Confidence Interval of the Residual of Diagnostic Parameter

REIJI MEZAKI

Department of Chemical Engineering
New York University, Bronx, New York 10453

In previous articles Hunter and Mezaki (1964), Kittrell and et al. (1966), and Hill and Mezaki (1967) showed that a residual analysis of a diagnostic parameter can reveal the inadequacy of a proposed mechanistic kinetic model and can also suggest a systematic modification to lead to a more appropriate model. In this analysis the sign of residual is vital. Thus, it was felt that the presentation of residual values with their confidence intervals provides a more definitive result concerning model adequacy. This is particularly true when the absolute value of residual is determined to be quite small.

In this communication, a method is presented to calculate the confidence interval of the residual of the diagnostic parameter. Then it is shown how to use the confidence intervals with the residuals to ensure the adequacy of a proposed model.

KINETIC MODEL AND ITS DIAGNOSTIC PARAMETER

In an article of Hunter and Mezaki (1964), Hougen-Watson rate models were postulated to describe the total catalytic oxidation of methane. Based upon the result of residual analysis of a diagnostic parameter, the following model was selected as appropriate: the surface reaction between gaseous methane and adsorbed oxygen is rate-controlling and the reaction produces both adsorbed carbon dioxide and adsorbed water. For this reaction a Hougen-Watson model is given by

$$r = \frac{k\,K_2\,x_1\,(1-y)\,(x_2-2x_1y)^2}{[1+\sqrt{K_2}\,(x_2-2x_1y)+K_3\,(x_3+x_1y)+K_4\,(x_4+2x_1y)]^3} \tag{1}$$

and a diagnostic parameter C_1 is defined as

$$C_1 = \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \tag{2}$$

where

$$\beta_0 = \frac{1}{\sqrt[3]{kK_2}} \qquad \beta_2 = \frac{\sqrt{K_2}}{\sqrt[3]{kK_2}} \qquad \beta_3 = \frac{K_3}{\sqrt[3]{kK_2}} \qquad \beta_4 = \frac{K_4}{\sqrt[3]{kK_2}} \tag{3}$$

The estimates of the coefficients β_0, β_2, β_3, and β_4 and subsequently $\hat{C}_1$ are readily determined through a linear regression provided that the experimental values of C_1 are available. Combining Equation (1) with the following relationship

$$r = \left(\frac{\partial w}{\partial y}\right)^{-1} \tag{4}$$

we get

$$C_1 = \left\{ x_1 x_2^2 \frac{\partial w}{\partial y}\bigg|_{y=0} \right\}^{1/3} \tag{5}$$

Using Equation (5) one can readily obtain the observed value of C_1.

RESIDUALS OF THE DIAGNOSTIC PARAMETER AND ITS CONFIDENCE INTERVALS

Now a case is considered where several experimental observations are available. For this case Equation (2) may be written in a matrix form

$$\mathbf{C}_1 = \mathbf{X}\,\boldsymbol{\beta} \tag{6}$$

where $\mathbf{C}_1$ is the vector of observed values of C_1, $\mathbf{X}$ is the matrix of independent variables x_2, x_3, and x_4, and $\boldsymbol{\beta}$ is the vector of coefficients β_0, β_2, β_3, and β_4. If $\mathbf{X'X}$ is a nonsingular matrix, then the residuals of $\mathbf{C}_1$ are given by

$$\mathbf{C}_1 - \hat{\mathbf{C}}_1 = (\mathbf{I} - \mathbf{X}(\mathbf{X'X})^{-1}\,\mathbf{X'})\,\mathbf{C}_1 \tag{7}$$

Draper and Smith (1966) showed that the variance-covariance matrix of residuals $V(\mathbf{e})$ is

$$V(\mathbf{e}) = V(\mathbf{C}_1 - \hat{\mathbf{C}}_1) = (\mathbf{I} - \mathbf{X}\,(\mathbf{X'X})^{-1}\,\mathbf{X'})\,\sigma^2 \tag{8}$$

where σ^2 is the experimental error variance which can be estimated from replicated experiments. Thus the confidence interval of the residual of the ith experiment $C_{1i} - \hat{C}_{1i}$ is given by

$$(C_{1i} - \hat{C}_{1i}) \pm t_{\nu,\alpha/2}\,\{V(\hat{e}_i)\}^{1/2} \tag{9}$$

where $V(\hat{e}_i)$ is the ith diagonal element of symmetrical matrix $(\mathbf{I} - \mathbf{X}\,(\mathbf{X'X})^{-1}\,\mathbf{X'})\,\hat{\sigma}^2$; $t_{\nu,\alpha/2}$ is the $(1-\alpha/2)\,100\%$ point of the t distribution with ν degrees of freedom. This is true provided that the basic assumptions are satisfied (see Draper and Smith, 1966):

1. e_i is a normally distributed random variable with mean zero and variance σ^2 (unknown).
2. e_i and e_j are uncorrelated, $i \neq j$, and by (1), are independent.

For the catalytic oxidation of methane which was studied by Hunter and Mezaki (1964) the levels of experi-

mental variables x_1, x_2, x_3, and x_4 were selected according to the 2^{4-1}_{III} fractional factorial design of Table 1. This table has been abstracted from Hunter and Mezaki. The residuals of the diagnostic parameter were computed and the results were presented elsewhere (Hunter and Mezaki, 1964; Kittrell et al., 1966). Table 2 of this note was taken from Table 4 of Hunter and Mezaki (1964). Comparison of Table 1 and Table 2 indicates that the residuals of C_1 do not appear to be correlated with the levels of any experimental variables and the signs of the residuals do not exhibit any pattern. However, it should be noted that the residual for the sixth experiment is comparatively small. Hence for this particular experiment the possibility of the other sign should be considered. If this sign had been positive, the residuals of the diagnostic parameter C_1 would have been perfectly correlated with the concentration levels of methane (see Table 1). This could imply that the concentration of methane was not adequately taken into account. In case that the true value of the sixth residual is indeed negative, the above conclusion is quite misleading. With the sole presentation of signs and values of residuals, however, such a situation as described above will be encountered. In this connection the calculation of the confidence interval of the diagnostic parameter allows the elimination of the inherent uncertainty of the residual analysis.

For the example illustrated in this communication the confidence interval of the diagnostic parameter C_1 has been computed through Equation (9). The result is shown in Table 2. For the calculation we used $\hat{\sigma}^2$ of 0.4287×10^{-6}, the detailed computation of which is given in Hunter and Mezaki. As can be seen from the table, the 95% confidence interval of each residual will not include zero provided the statistical assumptions are satisfied. Although they are probably not satisfied exactly, for problems of practical interest the results will still be nearly correct. For the sixth run, for example, we expect

Table 1. Experimental Design

Run	ξ_1	ξ_2	ξ_3	ξ_4
1	−1	−1	−1	+1
2	+1	−1	−1	−1
3	−1	+1	−1	−1
4	+1	+1	−1	+1
5	−1	−1	+1	−1
6	+1	−1	+1	+1
7	−1	+1	+1	+1
8	+1	+1	+1	−1

$$\xi_1 = \frac{x_1 - 0.015}{0.005} \qquad \xi_2 = \frac{x_2 - 0.120}{0.060}$$

$$\xi_3 = \frac{x_3 - 0.065}{0.035} \qquad \xi_4 = \frac{x_4 - 0.095}{0.055}$$

Table 2. Residuals of Diagnostic Parameter

Run	$C_1 \times 10^3$	$\hat{C}_1 \times 10^3$	$(C_1 - \hat{C}_1) \times 10^3$ (with 95% confidence interval)
1	15.57	15.99	-0.42 ± 0.01
2	11.94	11.03	$+0.91 \pm 0.03$
3	25.99	27.04	-1.05 ± 0.07
4	31.53	30.92	$+0.61 \pm 0.05$
5	13.56	14.08	-0.52 ± 0.04
6	18.52	18.55	-0.03 ± 0.01
7	34.03	34.14	-0.11 ± 0.06
8	29.87	29.26	$+0.61 \pm 0.07$

the residual of the diagnostic parameter C_1 developed from Equation (1) to fall between −0.02 and −0.04, its calculated 95% confidence interval. Hence the adequacy of the rate model given by Equation (1) is indeed supported.

Through a Hougen-Watson rate model for the catalytic oxidation of methane, there has been illustrated the application of confidence interval of the diagnostic parameter. The supplement of the interval appears to be of considerable help in selecting an adequate model and it is recommended.

NOTATION

C_1 = diagnostic parameter
$\hat{C}_1$ = predicted value of the diagnostic parameter
$\mathbf{C}_1$ = $n \times 1$ vector of the values of the diagnostic parameter
$\hat{\mathbf{C}}_1$ = $n \times 1$ vector of the predicted value of the diagnostic parameter
$\mathbf{I}$ = $n \times n$ identity matrix with 1's down the diagonal and 0's elsewhere
k = rate constant g mole/(g.cat.)(hr)(atm)
K_1 = adsorption equilibrium constant for methane, atm.$^{-1}$
K_2 = adsorption equilibrium constant for oxygen, atm.$^{-2}$
K_3 = adsorption equilibrium constant for carbon dioxide, atm.$^{-1}$
K_4 = adsorption equilibrium constant for water, atm.$^{-1}$
n = number of observations
$t_{\nu,\alpha/2}$ = 100 $(1 - \alpha/2)$ % point of the t-distribution with ν degrees of freedom
r = reaction rate, g mole/(g.cat.)(hr)
$V(\mathbf{e})$ = error variance of the residual of diagnostic parameter
$V(\hat{\mathbf{e}})$ = estimated error variance of the residual of diagnostic parameter
w = space time, (hr)(g-cat.)/g mole
x_1 = initial partial pressure of methane, atm.
x_2 = initial partial pressure of oxygen, atm.
x_3 = initial partial pressure of carbon dioxide, atm.
x_4 = initial partial pressure of water vapor, atm.
y = fractional conversion of methane
α = confidence level of t-distribution
β_i = true values of the collections of rate and adsorption constants denoted by b_i, $i = 0, 1, 2, \ldots 4$
$\boldsymbol{\beta}$ = vector of parameter values β_i
ν = number of degrees of freedom
ξ_i = standardized independent variables as defined in Table 1
σ^2 = experimental error variance
$\hat{\sigma}^2$ = estimated experimental error variance

LITERATURE CITED

Draper, N. R., and H. Smith, Applied Regression Analysis, New York (1966).

Hill, W. J., and Reiji Mezaki, "The Use of Residual Analysis for Building Mechanistic Models," *AIChE J.*, 13, 611 (1967).

Hunter, W. G., and Reiji Mezaki, "A Model Building Technique for Chemical Engineering Kinetics," *ibid.*, 10, 315 (1964).

Kittrell, J. R., W. G. Hunter, and Reiji Mezaki, "The Use of Diagnostic Parameters for Kinetic Model Building," *ibid.*, 12, 1014 (1966).

Manuscript received February 16, 1972; revision received March 21, 1972; note accepted March 24, 1972.