11-1-2008

Least Squares Percentage Regression

Chris Tofallis
University of Hertfordshire, c.tofallis@herts.ac.uk

Least Squares Percentage Regression

Chris Tofallis
University of Hertfordshire
United Kingdom

In prediction, the percentage error is often felt to be more meaningful than the absolute error. We therefore extend the method of least squares to deal with percentage errors, for both simple and multiple regression. Exact expressions are derived for the coefficients, and we show how such models can be estimated using standard software. When the relative error is normally distributed, least squares percentage regression is shown to provide maximum likelihood estimates. The *multiplicative* error model is linked to least squares percentage regression in the same way that the standard additive error model is linked to ordinary least squares regression.

Key words: Regression, error measures, relative error, percentage regression, weighted least squares, multiplicative error, heteroscedasticity.

Introduction

When a regression model is used for prediction the size of the error is of interest. The magnitude of an error is not meaningful in isolation – it needs to be viewed in relation to the size of the observed or actual value. Percentage errors are often used for this purpose. Our definition of percentage error is $100 \times$ (observed value − predicted value)/(observed value), as used in the fields of forecasting and time series analysis. In traditional least squares regression, an error of one unit is treated equally whether the dependent variable has a value of ten or a hundred, even though in percentage terms an error of one in ten would usually be considered more serious than an error of one in a hundred. In this article the method of least squares regression will be adapted to deal with percentage errors. There is a separate body of literature dealing with minimizing the mean absolute percentage error (MAPE), e.g. Narula & Wellington, 1977. This suffers from at least two deficiencies: (1) there is

no formula for the coefficients (one must solve a linear programming problem to find them), and (2) the resulting parameter estimates may not be unique. The method presented in this article does not have these drawbacks.

It is important to highlight a difference between the above definition of relative error vs. (observed value − predicted value)/(predicted value). The latter was used by Book and Lao (1999) and Goldberg and Touw (2003). The question is: Should we compare the error with the actual observed value or with the value predicted from the model? The following may be one way of choosing. When dealing with a controlled scientific situation where the functional form of the underlying theoretical model is known, then any departures from the predictions may be due to measurement error; in this case, it may make sense to consider the error relative to the predicted value. If however, the 'true' underlying model or all its constituent variables are unknown then the 'true' value is also unknown and we recommend the approach taken here.

This is the usual situation in finance, economics, psychology and the other social sciences. For example, when forecasting the value of investments traded on the stock market it makes sense to relate prediction errors to the observed values. The same argument usually applies in the area of cost estimation. The people

Chris Tofallis is a Senior Lecturer in Operational Research at The Business School. He is a member of The Royal Statistical Society, and the Operational Research Society. Email: C.Tofallis@herts.ac.uk

paying the costs will find it more meaningful to assess the predictive ability of a cost-estimating relation (CER) using the error relative to what they actually paid, not relative to what the model predicted. Similarly, a prediction that a salary bonus would be $10k, but which actually turned out to be $5k corresponds to an error of 100% by the definition used in this article, whereas the other definition would rate this as only a 50% error in prediction.

The definition of relative error used here also has computational advantages over the other form. The minimization of the sum of squares of the other form cannot be solved exactly because the normal equations are a nonlinear system. Book and Lao (1999) noted numerical optimization techniques are usually necessary to find the coefficients; they pointed out that due to multiple local minima unreasonable solutions must be excluded, and that the is most plausible solution physically selected. Moreover, the resulting estimators are inconsistent. Goldberg and Touw (2003) explained the reason for this: "simply inflating the predictions in the denominator [of the relative error] will tend to deflate the percentage errors, at the expense of worsening the fit" (p. 62). This problem does not arise if the standard definition of relative error is used.

Before deriving the necessary equations for the coefficients, alternative approaches will be considered. Consider the simple case where a scatter plot of the data indicates that fitting a straight line ($y = a + bx$) is appropriate. One suggestion might be to use logarithms in the following way: regress $\ln(y)$ against x. The trouble with this is that the resulting model would <u>not</u> be a linear relationship between y and x, instead it would have $\ln(y)$ linearly related to x, and so y would be exponentially related to x. Although this does correspond to a straight line when the exponent is zero, the slope of the line is forced to be zero.

It is in fact a common misconception that regressing $\ln(y)$ is equivalent to minimizing the squared relative errors; it is approximately true only if all the errors are small, as then $\ln(\hat{y}/y) \approx (\hat{y}/y) - 1$. The difference in these regression models will be illustrated with a numerical example below.

Regressing $\ln(y)$ on $\ln(x)$, the fitted model is:

$$\ln(y) = A + B \ln(x),$$

hence,

$$y = \exp[A + B \ln(x)] = \exp(A)\, x^B,$$

which is a power law. For the case $B = 1$ this does correspond to a line, but it is forced to have a zero intercept and so passes through the origin.

Thus, both of these approaches involving log transformations are inadequate because they depart from a linear model in the original variables, which is our assumed starting point. Another suggestion might be to regress $\ln(y)$ on $\ln(a + bx)$. This is a non-linear problem requiring iterative computational procedures. By contrast, in the proposed approach exact expressions for the coefficients are available.

Derivation of Formulae for the Coefficients

An exact expression is now derived for the coefficients for percentage least squares regression. Let X be a matrix in which each column contains the data for one of the explanatory variables, and the first column contains the value 1 in each position. The aim is to obtain a coefficient b_i for each column variable, and the coefficient associated with the first column will be the constant.

The values of the dependent variable are contained in a column vector y, which is assumed strictly positive. The data in the ith row of the matrix is associated with the ith element of the y vector.

Traditionally, the sum of squared errors would be minimized, $e^T e$, where e denotes the vector of errors, $y - Xb$. (Superscript T denotes the transpose.) However, the primary interest is in the relative errors r (percentage error = 100 times relative error), so each error e_i needs to be divided by y_i, so $r_i = e_i/y_i$. Carrying out this division on the form $y - Xb$ requires that the ith row of X be divided by y_i. This is achieved using the form $r = Dy - DXb$, where D is an *n by n* diagonal division matrix containing the value $1/y_i$ in the ith diagonal position and zeros elsewhere. D can be viewed as a matrix of weights.

Minimizing the sum of squares of relative errors $\sum r_i^2$, which, in vector notation, becomes

$$r^T r = (Dy - DXb)^T (Dy - DXb)$$
$$= (Dy)^T Dy - (Dy)^T DXb$$
$$- (DXb)^T Dy + b^T X^T D^2 Xb$$

To find the minimum, differentiate this with respect to b and equate to zero:

$$-(Dy)^T Dx + X^T D^2 Xb = 0$$

This is the matrix equivalent of the normal equations of ordinary least squares regression. Notice that these equations have the great convenience of being linear in *b* and so can be easily solved.

Rearranging the previous equation:
$$X^T D^2 Xb = (DX)^T Dy,$$
and thus
$$b = (X^T D^2 X)^{-1} (DX)^T Dy$$
$$= (X^T D^2 X)^{-1} X^T D^2 y \qquad (1)$$

It seems that this formula for the coefficients has not previously appeared as a solution for relative least squares.

If a spreadsheet is used for the calculations, the vector *b* can be computed directly using the matrix functions MINVERSE, MMULT (to multiply) and TRANSPOSE.

To satisfy the second order condition for a minimum, the second derivative of $r^T r$ with respect to b must be positive definite. This derivative equals $X^T D^2 X$ or $(DX)^T DX$. This square matrix will be positive definite if the columns of DX are linearly independent. Thus, the required *unique* minimum is obtained provided that no column of DX is expressible as a linear combination of the remaining columns.

If (1) is compared with the expression for ordinary least squares coefficients: $(X^T X)^{-1} X^T y$, observe that X has been replaced by DX, and y has been replaced by the vector Dy. Thus, D acts as a matrix of weights, as discussed further below.

In Ferreira et al. (2000)'s important article on relative least squares regression, expressions are derived for the coefficients, and also for their variance. They pointed out the connection between weighted least squares and relative least squares. Their formulae for the coefficients are in terms of ratios of determinants. These are less compact and less computationally convenient than the above formula (1), because a separate matrix has to be set up for each coefficient. A more practical computational method will be shown that can be applied using any standard software regression routine.

The consistency properties of relative least squares coefficients have been studied by Khoshgoftaar, et al. (1992). Using mild non-distributional assumptions such as independent error terms, a finite value for the expected measure of goodness of fit, and compact coefficient space, they prove that the coefficients are strongly consistent. That is, apart from a set of probability-measure zero, the coefficients will converge to the true values as the sample size increases.

Park and Stefanski (1998) also studied the best mean squared relative error prediction of y given x. Rather than provide formulae for coefficients, they assumed that some underlying distribution for y is given, and derive an expression for the predictor in terms of conditional inverse moments:

$$\hat{y} = E[y^{-1} \,|\, x] / E[y^{-2} \,|\, x].$$

They applied this using the lognormal and gamma distributions. They also showed that the mean squared relative prediction error is

$$\text{var} \, (y^{-1} \,|\, x) / E[y^{-2} \,|\, x].$$

Observe that in their experience "engineers often think in terms of relative error" (p. 227), and that they were motivated to explore relative least squares by a consulting problem with environmental engineers, who "citing engineering and political reasons, were steadfast in their dissatisfaction with the usual prediction methods, that too frequently resulted in unacceptably large relative errors. They wanted a "simple, easily implemented, and generally applicable approach to predicting" (p. 228). Park and Shin (2005) applied this to stationary ARMA time series.

Returning to (1) for b and focusing on the simple straight-line case, it follows from the above that the slope for percentage regression is given by

$$b = \frac{\Sigma \frac{x}{y} \Sigma \frac{1}{y^2} - \Sigma \frac{1}{y} \Sigma \frac{x}{y^2}}{\Sigma \frac{1}{y^2} \Sigma \frac{x^2}{y^2} - (\Sigma \frac{x}{y^2})^2} \qquad . \qquad (2)$$

(Note: all summations are from 1 to n, where n is the number of data points.) The intercept is given by

$$a = \frac{\Sigma \frac{1}{y} - b\Sigma \frac{x}{y^2}}{\Sigma \frac{1}{y^2}} \qquad . \qquad (3)$$

The normal equation arising from differentiating $r^T r$ with respect to the intercept can be written in the form

$$\Sigma \frac{e}{y^2} = 0 \qquad . \qquad (4)$$

This expression informs that the mean weighted error is zero if the weights are $1/y^2$. In vector terms this corresponds to $E[D^2 e] = 0$. From (3)

$$\Sigma \frac{1}{y_i} = a\Sigma \frac{1}{y_i^2} + b\Sigma \frac{x_i}{y_i^2}$$

it follows that there is a point through which the line will always pass (this would be the centroid of the data when using the ordinary least squares line). This is the point with coordinates given by

$$x = \frac{\Sigma \frac{x_i}{y_i^2}}{\Sigma \frac{1}{y_i^2}} \qquad\qquad y = \frac{\Sigma \frac{1}{y_i}}{\Sigma \frac{1}{y_i^2}}$$

Easy Computation by Transforming the Model Equation

Consider the model equation $y_i = a + bx_i + e_i$ and divide through by y_i, this yields

$$1 = \frac{a}{y_i} + b\frac{x_i}{y_i} + \frac{e_i}{y_i}. \qquad (5)$$

If ordinary least squares is used to regress the constant left hand side on the first two terms on the right, (notice there is now no constant term), then once again we are minimizing the sum of squared relative errors $\Sigma(e_i/y_i)^2$. Therefore, the same coefficients are derived, and the residuals will be the relative errors. This is a more convenient method of estimation, as even the Excel spreadsheet regression tool (part of the Analysis Toolpack) has the option to hold the constant to zero. Naturally, the above estimation approach carries over to the case of multiple explanatory variables.

The regression represented by (5) can be viewed as a novel form of weighted least squares with weights $1/y$. Weighted least squares is a standard way of dealing with unequal variances (heteroscedasticity). In econometrics, for example, the heteroscedasticity problem has been dealt with by using weights which are a function of *one* of the explanatory variables and so some element of trial and error has been required to select this variable. (See, for example Greene, 2003, section 11.5). However, in this treatment it is not necessary to be concerned with choosing from the explanatory variables for the transformation, because the single dependent variable is used instead.

Saez and Rittmann (1992) carried out Monte Carlo investigations of relative least squares regression where the y-data does not have constant variance but does have constant *relative* variance. By using generated data they could compare estimated parameters with the known values from the generating model. They found that the 90% confidence regions for the coefficients were approximately centered on the true values, whereas this was not the case for ordinary least squares. The OLS confidence regions did not even always include the true values. The relative least squares confidence regions were also much smaller than those for

OLS. They concluded that relative least squares is superior to OLS for such heteroscedastic data.

Analysis of Relative Variance and Goodness of Fit

In ordinary least squares the disturbance term is orthogonal to each of the explanatory variables. From (5) the equivalent orthogonal relations for our weighted regression are:

$$\Sigma \frac{e_i}{y_i^2} = 0 \quad \text{and} \quad \Sigma \frac{e_i \, x_i}{y_i^2} = 0$$

The disturbance term is also orthogonal to the predicted dependent variable, which in this case corresponds to $\hat{y}_i/y_i$. Therefore

$$\Sigma \frac{e_i \, \hat{y}_i}{y_i^2} = 0 \quad \text{i.e.} \quad \Sigma \frac{\hat{y}_i}{y_i}\left(1 - \frac{\hat{y}_i}{y_i}\right) = 0 \quad (6)$$

Define the relative variance as:

$$\frac{1}{n}\Sigma \; (\frac{y - \bar{y}}{y})^2$$

Ignoring the 1/n , this can be written as

$$\Sigma \frac{(\hat{y} - \bar{y} + y - \hat{y})^2}{y^2} =$$

$$\Sigma \frac{(\hat{y} - \bar{y})^2}{y^2} + \Sigma \; \frac{(y - \hat{y})^2}{y^2}$$

$$+ \Sigma \frac{(\hat{y} - \bar{y})(y - \hat{y})}{y^2}$$

The final term in the previous expression is zero as a consequence of the normal equations above.

Total relative variation =
Explained relative variation + Unexplained relative variation,

which is a decomposition of the relative variance.

A statistic can now be defined to measure the goodness of fit of our model, akin to r^2. The coefficient of relative determination is the ratio

$$\frac{\text{Explained relative variation}}{\text{Total relative variation}}$$

This ratio gives the proportion of the relative variation that is explained by the model. It will have a value in the range zero to one.

A Note on Measurement Scale

If all values of the dependent variable are re-scaled by multiplying by a positive constant, then the percentage errors remain unchanged. Consequently the resulting percentage least squares model will be equivalent to the original model, and it will provide equivalent predictions. For example if the y-variable is multiplied by 10 (e.g. due to conversion from centimeters to millimeters), then all coefficients in the fitted model equation will also be multiplied by 10.

If however, a constant is added to each value of the dependent variable then the percentage errors will not be the same as before. In this case the model fitted using percentage least squares will not be equivalent to the previously estimated model. The situation is exemplified when speaking of percentage changes in Fahrenheit temperature and percentage changes measured on the Celsius scale. The two are not the same because these scales do not share a common zero point. The dependent variable needs to be measured on a ratio scale when using percentage regression. This is because a percentage is not meaningful if one is permitted to shift the zero of the scale.

Maximum Likelihood

Is there a distribution for which the above estimators are maximum likelihood estimators? Consider the following multiplicative representation

$$y = X\beta u \qquad (7)$$

where u is multiplicative error factor, as opposed to an additive error term. Obviously, the expected value of u is desired to be unity, and thus the choice of the symbol *u*. E[y] = Xβ is desirable, so assume that the error factor is independent of the explanatory variables so that

$$E[y] = E[X\beta] \; E[u] = E[X\beta] = y$$

so that the estimate of the mean response will be unbiased.

Define $v_i = 1/u_i$. Once there is an estimator b then the conditional estimate of the mean of y is $\hat{y} = Xb$, then

$$v_i = E[y_i]/y_i \qquad (8)$$

An error is indicated by this accuracy ratio differing from unity. Notice that $1 - v_i = r_i$, which is the relative error. Assume that the relative error is normally distributed with mean zero and constant variance (σ^2). This implies that v is normally distributed with mean value unity and constant variance (σ^2). [See the Appendix for the implications regarding the conditional distribution of y.] From (8), for any given x_i there is a one to one relationship between y and v. For a given data sample the likelihood function in terms of v is given by

$$\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{\sum_{i=1}^{n}(v_i - 1)^2}{2\sigma^2} \right]$$

and in terms of y, the negative of the log likelihood becomes

$$\frac{\sum_{i=1}^{n}(\frac{X\beta}{y_i} - 1)^2}{2\sigma^2} + \frac{n}{2}\ln\sigma^2 + \frac{n}{2}\ln(2\pi) \qquad (9)$$

The summand is the square of the relative error, so it is now apparent that if the coefficient values are chosen to maximize the log likelihood, the same estimates for the coefficients as in (1) are obtained. The result is that when the relative error is normally distributed $N(1,\sigma^2)$ then the least squares percentage regression estimators are maximum likelihood estimators.

It is also possible to estimate σ^2 in the same way by differentiating the log likelihood

with respect to σ^2 and setting the derivative to zero:

$$\frac{\sum_{i=1}^{n}(\frac{X\beta}{y_i} - 1)^2}{2\sigma^4} - \frac{n}{2\sigma^2} = 0$$

If the estimators are substituted for β, the following is obtained as the estimator for σ^2

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(\frac{Xb}{y_i} - 1)^2$$

From (9) the log likelihood contains the sum of squared relative errors. If these are independent and identically distributed then for large n, the central limit theorem can be applied. This can be used to deduce a confidence interval for the coefficients.

Unbiasedness.
The estimator for β can be shown to be unbiased as follows. From (1)

$$\begin{aligned} E[b] &= E[(X^T D^2 X)^{-1} X^T D^2 \, y] \\ &= E[(X^T D^2 X)^{-1} X^T D^2 \, X \, \beta u] \\ &= E[\beta u] \qquad . \end{aligned}$$

Assuming that the error factor is independent of β, we have: $E[b] = E[\beta] \; E[u] = E[\beta] = \beta$. Hence b is an unbiased estimator of β.

Example. The following table gives the sales figures from 18 different US industries, as well as the expenditure on research and development (millions of dollars). The sales variable has a wide range, and so it is likely that observations near the upper end will dominate over those at the lower end in positioning the regression line; this is because residuals for high sales are likely to be much larger. The correlation between the variables is 0.69 and a scatter plot shows evidence of heteroscedasticity.

Table 1. Sales and research & development expenditure in millions of dollars for 18 US industries.

Sales	R & D Expenses
6375	62.5
11626	92.9
14655	178.3
21869	258.4
26408	494.7
32406	1083
35108	1620.6
40295	421.7
70762	509.2
80553	6620.1
95294	3918.6
101314	1595.3
116141	6107.5
122316	4454.1
141650	3163.8
175026	13210.7
230614	1703.8
293543	9528.2

Source: Gujarati, 2003, page 424. Originally published in *Business Week* 1989.

If ordinary least squares is applied with sales as the dependent variable, the following model is obtained:

$$Sales = 43942 + 15.00 \ R\&D,$$

with p-values of 0.03 and 0.0015 for the intercept and slope respectively.

Consider the absolute percentage error (APE), defined as the residual expressed as a percentage of the observed value. The above model has a mean absolute percentage error (MAPE) of 105%, which is very poor. In fact three of the 18 industries have APEs exceeding 200%. The largest APEs occur for those industries which have low sales.

Some analysts advise taking logs of the dependent variable if one is interested in reducing percentage errors. If ordinary least squares is conducted, the following model is obtained:

$$Ln(Sales) = 10.341 + 0.000198 \ R\&D$$

with p-values of 0.002 for the slope and essentially zero for the intercept. If the exponential is taken, it is possible to predict sales and calculate percentage errors. The MAPE is then 76%, which is an improvement. However, there are four industries with an APE exceeding 100%, three of these are at the lower end of the sales range.

Finally, consider our approach of minimizing the squared percentage residuals. The resulting model is found to be:

$$Sales = 8817 + 17.88 \ R\&D$$

with p-values of 0.002 and 5×10^{-5} for the slope and intercept respectively.

The MAPE is now 38.5%. This is a large improvement as it is actually half of the percentage error from the log model. No residuals exceeded 100%, in fact the largest residual was 83%. The differences with the log model are worth emphasizing because it is a common misconception among statisticians that taking logs is equivalent to minimizing percentage errors. As mentioned in the introduction, this is true only in the limit as the residuals tend to zero.

Conclusion

Percentage error (relative to the observed value) is often felt to be more meaningful than the absolute error in isolation. The mean absolute percentage error (MAPE) is widely used in forecasting as a basis of model comparison, and regression models can be fitted which minimize this criterion. Unfortunately, no formula exists for the MAPE coefficients, and models for a given data set may not be unique. I have instead explored least squares regression based on the percentage error. I was able to derive exact expressions for the regression coefficients when the model is linear in these coefficients. Another advantage over MAPE is that this solution is unique.

The percentage errors are defined relative to the observed values. This is the standard definition of percentage error used in forecasting. When making predictions it usually

makes more sense to relate the size of the error to the actual observation to measure its relative size. This is a departure from some of the existing literature on relative error least squares regression (e. g., Book & Lao (1999), & Goldberg & Touw, 2003), where the error relative to the *predicted* value has been used. The latter approach suffers on two counts. First, because the predicted values appear in the denominator of the fitting criterion, the latter value can be improved by inflating the predicted values, despite the fact that this worsens the fit (i.e., it gives biased estimates). Second, even for a linear model, estimation requires iteratively re-weighted least squares, which is computationally more demanding.

It has been shown that the proposed method is equivalent to a form of weighted least squares where, unusually, the weights depend on the dependent variable. This connection allowed us to develop a form which has great ease of computation. Indeed the models are attractive to the practitioner because they can easily be fitted using standard spreadsheet software. In comparing ordinary least squares with percentage least squares, the key difference is that the former ignores how large the residual is relative to the quantity being predicted, whereas the latter takes this into account. I believe that this method will be of use when the dependent variable has a wide range, as then the residuals at the upper end would dominate if ordinary least squares were used, unless the error variance is constant, which is often not the case in such situations.

It has also been shown that for a normally distributed multiplicative error model the least squares percentage estimators are maximum likelihood estimators. In short, the multiplicative error model is linked to least squares percentage regression in the same way that the standard additive error model is linked to ordinary least squares regression.

References

Book, S. A., & Lao, N. Y. (1999). Minimum-percentage-error regression under zero-bias constraints. *Proceedings of the 4th US Army Conference on Applied Statistics 1998*, US Army Research Lab report no. ARL-SR-84, pages 47-56.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis*. 3rd edition. New York: Wiley.

Ferreira, J. M., Caramelo, L., & Chhabra, R. P. (2000). The use of relative residues in fitting experimental data: an example from fluid mechanics. *International Journal of Mathematical Education in Science and Technology*, 31(4), 545-552.

Goldberg, M. S., & Touw, A. E. (2003). *Statistical methods for learning curves and cost analysis*. Institute for Operations Research and the Management Sciences, Linthicum, MD.

Greene, W. H. (2003). *Econometric analysis*. 5th edition. New Jersey: Prentice Hall.

Gujarati, DN. (2003). *Basic econometrics*. 4th edition. NY: McGraw Hill

Khoshgoftaar, T. M., Bhattacharyya, B. B., & Richardson, G. D. (1992). Predicting software errors, during development, using nonlinear regression models: A comparative study. *IEEE Transactions on Reliability*, *41*(3) 390-395.

Narula, S. C., & Wellington, J. F. (1977). Prediction, linear regression and the minimum sum of relative errors. *Technometrics*, *19*(2), 185-190.

Park, H., & Shin, K-I. (2005). A shrinked forecast in stationary processes favouring percentage error. *Journal of Time Series Analysis*, *27*(1), 129-139.

Park, H., & Stefanski, L. A. (1998). Relative-error prediction. *Statistics and Probability Letters, 40*, 227-236.

Saez, P. B., & Rittmann, B. E. (1992). Model-parameter estimation using least squares. *Water Research*, *26*(6), 789-796.

Appendix: The distribution of y when the relative error is normally distributed

In deducing the maximum likelihood estimates, assume for a given x-value that the *relative* error ($r_i = 1 - \mu_y/y_i$) is normally distributed, $N(0, \sigma^2)$. Consider the implication for the conditional distribution of y; from (8) $r_i = 1 - v_i = 1 - \mu_y/y_i$ and $v_i \sim N(1, \sigma^2)$. The conditional value of y should therefore follow the reciprocal normal distribution (not to be confused with the inverse normal). Specifically, use the change of variable rule to deduce the distribution of y_i for a given x_i (Greene, 2003, Appendix B6). This gives the following distributional form:

$$\frac{\mu_y}{y^2 \sigma \sqrt{2\pi}} \exp\left[-\frac{\left(\dfrac{\mu_y}{y} - 1\right)^2}{2\sigma^2} \right] ,$$

where σ is the standard deviation of the relative error, here assumed to have mean value unity. Figure 1 charts this density function for two values of σ.
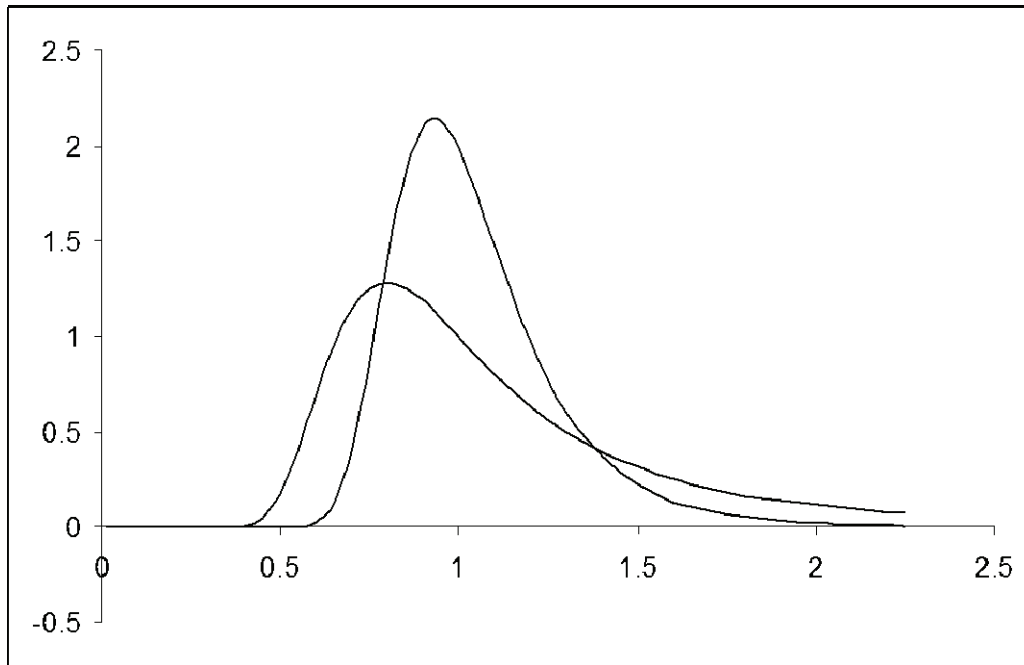

Figure 1. Probability density of y when the relative error is normally distributed with mean unity and $\sigma = 20\%$ (taller curve) and $\sigma = 40\%$ (shorter curve) .