

Least Squares Percentage Regression

University of Hertfordshire Business School Working Paper
UHBS 2006: 6 (revised 2007)

Dr. Chris Tofallis
The Business School
University of Hertfordshire
College Lane
Hatfield
Herts
AL10 9AB
United Kingdom

Email: C.Tofallis@herts.ac.uk

ABSTRACT

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 comparison, and regression models can be fitted which minimize this criterion. Unfortunately, no formula exists for the coefficients, and models for a given data set may not be unique. We instead explore least squares regression based on the percentage error. We are able to derive exact expressions for the regression coefficients when the model is linear in these coefficients. Another advantage of this approach over MAPE is that the solution is unique. Furthermore, it has been shown that this approach provides strongly consistent coefficient estimates, and is superior to ordinary least squares when the data does not possess constant variance.

For the practitioner we demonstrate that percentage regression models can easily be fitted using ordinary regression software, as well as by spreadsheets, using simple transformation of the data.

Finally, we show that when the relative error is normally distributed, least squares percentage regression provides maximum likelihood estimates.

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

1. INTRODUCTION

When a regression model is to be used for prediction, the size of the error is obviously of interest. However, 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. We shall use the following definition of percentage error: $100 \times (\text{observed value} - \text{predicted value}) / (\text{observed value})$. This is the standard definition that is used in the field of forecasting/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 - despite the fact that in percentage terms an error of one in ten would usually be considered more serious than an error of one in a hundred. For this reason we intend to adapt the method of least squares regression so that it deals with percentage errors. We note that there is a separate body of work which deals with minimizing the mean absolute percentage error (MAPE), (see for example Narula and Wellington, 1977). This suffers from at least two deficiencies: there is no formula for the coefficients (one must solve a linear programming problem to find them), and the resulting parameter estimates may not be unique. The work that we present here does not have these drawbacks.

It is important to highlight a difference between the standard definition of relative error that we shall use here, and another, namely: $(\text{observed value} - \text{predicted value}) / (\text{predicted value})$. The latter has been 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 the value predicted from the model? The following may be one way of choosing. If one is 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 instead, we are unsure what the underlying model is, or even whether we have included all explanatory variables, then we cannot know what the ‘true’ value should be. This is the norm 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 who are 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, on an individual level, 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 paper,

whereas the other definition would rate this as only a 50% error in prediction.

The definition of relative error that we use also has computational advantages over the other form. The minimisation of the sum of squares of the other form cannot be solved exactly. Indeed the normal equations now become a nonlinear system. Book and Lao (1999) have studied this, and note that numerical optimization techniques are usually necessary to find the coefficients; they point out that due to multiple local minima ‘unreasonable solutions must be excluded, and the solution that is most plausible “physically” selected’. Moreover, the resulting estimators are inconsistent. Goldberg and Touw (2003, p.62) explain 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’. This problem does not arise if we use the standard definition of relative error.

Before proceeding to derive the necessary equations for the coefficients, we shall look at approaches which some might think would achieve the same effect as what we propose to do.

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 *not* 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 . Admittedly this does correspond to a straight line when the exponent is zero, but the slope of the line is forced to be zero. It is in fact a common misconception that regressing $\ln(y)$ is equivalent to minimising the squared relative errors; it is approximately true only if all the errors are small, as then $\ln(\hat{y}/y) \approx (\hat{y}/y) - 1$.

Now let’s instead try regressing $\ln(y)$ on $\ln(x)$. The fitted model will be:

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

$$\text{hence } y = \exp[A + B \ln(x)] = \exp(A) x^B$$

which is again not a straight line but a power law. For the case $B = 1$ this does correspond to a line, but this time it is forced to have a zero intercept and so passes through the origin.

Hence, 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, what we shall propose is much more elegant because exact expressions for the coefficients are available.

2. DERIVATION OF FORMULAE FOR THE COEFFICIENTS

We now derive exact expressions for the coefficients for percentage least squares regression. Let \mathbf{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. We aim 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 \mathbf{y} , which is assumed strictly positive. The data in the i 'th row of the matrix is associated with the i 'th element of the \mathbf{y} vector.

Traditionally we would minimise the sum of squared errors, $\mathbf{e}^T \mathbf{e}$, where \mathbf{e} denotes the vector of errors, $\mathbf{y} - \mathbf{X}\mathbf{b}$. (Superscript T denotes the transpose.) However, we are interested in the relative errors \mathbf{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 $\mathbf{y} - \mathbf{X}\mathbf{b}$ requires that the i 'th row of \mathbf{X} be divided by y_i ; this is achieved using the form $\mathbf{r} = \mathbf{D}\mathbf{y} - \mathbf{D}\mathbf{X}\mathbf{b}$ where \mathbf{D} is an n by n diagonal 'division' matrix containing the value $1/y_i$ in the i 'th diagonal position and zeros elsewhere. \mathbf{D} can be viewed as a matrix of weights.

Now we wish to minimise the sum of squares of relative errors $\sum r_i^2$ which, in vector notation, becomes

$$\begin{aligned} \mathbf{r}^T \mathbf{r} &= (\mathbf{D}\mathbf{y} - \mathbf{D}\mathbf{X}\mathbf{b})^T (\mathbf{D}\mathbf{y} - \mathbf{D}\mathbf{X}\mathbf{b}) \\ &= (\mathbf{D}\mathbf{y})^T \mathbf{D}\mathbf{y} - (\mathbf{D}\mathbf{y})^T \mathbf{D}\mathbf{X}\mathbf{b} - (\mathbf{D}\mathbf{X}\mathbf{b})^T \mathbf{D}\mathbf{y} + \mathbf{b}^T \mathbf{X}^T \mathbf{D}^2 \mathbf{X}\mathbf{b} \end{aligned}$$

To find the minimum we differentiate this with respect to \mathbf{b} and equate to zero:

$$-(\mathbf{D}\mathbf{y})^T \mathbf{D}\mathbf{x} + \mathbf{X}^T \mathbf{D}^2 \mathbf{X}\mathbf{b} = 0$$

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 \mathbf{b} and so can be easily solved.

Rearranging the previous equation: $\mathbf{X}^T \mathbf{D}^2 \mathbf{X}\mathbf{b} = (\mathbf{D}\mathbf{X})^T \mathbf{D}\mathbf{y}$, for which the solution is given by:

$$\mathbf{b} = (\mathbf{X}^T \mathbf{D}^2 \mathbf{X})^{-1} (\mathbf{D}\mathbf{X})^T \mathbf{D}\mathbf{y} = (\mathbf{X}^T \mathbf{D}^2 \mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}^2 \mathbf{y} \quad (1)$$

To our knowledge this formula has not previously appeared as a solution for relative least squares.

For those who use spreadsheets for their calculations, the vector \mathbf{b} can be easily computed directly using the matrix functions MINVERSE, MMULT (to multiply) and TRANSPOSE.

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

If we compare (1) with the expression for ordinary least squares coefficients: $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$, we observe that \mathbf{X} has been replaced by $\mathbf{D}\mathbf{X}$, and \mathbf{y} has been replaced by the vector $\mathbf{D}\mathbf{y}$. Thus \mathbf{D} acts as a matrix of weights – we discuss this in a later section.

In the important paper by Ferreira et al (2000) on relative least squares regression, expressions are derived for the coefficients, and also for their variance. They point 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. In any case, we shall show a more practical computational method than either of these in a later section – one 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.

In a different vein, Park and Stefanski (1998) have also studied the best mean squared relative error prediction of y given x . Rather than provide formulae for coefficients, they assume 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 then apply this using the lognormal and gamma distributions. They also show that the mean squared relative prediction error is:

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

It is worth observing that in their experience “engineers often think in terms of relative error”, 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”. In a separate paper, Park and Shin (2005) apply the above to stationary ARMA time series.

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

$$\mathbf{b} = \frac{\sum \frac{x}{y} \sum \frac{1}{y^2} - \sum \frac{1}{y} \sum \frac{x}{y^2}}{\sum \frac{1}{y^2} \sum \frac{x^2}{y^2} - (\sum \frac{x}{y^2})^2} \quad (2)$$

(Note: all summations in this paper are from 1 to n , where n is the number of data points.)

The intercept is given by

$$\mathbf{a} = \frac{\sum \frac{1}{y} - b \sum \frac{x}{y^2}}{\sum \frac{1}{y^2}} \quad (3)$$

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

$$\sum \frac{e}{y^2} = 0 \quad (4)$$

This expression tells us that the mean weighted error is zero if the weights are $1/y^2$. In vector terms this corresponds to $E[\mathbf{D}^2 \mathbf{e}] = 0$.

From (4)

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

from which 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

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

3. EASY COMPUTATION BY TRANSFORMING THE MODEL EQUATION

Suppose we take the model equation $y_i = a + bx_i + e_i$ and divide through by y_i , we obtain

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

If we then use ordinary least squares to regress the constant left hand side on the first two terms on the right, (notice there is now no constant term), then we shall once again be minimising the sum of squared relative errors $\Sigma(e_i/y_i)^2$. Therefore, we shall again obtain the same coefficients. This is a much 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). Traditionally in some disciplines (e.g. econometrics), 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 variable has been required to select this variable. (See, for example, Greene 2003, section 11.5). However, in our treatment, one no longer needs to be concerned with choosing from the explanatory

variables for the transformation – the single dependent variable is used instead.

Saez and Rittmann (1992) have 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 centred on the true values, whereas this was not the case for ordinary least squares – in fact 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 was superior to OLS for such heteroscedastic data.

4. 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:

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

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

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

Let us define the relative variance as:

$$\frac{1}{n} \sum \left(\frac{y - \bar{y}}{y} \right)^2$$

Ignoring the $1/n$, this can be written as

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

The final term in the previous expression is zero as a consequence of the normal equation (6). We thus have:

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

which is a decomposition of the relative variance.

From this we can now define a statistic to measure the goodness of fit of our model, akin to r^2 :

thus 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. One can see that this will have a value in the range zero to one.

5. 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 centimetres to millimetres), 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 akin to 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. In short, 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.

6. MAXIMUM LIKELIHOOD

We now deal with the following question: is there a distribution for which the above estimators are maximum likelihood estimators?

Consider the following multiplicative representation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}\mathbf{u} \quad (7)$$

where \mathbf{u} is *multiplicative error factor*, as opposed to an additive error term. Obviously, we want the expected value of \mathbf{u} to be unity – hence the choice of the symbol u . We would like to have $E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$, so we assume that the error factor is independent of the explanatory variables so that $E[\mathbf{y}] = E[\mathbf{X}\boldsymbol{\beta}] E[\mathbf{u}] = E[\mathbf{X}\boldsymbol{\beta}] = \mathbf{y}$ so that the estimate of the mean response will be unbiased.

Let us define $v_i = 1/u_i$

Once we have an estimator \mathbf{b} , our conditional estimate of the mean of y is $\hat{\mathbf{y}} = \mathbf{X}\mathbf{b}$, then

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

An error is indicated by this ratio differing from unity. Notice that $1 - v_i = r_i$, which is the relative error. We now 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 \mathbf{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

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

The summand is the square of the relative error, so it is now apparent that if we choose the coefficient values which maximise the log likelihood we shall obtain the same estimates for the coefficients as we had previously in (1).

Thus we have the important result that when the relative error is normally distributed $N(1, \sigma^2)$ the least squares percentage regression estimators are maximum likelihood estimators.

We can estimate σ^2 in the same way by differentiating the log likelihood with respect to σ^2 and setting the derivative to zero:

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

If we also substitute our estimators for β , we obtain the following as our estimator for σ^2

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

From (9) the log likelihood contains the sum of squared relative errors. If these are independent and identically distributed then for large n we can apply the central limit theorem. This can be used to deduce a confidence interval for the coefficients (see Lloyd 1984, page

7. UNBIASEDNESS

We now show that our estimator for β is unbiased. From (1)

$$E[b] = E[(\mathbf{X}^T \mathbf{D}^2 \mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}^2 \mathbf{y}] = E[(\mathbf{X}^T \mathbf{D}^2 \mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}^2 \mathbf{X} \beta u] = E[\beta u]$$

Assuming that the error factor is independent of β , we have:

$$E[b] = E[\beta] E[u] = E[\beta] = \beta. \text{ Hence } b \text{ is an unbiased estimator of } \beta.$$

8. CONCLUSION

This paper is aimed at those who wish to construct predictive models based on least squares regression where the user feels that reducing percentage errors is more important, more useful, or more meaningful than reducing absolute errors. We have derived, from first principles, exact expressions for the regression coefficients based on percentage errors for a linear model. These percentage errors are 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 and Lao (1999), and Goldberg and Touw (2003)) where the error relative to the *predicted* value has been used. The latter approach suffers on two counts. Firstly, because the predicted values appear in the denominator of the fitting criterion, its value can be improved by inflating the predicted values – despite the fact that this worsens the fit i.e. it gives biased estimates. Secondly, even for a linear model, estimation requires iteratively re-weighted least squares.

We have 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 had great ease of computation. Indeed the models can easily be fitted using standard spreadsheet software. In comparing ordinary least squares with percentage least squares, the key difference is that the latter places greater weight on smaller values of the dependent variable, whereas the former places greater weight on the larger values.

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

Appendix

The distribution of y when the relative error is normally distributed

In deducing the maximum likelihood estimates we assumed that the relative error ($r_i = 1 - \mu_y/y_i$) was normally distributed, $N(0, \sigma^2)$. What does this imply about the conditional distribution of y? From (8) we have $r_i = 1 - v_i = 1 - \mu_y/y_i$ and thus $v_i \sim N(1, \sigma^2)$. The conditional value of y should therefore follow the reciprocal normal distribution (not to be confused with the inverse normal). Specifically, we can use the change of variable rule to deduce the distribution of y_i for a given x_i (Greene, 2003, Appendix B6):

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

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 .

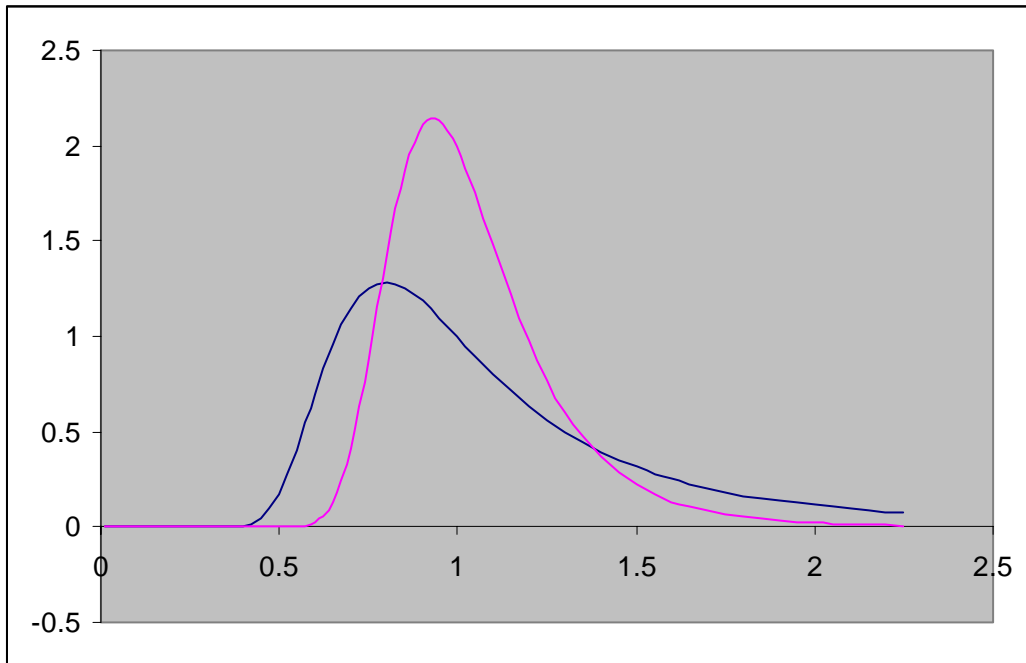


Figure 1. Probability density of y when the relative error is normally distributed with mean unity and $\sigma = 20\%$ (upper curve) and $\sigma = 40\%$.

Acknowledgements: I am grateful to Emeritus Professor MJR Healy, Dr. Paul Taylor, and Dr. Neil Spencer for sharing their statistical knowledge. Their comments on this work were most helpful. All errors remain my own responsibility.

REFERENCES

Book, SA and Lao, NY. (1999). Minimum-percentage-error regression under zero-bias constraints. *Procs. of the 4th US Army Conference on Applied Statistics 1998*, US Army Research Lab report no. ARL-SR-84, pages 47-56.

Draper, NR, and Smith, H. (1998). *Applied regression analysis*. 3rd edition. Wiley, New York.

Ferreira, JM, Caramelo, L, Chhabra, RP. (2000). The use of relative residues in fitting experimental data: an example from fluid mechanics. *Int. J. Math. Edu. Sci. Technol.* 31(4), 545-552.

Goldberg, MS and Touw, AE. (2003). *Statistical Methods for Learning Curves and Cost Analysis*. Inst. for Operations Research and the Management Sciences, Linthicum, MD.

Greene, WH. (2003). *Econometric Analysis*. 5th edition. Prentice Hall, New Jersey.

Khoshgoftaar, TM, Bhattacharyya, BB and Richardson, GD. (1992). Predicting software errors, during development, using nonlinear regression models: A comparative study. *IEEE Transactions on Reliability*, 41(3) 390-395.

Narula, SC and Wellington, JF. (1977). Prediction, linear regression and the minimum sum of relative errors. *Technometrics*, 19(2), 185-190.

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

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

Saez, PB and Rittmann, BE. (1992). Model-parameter estimation using least squares. *Water Research*, 26(6), 789-796.