JOURNAL OF CATALYSIS 87, 530–535 (1984)

On the Estimation of Catalytic Rate Equation Parameters

The analysis of reaction rate data to determine the parameters of a suitable rate model is a familiar exercise to the catalytic reaction engineer. Usually a method of analysis, hopefully based on a firm statistical foundation, is selected and applied to a variety of models using appropriate computational techniques in order to determine which model provides the "best" fit to the available experimental data. However, there are many possible pitfalls in this process, especially in the selection of the analytical method. The purpose of this note is to detail our experiences in just such a situation.

In a recent paper (*1*) we demonstrated that cumene disproportionation was a feasible probe reaction for characterization of the acidic function of a dual-functional hydrocracking catalyst. Catalyst deactivation was slow enough in the experiments reported to permit reproducible extrapolation to zero time-on-stream, and the resultant initial rate data agreed well with a Langmuir–Hinshelwood kinetic model. The form of this model is indicated by

$$r = \frac{k_1 K_2^2 P_C^2}{[1 + K_2 P_C]^2} \qquad [1]$$

where

$$k_1 = k_1^0 \exp(-E/RT) \qquad [2]$$

$$K_2 = K_2^0 \exp(Q/RT) \qquad [3]$$

The four parameters of this model, k_1^0, E, K_2^0, and Q, were determined by analysis of the initial rate data through application of an unweighted linear least-squares method in a two-step procedure, as follows. Equation [1] was first rewritten in a linear form in terms of the variables $r^{-0.5}$ and P_C^{-1}, as shown in

$$\frac{1}{r^{0.5}} = \frac{1}{k_1^{0.5} K_2 P_C} + \frac{1}{k_1^{0.5}} \qquad [4]$$

and the values of k_1 and K_2 were determined from data at each of three temperatures using linear least-squares. Equations [2] and [3] were then transformed as shown in

$$\ln k_1 = \ln k_1^0 + (-E/RT) \qquad [5]$$

$$\ln K_2 = \ln K_2^0 + (Q/RT) \qquad [6]$$

and the linear least-squares method was used again to find the constants in these relations between $\ln k_1$, $\ln K_2$, and $1/T$. The resultant estimates of these parameters are listed for reference in the column labeled "Linear regression" in Table 1 (*1*). Also listed are some measures of the goodness of fit of these equations (using the indicated parameter estimates) to the experimental data employed. These results indicate that a generally good fit was obtained, considering the usual scatter in experimental reaction rate data.

Despite the apparent success of this procedure, we decided to investigate the analysis of the rate data further in hopes of uncovering a more efficient and straightforward method for parameter estimation, since the use of the two-step process described above was motivated more by expediency than statistical justification. Indeed, the theoretical bases for use of unweighted linear least-squares analysis were not satisfied in this prior work. In particular, linearizing the rate equation violates the requirement of constant variance in ran-

0021-9517/84 $3.00

TABLE 1

Kinetic Parameters of Cumene Disproportionation Rate Expression Based on Rates of Reaction Extrapolated to Undeactivated Catalyst by Hyperbolic Correlation

Parameters	Linear regression	Nonlinear regression
E (kcal/mol)	20.8	22.5
Q (kcal/mol)	19.7	7.7
k_1^0 (mol/g · cat · min)	8.36×10^5	6.25×10^6
K_2^0 (atm^{-1})	9.44×10^{-9}	5.20×10^{-3}
$\langle e \rangle^a$	10.0	10.5
e_{max}	20.5	24.7
$\langle e^* \rangle^b$	5.3	
e^*_{max}	12.2	
Sum of squares of the residuals	1.04×10^{-9}	5.90×10^{-10}

[a] Average value of e_i, where $e_i = [(r_{obs,i} - r_{prcd,i})/r_{obs,i}] \times 100$.

[b] Average value of e_i^*, where $e_i^* = [(r_{obs,i}^{-0.5} - r_{pred,i}^{-0.5})/r_{obs,i}^{-0.5}] \times 100$.

dom error for the unweighted linear least-squares method (*2–7*).

One way to circumvent this difficulty is to substitute Eqs. [2] and [3] into the rate equation (Eq. [1]) and then to perform a nonlinear regression on *all* of the rate data simultaneously in a one-step procedure (*4, 8*). This was done using the results of the previous two-step linearized procedure as initial estimates and employing the Marquardt iteration technique and NREG, a nonlinear regression subroutine provided by Vogelback Computing Center at Northwestern University. To ensure rapid convergence, the rate parameters were redefined prior to analysis as (9)

$$k_1 = \bar{k}_1^0 \exp[-E/R(1/T - 1/\bar{T})] \quad [7]$$

$$\bar{k}_1^0 = k_1^0 \exp[-E/(R\bar{T})] \quad [8]$$

$$K_2 = \bar{K}_2^0 \exp[Q/R(1/T - 1/\bar{T})] \quad [9]$$

$$\bar{K}_2^0 = K_2^0 \exp[Q/(R\bar{T})] \quad [10]$$

where $\bar{T}$, the average absolute temperature of the set of data, has been introduced as a convenient scaling factor. The results of this analysis are also listed in Table 1 in the column labeled "Nonlinear regression."

Examination of the sum of the squares of the residuals and the average deviations as determined by the nonlinear regression method indicates once again that the proposed rate model (Eq. [1]) provides a good fit to the available data. Furthermore, the nonlinear regression method apparently leads to parameters which offer a somewhat better overall fit than those found by the linearized two-step approach. However, there are major differences in the parameter values obtained by these two different methods of data analysis, particularly in the estimates of K_2^0 and Q. Because the two-step linear least-squares method is not fully justified in the present case, we believe the parameter values found by nonlinear regression are more appropriate.

It has been shown previously (*4, 5*) that although the simpler linearized approach is frequently satisfactory, it may indeed be inadequate for the estimation of kinetic parameters under some conditions. The balance of this note is aimed at exploring the range of validity and robustness of this simpler approach as well as the sensitivity of parameter estimates made by both methods to random errors in the data.

To accomplish these goals, a computer program was written to generate data comparable to those generated in our experiments and, subsequently, to analyze these data using the two techniques mentioned above.

Simulated kinetic data were generated using the Langmuir–Hinshelwood model of Eq. [1] but incorporating a random error contribution determined from a Gaussian distribution typical of most experimentally based parameters (*10*). Consequently, r_i, the simulated rate of reaction at temperature T_i and cumene partial pressure P_i, was determined by generating a random number from a Gaussian distribution $g(\bar{r}_i, \sigma)$ with mean of $\bar{r}_i$ and standard deviation σ. The mean rate of reaction, $\bar{r}_i$, is defined as

$$\bar{r}_i = f(T_i, P_i, \boldsymbol{\beta}) = \frac{\beta(1)\exp[-\beta(3)/T_i](\beta(2)\exp[\beta(4)/T_i])^2 P_i^2}{(1 + \beta(2)\exp[\beta(4)/T_i]P_i)^2} \quad [11]$$

where for convenience the vector $\boldsymbol{\beta}$ has been introduced to represent the kinetic parameters. Specifically,

$$\beta(1) = k_1^0 \qquad \beta(2) = K_2^0$$
$$\beta(3) = E/R \qquad \beta(4) = Q/R \quad [12]$$

For present purposes, the standard deviation σ was assumed to be homogeneous, i.e., the variance in random error over the range of independent variables examined was constant. (In this case the kinetic data are said to be homoscedastic. The effect of heteroscedasticity on parameter estimation in nonlinear models has been discussed elsewhere (*7*).) Calculations were performed using the two different sets of kinetic parameters found previously by linear and nonlinear regression methods and reported in Table 1. These are listed for reference in the first rows of Tables 2 and 3.

Ten sets of numerical experiments were carried out. Each set consisted of 30 simulated rate measurements with two calculated at each point of the following temperature–pressure grid: temperatures of 150, 175, and 200° C; pressures of 0.05, 0.1, 0.25, 0.50, and 0.75 atm. Each of these 30 point sets was then analyzed by the two-step linear regression method and by nonlinear regression to yield vectors of the model parameters. These are designated as $\boldsymbol{\beta}_L$ and $\boldsymbol{\beta}_{NL}$, respectively. In addition, $\alpha(I)$, the ratios of the final estimates of each parameter

TABLE 2

Comparison of Parameter Estimates Using Linear Regression Data Set

Std. Devn.	Item	$\beta(1) \times 10^{-6}$	$\beta(2) \times 10^{8}$	$\beta(3) \times 10^{-5}$	$\beta(4) \times 10^{-4}$
Base data set		.836	.944	.104	.983
	$\langle\beta_L(I)\rangle$	.835 ± .002	.942 ± .014	.104 ± .000	.983 ± .001
1×10^{-8}	$\langle\beta_{NL}(I)\rangle$	.836 ± .001	.944 ± .003	.104 ± .000	.983 ± .000
	$\langle\alpha(I)\rangle$	.999 ± .002	.998 ± .014	1.000 ± .000	1.000 ± .000
	$\langle\beta_L(I)\rangle$	.831 ± .017	.926 ± .128	.104 ± .000	.984 ± .006
1×10^{-7}	$\langle\beta_{NL}(I)\rangle$	.837 ± .010	.945 ± .028	.104 ± .000	.983 ± .001
	$\langle\alpha(I)\rangle$	.991 ± .019	.974 ± .129	1.000 ± .001	1.002 ± .006
	$\langle\beta_L(I)\rangle$	.835 ± .145		.103 ± .001	1.003 ± .062
1 ± 10^{-6}	$\langle\beta_{NL}(I)\rangle$	.851 ± .092	[a]	.104 ± .001	.982 ± .013
	$\langle\alpha(I)\rangle$	.949 ± .166		.997 ± .008	1.021 ± .060
	$\langle\beta_L(I)\rangle$	.862 ± .325		.103 ± .002	1.033 ± .132
2×10^{-6}	$\langle\beta_{NL}(I)\rangle$	.884 ± .176	[a]	.104 ± .001	.983 ± .026
	$\langle\alpha(I)\rangle$	.984 ± .321		.995 ± .017	1.051 ± .129
	$\langle\beta_L(I)\rangle$	1.037 ± .572		.103 ± .029	1.088 ± .225
3×10^{-6}	$\langle\beta_{NL}(I)\rangle$	.934 ± .254	[a]	.104 ± .002	.983 ± .039
	$\langle\alpha(I)\rangle$	1.126 ± .504		.994 ± .027	1.105 ± .218
	$\langle\beta_L(I)\rangle$			.103 ± .044	1.642 ± .781
4×10^{-6}	$\langle\beta_{NL}(I)\rangle$	[a]	[a]	.104 ± .021	.983 ± .052
	$\langle\alpha(I)\rangle$			.989 ± .032	1.673 ± .806

[a] Differences between linear least-squares estimate and nonlinear regression estimate larger than 1 order of magnitude.

TABLE 3

Comparison of Parameter Estimates Using Nonlinear Regression Data Set

Std. Devn.	Item	$\beta(1) \times 10^{-6}$	$\beta(2) \times 10^{8}$	$\beta(3) \times 10^{-5}$	$\beta(4) \times 10^{-4}$
Base data set		.625	.520	.113	.385
	$\langle\beta_L(I)\rangle$	.625 ± .002	.518 ± .005	.113 ± .000	.385 ± .000
1×10^{-8}	$\langle\beta_{NL}(I)\rangle$	.625 ± .001	.520 ± .001	.113 ± .000	.385 ± .000
	$\langle\alpha(I)\rangle$	.999 ± .002	.998 ± .009	1.000 ± .000	1.000 ± .000
	$\langle\beta_L(I)\rangle$	.624 ± .014	.510 ± .053	.113 ± .000	.386 ± .005
1×10^{-7}	$\langle\beta_{NL}(I)\rangle$	.627 ± .006	.517 ± .012	.113 ± .000	.385 ± .001
	$\langle\alpha(I)\rangle$	.994 ± .018	.985 ± .090	1.000 ± .001	1.003 ± .012
	$\langle\beta_L(I)\rangle$	.641 ± .148	1.103 ± 1.412	.112 ± .001	.397 ± .047
1×10^{-6}	$\langle\beta_{NL}(I)\rangle$	.651 ± .057	.511 ± .127	.113 ± .000	.388 ± .011
	$\langle\alpha(I)\rangle$	.974 ± .179	1.858 ± 2.010	.998 ± .008	1.021 ± .110
	$\langle\beta_L(I)\rangle$	.738 ± .336		.112 ± .002	.405 ± .101
2×10^{-6}	$\langle\beta_{NL}(I)\rangle$	.687 ± .116		.113 ± .001	.392 ± .022
	$\langle\alpha(I)\rangle$	1.027 ± .379		.995 ± .017	1.027 ± .237
	$\langle\beta_L(I)\rangle$			.111 ± .008	
4×10^{-6}	$\langle\beta_{NL}(I)\rangle$			.113 ± .002	
	$\langle\alpha(I)\rangle$			.977 ± .071	

[a] Differences between linear-least-squares estimate and nonlinear regression estimate larger than 1 order of magnitude.

from the two methods, were calculated. Finally, the parameter estimates and the values of $\alpha(I)$ were averaged and the corresponding 95% confidence intervals were calculated. The results of these calculations are indicated in Tables 2 and 3 for several different values of the assumed standard deviation used in the data generation process.

Although the use of unweighted linear least-squares in place of weighted linear least-squares or nonlinear regression is not really justified at present, this method should not introduce any bias in the parameter estimates (*11*). Therefore, analysis of a sufficiently large number of observations by the two-step procedure should yield valid estimates. Furthermore, $V(y(x))$, the variance of the transformed dependent variable $y(x)$, is related to $V(x)$, the variance of the dependent variable, x, as (*12*)

$$V(y(x)) = \left\{\frac{\partial}{\partial x}(y(x))\right\}^2 V(x) \qquad [13]$$

Consequently, in the limit as $V(x)$ approaches zero, $V(y(x))$ also approaches zero and both techniques yield "exact" fits. In the present case, the roles of x and $y(x)$ are played by the following pairs: r and $r^{-0.5}$; k_1 and $\ln k_1$; K_2 and $\ln K_2$. Thus, at low standard deviations we expect close agreement between the two sets of estimates presented in Tables 2 and 3. As the standard deviation increases, however, 10 sets of 30 experiments or observations may not be sufficient and the averaged estimates of the parameters from either method are likely to differ from the true values.

Careful examination of the results reveals the following: At standard deviations less than 10^{-7}, both the linear and nonlinear regression techniques yield accurate and precise estimates of the parameters of the model. This is shown by the magnitude as well as the small confidence limits of the average parameter estimates. As the standard deviation of the distribution from which the rates of reaction are calculated increases, the 95% confidence interval associated with each kinetic parameter increases. In addition, the confidence intervals of the average estimates of the

TABLE 4

Scatter in Alternative Parameter Estimates Using Linear Regression Data Set ($\sigma = 4.0 \times 10^{-6}$)

Item	$\beta(1) \times 10^{-6}$	$\beta(2) \times 10^{8}$	$\beta(3) \times 10^{-5}$	$\beta(4) \times 10^{-4}$
Base data set	.836	.944	.104	.983
(a) Case 1				
$\beta_L(I)$	1.701	$.132 \times 10^{-10}$	.107	2.140
$\beta_{NL}(I)$	1.412	.447	.106	1.016
(b) Case 2				
$\beta_L(I)$	1.226	.176	.105	1.060
$\beta_{NL}(I)$	1.098	1.565	.105	0.960

parameters determined by the two-step technique are always larger than those determined by nonlinear regression, as expected (*5*). Furthermore, the preexponential factors show greater variation than the exponential parameters while $\beta(2)$ and $\beta(4)$, which enter the rate equation in a more nonlinear fashion, exhibit greater variation than $\beta(1)$ and $\beta(3)$, respectively.

Some indication of the variability of individual estimates of the kinetic parameters is shown by the example listed in Table 4, where two sets of estimates made with both calculational models and with a value of standard deviation σ comparable to that of our original experimental measurements are presented. Case 1 illustrates a situation in which the estimates from the two different approaches are distinctly different. In particular, the two-step and the nonlinear regression estimates of $\beta(2)$ and $\beta(4)$ differ by factors of 10^{10} and 2, respectively. Similar behavior was observed in the linearized and nonlinear regression estimates based on the experimental data (see Table 1). Case 2, on the other hand, indicates another set of estimates that are considerably closer to one another, despite the large value of σ. The estimates of $\beta(2)$ are clearly much more variable than those for the other parameters of the model. In the analysis of kinetic data with large values of σ, the occurrence of behavior similar to either of these two cases is a matter of chance.

These numerical studies have demonstrated that disagreement between the estimates of the parameters in Langmuir–Hinshelwood rate models made by the linear and the nonlinear regression techniques are not only possible, but quite likely, depending upon the expected standard deviation in the original data set as well as the total number of observations available for analysis. Clearly, reducing the former and increasing the latter will result in much better estimates of the model parameters as well as convergence between estimates made from different techniques. In the face of experimental realities with a limited set of results, it is clear that use of a nonlinear regression analysis is more reliable.

As Kittrell *et al.* [11] have pointed out, one important problem in applications of nonlinear regression analysis is the determination of reasonable initial estimates of the unknown parameters required to initiate the iteration procedure. These initial estimates influence significantly the convergence of the estimation, the parameter values so estimated, and the sum of the squares of the residuals at convergence. They have suggested several methods of obtaining these initial estimates that include: use of prior information, cyclic parameter estimation, linearization, grid search, and analog simulation. In the numerical studies presented here the initial parameter estimates were obtained by linearization and convergence was obtained for all cases examined. We thus conclude that linearized estimates of parameter values provide a reasonable beginning point for the four parameter models so often involved in catalytic correlations.

APPENDIX: NOTATION

e_i percentage deviation of model reaction rate from observed value for experiment i (see Table 1)

e_i^* percentage deviation of inverse square root of model reaction rate from observed value for experiment i (see Table 1)

E activation energy, kcal/mol

$f(T_i, P_i, \boldsymbol{\beta})$ simulated rate function (see Eq. [11]), mol/(g · cat · min)

$g(\bar{r}_i, \sigma)$ Gaussian distribution with mean value $\bar{r}_i$ and standard deviation σ

k_1 rate constant (see Eqs. [1] and [2]), mol/(g · cat · min)

k_1^0 preexponential factor (see Eq. [2])

$\bar{k}_1^0$ reparameterized preexponential factor (see Eqs. [7] and [8])

K_2 adsorption equilibrium constant, atm^{-1} (see Eqs. [1] and [3]

K_2^0 preexponential factor (see Eq. [3])

$\bar{K}_2^0$ reparameterized preexponential factor (see Eqs. [9] and [10])

P_C cumene partial pressure, atm

P_i cumene partial pressure in ith experiment, atm

Q heat of adsorption, kcal/mol

r initial rate of benzene formation via cumene disproportionation, mol/(g · cat · min)

$\bar{r}_i$ mean simulated rate of reaction predicted by Eq. [11], mol/(g · cat · min)

r_i simulated reaction rate determined from Gaussian distribution, mol/(g · cat · min)

R gas constant, kcal/(mol °K)

T reaction temperature, °K

T_i reaction temperature in ith experiment, °K

$\bar{T}$ average reaction temperature for data set, °K

$V(x)$ variance of variable x

$\alpha(I)$ ratio of estimates of parameter $\beta(I)$ as predicted by the two-step approach and by nonlinear regression

$\boldsymbol{\beta}$ vector of model kinetic parameters (see Eq. [11] and [12])

$\boldsymbol{\beta}_L$ vector of the kinetic parameters determined by the two-step (linear) approach

$\boldsymbol{\beta}_{NL}$ vector of the kinetic parameters determined by nonlinear regression

σ standard deviation in Gaussian distribution

REFERENCES

1. Absil, R. P. L., Butt, J. B., and Dranoff, J. S., *J. Catal.* **85,** 415 (1984).
2. Chou, Chan-Hui, *Ind. Eng. Chem.* **50,** 799 (1958).
3. White, R. R., and Churchill, S. W., *AIChE J.* **5,** 354 (1959).
4. Blakemore, J. W., and Hoerl, A. E., *Chem. Eng. Progr. Symp. Ser.* No. 42, **59,** 14 (1963).
5. Kittrell, J. R., Hunter, W. G., and Watson, C. C., *AIChE J.* **11**(6), 1051 (1965).
6. Bard, Y., and Lapidus, L., *Catal. Rev.* **2**(1), 67 (1968).
7. Pritchard, D. J., Downie, J., and Bacon, D. W., *Technometrics* **19**(3), 227 (1977).
8. Froment, G. F., *AIChE J.* **21**(6), 1041 (1975).
9. Mezaki, R., and Kittrell, J. R., *Ind. Eng. Chem.* **59**(5), 63 (1967).
10. Anderson, T. F., Abrams, D. S., and Grens, E. A., *AIChE J.* **24**(1), 20 (1978).
11. Kittrell, J. R., Mezaki, R., and Watson, C. C., *Ind. Eng. Chem.* **57**(12), 19 (1965).
12. Kittrell, J. R., *Advan. Chem. Eng.* **8,** 97 (1970).

R. P. L. ABSIL
J. B. BUTT
J. S. DRANOFF

Department of Chemical Engineering and the Ipatieff Laboratory
Northwestern University
Evanston, Illinois 60201

Received October 18, 1983