Evaluating the Uncertainty of Polynomial Regression Models Using Excel

Sheldon M. Jeter

**The George W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology**

Introduction. Regression analysis is widely used in experimental mechanical engineering to model physical properties and processes and to calibrate instruments. Standard textbooks address linear models and the uncertainties of linear models quite well. Usually, the texts adequately address multiple regression models based on several truly independent variables. The situation is more problematical with respect to multiple regression models that are polynomials in one variable such as the general quadratic model,

$$y_{\mathrm{est}} = y_{\mathrm{ave}} + b_1\left(x - x_{\mathrm{ave}}\right) + b_2\left(x^2 - x_{\mathrm{ave}}^2\right) \tag{1}$$

Surprisingly, the uncertainly of such a polynomial model is not addressed directly in the standard texts on regression analysis. In this model, the independent variables are linearly independent; so linear regression will determine the coefficients correctly. After evaluating the model the next obvious and needed step is finding the uncertainty of the model.

An indirect measurement is a value calculated from more direct measurements. A regression model is obviously an indirect measurement. Within the minimal restrictions that are usually satisfied in experimental engineering, the uncertainty of any indirect measurement, y, may be determined from the uncertainties of the more direct measurements, the x_i. This calculation is done with the familiar formula from Error Propagation Analysis (EPA),

$$u_y^2 = \left(\frac{\partial y}{\partial x_1}u_{x_1}\right)^2 + \left(\frac{\partial y}{\partial x_2}u_{x_2}\right)^2 + \cdots + \left(\frac{\partial y}{\partial x_m}u_{x_m}\right)^2 + \cdots \tag{2}$$

Where,

$u_y =$ the Standard Uncertainty of the indirect measurement, y

$u_{x_i} =$ the Standard Uncertainty of the i-th more direct measurement

In general, the preceding formula can be used to find the uncertainty of the systematic dependence of y on x. A complication in the polynomial case is that the two variables and therefore the two coefficients are strongly correlated. Because of this correlation, the coefficients cannot be considered to be independent sources of error in uncertainty analysis. Consequently, it will be shown that the familiar EPA formula must be modified. For example, the uncertainty of the quadratic model should be written using conditional uncertainties as

$$u^2_{\text{poly-model}} = \left(\frac{\text{SEE}}{\sqrt{n}}\right)^2 + (x - x_{\text{ave}})^2 u(b_1|b_2)^2 + (x^2 - x^2_{\text{ave}})^2 u(b_2|b_1)^2 \tag{3}$$

Here SEE is the conventional Standard Error of Estimate, $u(b_1|b_2)$ is the conditional uncertainty of b_1 given a value for b_2 , and similarly $u(b_2|b_1)$ is the conditional uncertainty of b_2 given b_1. This formula seems easy enough to apply. However, when a numerical spreadsheet program such as Excel is used, extra regression steps are necessary to compute the conditional uncertainties. In the following, the required auxiliary calculations will be explained and illustrated and a convenient spreadsheet block to implement the needed analysis and plot the results will be presented.

In the following, linear regression will first be briefly reviewed. Then the formula for the uncertainty of a simple linear model will be developed. The complications encountered with a polynomial model will be addressed before the modified formula for a polynomial model is presented. An exhaustive numerical experiment has been conducted to verify this uncertainty and the results of this investigation are presented below. Finally the technique is demonstrated for two commonplace and representative applications. One example is application to a quadratic form of the Clausius-Clapeyron model commonly used to represent vapor pressure data. Another example is the evaluation of the quartic calibration function typically used for a thermal anemometer. When the needed auxiliary calculations have been made, the spreadsheet block presented in this paper makes it easy to compute and plot the uncertainty for any familiar regression model, whether linear, multiple, or polynomial.

<u>Synopsis of Linear Regression.</u> A few results from linear regression analysis are needed to support the development of the uncertainty calculations that are presented below. For a linear regression model with one independent variable, the regression model is

$$y_{\text{est}} = c + b\,x \tag{4}$$

This model has two parameters, the constant and the coefficient. In general the Residual Variation, RSS, is the squared deviation between the model values and the experimental data, or

$$\text{RSS} = \sum_{i=1}^{n} (y_i - y_{\text{est}})^2 = \sum_{i=1}^{n} (y_i - c - b\,x_i)^2 \tag{5}$$

Regression procedures or software determine the parameters of the model that minimize the RSS. In particular, it is relatively straightforward to show that the for the linear model above the explicit formula for the coefficient, b, that minimizes the RSS is

$$b = \frac{n\sum_{i=1}^{n} x_i\, y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i} = \frac{\sum_{i=1}^{n} (x_i - x_{\text{ave}})y_i}{\sum_{i=1}^{n} x_i^2 - n\, x_{\text{ave}}^2} \tag{6}$$

Where

$$x_{\text{ave}} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{7}$$

With b known, it is easy enough to solve for the constant, c, that minimizes the RSS, so

$$c = y_{\text{ave}} - b\, x_{\text{ave}} \tag{8}$$

Where, of course

$$y_{\text{ave}} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{9}$$

The equations in this section give the parameters for a least RSS linear model, and the uncertainty of the model can be addressed next after reviewing error propagation, which is the basis for finding the uncertainty.

Recap of Error Propagation Analysis. The basic idea in error propagation analysis (EPA) is the recognition of direct and indirect measurements. An indirect measurement is calculated from a direct measurement. Assume that m independent direct measurements, the set of w_i s below, contribute to an indirect measurement, z. The measurement formula is then merely the function used to calculate the z,

$$z = z(w_1, w_2, \cdots w_m) \tag{10}$$

The operational concepts of EPA are incorporated in two equations. The first addresses how the uncertainty in the indirect measurement z (*i. e.*, the dependent variable) is caused by the uncertainty in some more directly measured variable, w (*i. e.*, an independent variable). Call this uncertainty $u_{z,w}$. A truncated Taylor Series representation relates the resulting uncertainty $u_{z,w}$ to the uncertainty in w as follows,

$$u_{z,w} = \frac{\partial z}{\partial w} u_w \tag{11}$$

Proceedings of the 2003 American Society for Engineering Education Annual Conference & Exposition

The other basic operational formula shows how independent uncertainties are combined when several direct measurements contribute to an indirect measurement. Analysis shows that, given some nonrestrictive assumptions, the squares of the contributing uncertainties sum to give the squared combined uncertainty, or

$$u_z^2 = u_{z,1}^2 + u_{z,2}^2 + \Lambda\ u_{z,m}^2 = \left(\frac{\partial z}{\partial w_1} u_{w_1}\right)^2 + \left(\frac{\partial z}{\partial w_2} u_{w_2}\right)^2 + \cdots \left(\frac{\partial z}{\partial w_n} u_{w_m}\right)^2 \tag{12}$$

Each partial derivative is recognized to be the influence factor showing how each w_i influences the z. This equation can be called the combining rule for uncertainties. The two formulas are used together to formulate the uncertainty of a regression model.

<u>Uncertainty of the Coefficient.</u> For the coefficient b, the relationship that yields the influence factor is contained in Equation (6) above, so the influence coefficient is

$$\frac{\partial b}{\partial y_i} = \frac{(x_i - x_{\text{ave}})}{\sum_{i=1}^{n} x_i^2 - n\, x_{\text{ave}}^2} \tag{13}$$

The data set contains n pairs of data, x_i and y_i. By convention, every x_i is assumed to be known without any uncertainty, but every y_i is assumed to have some uncertainty. Since the variation in every y_i value effects b, the uncertainty of the coefficient is written as follows using the influence coefficients from the previous equation. The pertinent result is

$$u_b^2 = \sum_{i=1}^{n}\left(\frac{\partial b}{\partial y_i} u_y\right)^2 = \sum_{i=1}^{n}\left(\frac{(x_i - x_{\text{ave}})}{\sum_{i=1}^{n} x_i^2 - n\, x_{\text{ave}}^2} u_y\right)^2 = \frac{u_y^2 \sum_{i=1}^{n}(x_i - x_{\text{ave}})^2}{\left(\sum_{i=1}^{n} x_i^2 - n\, x_{\text{ave}}^2\right)^2} \tag{14}$$

Expanding the summed term in the numerator and regrouping gives

$$u_b^2 = u_y^2 \frac{\sum_{i=1}^{n}\left(x_i^2 - 2x_i\, x_{\text{ave}} + x_{\text{ave}}^2\right)}{\left(\sum_{i=1}^{n} x_i^2 - n\, x_{\text{ave}}^2\right)^2} = u_y^2 \frac{\sum_{i=1}^{n} x_i^2 - 2x_{\text{ave}} \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_{\text{ave}}^2}{\left(\sum_{i=1}^{n} x_i^2 - n\, x_{\text{ave}}^2\right)^2} \tag{15}$$

Then simplifying the numerator gives

Proceedings of the 2003 American Society for Engineering Education Annual Conference & Exposition

$$u_b^2 = u_y^2 \frac{\sum_{i=1}^{n} x_i^2 - 2x_{\text{ave}}\left(n\,x_{\text{ave}}\right) + n\,x_{\text{ave}}^2}{\left(\sum_{i=1}^{n} x_i^2 - n\,x_{\text{ave}}^2\right)^2} = u_y^2 \frac{\sum_{i=1}^{n} x_i^2 - n\,x_{\text{ave}}^2}{\left(\sum_{i=1}^{n} x_i^2 - n\,x_{\text{ave}}^2\right)^2} \tag{16}$$

Simplifying the ratio gives the essential result

$$u_b^2 = \frac{u_y^2}{\sum_{i=1}^{n} x_i^2 - n\,x_{\text{ave}}^2} \tag{17}$$

In practice, the uncertainty in y is not and cannot known *a priori*, so it is estimated by the Standard Error of y Estimate. The resulting computational formula becomes

$$u_b^2 = \frac{\text{SEE}^2}{\sum_{i=1}^{n} x_i^2 - n\,x_{\text{ave}}^2} \tag{18}$$

The uncertainty of the coefficient is an important statistic used routinely in significance testing, and it is a critical feature in the uncertainty of the regression model itself.

<u>Uncertainty of a Simple Linear Model.</u> EPA is also readily applied to the model itself. A quick look at Equation 4 might lead one to think that the uncertainty in a simple one variable linear model can be determined by direct application of the combining rule since the uncertainties in the constant and the coefficient are known. If this were true, the squared uncertainty in the model would be sum of the squared uncertainty in the constant and the squared uncertainty in the model due to the uncertainty in the coefficient, or

$$u_{\text{model}}^2 = u_c^2 + x^2\,u_b^2 \qquad \text{erroneous!} \tag{19}$$

This result is intuitively unsatisfactory because it increases monotonically with x while one would expect the uncertainty to increase toward the end points of the range of x. The conjecture in the preceding equation is wrong because c and b are not independent. The proper approach is to first eliminate c using Equation 8. After eliminating c, the model equation becomes

$$y_{\text{est}} = c + b\,x = \left(y_{\text{ave}} - b\,x_{\text{ave}}\right) + b\,x \tag{20}$$

so

$$y_{\text{est}} = y_{\text{ave}} + b\left(x - x_{\text{ave}}\right) \tag{21}$$

Both of the terms in the preceding equation can be assumed to be independent. Now, the uncertainty in the model is easy to represent by applying the combining rule to the preceding relationship as

$$u_{\text{model}}^2 = u_{y-\text{ave}}^2 + \left(x - x_{\text{ave}}\right)^2 u_b^2 \tag{22}$$

The SEE is used as the estimate of the uncertainty in any individual y datum. Then the uncertainty in the average of y, which is the average of n individual y data, is

$$u_{y-\text{ave}} = \frac{\text{SEE}}{\sqrt{n}} \tag{23}$$

Since the uncertainty in the coefficient has been determined in Equation 6, then an expanded formula for the uncertainty in the model is

$$u_{\text{model}}^2 = \left(\frac{\text{SEE}}{\sqrt{n}}\right)^2 + \left(x - x_{\text{ave}}\right)^2 \frac{\text{SEE}^2}{\left(\sum_{i=1}^{n} x_i^2 - n\, x_{\text{ave}}^2\right)^2} \tag{24}$$

Simplifying slightly gives the more concise result handy for calculations,

$$u_{\text{model}}^2 = \text{SEE}^2 \left(\frac{1}{n} + \frac{\left(x - x_{\text{ave}}\right)^2}{\left(\sum_{i=1}^{n} x_i^2 - n\, x_{\text{ave}}^2\right)^2} \right) \tag{25}$$

This result is intuitively satisfactory for at least two reasons. First, u is a minimum where x is at its average value. This is obviously the point where the estimate of the actual trend in y should be the best. Second, u increases monotonically and approximately quadratically toward the ends of the range in x. This is just the expected and observed behavior for uncertainty of a model. As usual, the uncertainty calculated in the preceding equation is the Standard Uncertainty of the model. A Standard Uncertainty is analogous to a Standard Deviation. Note that the uncertainty under here is due to random error. This is the Uncertainty A for the model, which is the uncertainty that can be evaluated by statistical analysis of repeated measurements. To plot the 95 % error band, the appropriate coverage factor, k_c, must be used to compute the Expanded Uncertainty A, which is

$$U_{\text{model}} = k_{\text{c}}\, u_{\text{model}} \tag{26}$$

Where u_{model} is given by the slightly complicated formula in Equation 25. For a large number of data, the coverage factor will be 2. For smaller samples, the rigorous coverage factor should be computed from the t-distribution using the appropriate Degrees of Freedom (*DF*). Recall that *DF* is the number of data less the number of parameters in the model.

An example application of the uncertainty of a simple linear model is plotted in Figure 1. The figure displays the relationship in a log-log model between the logarithm of the Nusselt number as the dependent or *y* variable against the logarithm of the Reynolds number as the independent or *x* variable. The regression parameters and the SEE for the data in Figure 1 were computed by the Excel regression package. The error envelopes, which are the curves showing the 95 % error bounds, were computed as described below in a compact block of the Excel workbook that is illustrated below as Attachment 1.

An appropriate regression model should represent the systematic variation in the dependent variable. Assuming this systematic representation is adequate, the SEE should represent the residual random scatter in the data with respect to the model. Using the SEE as the Standard Uncertainty in the individual data, the expanded uncertainty of the data with respect to the model is

$$U_{\text{data}} = k_{\text{c}}\,\text{SEE} \tag{27}$$

This uncertainty is the Uncertainty A due to random variation in the data. The corresponding error envelope is also plotted in the following figure. Note that the error band for the data is much wider than the error bound on the model, reflecting the averaging effect of the regression model.

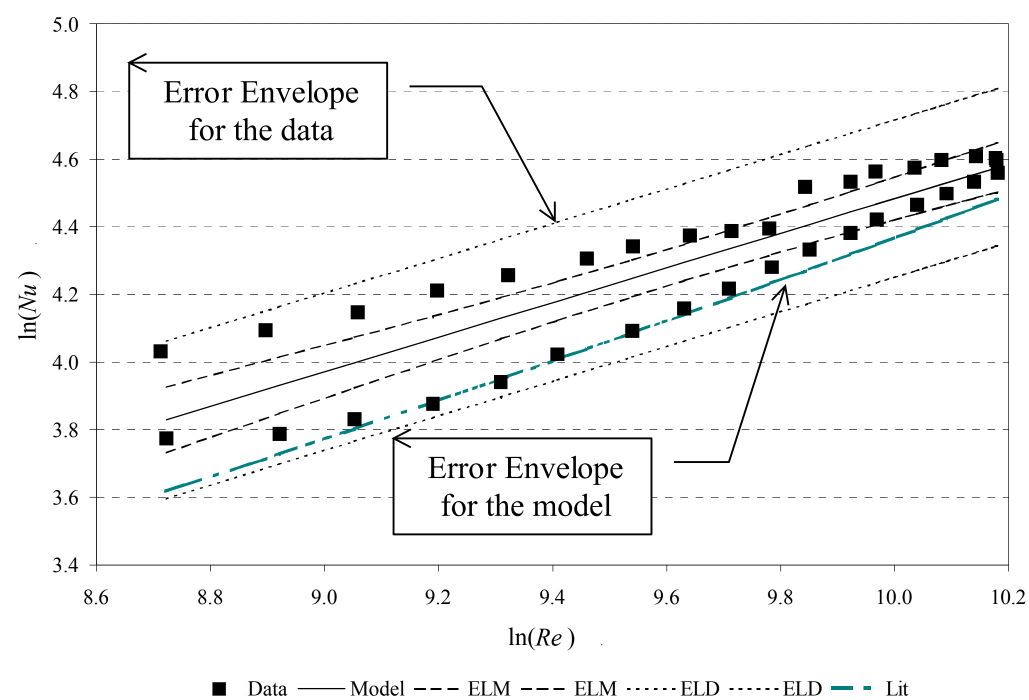

Figure 1 Forced Convection Data, Literature Model, and Error Bounds.
This model is an example of simple linear regression.

Uncertainty of a Multiple Regression Model. EPA is also readily applied to the uncertainty of a more complex model with multiple independent variables so long as the model is linear in its parameters. Standard texts show that centering the data by subtracting their averages from the dependent variable and the independent variables always eliminates the constant, so the model can always be written as

$$y_{\text{est}} = y_{\text{ave}} + b_1\left(x_1 - x_{1,\text{ave}}\right) + b_2\left(x_2 - x_{2,\text{ave}}\right) + \cdots + b_m\left(x_m - x_{m,\text{ave}}\right) \tag{28}$$

As before, the uncertainty in the model is easy to formulate by applying the combining rule to the preceding relationship, so

$$u_{\text{model}}^2 = u_{y-\text{ave}}^2 + \left(x_1 - x_{1,\text{ave}}\right)^2 u_{b1}^2 + \cdots + \left(x_m - x_{m,\text{ave}}\right)^2 u_{bm}^2 \tag{29}$$

The uncertainty in the average of y, is again computed using the SEE as the standard deviation in the formula for an average of n data, so

$$u_{y-\text{ave}} = \frac{\text{SEE}}{\sqrt{n}} \tag{30}$$

The uncertainty in a more complicated multiple regression model is then

$$u_{\text{model}}^2 = \left(\frac{\text{SEE}}{\sqrt{n}}\right)^2 + \left(x_1 - x_{1,\text{ave}}\right)^2 u_{b1}^2 + \cdots + \left(x_m - x_{m,\text{ave}}\right)^2 u_{bm}^2 \tag{31}$$

The previous model applies to a case like modeling the heat capacity of a dense vapor that is a function of two independent variables, temperature and pressure.

Polynomial Models. The situation is more subtle with respect to a more complex model that is a polynomial in one variable such as this quadratic model,

$$y_{\text{est}} = y_{\text{ave}} + b_1\left(x - x_{\text{ave}}\right) + b_2\left(x^2 - x_{\text{ave}}^2\right) \tag{32}$$

Surprisingly, the polynomial case is not addressed in the standard texts on regression analysis. Since polynomial models are very common in experimental engineering, the associated uncertainty should be investigated. The model itself presents no special problems. In particular, the independent variables, although related, are linearly independent; therefore, linear regression can find the correct coefficients easily.

The coefficients are, however, likely to be strongly correlated. For example, a typical quadratic model has been investigated numerically. Simulated experimental data was generated according to the following model,

$$y_{\mathrm{exp}} = y_{\mathrm{ave}} + b_1\left(x - x_{\mathrm{ave}}\right) + b_2\left(x^2 - x_{\mathrm{ave}}^2\right) + \mathrm{error} \tag{33}$$

Where the artificial error has a normal distribution with a mean of zero and a variance of 0.01 unit. The results of a numerical experiment based on the simulated data are illustrated in Figure 2 below for the case when $b_1 = 0.9$ and $b_2 = 0.1$. Coefficients from 500 cases with 11 data pairs per case are plotted. The plot shows that the parameters are indeed highly correlated. Specifically b_2 decreases as b_1 increases to follow the systematic trend in the data.

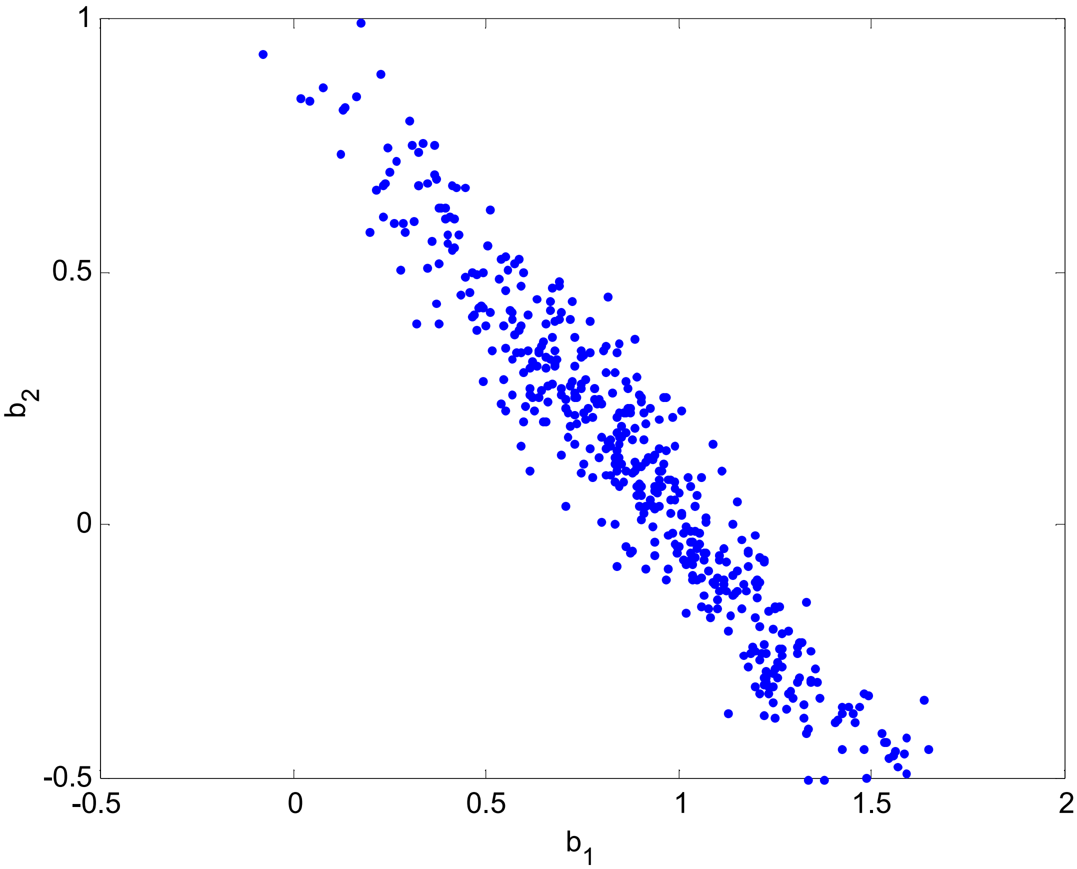

Figure 2 Variation of b_2 with Respect to b_1 in a Typical Quadratic Model.
Note that the uncertainty in b_2 given b_1 is relatively small.
If plotted inversely, the uncertainty in b_1 given b_2 would also be seen to be relatively small.

Since the coefficients are correlated, then the uncertainties of the coefficients cannot be considered to be independent sources of error, and the EPA formula must be modified. Obviously, b_2 depends on b_1 , and inversely b_1 depends on b_2. Consequently, the uncertainty of the quadratic model should be written using conditional uncertainties as

Proceedings of the 2003 American Society for Engineering Education Annual Conference & Exposition

$$u^2_{\text{poly-model}} = \left(\frac{\text{SEE}}{\sqrt{n}}\right)^2 + (x - x_{\text{ave}})^2 u(b_1|b_2)^2 + (x^2 - x^2_{\text{ave}})^2 u(b_2|b_1)^2 \qquad (34)$$

Here $u(b_1|b_2)$ is the conditional uncertainty of b_1 given a value for b_2 , and similarly $u(b_2|b_1)$ is the conditional uncertainty of b_2 given b_1.

The modified formula is easy enough to apply, but extra regression steps are necessary to compute the conditional uncertainties. For example, compute $u(b_1|b_2)$ as follows: First subtract the product term, $b_2 x^2$, from the dependent variable. This step creates a new dependent variable corrected for the variation with respect to x^2 that depends on b_2.

$$y_{\text{CORR}} = y - b_2 x^2 \qquad (35)$$

Then execute the regression analysis with x as the only variable. The resulting uncertainty of the coefficient is obviously the uncertainty of b_1 given a specific value of b_2. Similarly correct for the variation related to b_1 by subtracting the term $b_1 x$. Then run the regression with x^2 as the only independent variable. This result will give the uncertainty in b_2 given a value for b_1.

Verification. The development for the uncertainty of the model presented here appears to be entirely reasonable; however, it cannot be asserted to be a rigorous development because no attempt was made to develop the formula from first principles. In addition no attempt was made to ensure that the conditional uncertainties exist and are unique. Indeed an entirely rigorous mathematical proof does not even appear to be feasible. Instead a numerical verification was conducted.

In the numerical verification, simulated experimental data was generated for a quadratic dependence contaminated by normally distributed errors. Since the constant only represents a fixed shift in origin of the y-axis, it was set to zero, and the formula used to generate the simulated is

$$y_{\text{exp}} = c + b_1 x + b_2 x^2 + \text{error} = b_1 x + b_2 x^2 + \text{error} \qquad (36)$$

The errors are normally distributed with a standard deviation of 0.1 unit. The data were taken to be normalized with both x and the error-free value of y restricted to range from 0 to 1. For this range and domain the coefficients must sum to unity. Three cases appear to represent the range of relative values of the coefficients. The case when both coefficients are similar in magnitude is represented by b_1 and b_2 equal 0.5. The case when b_1 is dominant is represented by $b_1 = 0.9$ and $b_2 = 0.1$. The case when b_2 is dominant is represented by $b_1 = 0.1$ and $b_2 = 0.9$. For all three models, 500 data sets of 11 data pairs each were generated. A Matlab program was used to generate the resulting regression models and the associated uncertainties. The pertinent features of the numerical experiment are presented in the next figures.

Figure 3 is probably the more interesting presentation. It shows the simulated data points, dotted lines for all of the 500 regression models, the exact value of y, and the average error envelopes of the model generated using the mean values of the conditional uncertainties of the coefficients. This average error envelope should represent the typical predictive power of the uncertainty of the model. Inspection of the figure shows immediately that the average error envelope does indeed bound the main body of the simulated variation in the regression models. Indeed, when a point by point comparison is made 95.6 % of the points in the models fall within the error envelope in almost exact agreement with the expected 95 % range for an expanded uncertainty. Similar results are obtained with the other two representative cases as summarized in Table 1.

Of course any particular regression model can deviate significantly from the exact y relation due only to unfortunately extreme variation in the data. Consequently, the variation in the range of the error envelopes is at least as much as the variation in the range of the models. The range on the error envelopes is illustrated in Figure 4. Note the considerable range about the average values.

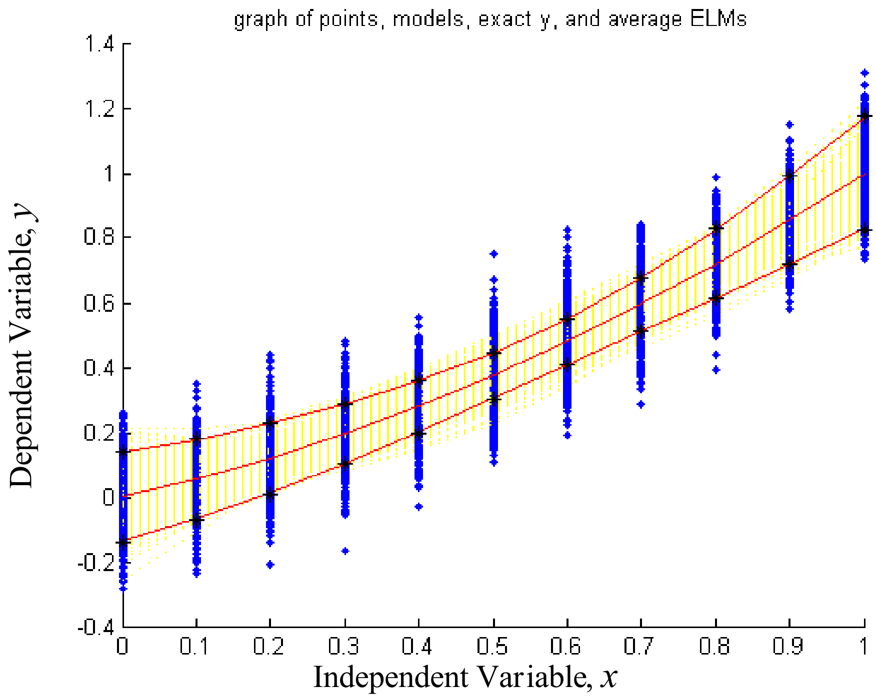

Figure 3. Plot of the Simulated Data, Regression Models, Exact y Relation, and the Average Error Envelopes for the Case where b_1 and b_2 Equal 0.5. Data are the individual markers in vertical strips. Regression models are plotted as faint dotted lines, and the average error envelopes are the solid hourglass shaped lines.

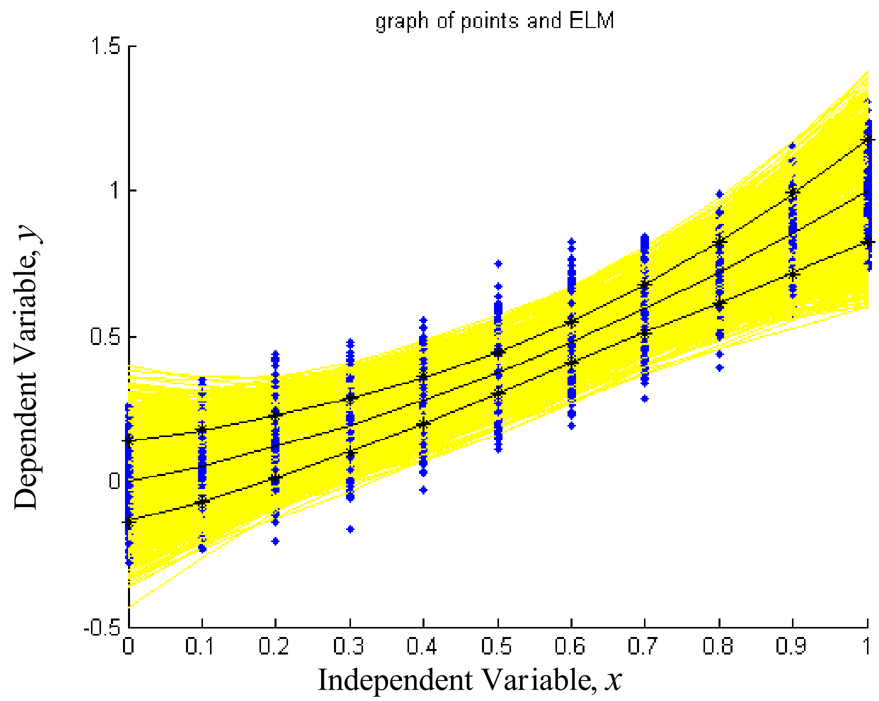

Figure 4. Plot of the Simulated Data and the Error Envelopes for the Case where b_1 and b_2 Equal 0.5.Data are the individual markers in vertical strips. The error envelopes are plotted as faint solid lines.

The extra steps to define the corrected dependent variables and calculate the conditional uncertainties will require only a little extra time in the very frequent cases when polynomial models are required. The extra time is well rewarded by an unambiguous result for the uncertainty of the model.

Note that the error envelopes described above only represent the expected random uncertainty in the model. In addition the Uncertainty B due to possible bias in the data should be computed by Error Propagation Analysis considering all significant sources of error in the measurement system. The regression model will surely share the Uncertainty B in the underlying data, and the Combined Uncertainty can be computed by the usual formula,

$$U_{\mathrm{C,model}}^2 = U_{\mathrm{A,model}}^2 + U_{\mathrm{B}}^2 \tag{37}$$

This Combined Uncertainty is the error envelope that should be plotted in experimental engineering.

Quadratic Example. A first example application of polynomial uncertainty analysis is the quadratic Clausius-Clapeyron model developed in a vapor pressure experiment that is part of an undergraduate lab course. In this lab, linear and a quadratic models for the logarithm of the vapor pressure with the inverse absolute temperature as the independent variable are developed. Often the quadratic model is not statistically significant; however, in the case illustrated in Figure 5, it is barely significant with an alpha risk of 4.1 %. The minimal contribution and significance of the quadratic term is evident since the model has almost no curvature. However when the random scatter in the data is small as in this case, a quadratic term with a rather small coefficient can still be statistically significant. When such a term is significant, there is no principled reason for rejecting it. If the quadratic model is retained as it should be in this particular case, the uncertainty should be calculated and the error envelopes should be plotted along with the model as has been done in Figure 5. In this figure, the coverage factor has been multiplied by 10 to make the error envelopes visible. The spreadsheet block illustrated in Attachment 1 was used to compute and plot the error envelope. The spreadsheet is available online for interested teachers and researchers (Jeter, 2003). Calculating the model does require some minimal individual spreadsheet manipulation to compute the two needed conditional uncertainties as described above, but this task is straightforward Since the spreadsheet is designed for applications with as many as four completely independent variables, it can easily accommodate this minimally complicated application. In this particular case, the only significant source of Uncertainty B was the imperfection in the calibration of the Bourdon tube gage used to measure the pressure. In this application the Uncertainty B was not added to the uncertainty of the data with respect to the model because this uncertainty is typically used for screening the data for possible outliers. Since the data and the model share the same bias, it is not reasonable to include the Uncertainty B for that application. A block is available in the spreadsheet for entering the user-defined Uncertainty B data and calculating the Combined Uncertainty of the model. It is the Combined Uncertainty of the model that should be plotted and is plotted in Figure 5.

<u>Quartic Model.</u> The second example that is posted is calibration data for a typical constant temperature thermal anemometer. A multiple point regression was done using the recommended

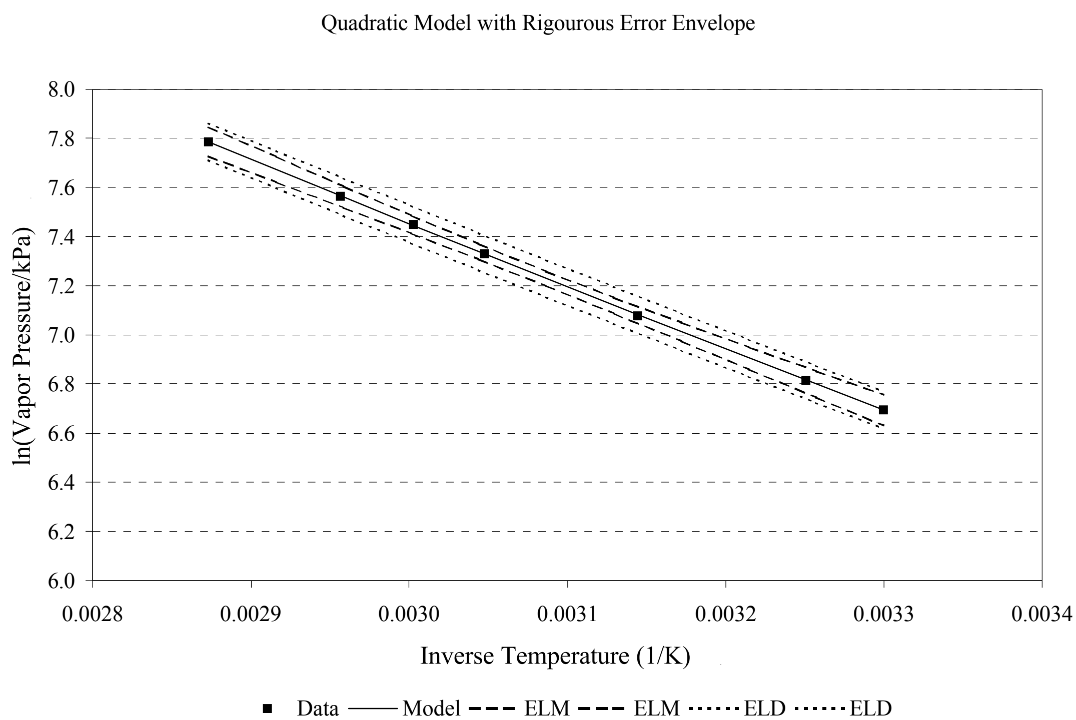

Figure 5. Plot of Vapor Pressure Model with Error Envelopes for Model and Data.

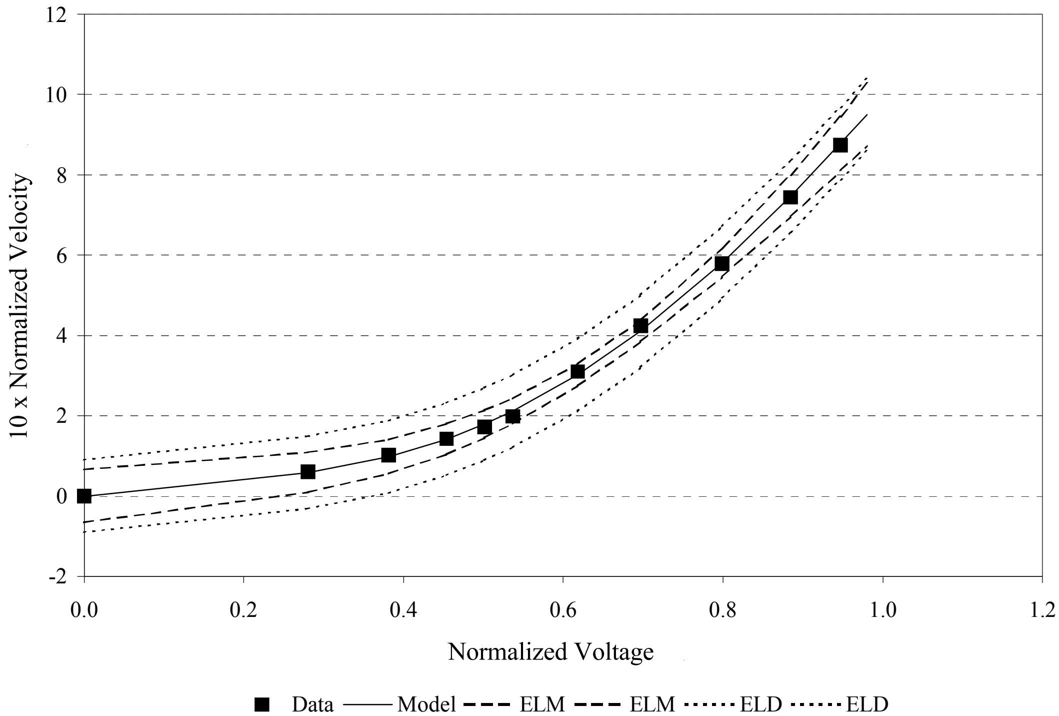

Figure 6. Plot of Calibration Curve for Constant Temperature Thermal Anemometer along with Error Envelopes for the Calibration Curve and Data.

homogeneous model that is fourth degree in the normalized voltage. All four coefficients were found to be statistically significant in this case. The conditional uncertainties were then computed for each of the four coefficients by correcting for the influence of the three other coefficients in turn. In this case the Uncertainty B of the calibration wind tunnel was combined with the Uncertainty A of the model to yield the Combined Uncertainty that is plotted in Figure 5. Even though four auxiliary calculations were necessary, the uncertainty could be computed and plotted in only a few minutes using the spreadsheet block designed for this purpose. Instructions for using the spreadsheet are included in the document, and a video example is available to interested teachers and researchers who contact the author. Indeed this spreadsheet should make it easy to compute and plot the uncertainties for any regression model.

Table 1. Averaged Regression Results for Three Values of the Exact Coefficients
Each case is represented by 500 simulated sets of 11 data each.

quantity			
exact b_1	0.5	0.1	0.9
average $u(b_1 \vert b_2)$	.0879	.0891	.0873
average $u(b_2 \vert b_1)$	.0847	.0858	.0841
average value of $u_{y-\text{ave}} = \frac{\text{SEE}}{\sqrt{n}}$	.0295	.0299	.0293
Percent of Model Points within Error Envelope	95.6	96.2	95.2

Reference

Jeter, S. M., 2003, “Spreadsheet ELM for Computing the Error Limits of a Model”, ME 4053 Engineering Systems Laboratory, the George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, 10 January 2003, available on line at <me.gatech.edu/~sjeter>.

Sheldon M. Jeter is Associate Professor of Mechanical Engineering at the George W. Woodruff School of Mechanical Engineering at Georgia Tech. He has degrees from Clemson University, the University of Florida, and Georgia Tech. He has been on the academic faculty at Georgia Tech since 1979. His research interests are thermodynamics, experimental engineering, heat and mass transfer, and energy systems

Attachment 1. Block of Excel Workbook used to Compute and Plot the Uncertainties of a Regression Model

Sheet to Compute and Plot the Uncertainty of a Model, SMJ Nov 2001, updated 14 Jan 2003

User must insert and/or update data coded with yellow. User should format these and other cells for neatness or legibility as needed.

Instructions:

(1) Copy this sheet to your experimental *.xls workbook.
(2) Insert the experimental data into block 4.
(3) Insert the regression results into block 2.
(4) Select the desired coverage factor in Block 3.
(5) Update the cell ranges to compute the averages in Block 3.
(6) Identify by cell formula the max and min X1 values in Block 5.
(7) Insert data or formulas for other XN data in Block 5.
(8) Insert optional data for U_b in Blocks 4 and 5.
(9) Plot experimental data points with green block in Block 4.
(10) Plot model and limits with green block in Block5.

1. Summary Data:

The averaged U_a of model = 0.043 The averaged U_c of model = 0.044
The constant U_a of the data = 0.076

2. Block of Data from Regression

User must insert the following data from from t he regression block.

Constant =	17.5		
Coefficient B1 =	-4064.6	6.4	= Std error of B1
Coefficient B2 =	243635.5	1028.7	= Std error of B2
Coefficient B3 =	0.00000	0.00000	= Std error of B3
Coefficient B4 =	0.00000	0.00000	= Std error of B4
Std Error of y Est =	0.00273		
n, number of data =	7		
p, number of parameters =	3		
coverage factor, kc, by t-dist =	2.78		

3. Block of Calculations

User must specify the desired coverage factor to be used. Rigorous value is in cell D23.

coverage factor, kc, used = 27.76 **<-----User must select.** 10 = multiplier

User must update the cell ranges in the following four formulas to calculate the correct averages.

Average value of X1 =	0.00308
Average value of X2 =	0.00001
Average value of X3 =	0.00000
Average value of X4 =	0.00000

4. Block of Experimental Data and Results, plot the data with the block in green

User must insert the complete set of y and x data from the experim ental data set into following block. Add rows as required.
User may insert column of data for Expanded Uncertainty B, uncertainty due to possible bias, if desired.

Experimental Data, i nsert at least one zero in every otherwise unused XN column

Y data	X1 data X1	X2 data X2	X3 data X3	X4 data X4	*U* b of model	*U* a of model	*U* c of model	X1 data	Y data	regress model	Lower E-limit on model	Upper E-limit on model	Lower E-limit on data	Upper E-limit on data
6.696	0.00330	0.000011	0.000	0.000	0.006	0.062	0.062	0.003	6.696	6.694	6.632	6.756	6.618	6.770
6.815	0.00325	0.000011			0.005	0.051	0.051	0.003	6.815	6.816	6.765	6.867	6.740	6.892
7.079	0.00314	0.000010			0.004	0.032	0.033	0.003	7.079	7.082	7.050	7.115	7.007	7.158
7.332	0.00305	0.000009			0.003	0.030	0.030	0.003	7.332	7.330	7.299	7.360	7.254	7.405
7.448	0.00300	0.000009			0.003	0.035	0.035	0.003	7.448	7.446	7.411	7.481	7.370	7.522
7.566	0.00296	0.000009			0.003	0.043	0.043	0.003	7.566	7.566	7.524	7.609	7.491	7.642
7.785	0.00287	0.000008			0.002	0.059	0.059	0.003	7.785	7.787	7.727	7.846	7.711	7.862

5. Block of Uniformly Spaced Data for Plotting wrt X1, plot model and limits with the block in green

User must insert cell references to identify the maximum and mi numum X1 values in the cells on the next row.

max X1 = 0.00330 min X1 = 0.00287 4.27E-05 = computed delta X1

Spreadsheet will compute uniform ly spaced X1 values. User must code columns for corresponding values of X2, X3, and X4.
User may insert column of data for Expanded Uncertainty B, uncertainty due to possible bias, if desired.

X1	X2	X3	X4	*U* b of model	*U* a of model	*U* c of model	X1 data	regress model	Lower E-limit on model	Upper E-limit on model	Lower E-limit on data	Upper E-limit on data
2.87E-03	0.00			0.002	0.059	0.059	2.9E-03	7.787	7.727	7.846	7.711	7.862
2.92E-03	0.00			0.010	0.050	0.051	2.9E-03	7.673	7.622	7.725	7.598	7.749
2.96E-03	0.00			0.010	0.042	0.043	3.0E-03	7.561	7.518	7.604	7.485	7.637
3.00E-03	0.00			0.010	0.035	0.037	3.0E-03	7.449	7.413	7.486	7.374	7.525
3.04E-03	0.00			0.010	0.030	0.032	3.0E-03	7.339	7.307	7.371	7.263	7.415
3.09E-03	0.00			0.010	0.029	0.030	3.1E-03	7.229	7.199	7.260	7.154	7.305
3.13E-03	0.00			0.010	0.031	0.032	3.1E-03	7.120	7.088	7.153	7.045	7.196
3.17E-03	0.00			0.010	0.036	0.038	3.2E-03	7.013	6.975	7.050	6.937	7.088
3.21E-03	0.00			0.010	0.044	0.045	3.2E-03	6.905	6.861	6.950	6.830	6.981
3.26E-03	0.00			0.010	0.052	0.053	3.3E-03	6.799	6.746	6.853	6.724	6.875
3.30E-03	0.00			0.010	0.062	0.063	3.3E-03	6.694	6.631	6.757	6.618	6.770

(This page intentionally left blank.)