	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
17	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
18	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
19	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*
20	1.1131261408E+05	-9.4542429E-17	3.5949255E-05	1.1131261408E+05	0.000E+00	4.0000000000E+00	-1.11E-16	*

References

[1] ASTM. ASTM E 74-13a, *Practice of Calibration of Force-Measuring Instruments for Verifying the Force Indication of Testing Machines.* West Conshohocken, PA, 2013.

[2] ISO. *Metallic materials – calibration of force-proving instruments used for the verification of uniaxial testing machines.* International Organization for Standardization (ISO), Geneva, Switzerland, 2011. International Standard ISO 376:2011(E).

[3] T. Bartel. Uncertainty in NIST force measurements. *Journal of Research of the National Institute of Standards and Technology*, **110**(6):589-603, 2005.

[4] T. Bartel, S. Stoudt and A. Possolo. Force Calibrations using Errors-in-Variables Regression and Monte Carlo Uncertainty Evaluations. *Metrologia* **53**:965-980, 2016.

[5] R Core Team. R: *A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2015. URL www.R-project.org/.