DEPARTMENT OF STATISTICS

University of Wisconsin
Madison, Wisconsin 53706

TECHNICAL REPORT NO. 323

November 1972

TOPICS IN MODEL BUILDING

PART IV

SOME PROBLEMS IN MODEL DISCRIMINATION

By

H. Kanemasu

Typist: Jacquelyn R. Forer

This research was supported by the Air Force Office of Scientific Research under Grant AFOSR-72-2363.

TOPICS IN MODEL BUILDING

PART IV. SOME PROBLEMS IN MODEL DISCRIMINATION

Certain authors ([5],[10]) have remarked on what they have believed was the instability of the posterior probabilities of the models calculated using Box and Hill's discrimination technique [1]. We show in this chapter that this instability arose not for any inappropriateness of the technique itself but because it was used under the conditions which violated critical assumptions.

4.1 The posterior model probability $P(M_i|\underset{\sim}{y})$ when the experimental error variance σ^2 is not known.

In the preceding part, methods are given for obtaining posterior model probabilities $P(M_i|\sigma,\underset{\sim}{y})$ for the case where the experimental error variance σ^2 is known. There are, however, many instances in practice where σ^2 is not known. In some cases, a limited knowledge of σ^2 is available from some replicated runs in a preparatory investigation or, in other cases, σ^2 is not known because of difficulty in repeating the experiment at the same experimental condition. We first present the procedure for computing the posterior model probability when σ is not known and then consider the situation where some previous information is available.

The posterior probability for the ith model M_i is

$$P(M_i|\underset{\sim}{y}) = \int p(M_i,\sigma|\underset{\sim}{y})\,d\sigma$$

$$= \int P(M_i|\sigma,\underset{\sim}{y})p(\sigma|\underset{\sim}{y})\,d\sigma;\quad i=1,2,\ldots,m \qquad (4.1.1)$$

However

$$p(\sigma|\underset{\sim}{y}) = \sum_{j=1}^{m} p(\sigma|M_j,\underset{\sim}{y})P(M_j|\underset{\sim}{y}) \qquad (4.1.2)$$

that is to say, the posterior density of σ is the weighted sum of the $p(\sigma|M_i,\underset{\sim}{y})$ with weights $P(M_i|\underset{\sim}{y})$. Unfortunately, the latter probabilities $P(M_i|\underset{\sim}{y})$ are precisely the quantities we wish to compute. The dilemma can be resolved in the following way.

Substituting (4.1.2) into (4.1.1) we obtain

$$P(M_i|\underset{\sim}{y}) = \int P(M_i|\sigma,\underset{\sim}{y})\left[\sum_{j=1}^{m} p(\sigma|M_j,\underset{\sim}{y})P(M_j|\underset{\sim}{y})\right]d\sigma$$

$$= \sum_{j=1}^{m} P(M_j|\underset{\sim}{y})\int P(M_i|\sigma,\underset{\sim}{y})p(\sigma|M_j,\underset{\sim}{y})d\sigma. \qquad (4.1.3)$$

Making use of the notation

$$Q_{ij} = \int P(M_i|\sigma,\underset{\sim}{y})p(\sigma|M_j,\underset{\sim}{y})d\sigma, \qquad (4.1.4)$$

we have

$$P(M_i|\underset{\sim}{y}) = \sum_{j=1}^{m} P(M_j|\underset{\sim}{y})Q_{ij};\ i=1,2,\ldots,m. \qquad (4.1.5)$$

Notice here that we have the relationship

$$\sum_{i=1}^{m} Q_{ij} = \int \left(\sum_{i=1}^{m} p(M_i|\sigma,\underset{\sim}{y})\right)p(\sigma|M_j,\underset{\sim}{y})d\sigma$$

$$= \int p(\sigma|M_j,\underset{\sim}{y})d\sigma = 1;\ j=1,2,\ldots,m. \qquad (4.1.6)$$

Eliminating $P(M_m|\underset{\sim}{y})$ from (4.1.5) by $P(M_m|\underset{\sim}{y}) = 1 - \sum_{j=1}^{m-1} P(M_j|\underset{\sim}{y})$, we obtain

$$P(M_i|\tilde{y}) = \sum_{j=1}^{m-1} P(M_j|\tilde{y})Q_{ij} + (1 - \sum_{j=1}^{m-1} P(M_j|\tilde{y}))Q_{i,m};$$

$$i=1,2,\ldots,m. \quad (4.1.7)$$

Thus we have the m-1 simultaneous linear equations

$$(Q_{i,m}-Q_{i,1})P(M_1|\tilde{y}) + (Q_{i,m}-Q_{i,2})P(M_2|\tilde{y}) + \ldots$$

$$+ (Q_{i,m}-Q_{i,i-1})P(M_{i-1}|\tilde{y}) + (Q_{i,m}-Q_{i,i}+1)P(M_i|\tilde{y})$$

$$+ (Q_{i,m}-Q_{i,i+1})P(M_{i+1}|\tilde{y}) + \ldots + (Q_{i,m}-Q_{i,m-1})P(M_{m-1}|\tilde{y}) = Q_{i,m}$$

$$i=1,2,\ldots,m-1 \quad (4.1.8)$$

Using the matrix notation

$$\tilde{Z}\tilde{P} = \tilde{w} \quad (4.1.9)$$

where

$$\tilde{P}' = \left(P(M_1|\tilde{y}),P(M_2|\tilde{y}),\ldots,P(M_{m-1}|\tilde{y})\right), \quad (4.1.10)$$

$$\underset{\sim}{Z} = \begin{bmatrix} Q_{1,m}-Q_{1,1}+1 & Q_{1,m}-Q_{12} & \cdots & Q_{1,m}-Q_{1,m-1} \\ Q_{2,m}-Q_{2,1} & Q_{2,m}-Q_{2,2}+1 & \cdots & Q_{2,m}-Q_{2,m-1} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{m-1,m}-Q_{m-1,1} & Q_{m-1,m}-Q_{m-1,2} & & Q_{m-1,m}-Q_{m-1,m-1}+1 \end{bmatrix} \tag{4.1.11}$$

and

$$\underset{\sim}{w}' = (Q_{1,m}, Q_{2,m}, \ldots, Q_{m-1,m}). \tag{4.1.12}$$

Solving equation (4.1.9) for $P(M_i|\underset{\sim}{y})$; $i=1,2,\ldots,m$ we have

$$\underset{\sim}{P} = \underset{\sim}{Z}^{-1}\underset{\sim}{w}. \tag{4.1.13}$$

For the case where we have only two models in consideration

$$P(M_1|\underset{\sim}{y}) = \frac{Q_{12}}{Q_{12}-Q_{11}+1}\,. \tag{4.1.14}$$

However, from (4.1.6), $Q_{11}+Q_{21} = 1$. Using this relationship in (4.1.14) we have

$$P(M_1|\underset{\sim}{y}) = \frac{Q_{12}}{Q_{12}+Q_{21}} = \frac{1-Q_{22}}{Q_{12}+Q_{21}} \qquad (4.1.15a)$$

and so

$$P(M_2|\underset{\sim}{y}) = 1 - P(M_2|\underset{\sim}{y})$$

$$= \frac{Q_{21}}{Q_{12}+Q_{21}} = \frac{1-Q_{11}}{Q_{12}+Q_{21}} \qquad (4.1.15b)$$

Now we recall that $Q_{22} = \int p(M_2|\sigma,\underset{\sim}{y})p(\sigma|M_2,\underset{\sim}{y})d\sigma$. Consider the range of σ which the model M_2 indicates is likely to contain the true value of σ. If the probability $P(M_2|\sigma,\underset{\sim}{y})$ is low in this region, then Q_{22} will be small relative to unity and thus the rival model M_1 will have a high posterior probability.

For the case where we have three models M_1, M_2 and M_3, the equation (4.1.13) gives

$$\begin{bmatrix} P(M_1|\underset{\sim}{y}) \\ P(M_2|\underset{\sim}{y}) \end{bmatrix} = \begin{bmatrix} Q_{13}-Q_{11}+1 & Q_{13}-Q_{12} \\ Q_{23}-Q_{21} & Q_{23}-Q_{22}+1 \end{bmatrix}^{-1} \begin{bmatrix} Q_{13} \\ Q_{23} \end{bmatrix} \qquad (4.1.16)$$

which gives

$$P(M_1|\underset{\sim}{y}) \propto Q_{12}Q_{13} + Q_{12}Q_{23} + Q_{13}Q_{32}$$

$$P(M_2|\underset{\sim}{y}) \propto Q_{23}Q_{21} + Q_{23}Q_{31} + Q_{21}Q_{13}$$

$$P(M_3|\underset{\sim}{y}) \propto Q_{31}Q_{32} + Q_{31}Q_{12} + Q_{32}Q_{21} \qquad (4.1.17)$$

It is noted that when M_3 is not included, that is, $P(M_3|\sigma, y_n) = 0$ for all σ, this last expression reduces to equation (4.1.15) apart from the common factor, since $Q_{3i} = 0$.

To compute Q_{ij} we have to integrate the product of $P(M_i|\sigma, y_n)$ and $p(\sigma|M_j, y_n)$ with respect to σ. When there is practically no information on σ we may use the non-informative prior $p(\log \sigma) \propto$ constant ([3]) and then it can be shown that

$$p(\sigma|M_i, y_n) = \frac{2\left(\frac{S_{i,n}}{2}\right)^{\frac{n-p_i}{2}}}{\Gamma\left(\frac{n-p_i}{2}\right)} e^{-\frac{S_{i,n}}{2\sigma^2}} \sigma^{-(n-p_i+1)} \qquad (4.1.18)$$

assuming that the vector ϵ_n of the experimental errors $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ follows $N(0, I_n\sigma^2)$. Also this is the result based on the linear approximation already mentioned in the preceding sections.

When some information on σ^2 is available, for example, from replicated runs in a preparatory investigation, the prior density of σ given by

$$p(\sigma) = \frac{2\left(\frac{\nu s^2}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \sigma^{-(\nu+1)} e^{-\frac{\nu s^2}{2\sigma^2}} \qquad (4.1.19)$$

should be used instead of the non-informative prior density. In this expression, s^2 is the mean square (with ν degrees of freedom) from the replicated runs. In this situation the posterior density of σ under M_i becomes

$$p(\sigma|\underset{\sim}{y},M_i) = \frac{2\left(\frac{\nu s^2+S_{i,n}}{2}\right)^{\frac{\nu+(n-p_i)}{2}}}{\Gamma\left(\frac{\nu+n-p_i}{2}\right)} \sigma^{-(\nu+n-p_i+1)} e^{-\frac{\nu s^2+S_{i,n}}{2\sigma^2}} \tag{4.1.20}$$

When we have some replicated runs in the current experiment which yield the sum of squares νs^2 with ν degrees of freedom, we have the same $p(\sigma|\underset{\sim}{y},M_i)$ as above except that $S_{i,n}$ becomes the lack of fit sum of squares in this case. For a given value of σ, the methods described in Part III can be used to compute $P(M_i|\sigma,\underset{\sim}{y}_n)$.[1] Unfortunately it is not possible to evaluate the integral Q's analytically. The numerical integration to obtain Q's, however, is quite feasible.

It is of interest to note that, as ν becomes larger, νs^2 will overwhelm $S_{i,n}$ in (4.1.20) and $p(\sigma|\underset{\sim}{y},M_j)$ will tend to become more sharply concentrated about $\sigma = \sigma^*$ (the true value of σ), whence Q_{ij} will tend to $P(M_i|\underset{\sim}{y},\sigma^*)$ for all j. Thus the matrix Z given by (4.1.11) will look more and more

---

[1] In Part III, for convenience the notation $P(M_i|\underset{\sim}{y}_n)$ instead of $P(M_i|\sigma,\underset{\sim}{y}_n)$ was used for the posterior probabilities of the model M_i given the exact knowledge of the experimental error variance σ^2.

like I_{m-1}, the $(m-1)\times(m-1)$ identity matrix so that $P(M_i|\underset{\sim}{y})$ given by the equation (4.1.13) will approach $P(M_i|\underset{\sim}{y},\sigma^*)$.

In Part III an example was studied where σ was supposed to be known and equal to 0.05. This same example is now reconsidered but supposing σ to be unknown. The posterior probabilities obtained from (4.1.13) above are 0.001, 0.675, 0.233 and 0.091 respectively for the models M_1, M_2, M_3 and M_4. These probabilities, interestingly enough, are not much different from those obtained with $\sigma = 0.05$. The reason clearly is that the correct model included gives a mean square of residuals which is quite close to the value of σ assumed.

4.2 Sequential computation of the posterior probabilities of models.

Once the posterior probabilities are established after the preliminary runs, we can recompute them sequentially as new observations become available.[2]

[2] It is not recommended to use the procedures described in Part III to obtain the posterior probabilities in this sequential situation because, strictly speaking, this would amount to changing the initial prior information on parameters at different stages of experimentation.

4.2.1 σ known case.

Via Bayes' theorem,

$$P(M_i|\underset{\sim}{y}_{n+1},\sigma) \propto P(M_i|\underset{\sim}{y}_n,\sigma)p(\underset{\sim}{y}_{n+1}|M_i,\underset{\sim}{y}_n,\sigma) \qquad (4.2.1)$$

where the second term on the right hand side is given by

$$p(y_{n+1}|M_i,\underset{\sim}{y}_n,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2(1+b_i)}} \exp\left\{-\frac{1}{2\sigma^2(1+b_i)}(y_{n+1}-\hat{y}^{(i)}_{n+1})^2\right\} \qquad (4.2.2)$$

This was derived by Box and Hill [1] (and is p_i in their notation) under the assumption that the vector of experimental errors $\underset{\sim}{\varepsilon}$ is distributed $N(\underset{\sim}{0},I\sigma^2)$, by making use of a locally uniform prior for $\underset{\sim}{\theta}_i$ and a Taylor series linear approximation of the model

$$f_i(\underset{\sim}{\xi}_u,\underset{\sim}{\theta}_i) \doteq f_i(\underset{\sim}{\xi}_u,\hat{\underset{\sim}{\theta}}_i) + \sum_{i=1}^{p_i}\left[\frac{\partial f_i(\underset{\sim}{\xi}_u,\underset{\sim}{\theta}_i)}{\partial\theta_{ij}}\right]_{\underset{\sim}{\theta}_i=\hat{\underset{\sim}{\theta}}_i}(\theta_{ij}-\hat{\theta}_{ij}). \qquad (4.2.3)$$

In equation (4.2.2),

$$b_i = \underset{\sim}{x}^{(i)}_{n+1}(X'_{i,n}X_{i,n})^{-1}\,\underset{\sim}{x}^{(i)'}_{n+1} \qquad (4.2.4)$$

where, denoting $[\partial f_i(\underset{\sim}{\xi}_u,\underset{\sim}{\theta}_i)/\partial\theta_{ij}]_{\underset{\sim}{\theta}_i=\hat{\underset{\sim}{\theta}}_i}$ by $x^{(i)}_{uj}$,

$$\underset{\sim}{x}_{n+1}^{(i)} = (x_{n+1,1}^{(i)}, x_{n+1,2}^{(i)}, \ldots, x_{n+1,p_i}^{(i)}) \qquad (4.2.5)$$

and

$$X_{i,n} = \begin{bmatrix} x_{11}^{(i)} & x_{12}^{(i)} & \cdots & x_{1p_i}^{(i)} \\ \vdots & \vdots & & \\ x_{n1}^{(i)} & x_{n2}^{(i)} & \cdots & x_{np_i}^{(i)} \end{bmatrix}. \qquad (4.2.6)$$

Sometimes it is inconvenient to make computations after each run and it may be desired to carry out a group of experiments at each stage. In this case we proceed as follows. Denoting the vector of ℓ additional observations by $\underset{\sim}{y}_\ell$ and that of combined $n+\ell$ observations by $\underset{\sim}{y}_{n+\ell}$, from Bayes' theorem we have

$$P(M_i|\underset{\sim}{y}_{n+\ell},\sigma) \propto P(M_i|\underset{\sim}{y}_n,\sigma)p(\underset{\sim}{y}_\ell|M_i,\underset{\sim}{y}_n,\sigma). \qquad (4.2.7)$$

In a similar manner to Box and Hill's derivation of (4.2.7) $p(\underset{\sim}{y}_\ell|M_i,\underset{\sim}{y}_n,\sigma)$ is given by

$$p(\underset{\sim}{y}_\ell|M_i,\underset{\sim}{y}_n,\sigma) = \frac{|M_i|^{-\frac{1}{2}}}{(2\pi\sigma^2)^{\frac{\ell}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(\underset{\sim}{y}_\ell-\tilde{\underset{\sim}{y}}_\ell^{(i)})' M_i^{-1}(\underset{\sim}{y}_\ell-\tilde{\underset{\sim}{y}}_\ell^{(i)})\right\}, \qquad (4.2.8)$$

where

$$\tilde{\underset{\sim}{y}}_{\ell}^{(i)\prime} = (f_i(\underset{\sim}{\xi}_{n+1},\hat{\underset{\sim}{\theta}}_i), f_i(\underset{\sim}{\xi}_{n+2},\hat{\underset{\sim}{\theta}}_i), \ldots, f_i(\underset{\sim}{\xi}_{n+\ell},\hat{\underset{\sim}{\theta}}_i)) \tag{4.2.9}$$

$$M_i = I_\ell + X_\ell^{(i)}(X_{i,n}'X_{i,n})^{-1}X_\ell^{(i)\prime} \tag{4.2.10}$$

and

$$X_\ell^{(i)} = \begin{bmatrix} x_{n+1,1}^{(i)} & x_{n+1,2}^{(i)} & \cdots & x_{n+1,p_i}^{(i)} \\ \vdots & & & \\ x_{n+\ell,1}^{(i)} & x_{n+\ell,2}^{(i)} & \cdots & x_{n+\ell,p_i}^{(i)} \end{bmatrix}. \tag{4.2.11}$$

More convenient forms of (4.2.8) can be obtained by using the updating formula given by Box and Wilson [4]

$$(\underset{\sim}{y}_\ell - \tilde{\underset{\sim}{y}}_\ell^{(i)})' M_i^{-1} (\underset{\sim}{y}_\ell - \tilde{y}_\ell^{(i)}) = S_{i,n+\ell} - S_{i,n} \tag{4.2.12}$$

where $S_{i,n+\ell}$ is the sum of squares for the model M_i based on $n+\ell$ observations. Furthermore we have

$$\begin{aligned} |M_i| &= |I_{p_i} + (X_{i,n}'X_{i,n})^{-1}X_\ell^{(i)\prime}X_\ell^{(i)}| \\ &= |X_{i,n}'X_{i,n}|^{-1}|X_{i,n}'X_{i,n} + X_\ell^{(i)\prime}X_\ell^{(i)}| \\ &= |X_{i,n+\ell}'X_{i,n+\ell}|/|X_{i,n}'X_{i,n}|. \end{aligned} \tag{4.2.13}$$

Substituting (4.2.12) and (4.2.13) into (4.2.8)

$$p(\underset{\sim}{y}_\ell | M_i, \underset{\sim}{y}_n, \sigma) = \frac{1}{(2\pi\sigma^2)^{\frac{\ell}{2}}} \left\{ \frac{|X'_{i,n} X_{i,n}|}{|X'_{i,n+\ell} X_{i,n+\ell}|} \right\}^{\frac{1}{2}} \exp\left\{ -\frac{S_{i,n+\ell} - S_{i,n}}{2\sigma^2} \right\} . \quad (4.2.14)$$

Using this in (4.2.7), we obtain

$$P(M_i | \underset{\sim}{y}_{n+\ell}, \sigma) \propto P(M_i | \underset{\sim}{y}_n, \sigma) \left\{ \frac{|X'_{i,n} X_{i,n}|}{|X'_{i,n+\ell} X_{i,n+\ell}|} \right\}^{\frac{1}{2}} \exp\left\{ -\frac{S_{i,n+\ell} - S_{i,n}}{2\sigma^2} \right\} . \quad (4.2.15)$$

4.2.2 σ unknown case.

As in the σ known case, updating of posterior probabilities can be done by

$$P(M_i | \underset{\sim}{y}_{n+1}) \propto P(M_i | \underset{\sim}{y}_n) p(y_{n+1} | M_i, \underset{\sim}{y}_n) \quad (4.2.16)$$

where $p(y_{n+1} | M_i, \underset{\sim}{y}_n)$ is given by

$$p(y_{n+1} | M_i, y_n) = \int p(y_{n+1} | M_i, \underset{\sim}{y}_n, \sigma) p(\sigma | M_i, \underset{\sim}{y}_n) d\sigma . \quad (4.2.17)$$

Substituting (4.1.18) for $p(\sigma | M_i, \underset{\sim}{y}_n)$ and (4.2.2) for $p(y_{n+1} | M_i, \underset{\sim}{y}_n, \sigma)$

$$p(y_{n+1}|M_i,\underset{\sim}{y}_n)$$

$$= \frac{2\left(\frac{S_{i,n}}{2}\right)^{\frac{n-p_i}{2}}}{\Gamma\left(\frac{n-p_i}{2}\right)\sqrt{2\pi}(1+b_i)^{\frac{1}{2}}} \int_0^\infty \sigma^{-(n-p_i+2)} \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{(y_n-y_n^{(i)})^2}{1+b_i} + S_{i,n}\right)\right\}d\sigma. \tag{4.2.18}$$

Making use of the identity

$$\int_0^\infty \sigma^{-(k+1)} e^{-\frac{A}{2\sigma^2}} d\sigma = \Gamma(\tfrac{k}{2})2^{\frac{k}{2}-1} A^{-\frac{k}{2}} \tag{4.2.19}$$

in equation (4.2.18), we obtain

$$p(y_{n+1}|M_i,\underset{\sim}{y}_n)$$

$$= \frac{[(1+b_i)s_{i,n}]^{-\frac{1}{2}}}{B\left(\frac{n-p_i}{2},\frac{1}{2}\right)\sqrt{n-p_i}} \left\{1 + \frac{1}{n-p_i}\left(\frac{y_{n+1} - \tilde{y}_{n+1}^{(i)}}{s_{i,n}(1+b_i)^{\frac{1}{2}}}\right)^2\right\}^{-\frac{n-p_i+1}{2}}, \tag{4.2.20}$$

where $s_{i,n} = S_{i,n}/(n-p_i)$. That is to say, the quantity $(y_{n+1} - \tilde{y}_{n+1}^{(i)})/s_{i,n}(1+b_i)^{\frac{1}{2}}$ follows a student's t distribution with $n-p_i$ degrees of freedom.

In deriving a version of Box and Hill's discrimination criterion for the σ unknown case, Hunter and Hill [7] assumed that some replicated runs are available which give an independent estimate s^2 (with ν degrees of freedom) of σ^2 and, instead of (4.2.17), used

$$p(y_{n+1}|M_1,\underset{\sim}{y}_n) = \int_0^{\infty} p(y_{n+1}|\sigma,M_i,\underset{\sim}{y}_n)p(\sigma|s^2)d\sigma$$

where $p(\sigma|s^2)$ is the posterior density based only on s^2 and is given by

$$p(\sigma|s^2) = \frac{2\left(\frac{\nu s^2}{2}\right)^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})}\sigma^{-(\nu+1)}e^{-\frac{\nu s^2}{2\sigma^2}} . \qquad (4.2.21)$$

This, however, ignores the information about σ coming from the non-replicated runs.

When a group of ℓ additional observations, $\underset{\sim}{y}_\ell$ say, becomes available at a time, we have

$$P(M_i|\underset{\sim}{y}_{n+\ell}) \propto P(M_i|\underset{\sim}{y}_n)p(\underset{\sim}{y}_\ell|M_i,\underset{\sim}{y}_n), \qquad (4.2.22)$$

where $p(\underset{\sim}{y}_\ell|M_i,\underset{\sim}{y}_n)$ is obtained by

$$p(\underset{\sim}{y}_\ell|M_i,\underset{\sim}{y}_n) = \int p(\underset{\sim}{y}_\ell|M_i,\underset{\sim}{y}_n,\sigma)p(\sigma|M_i,\underset{\sim}{y}_n)d\sigma . \qquad (4.2.23)$$

Substituting (4.1.18) and (4.2.8) into (4.2.23) and carrying out the integration gives

$$p(\underset{\sim}{y}_\ell | M_i, \underset{\sim}{y}_n) = \frac{\Gamma\left(\frac{n-p_i+\ell}{2}\right) |M_i|^{-\frac{1}{2}} s_{i,n}^{-\frac{\ell}{2}}}{(n-p_i)^{\frac{\ell}{2}} \{\Gamma(\frac{1}{2})\}^\ell \Gamma\left(\frac{n-p_i}{2}\right)}$$

$$\times \left[1 + \frac{1}{n-p_i} \frac{(\underset{\sim}{y}_\ell - \tilde{\underset{\sim}{y}}_\ell^{(i)})' M_i^{-1} (\underset{\sim}{y}_\ell - \tilde{\underset{\sim}{y}}_\ell^{(i)})}{s_{i,n}^2}\right]^{-\frac{n-p_i+\ell}{2}} . \quad (4.2.24)$$

This shows that $(\underset{\sim}{y}_\ell - \tilde{\underset{\sim}{y}}_\ell^{(i)})/s_{i,n}$ follows the multivariate-t distribution $\underset{\sim}{t}(M_i, n-p_i)$. Again, using the updating formulae (4.2.12) and (4.2.13) in (4.2.24), we have

$$p(\underset{\sim}{y}_\ell | M_i, \underset{\sim}{y}_n) = \Gamma(\tfrac{1}{2})^{-\ell} \frac{\Gamma\left(\frac{n+\ell-p_i}{2}\right)}{\Gamma\left(\frac{n-p_i}{2}\right)} \frac{|X_{i,n}' X_{i,n}|^{\frac{1}{2}}}{|X_{i,n+\ell}' X_{i,n+\ell}|^{\frac{1}{2}}} \frac{S_{i,n}^{\frac{n-p_i}{2}}}{S_{i,n+\ell}^{\frac{n+\ell-p_i}{2}}} . \quad (4.2.25)$$

Substituting (4.2.25) into (4.2.22), we obtain the updating formula

$$P(M_i | \underset{\sim}{y}_{n+\ell}) \propto P(M_i | \underset{\sim}{y}_n) \frac{\Gamma\left(\frac{n+\ell-p_i}{2}\right)}{\Gamma\left(\frac{n-p_i}{2}\right)} \frac{|X_{i,n}' X_{i,n}|^{\frac{1}{2}}}{|X_{i,n+\ell}' X_{i,n+\ell}|^{\frac{1}{2}}} \frac{S_{i,n}^{\frac{n-p_i}{2}}}{S_{i,n+\ell}^{\frac{n+\ell-p_i}{2}}} . \quad (4.2.26)$$

We see from equations (4.2.15) and (4.2.26) that, with y_ℓ fixed, we would arrive at the same posterior probabilities whether we update them after each observation or after ℓ observations, exactly as we would expect.

4.3 On the asymptotic behavior of the posterior probabilities of models.

We examine the asymptotic behavior of posterior probabilities by applying the updating formula (4.2.15) obtained in Section 4.2 to the following simple example in which M_1 is a special case of M_2. Suppose the following two models are being considered.

$$M_1: \ \eta_1 = \theta_{11}$$

$$M_2: \ \eta_2 = \theta_{21} + \theta_{22}x \tag{4.3.1}$$

and σ is known. Using (4.2.15), the ratio of posterior probabilities for two models after $n+\ell$ observations is given by

$$R_{n+\ell} = \frac{P(M_1|y_{n+\ell},\sigma)}{P(M_2|y_{n+\ell},\sigma)}$$

$$= \frac{P(M_1|y_n,\sigma)}{P(M_2|y_n,\sigma)} \frac{|X_{1,n}'X_{1,n}|^{\frac{1}{2}}}{|X_{2,n}'X_{2,n}|^{\frac{1}{2}}} \frac{|X_{2,n+\ell}'X_{2,n+\ell}|^{\frac{1}{2}}}{|X_{1,n+\ell}'X_{1,n+\ell}|^{\frac{1}{2}}} \frac{\exp\left\{-\dfrac{S_{1,n+\ell}-S_{1,n}}{2\sigma^2}\right\}}{\exp\left\{-\dfrac{S_{2,n+\ell}-S_{2,n}}{2\sigma^2}\right\}} \tag{4.3.2}$$

and we are interested in the behaviour of $R_{n+\ell}$ as ℓ becomes larger. Since $|X_{1,h}'X_{1,h}| = h$ and $|X_{2,h}'X_{2,h}| = h \sum_{u=1}^{h} (x_u-\bar{x}_h)^2$ where $\bar{x}_h$ is the arithmetic mean of x_u, $u=1,2,...,h$, we have in the above expression

$$\frac{|X_{1,n}'X_{1,n}|^{\frac{1}{2}}}{|X_{2,n}'X_{2,n}|^{\frac{1}{2}}} \frac{|X_{2,n+\ell}'X_{2,n+\ell}|^{\frac{1}{2}}}{|X_{1,n+\ell}'X_{1,n+\ell}|^{\frac{1}{2}}} = \frac{v_{n+\ell}^{\frac{1}{2}}}{v_n^{\frac{1}{2}}} \qquad (4.3.3)$$

where $v_h = \sum_{u=1}^{h} (x_u-\bar{x}_u)^2$.

As ℓ gets larger, $S_{i,n+\ell}$ $(i=1,2)$ will dominate $S_{i,n}$ $(i=1,2)$ and also $v_{n+\ell}^{\frac{1}{2}}$ will grow larger since it is a monotone increasing function of ℓ, while $p(M_1|\underset{\sim}{y}_n,\sigma)/P(M_2|\underset{\sim}{y}_n,\sigma)$ and $v_n^{\frac{1}{2}}$ remain fixed. Therefore the asymptotic behavior of $R_{n+\ell}$ will be determined by

$$R_{n+\ell}^{*} = v_{n+\ell}^{\frac{1}{2}}\Big/\exp\left\{\frac{S_{1,n+\ell}-S_{2,n+\ell}}{2\sigma^2}\right\}. \qquad (4.3.4)$$

We now consider the two cases M_1 correct and M_2 correct.

Case I: M_1 is correct.

Since in this case $S_{1,n+\ell}-S_{2,n+\ell} = u\sigma^2$ where u is a χ^2 variable with one degree of freedom, we have

$$R_{n+\ell}^{*} = v_{n+\ell}^{\frac{1}{2}}\Big/\exp(\tfrac{u}{2}) \qquad (4.3.5)$$

As ℓ tends to infinity, $v_{n+\ell}$ will get larger. Therefore $R^*_{n+\ell}$ and so $R_{n+\ell}$ will on the average tend to infinity and establish the model M_1 as the correct model. Box and Hill [1] found in one of their examples where the models are nested and the data were generated by using one of the simpler models that this model gradually gained in its posterior probability as more and more observations became available. The asymptotic behavior given above may sound paradoxical since when M_1 is true M_2 may also be considered to be true with $\theta_{22} = 0$. However, what is meant by the second model M_2 is a conjecture that we may need an extra term $\theta_{22}x$ whatever value θ_{22} may assume.

Case II: M_2 is correct

$$S_{1,n+\ell} - S_{2,n+\ell} = (\hat{\underset{\sim}{\theta}}_{22,n+\ell})^2 v_{n+\ell} \qquad (4.3.6)$$

where $\hat{\theta}_{22,n+\ell}$ is the least square estimate of θ_{22} based on $n+\ell$ observations, and so

$$R^*_{n+\ell} = v^{\frac{1}{2}}_{n+\ell} \Big/ \exp\left\{\frac{(\hat{\theta}_{22,n+\ell})^2 v_{n+\ell}}{2\sigma^2}\right\} \qquad (4.3.7)$$

As $\ell \to \infty$, $v_{n+\ell} \to \infty$ while $\hat{\theta}_{22,n+\ell}$ will tend to the true value θ_{22}. Since the denominator in (4.3.7) will increase much faster than the numerator, $R^*_{n+\ell}$ and so $R_{n+\ell}$ will on the average tend to zero and establish M_2 as the right model.

It is of interest to note that the quantity $v_{n+\ell}^{\frac{1}{2}}/\exp\{\theta_{22}v_{n+\ell}/2\sigma^2\}$ decreases monotonically as $v_{n+\ell}$ increases in the range $v_{n+\ell} \geq \theta_{22}^2/\sigma^2$ or equivalently after $\mathrm{Var}(\hat{\theta}_{22,n+\ell})/\theta_{22}^2$ is less than unity.

It should be noted from equation (4.3.5) and (4.3.7) that, in Case I, where the simpler model M_1 is correct, $R_{n+\ell}^*$ contains a χ^2 variable u with only one degree of freedom however large ℓ may be, while in Case II where the more elaborate model M_2 is correct $\mathrm{Var}(\hat{\theta}_{22,n+\ell})$ is a monotone decreasing function of ℓ. Therefore, it is easy to imagine that in the former case some relatively large ripples in the probabilities may occur even for a large value of ℓ, although the general level of $P(M_1|\underset{\sim}{y}_{n+\ell},\sigma)$ will approach unity. On the other hand, in the latter case, the probability for the correct model M_2 will dominate over that for the wrong model M_1 in a quicker and smoother manner. Recently Sidik [8] carried out some simulations to study the behaviour of posterior probabilities in a situation where several linear models considered are nested. His results, although based on the use of multivariate normal prior densities of parameters which were chosen in an arbitrary manner, tends to confirm the points made above.

4.4 Box and Hill model discrimination design criterion.

In this section we give a version of Box and Hill model discrimination design criterion [1] for the σ unknown case which differs from the earlier result by Hunter and Hill [7] due to the difference in the derivation of $p(y_{n+1}|M_i, \underset{\sim}{y}_n)$ mentioned in the previous section.

Suppose n experiments have already been carried out. Box and Hill [1] proposed to maximize, over the operability region, the maximum expected entropy change

$$D = \sum_{i=1}^{m} \sum_{j\geq i+1}^{m} P(M_i|\underset{\sim}{y})P(M_j|\underset{\sim}{y}_n)$$
$$\times \left\{ \int p_i \, \ell n \frac{p_i}{p_j} dy_{n+1} + \int p_j \, \ell n \frac{p_j}{p_i} d_{yn+1} \right\} \quad (4.4.1)$$

where $p_k = p(y_{n+1}|M_k, \underset{\sim}{y}_n)$.

For the σ unknown case, substituting (4.2.20) into (4.4.1) we have

$$D = \sum_{i=1}^{m} \sum_{j\geq i+1}^{m} P(M_i|\underset{\sim}{y}_n)P(M_j|\underset{\sim}{y}_n)$$
$$\times \left[\int_{-\infty}^{\infty} H_i \left(1 + \frac{1}{\nu_i} \left(\frac{y_{n+1} - \tilde{y}_{n+1}^{(i)}}{s_{i,n}\sqrt{1+b_i}} \right)^2 \right)^{-\frac{\nu_i+1}{2}} \cdot \left\{ -\frac{\nu_i+1}{2} \ell n \, H_i \left(1 + \frac{1}{\nu_i} \left(\frac{y_{n+1} - \tilde{y}_{n+1}^{(i)}}{s_{i,n}\sqrt{1+b_i}} \right)^2 \right) \right. \right.$$
$$\left. + \frac{\nu_j+1}{2} \ell n \, H_j \left(1 + \frac{1}{\nu_j} \left(\frac{y_{n+1} - \tilde{y}_{n+1}^{(j)}}{s_{j,n}\sqrt{1+b_j}} \right)^2 \right) \right\} dy_{n+1}$$
$$+ \int_{\infty}^{\infty} H_j \left(1 + \frac{1}{\nu_j} \left(\frac{y_{n+1} - \tilde{y}_{n+1}^{(i)}}{s_{j,n}\sqrt{1+b_j}} \right)^2 \right)^{-\frac{\nu_j+1}{2}} \cdot \left\{ -\frac{\nu_j+1}{2} \ell n \, H_j \left(1 + \frac{1}{\nu_j} \left(\frac{y_{n+1} - \tilde{y}_{n+1}^{(j)}}{s_{j,n}\sqrt{1+b_j}} \right)^2 \right) \right.$$
$$\left. \left. + \frac{\nu_i+1}{2} \ell n \, H_i \left(1 + \frac{1}{\nu_i} \left(\frac{y_{n+1} - \tilde{y}_{n+1}^{(i)}}{s_{i,n}\sqrt{1+b_i}} \right)^2 \right) \right\} dy_{n+1} \right] \quad (4.4.2)$$

where $\nu_k = n - p_k$, $s^2_{k,n} = S_{k,n}/\nu_k$ and

$H_k = 1/B(\frac{\nu_k}{2},\frac{1}{2})\sqrt{\nu_k}[(1+b_k)s^2_{k,n}]^{\frac{1}{2}}$ for $k=1,2,\ldots,m$.

As in Hunter and Hill [7] we may approximate the log function in the above expression by expanding it via a Taylor series and truncating after the first order terms, whence

$$D \doteq \sum_{i=1}^{m} \sum_{j\geq i+1}^{m} P(M_i|\underset{\sim}{y}_n)P(M_j|\underset{\sim}{y}_n)$$

$$\times \left[\int H_i\left(1+\frac{1}{\nu_i}\left(\frac{y_{n+1}-\tilde{y}^{(i)}_{n+1}}{s_{i,n}\sqrt{1+b_i}}\right)^2\right)\cdot\left\{-\frac{\nu_i+1}{2}\frac{1}{\nu_i}\left(\frac{y_{n+1}-\tilde{y}^{(i)}_{n+1}}{s_{i,n}\sqrt{1+b_i}}\right)^2\right.\right.$$

$$\left.+\frac{\nu_j+1}{2}\frac{1}{\nu_j}\left(\frac{y_{n+1}-\tilde{y}^{(j)}_{n+1}}{s_{j,n}\sqrt{1+b_j}}\right)^2\right\}dy_{n+1}$$

$$+\int H_j\left(1+\frac{1}{\nu_j}\left(\frac{y_{n+1}-\tilde{y}^{(j)}_{n+1}}{s_{j,n}\sqrt{1+b_j}}\right)^2\right)\times\left\{-\frac{\nu_j+1}{2}\frac{1}{\nu_j}\left(\frac{y_{n+1}-\tilde{y}^{(j)}_{n+1}}{s_{j,n}\sqrt{1+b_j}}\right)^2\right.$$

$$\left.\left.+\frac{\nu_i+1}{2}\frac{1}{\nu_i}\left(\frac{y_{n+1}-\tilde{y}^{(i)}_{n+1}}{s_{i,n}\sqrt{1+b_i}}\right)^2\right\}dy_{n+1}\right]. \qquad (4.4.3)$$

Making use of the fact that $(y_{n+1}-\tilde{y}^{(k)}_{n+1})/s_{k,n}\sqrt{1+b_k}$ follows a Student's t distribution with ν_k degrees of freedom under the model M_k, we obtain after some straight-forward manipulation

$$D \doteq \frac{1}{2} \sum_{i=1}^{m} \sum_{j \geq i+1}^{m} P(M_i|\underset{\sim}{y}_n)P(M_j|\underset{\sim}{y}_n)$$

$$\times \left[\frac{\{(\nu_j-2)\nu_i s_{i,n}^2(1+b_i)-(\nu_i-2)\nu_j s_{j,n}^2(1+b_j)\} \cdot \{(\nu_j+1)\nu_i s_{i,n}^2(1+b_i)-(\nu_i+1)\nu_j s_{j,n}^2(1+b_j)\}}{(\nu_i-2)(\nu_j-2)\nu_i s_{i,n}^2(1+b_i) \cdot \nu_j s_{j,n}^2(1+b_j)} \right.$$

$$\left. + (\tilde{y}_n^{(i)}-\tilde{y}_n^{(j)})^2 \left\{ \frac{\nu_j+1}{\nu_i s_{i,n}^2(1+b_i)} + \frac{\nu_j+1}{\nu_j s_{j,n}^2(1+b_j)} \right\} \right] \qquad (4.4.4)$$

for $\min\{\nu_k\} > 2$.

As n becomes large, equation (4.4.4) tends to

$$D \doteq \frac{1}{2} \sum_{i=1}^{m} \sum_{j \geq i+1}^{m} P(M_i|\underset{\sim}{y}_n)P(M_j|\underset{\sim}{y}_n)$$

$$\times \left[\frac{\{s_{i,n}^2(1+b_i)-s_{j,n}^2(1+b_j)\}^2}{s_{i,n}^2(1+b_i) \cdot s_{j,n}^2(1+b_j)} + (\tilde{y}_n^{(i)}-\tilde{y}_n^{(j)})^2 \left\{ \frac{1}{s_{i,n}^2(1+b_i)} + \frac{1}{s_{j,n}^2(1+b_j)} \right\} \right]$$

$$(4.4.5)$$

It is noted that the last expression would be obtained by replacing $\sigma^2+\sigma_i^2$ in the Box and Hill discrimination design criterion (1.2.20) for the σ known case with $s_{i,n}^2(1+b_i)$. From the equations (4.2.2) and (4.2.20), it can be seen that, for the σ known case, $\sigma^2+\sigma_i^2$ is $Var(y_{n+1}|M_i,\underset{\sim}{y}_n,\sigma^2)$, the variance of the distribution of y_{n+1} given M_i and $\underset{\sim}{y}_n$ while, for the σ unknown case, this variance tends to $s_{i,n}^2(1+b_i)$ as n becomes large.

Suppose there is some independent information about σ^2 from the replicated runs either in the preparatory investigation or in the current runs which give a sum of squares νs^2 with ν degrees of freedom. Then $\nu_i s_{i,n}^2$ in (4.4.4) can be separated into two parts

$$\nu_i s_{i,n}^2 = \nu s^2 + S_{i,n}. \qquad (4.4.6)$$

As ν becomes large, the expression (4.4.4) approaches the discrimination design criterion for the σ known case, as we would anticipate.

4.5 Checking the constraint (H) that an adequate model is included.

We have shown, in Sections 4.1, 4.2 and 4.3, that we do not necessarily require replicated runs to carry out the discrimination procedure. It is important, however, to be able to check whether or not an adequate model is included among those considered. When the data contain replicated runs so that

$$\underset{\sim}{y}_n' = (y_{11}, y_{12}, \ldots, y_{1n_1} | y_{21}, y_{22}, \ldots, y_{2n_2} | \ldots | y_{g1}, y_{g2}, \ldots, y_{gn_g}) \qquad (4.5.1)$$

where $n = \sum_{i=1}^{g} n_i$, the sum of squares of residuals $S_{i,n}$ for the model M_i can be separated into two parts, the pure error sum of squares

$$\nu s^2 = \sum_j \sum_k (y_{jk}-\bar{y}_j)^2 \text{ with degrees of freedom } \nu = n-g, \tag{4.5.2}$$

and the lack of fit sum of squares

$$S_{\ell 0f}^{(i)} = \sum_j n_j(\bar{y}_j-\hat{y}_j)^2 \text{ with degrees of freedom } g-p_i. \tag{4.5.3}$$

The models may be checked individually by a lack of fit test comparing $(S_{\ell 0f}/(g-p_i))/s^2$ with a suitable percentage point of the F distribution with degrees of freedom $(g-p_i,\nu)$. In the Bayesian approach [9], residuals may be parameterized according to

$$\gamma_j = \mu_j - f_i(\underset{\sim}{\xi}_j,\underset{\sim}{\theta}_i) \tag{4.5.4}$$

where $\underset{\sim}{\xi}_j$ is the experimental condition at which the j-th group of the replicates $(y_{j1},y_{j2},\ldots,y_{jn_j})$ were obtained, and μ_j is the true value of the response at $\underset{\sim}{\xi}_j$. It can be shown that checking if the 100(1-α)% HPD region of the posterior distribution of the lack of fit parameters $\underset{\sim}{\Gamma}' = (\gamma_1,\gamma_2,\ldots,\gamma_g)$ contains the point $\underset{\sim}{\Gamma} = \underset{\sim}{0}$ is computationally equivalent to the classical lack of fit test mentioned above with a significance level of 100α%.

The overall checking of the constraint H may be done in the following way. We first reduce the data $\underset{\sim}{y}_n$ to $\underset{\sim}{y}_n^*$ by replacing the replicated observations $y_{j1},y_{j2},\ldots,y_{jn_j}$ with their arithmetic mean $\bar{y}_j$ (with variance σ^2/n_j). This reduction in the data affects neither $P(M_i|\sigma,\underset{\sim}{y}_n)$ in (4.1.4) nor $p(y_{n+1}|\sigma,M_i,\underset{\sim}{y}_n)$ in (4.2.2), and the residual sum of squares $S_{i,n}$ reduces to $S^{(i)}_{\ell 0f}$, the lack of fit sum of squares for M_i. Therefore by replacing $S_{i,n}$ with $S^{(i)}_{\ell 0f}$ in the equations (4.1.18) and (4.2.20), we can compute $P(M_i|\underset{\sim}{y}^*)$ and so the posterior density of σ based on $\underset{\sim}{y}^*$ is given by

$$p(\sigma|y^*) = \sum_{i=1}^{m} P(M_i|\underset{\sim}{y}^*)p(\sigma|M_i,\underset{\sim}{y}^*) \tag{4.5.5}$$

which is also <u>conditional on the constraint H that an adequate model has been included</u>. On the other hand, the information on σ^2 independent of the constraint H is available through

$$p(\sigma|s^2) = \frac{2\left(\frac{\nu s^2}{2}\right)^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})}\,\sigma^{-(\nu+1)}\,e^{-\frac{\nu s^2}{2\sigma^2}}. \tag{4.5.6}$$

Plotting and comparing the two densities (4.5.5) and (4.5.6) will give an overall check of the constraint H.

When σ^2 is known or some prior information on σ is available we may compute $P(M_i|\underset{\sim}{y})$ based on the noninformative prior density following the procedures given in (4.1.2). An overall check of the constraint H will be done by checking the location of known σ, or the prior information available on σ, with respect to $p(\sigma|\underset{\sim}{y}_n)$ obtained above.

4.6 Froment and Mezaki's problem.

Using the theory we have developed in this chapter, we now consider some examples to examine to what extent we can shed light on supposed difficulties which have been experienced in using the model discrimination procedure.

To examine the efficacy of Box and Hill's model discrimination method, Froment and Mezaki [5] simulated the sequential procedure making use of the data of Hosten [6] for the isomerization of n-pentane over Pt-$A\ell_2O_3$ catalyst in the presence of hydrogen with chlorine added as $CC\ell_4$ in order to maintain the catalyst activity. After eliminating other possible mechanisms, two models (M_1 and M_2) were chosen as worthy of further study. They are derived assuming the adsorption or desorption process within the isomerization step as rate controlling.

M_1 (Adsorption rate controlling):

$$\eta = \frac{1}{\theta_{11} A \frac{K+U}{KU}} (a + c\theta_{12}) \quad (4.6.1)$$

M_2 (Desorption rate controlling):

$$\eta = \frac{1}{\theta_{21} A \frac{K+U}{KU}} (a + b\theta_{22}) \quad (4.6.2)$$

where

$$a = [A\xi_2 + \frac{U(AK-B)}{K+U}(1 - \frac{1}{U})]$$

$$\times \ln[1 - \frac{A(K+U)}{U(AK-B)}\xi_1] + A(1 - \frac{1}{U})\xi_1 \quad (4.6.3)$$

$$b = (A - \frac{AK-B}{K+U})\ln[1 - \frac{A}{U}\frac{(K+U)}{(AK-B)}\xi_1] - \frac{A\xi_1}{U} \quad (4.6.4)$$

and

$$c = (B + \frac{U(AK-B)}{K+U})\ln[1 - \frac{A(K+U)}{U(AK-B)}\xi_1] + A\xi_1. \quad (4.6.5)$$

In these expressions, η is the space time (weight of catalyst/ molar feed rate of n-pentane), ξ_1 is the conversion of normal into iso-pentane, ξ_2 is the molar ratio of hydrogen to n-pentane at the reactor outlet. θ_{11} and θ_{21} are the forward rate constants, and θ_{12} and θ_{22} are the adsorption equilibrium constants for i-pentane and n-pentane, respectively. A, B, K and U are fixed known constants.

The output response was taken to be η, and ξ_1 and ξ_2 were regarded as the input variables. θ's are unknown parameters. The data used in their study were obtained at 425°C with 0.0121 mole per cent chlorine and are listed in Table 4.6.1. There were several replicated runs performed at the same temperature 425°C but with a different chlorine level 0.0242. Using these runs the experimental error variance was estimated to be 2.999×10^{-3} (g-cat./g-moles n-pentane/hr)^2. Froment and Mezaki therefore proceeded as though 2.999×10^{-3} were the exact known value of σ^2.

Froment and Mezaki first chose 3 observations from the 13 observations listed in Table 4.6.1, regarded them as from "preliminary" experiments, and computed the posterior model probabilities of M_1 and M_2 using Box and Henson's formula [2]. They decided the next best experimental condition by maximizing the Box and Hill discrimination design criterion over the experimental region, chose the observation from the data in Table 4.6.1 that had been taken at the condition nearest to this next best condition, and recomputed new model probabilities with all four observations combined by the sequential formula

$$P(M_i|\underset{\sim}{y}_{n+1},\sigma^2) \propto P(M_i|\underset{\sim}{y}_n,\sigma^2)p(y_{n+1}|M_i,\sigma^2). \qquad (4.6.6)$$

This simulation was carried out several times, each time starting with a different set of "preliminary" observations selected from the data. The results are listed in Table 4.6.2.

From this table, it looks as if the posterior probabilities of the models very much depend on the particular set of three "preliminary" observations chosen. In Cases 2 and 3, in particular, M_2 looks overwhelmingly superior to M_1 while in Cases 1, 4, 5 and 6, M_1 is given the posterior probability 1.000 after only five observations. Furthermore, in Case 4

Table 4.6.1 Data of n-pentane isomerization used in the study by Froment and Mezaki. Reaction temperature 425°C; chlorine level 0.0121 mole percent; A = 92.65, B = 6.37, K = 2.07, U = 0.9115

Run No.	ξ_1	ξ_2	η
105	0.4025	4.853	5.92
106	0.3500	5.253	3.84
107	0.2784	5.290	2.84
108	0.2001	5.199	1.75
119	0.3529	6.833	5.74
120	0.2728	7.330	3.84
121	0.2038	7.344	2.66
109	0.3248	7.638	5.28
110	0.2571	8.514	3.90
111	0.2011	8.135	2.65
114	0.3017	10.598	5.73
115	0.2423	11.957	4.37
116	0.1734	10.227	2.65

$P(M_2|\underset{\sim}{y},\sigma) = 0.814$ after 4 observations (Runs 109, 120, 121 and 108), heavily favoring M_2. However, with only one more observation (Run 105) added, the posterior probability for M_2 went down to 0.000! These erratic results certainly raise a serious anxiety about the posterior model probabilities. Froment and Mezaki concluded that care has to be taken in the choice of preliminary experimental conditions.

When closely examined, however, it is readily shown that this disturbing instability in the results which these authors believed they had found arises either because of a gross under-estimation of the experimental variance σ^2 or else because none of the models considered is really adequate. We can see this as follows. When all the thirteen observations are used in the least squares procedure, the sum of squares of residuals $S_1 = 0.70$ for M_1 and $S_2 = 1.05$ for M_2 are obtained. The residual variances are respectively 0.0636 and 0.0954 and these are respectively 21 and 32 times as high as the estimated variance. The thirteen residuals from the model M_1 which are plotted in Figures 4.6.1 and 4.6.2 do not reveal any obvious lack of fit, however.

When the posterior model probabilities are computed by the methods for the σ unknown situation in Sections 4.1 and 4.2 we obtain the results given in Table 4.6.3, which show no irregularities as encountered by Froment and Mezaki. In Cases 2 and 3 we are told that with only 4 or 5

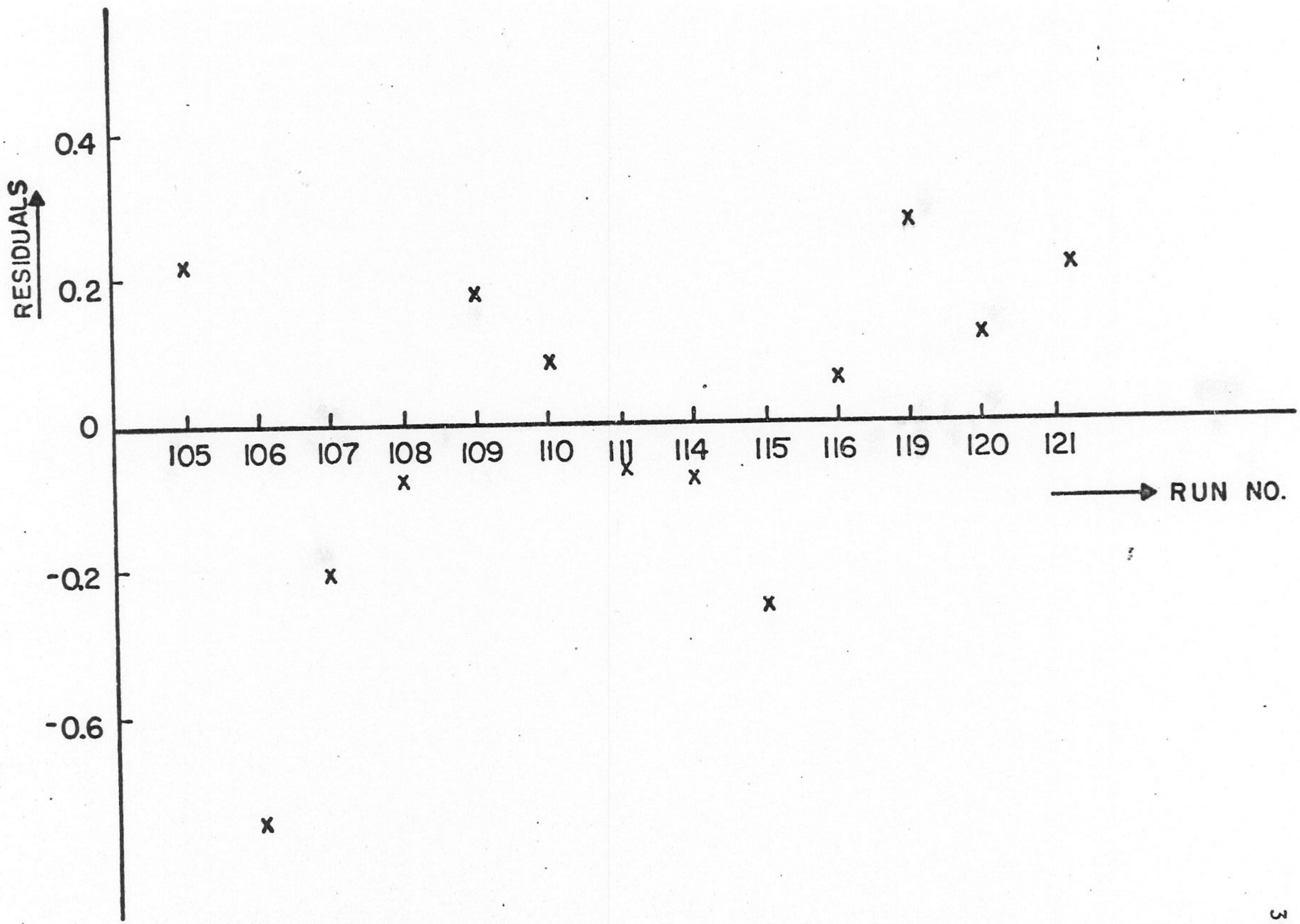

Figure 4.6.1. Thirteen residuals from the model M_1 against the run number.

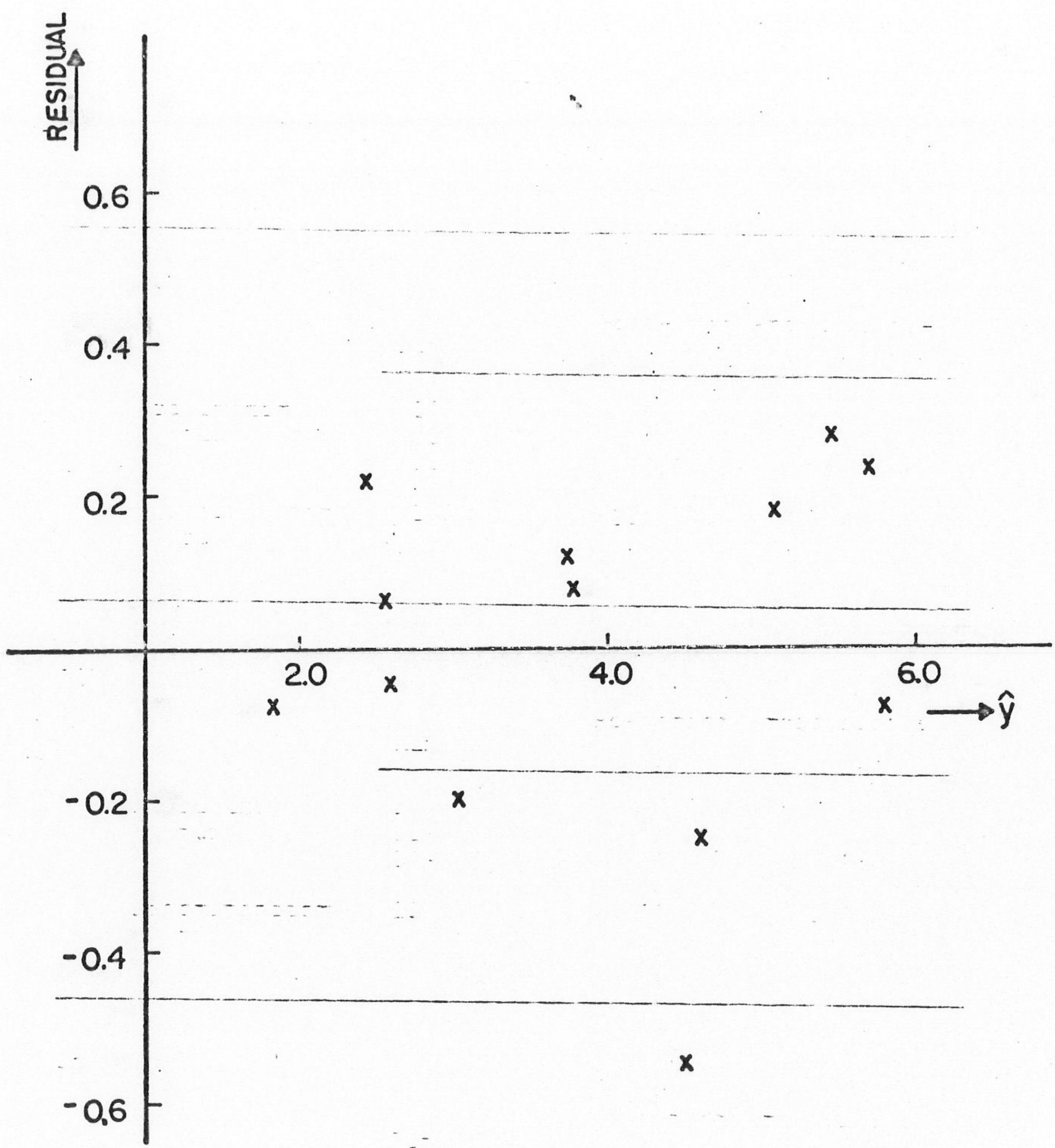

Figure 4.6.2. Thirteen residuals from the model M_1 against the estimated response $\hat{y}$.

Table 4.6.2 Results of the "simulation" by Froment and Mezaki with 6 different sets of "preliminary" runs.

Case 1

	Run No.	Posterior Probabilities M_1	Posterior Probabilities M_2
Preliminary runs	108 121 111	0.517	0.483
Discriminatory runs	105	1.000	0.000

Case 2

	Run No.	Posterior Probabilities M_1	Posterior Probabilities M_2
Preliminary runs	106 119 109	0.036	0.964
Discriminatory runs	120	0.000	1.000

Case 3

	Run No.	Posterior Probabilities M_1	Posterior Probabilities M_2
Preliminary runs	106 107 108	0.519	0.481
Discriminatory runs	109	0.147	0.853
	120	0.002	0.998

Table 4.6.2 (continued)

Case 4

	Run No.	Posterior Probabilities M_1	M_2
Preliminary runs	109		
	120		
	121	0.419	0.581
Discriminatory runs	108	0.186	0.814
	105	1.000	0.000

Case 5

	Run No.	Posterior Probabilities M_1	M_2
Preliminary runs	106		
	120		
	116	0.439	0.561
Discriminatory runs	109	0.968	0.032
	105	1.000	0.000

Case 6

	Run No.	Posterior Probabilities M_1	M_2
Preliminary runs	114		
	115		
	116	0.788	0.212
Discriminatory runs	108	1.000	0.000

observations there is practically no basis for prefering either model, while in Cases 1, 4, 5 and 6 the posterior probabilities of the models gradually point toward the superiority of the model M_1.

The posterior model probabilities computed from the equation (4.1.13) using all the thirteen observations assuming that σ is not known are $P(M_1|\underset{\sim}{y}_{13}) = 0.966$ and $P(M_2|\underset{\sim}{y}_{13}) = 0.034$. The posterior density of σ, $p(\sigma|\underset{\sim}{y}_{13})$, obtained by (4.1.2) is plotted in Figure 4.6.3. It should be noted that the estimate of σ^2 used by Froment and Mezaki is located at the point at which the posterior density is nearly zero.

Incidentally, as far as this example is concerned, whether we use Box and Henson's result or the method described in Section 3.4 for $P(M_i|\underset{\sim}{y},\sigma)$ in the equation (4.1.4), only slight differences occur in the posterior probabilities of the models as is shown in Table 4.6.4.

The Froment and Mezaki example does not, unfortunately, seem to be an isolated one. Wentzheimer [10] also experienced severe instability of posterior probabilities when he used the model discrimination method in studying the gas-phase catalytic methanation of four different temperatures. The eleven models he considered are listed in the appendix A4.1. As in the example given above, the estimate (0.00111×10^{-10}) of the experimental error variance was obtained from several replicated runs at 1100°F that had

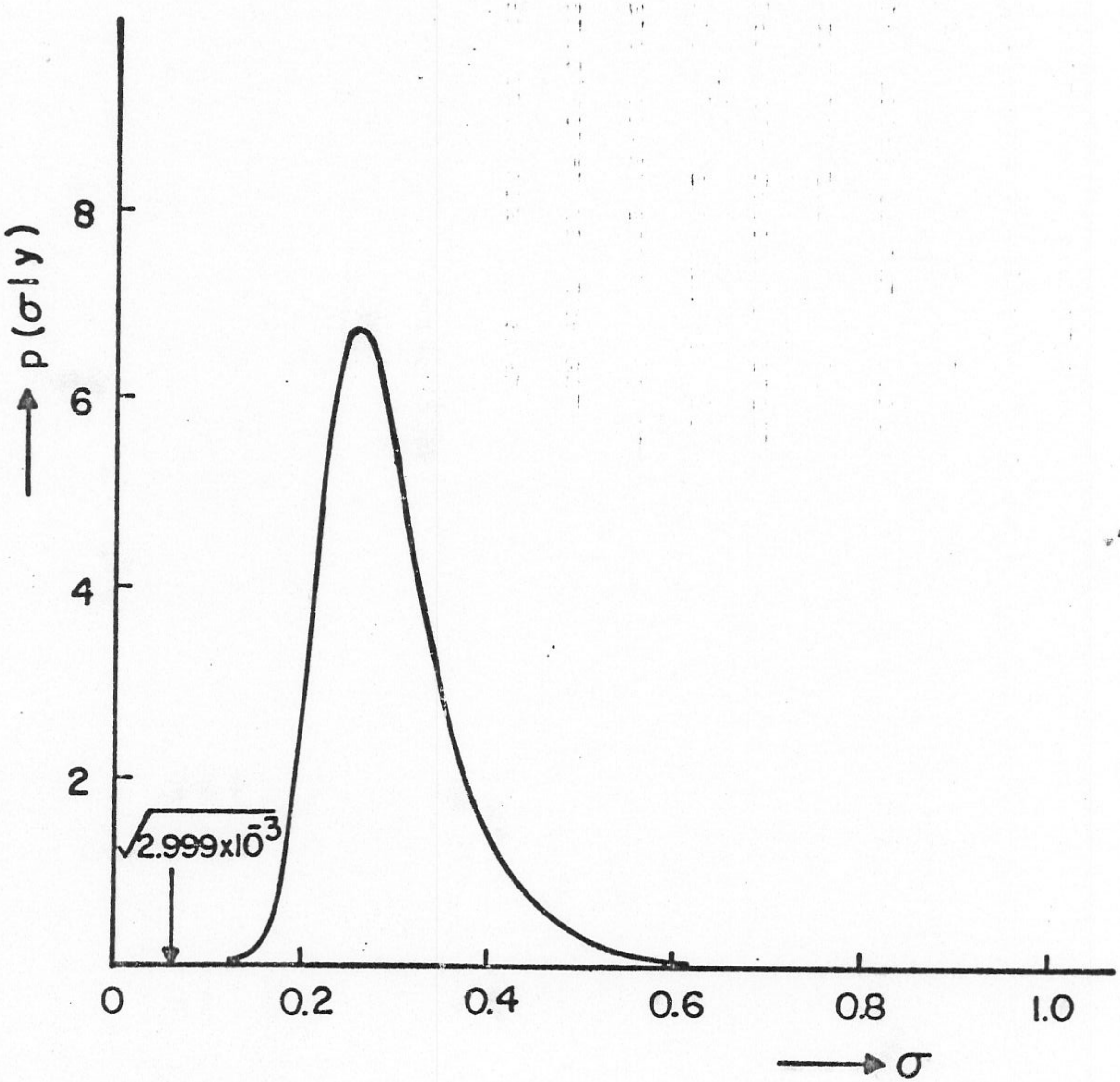

Figure 4.6.3. The posterior density of σ based on the thirteen observations of Froment and Mezaki assuming no prior information on σ.

been carried out in the preparatory investigation and Wentzheimer proceeded as if this estimate were the exact known value of σ^2. Table 4.6.5 shows the posterior probabilities obtained at various stages of the experimentation at 1040°F. However, by checking the residual sum of squares after 21 runs, listed in Table 4.6.6, we find that, even for the model M_2 which has the smallest residual sum of squares, the residual variance = $5.744\times10^{-10}/(21-5)$ = 0.359×10^{-10} and this is 323 times as high as the estimated variance given above. Obviously the instability encountered by Wentzheimer has the same cause as that encountered in Froment and Mezaki's example.

Table 4.6.3 Posterior model probabilities assuming no prior knowledge of σ

Case 1

	Run No.	M_1	M_2
Preliminary runs	108, 121, 111	.506	.494
	105	.742	.258

Case 2

	Run No.	M_1	M_2
Preliminary runs	106, 119, 109	.515	.485
	120	.486	.514

Case 3

	Run No.	M_1	M_2
Preliminary runs	106, 107, 108	.494	.506
	109	.528	.472
	120	.503	.497

Table 4.6.3 (continued)

Case 4

	Run No.	M_1	M_2
Preliminary runs	109		
	120		
	121	.597	.403
	108	.608	.392
	105	.936	.064

Case 5

	Run No.	M_1	M_2
Preliminary runs	106		
	120		
	116	.492	.508
	109	.549	.451
	105	.661	.339

Case 6

	Run No.	M_1	M_2
Preliminary runs	114		
	115		
	116	.527	.473
	108	.838	.162

Table 4.6.4 Comparison of the posterior probabilities of the models after the "preliminary runs" obtained by Box and Henson's formula and the methods given in Section 3.4 for $P(M_i|\underset{\sim}{y}_n,\sigma)$ in equation (4.1.4). (Data from Froment and Mezaki [5])

	$P(M_i\|\underset{\sim}{y})$ using Box and Henson's formula for $P(M_i\|\underset{\sim}{y},\sigma)$ in the equation (4.1.4)		$P(M_i\|\underset{\sim}{y})$ using the procedure given in Section 3.4 for $P(M_i\|\underset{\sim}{y},\sigma)$ in the equation (4.1.4)	
	M_1	M_2	M_1	M_2
Case 1	.502	.498	.506	.494
Case 2	.500	.500	.515	.485
Case 3	.501	.499	.494	.506
Case 4	.586	.414	.598	.402
Case 5	.494	.506	.492	.508
Case 6	.588	.412	.567	.433

Table 4.6.5 Posterior probabilities obtained by Wentzheimer at 1040°F.

Model No.	Run No.			
	6	9	12	15
1	0.163	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0
5	0.464	0.0	0.0	0.0
6	0.350	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0
9	0.012	0.001	0.0	1.0
10	0.008	0.999	1.0	0.0
11	0.003	0.0	0.0	0.0

Table 4.6.6 Residual sum of squares after 21 runs at 1040°F. (Wentzheimer data)

Model No.	Residual sum of squares$\times 10^{10}$
1	6.699
2	5.744
3	6.968
4	6.957
5	6.551
6	6.482
7	14.73
8	13.34
9	13.61
10	13.09
11	7.436

4.7 Effect of different methods for evaluating the derivatives x_{uj}.

In the process of model discrimination, it becomes necessary to compute the derivatives

$$x_{uj} = \left[\frac{\partial f(\underset{\sim}{\xi}_u;\underset{\sim}{\theta})}{\partial \theta_j}\right]_{\underset{\sim}{\theta}=\hat{\underset{\sim}{\theta}}} \qquad (4.7.1)$$

for each model where $\hat{\underset{\sim}{\theta}}$ is the estimate of parameters based on the available data. These derivatives are very often approximated by

$$x_{uj} = \left\{f(\underset{\sim}{\xi}_u;\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_j+\Delta\theta_j,\ldots,\hat{\theta}_p) - f(\underset{\sim}{\xi}_u;\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_j,\ldots,\hat{\theta}_p)\right\}/\Delta\theta_j \qquad (4.7.2)$$

either because this is computationally more convenient or sometimes because it is impossible to obtain the response function analytically, for example when the models are given in terms of differential equations that cannot be integrated analytically.

In the above approximation, $\Delta\theta_j$ is usually set to $\hat{\theta}_j \times d$ where d is a small number such as 0.01. It frequently happens, however, that the mechanistic models contain five or six parameters and some of the least squares estimates are very close to zero (for example, see [10]), and so the $\Delta\theta$'s will become even smaller. When this happens, the round-off error in x_{uj} becomes serious since we are dividing one number very close to zero by another. Even with models containing only two parameters such as in Froment and Mezaki's example mentioned previously, this effect can be substantial. Listed in Table 4.7.1 are the posterior probabilities of models (σ unknown) for Froment and Mezaki's cases 3, 4, 5 and 6 computed using analytically evaluated derivaties and also using the approximate derivatives obtained in the manner described above. Also shown are the estimates of the parameters. In Cases 3 and 4, where the second parameters θ_{12} and θ_{22} are sometimes estimated to be very close to zero, the effect of the different procedures to evaluate the derivatives is considerable while, in Cases 5 and 6, where none of the parameter estimates are close to zero, practically identical probabilities are obtained for both procedures.

It is not difficult to imagine that this effect, when coupled with an even moderately under-estimated error variance, will be large enough to cause instability in the posterior probabilities. This is because, in formula (4.2.2), the contribution of the derivatives comes through the term $\sigma^2(1+b_i)$ where $b_i = \underset{\sim}{x}_{n+1}^{(i)}(X_{i,n}'X_{i,n})^{-1}\underset{\sim}{x}_{n+1}^{(i)'}$.

Table 4.7.1 **Posterior model probabilities obtained using analytically computed derivatives and numerically computed derivatives**

	Run No.	Posterior Probabilities with analytically computed derivatives M_1	M_2	with numerically computed derivatives M_1	M_2	Estimates of Parameters M_1: $\hat{\theta}_{11}$	$\hat{\theta}_{12}$	M_2: $\hat{\theta}_{21}$	$\hat{\theta}_{22}$
Case 3	106, 107, 108	.494	.506	.500	.500	7.278×10^{-1}	1.972×10^{-5}	7.278×10^{-1}	5.917×10^{-4}
	109	.528	.472	.447	.553	7.103×10^{-1}	1.925×10^{-6}	7.104×10^{-1}	1.309×10^{-5}
	120	.503	.497	.498	.502	7.103×10^{-1}	4.995×10^{-5}	7.110×10^{-1}	1.182×10^{-5}
Case 4	109, 120, 121	.597	.403	.588	.412	7.750×10^{-1}	3.497	6.999×10^{-1}	8.976×10^{-5}
	108	.608	.392	.711	.289	7.436×10^{-1}	1.898	7.031×10^{-1}	1.491×10^{-5}
	105	.936	.064	.998	.001	9.362×10^{-1}	1.083×10^{1}	1.212	6.863
Case 5	106, 120, 116	.492	.508	.492	.508	7.966×10^{-1}	2.129	8.710×10^{-1}	1.434
	109	.549	.451	.549	.451	7.630×10^{-1}	1.726	7.466×10^{-1}	3.147×10^{-1}
	105	.661	.339	.661	.339	9.246×10^{-1}	9.143	1.194	6.261
Case 6	114, 115, 11[illegible]	.527	.473	.527	.473	1.078	1.586×10^{1}	1.373	8.746
	108	.838	.162	.838	.162	1.022	1.226×10^{1}	9.234×10^{-1}	1.578

References

[1] Box, G. E. P. and Hill, W. J. (1967), Discrimination among mechanistic models, Technometrics, 9, 57.

[2] Box, G. E. P. and Henson, T. L. (1969), Model Fitting and Discrimination, Technical Report No. 211, Department of Statistics, University of Wisconsin, Madison, Wisconsin.

[3] Box, G. E. P. and Tiao, G. C. (1962), A further look at robustness via Bayes' theorem, Biometrika, 49, 419.

[4] Box, G. E. P. and Wilson, K. B. (1951), On the experimental attainment of optimum conditions, J.R.S.S., B, 13, 1.

[5] Froment, G. F. and Mezaki, R. (1970), Sequential discrimination and estimation procedures for rate modeling in heterogeneous catalysis, Chemical Engineering Science, 25, 293.

[6] Hosten, L. (1967), Ph.D. Dissertation, Rijksuniversiteit te Gent, Gent, Belgium.

[7] Hunter, W. G. and Hill, W. J. (1969), A note on designs for model discrimination: Variance Unknown Case, Technometrics, 11, 396.

[8] Sidik, S. M. (1972), Ph.D. Dissertation, Department of Mathematics and Statistics, Case Western Reserve University.

[9] Sredni, J. (1970), Ph.D. Dissertation, Department of Statistics, University of Wisconsin, Madison, Wisconsin.

[10] Wentzheimer, W. W. (1969), Ph.D. Dissertation, Department of Chemical Engineering, University of Pennsylvania.

Appendix A4.1 Wentzheimer's models for methanation reaction

Model 1

$$r = \frac{k(P_{H_2} - (P_W P_M / P_{CO} K_{EQ})^{0.33})}{(1 + K_{H_2}^{0.5}(P_W P_M / P_{CO} K_{EQ})^{0.16} + K_{CO} P_{CO} + K_W P_W + K_M P_M)^2}$$

Model 2

$$r = \frac{k(P_{H_2} - (P_M P_W / P_{CO} K_{EQ})^{0.33})}{(1 + K_{H_2}(P_W P_M / P_{CO} K_{EQ})^{0.33} + K_{CO} P_{CO} + K_W P_W + K_M P_M)}$$

Model 3

$$r = \frac{k(P_{CO} - P_W P_M / P_{H_2}^3 K_{EQ})}{(1 + K_{H_2}^{0.5} P_{H_2}^{0.5} + K_{CO} P_W P_M / P_{H_2}^3 K_{EQ} + K_W P_W + K_M P_M}$$

Model 4

$$r = \frac{k(P_{CO} - P_W P_M / P_{H_2}^3 K_{EQ})}{(1 + K_{H_2} P_{H_2} + K_{CO} P_W P_M / P_{H_2}^3 K_{EQ} + K_W P_W + K_M P_M)}$$

Model 5

$$r = \frac{k(P_{H_2}^3 P_{CO} K_{EQ}/P_W - P_M)}{(1 + K_{H_2}P_{H_2} + K_{CO}P_{CO} + K_W P_W + K_M P_{H_2}^3 P_{CO} K_{EQ}/P_W)}$$

Model 6

$$r = \frac{k(P_{H_2}^3 P_{CO} K_{EQ}/P_W - P_M)}{(1 + K_{H_2}^{0.5} P_{H_2}^{0.5} + K_{CO}P_{CO} + K_W P_W + K_M P_{H_2}^3 P_{CO} K_{EQ}/P_W)}$$

Model 7

$$r = \frac{k(P_{H_2}^3 P_{CO} K_{EQ}/P_M - P_W)}{(1 + K_{H_2}P_{H_2} + K_{CO}P_{CO} + K_W P_{H_2}^3 P_{CO} K_{EQ}/P_M + K_M P_M)}$$

Model 8

$$r = \frac{k(P_{H_2}^3 P_{CO} K_{EQ}/P_M - P_W)}{(1 + K_{H_2}^{0.5} P_{H_2}^{0.5} + K_{CO}P_{CO} + K_W P_{H_2}^3 P_{CO} K_{EQ}/P_M + K_M P_M)}$$

Model 9

$$r = \frac{k(P_{H_2} P_{CO} - P_M P_W / K_{EQ} P_{H_2}^2)}{(1 + K_{H_2}P_{H_2} + K_{CO}P_{CO} + K_W P_W + K_M P_M)^2}$$

Model 10

$$r = \frac{k(P_{H_2}P_{CO} - P_M P_W / P_{H_2}^2 K_{EQ})}{(1 + K_{H_2}^{0.5} P_{H_2}^{0.5} + K_{CO}P_{CO} + K_W P_W + K_M P_M)^3}$$

Model 11

$$r = \frac{k(P_{H_2}^{0.5}P_{CO} - P_M P_W / P_{H_2}^{2.5} K_{EQ})}{(1 + K_{H_2}^{0.5} P_{H_2}^{0.5} + K_{CO}P_{CO} + K_W P_W + K_M P_M)^2}$$

The output response was taken to be the reaction rate r and the partial pressures of the reactants P_{H_2}, P_{CO}, P_W, P_M were the input variables. The unknown parameters were k, K_{H_2}, K_{CO}, K_W, and K_M. K_{EQ} was the known equilibrium constant.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) University of Wisconsin	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

TOPICS IN MODEL BUILDING, Part IV

SOME PROBLEMS IN MODEL DISCRIMINATION

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific Interim

5. AUTHOR(S) (First name, middle initial, last name)

H. Kanemasu

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
November 1972	50	10

8a. CONTRACT OR GRANT NO. AFOSR-72-2363 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Air Force Office of Scientific Research 1400 Wilson Boulevard Arlington, Virginia

13. ABSTRACT

Certain authors have remarked on what they have believed was the instability of the posterior probabilities of the candidate models calculated using the Box and Hill discrimination procedure. We first present several results related to the discrimination procedure and then, making use of these results, show that this instability arose not for any inappropriateness of the discrimination procedure itself but because it was used under the conditions which violated critical assumptions.

DD FORM 1 NOV 65 1473

Security Classification

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Model discrimination						
Posterior probabilities of models						
Design of experiments						
Chemical engineering examples						

Security Classification