



**Discrimination between Regression Models to Determine the Pattern of Enzyme Synthesis in Synchronous Cell Cultures**

D. A. Williams

*Biometrics*, Volume 26, Issue 1 (Mar., 1970), 23-32.

Stable URL:

<http://links.jstor.org/sici?sici=0006-341X%28197003%2926%3A1%3C23%3ADBRTD%3E2.0.CO%3B2-1>

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

*Biometrics* is published by International Biometric Society. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ibs.html>.

---

*Biometrics*

©1970 International Biometric Society

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).

©2003 JSTOR

# DISCRIMINATION BETWEEN REGRESSION MODELS TO DETERMINE THE PATTERN OF ENZYME SYNTHESIS IN SYNCHRONOUS CELL CULTURES

D. A. WILLIAMS

*Department of Statistics, University of Edinburgh, Scotland*

## SUMMARY

Alternative theories for the synthesis of enzyme during the cell cycle lead to two regression models for the increase in enzyme concentration in synchronous cultures of yeast cells. One model is a segmented linear regression with three segments; the alternative is a smooth exponential. The analysis of observations from a series of experiments designed to discriminate between the alternative models is discussed. Methods are described for estimating the segmented linear regression and for determining by simulation the distributions under the two models of the discriminating criterion, the ratio of maximised likelihoods. The consequences of non-validity of some of the assumptions made in setting up the regression models are investigated.

## 1. INTRODUCTION

The pattern of enzyme synthesis during the cell cycle is of considerable interest because of the light it throws on the regulation and sequence of chemical changes throughout the cycle. Many basic processes in the cell have to double in capacity during the cell lifetime and there is evidence that many processes do so continuously and exponentially. The majority of individual enzymes are synthesised only for a limited period of the cell cycle, the period varying from enzyme to enzyme. Some enzymes, however, are synthesised continuously throughout the cell cycle and previous work with bacteria has suggested that with such enzymes a pattern of synthesis alternative to the exponential is followed in which the rate of synthesis doubles sharply at some characteristic point in the cell cycle, remaining at a constant level both before and after this point.

This paper describes the methods used to analyse the data from a series of experiments which were carried out to determine whether this alternative pattern of synthesis applied to 3 enzymes, sucrase, alkaline phosphatase and acid phosphatase, which are synthesised continuously during the cell cycle of the yeast *Schizosaccharomyces pombe*.

The biological implications of this work and the experimental details are discussed more fully by Mitchison and Creanor [1969].

## 2. THE EXPERIMENTS

Information on the pattern of enzyme synthesis is obtained by studying the changes in concentration of the enzyme within a synchronous cell culture.

A perfectly synchronous culture is one in which, at any time, all cells are at the same growth stage. In particular the cells will divide in synchrony. In practice a synchronous culture is obtained by centrifuging a tube of cells, which causes the cells to arrange themselves along the tube according to cell size, and selecting a cross section of cells of uniform size. This produces a culture which is approximately but not perfectly synchronous. Moreover the culture will become less synchronous during its growth because of the natural variation in the generation time of the cells. The model proposed in section 3 relates strictly to cultures that remain perfectly synchronous, but the consequences of imperfect synchrony have been investigated and are described briefly in section 9.

Each culture was allowed to grow in a suitable medium for approximately 6 hours. At 5-minute intervals, small samples were taken; between successive samplings the culture was stirred to ensure that each sample was a random portion. The cells from each sample were freeze-dried and were assayed independently to determine the enzyme concentration. In any experiment only one of the 3 enzymes was investigated.

The generation time of these cells is about 2.4 hours. The observations, approximately 70 in number, therefore extended over 3 cycles covering about half of the first cycle, all of the second cycle, and most of the third cycle.

### 3. MODELS FOR THE OBSERVATIONS

If the pattern of enzyme synthesis is such that the rate of synthesis doubles at a characteristic point during the cell cycle, then the rate of increase of enzyme in a synchronous culture will follow a step function, with one step per generation. The observed enzyme concentrations which measure the cumulative production of enzyme will follow the integral of the step function, that is, a curve formed by a series of linear segments.

If enzyme synthesis increases smoothly and exponentially during the cell cycle then the observed enzyme concentration will follow an exponential curve.

Variation about the true curve will arise from three main sources:

- (1) Variations over time due to disturbance of the culture, deterioration of the medium, etc.
- (2) Sampling errors due to lack of uniformity of the culture.
- (3) Assay errors in determining the enzyme concentrations of the samples.

Variation due to (1) will be time-dependent. Variation due to (2) and (3) should be independent in successive samples but variances may increase with increase in the enzyme concentration. The errors about the true curve could therefore be heteroscedastic and not completely independent. Nevertheless I have followed the common practice of assuming that the errors are normally and independently distributed with constant variance. The effect of the non-validity of these assumptions on the results of the analysis is discussed in section 8.

The problem is thus regarded as one of discriminating between two regres-

sion models. If  $y_i$  is the observed enzyme concentration at time  $t_i$ , the two models are

(1) Segmented Model

$$y_i = f(\theta, t_i) + \epsilon_i, \quad \theta = (\alpha, \beta_1, \beta_2, \beta_3, T_1, T_2),$$

where

$$\begin{aligned} f(\theta, t) &= \alpha + \beta_1 t \quad \text{for } t \leq T_1 \\ &= \alpha + \beta_1 T_1 + \beta_2(t - T_1) \quad \text{for } T_1 \leq t \leq T_2 \\ &= \alpha + \beta_1 T_1 + \beta_2(T_2 - T_1) + \beta_3(t - T_2) \quad \text{for } T_2 \leq t, \end{aligned}$$

and the  $\epsilon_i$  are independently distributed  $N(0, \sigma_i^2)$ .  $T_1$  and  $T_2$  are the rate doubling times,  $T_2 - T_1$  equals the generation time. If the medium in which the cell culture grows does not change materially over the duration of the experiment, then  $\beta_1 : \beta_2 : \beta_3$  should be in the ratios 1 : 2 : 4.

(2) Smooth Model

$$y_i = g(\psi, t_i) + \epsilon_i, \quad \psi = (a, b, c),$$

where  $g(\psi, t) = a + b \exp(ct)$  and the  $\epsilon_i$  are independently distributed  $N(0, \sigma_i^2)$ .

The unusual feature of this problem is the presence of discontinuities in the derivative of the segmented model occurring at unknown points,  $T_1$  and  $T_2$ , in the range of the independent variate  $t$ . This raises difficulties in fitting the segmented model to the data and in discriminating between the alternative regression models.

#### 4. ESTIMATION OF THE SEGMENTED MODEL

The estimation of segmented regressions has been discussed by Hudson [1966]. The difficulties arise because at each of the observation times  $t_i$  the likelihood function is not differentiable with respect to  $T_1$  and  $T_2$ . In any of the intervals  $t_i \leq T_1 \leq t_{i+1}$ ,  $t_k \leq T_2 \leq t_{k+1}$  there may exist a local maximum (or supremum) of the likelihood function. Standard iterative procedures for obtaining maximum likelihood estimates are not suitable and it becomes necessary to evaluate the local supremum for each pair of intervals in turn. Hudson develops algorithms for this but discusses in detail only the case of 2 linear segments. The following procedure for fitting 3 linear segments merely extends Hudson's approach.

Let there be  $n$  observations  $(t_i, y_i)$  ordered so that the  $t_i$  increase with  $i$ . Consider the rectangle in the  $(T_1, T_2)$  plane defined by  $t_1 < t_i \leq T_1 \leq t_{i+1} < t_k \leq T_2 \leq t_{k+1} < t_n$ . Let the 3 segments  $t_1$  to  $t_i$ ,  $t_{i+1}$  to  $t_k$ , and  $t_{k+1}$  to  $t_n$  contain  $n_1$ ,  $n_2$ , and  $n_3$  points, respectively. Let the means of the  $t_i$  and their corrected sum of squares for the 3 segments be  $\bar{t}_1, \bar{t}_2, \bar{t}_3$  and  $S_1, S_2$ , and  $S_3$ .

Suppose that unconstrained lines fitted to the 3 segments by least squares

have slopes  $\beta_1^*$ ,  $\beta_2^*$ , and  $\beta_3^*$ , that they intersect at  $T_1^*$  and  $T_2^*$ , and that the total residual sum of squares is  $R^*$ .

It is then a straightforward exercise in least squares estimation under linear constraints to show that the residual sum of squares  $R(T_1, T_2)$  about fitted lines constrained to meet at  $(T_1, T_2)$  is given by

$$R(T_1, T_2) = R^* + \mathbf{m}'\mathbf{A}^{-1}\mathbf{m},$$

where

$$\mathbf{m} = \begin{bmatrix} (\beta_1^* - \beta_2^*)(T_1 - T_1^*) \\ (\beta_2^* - \beta_3^*)(T_2 - T_2^*) \end{bmatrix}$$

$$a_{11} = \frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{t}_1 - T_1)^2}{S_1} + \frac{(\bar{t}_2 - T_1)^2}{S_2},$$

$$a_{12} = a_{21} = -\frac{1}{n_2} - \frac{(\bar{t}_2 - T_1)(\bar{t}_2 - T_2)}{S_2},$$

$$a_{22} = \frac{1}{n_2} + \frac{1}{n_3} + \frac{(\bar{t}_2 - T_2)^2}{S_2} + \frac{(\bar{t}_3 - T_2)^2}{S_3}.$$

If  $t_i \leq T_1^* \leq t_{i+1}$  and  $t_k \leq T_2^* \leq t_{k+1}$ , then  $R(T_1, T_2)$  is a minimum, equal to  $R^*$ , at  $(T_1^*, T_2^*)$ . If not, then  $R(T_1, T_2)$  achieves its least value at some point on the boundary of the rectangle whose corners are  $(t_i, t_k)$ ,  $(t_{i+1}, t_k)$ ,  $(t_i, t_{k+1})$ , and  $(t_{i+1}, t_{k+1})$ . I have been unable to develop a simple method for determining where on the boundary the least value of  $R(T_1, T_2)$  is attained. For the enzyme concentration data, however,  $n$  is large, about 70, and it suffices to determine the maximum likelihood (ML) estimates of  $(T_1, T_2)$  to the nearest  $(t_i, t_k)$ . Values of  $R$  have therefore to be calculated only for the 4 corners of the rectangle.

To find the overall ML estimate of  $(T_1, T_2)$  the above investigation is repeated for all  $j, k$ . The calculations can be made more efficient by the following measures:

- (1) First obtain the value of  $R(T_1, T_2)$  for some initial estimate of  $(T_1, T_2)$ , from a graphical plot of the  $(t_i, y_i)$  or otherwise. Call this  $R_{\min}$ .
- (2) The range of values of  $j$  and  $k$  investigated can be restricted after study of the graph. If it were necessary to investigate values of  $j$  and  $k$  for which  $j < 2$  or  $k > n - 2$  or  $k - j < 2$ , the calculations described above would require modification.
- (3) Cycle increases of  $k$  within increases of  $j$ . Means and sums of squares and products of  $t$  and  $y$  do not then have to be recalculated completely but are merely adjusted for the addition or subtraction of a single observation.
- (4) For each  $j, k$  first calculate  $R^*$ . If this is greater than  $R_{\min}$  proceed to next interval.
- (5) If  $R^*$  is less than  $R_{\min}$  calculate  $(T_1^*, T_2^*)$ . If this lies within the rectangle, then  $(T_1^*, T_2^*)$  is the new best estimate, and  $R^*$  becomes the new  $R_{\min}$ . Proceed to the next interval.

(6) If  $(T_1^*, T_2^*)$  does not lie within the rectangle calculate  $R(t_i, t_k)$ . If  $R(t_i, t_k)$  is less than  $R_{\min}$ , then  $(t_i, t_k)$  is the new best estimate and  $R(t_i, t_k)$  becomes the new  $R_{\min}$ . Proceed to the next interval.

Note that the fitting of segmented regressions to a large number of observations requires considerable computing time. This becomes a severe restraint on the number of simulations one can reasonably perform when discriminating between regression models by the method discussed in the following section.

##### 5. DISCRIMINATION BETWEEN REGRESSION MODELS

The two regression models are separate in the sense that neither model is a special case of the other. Cox [1961] has discussed the general problem of testing separate families of hypotheses. He uses as discriminating criterion the ratio of maximised likelihoods  $\lambda$ , which here reduces to the ratio of the residual sums of squares, i.e.

$$\lambda = \frac{\text{residual sum of squares about fitted segmented model}}{\text{residual sum of squares about fitted smooth model}}.$$

To use this criterion knowledge is needed of the distributions of  $\lambda$  under the assumption in turn that each of the two models is true. The distribution of  $\lambda$  depends not only on the form of the true model but also on the parameters which are unknown. If minimal sets of sufficient statistics for the parameters exist an exact test would use the conditional distributions of  $\lambda$  given the observed values of the sufficient statistics. In the absence of sufficient statistics the distributions of  $\lambda$  could be determined conditional on the ML estimators but this approach would be intractable both analytically and by simulation. The alternative adopted here follows Cox ([1961] §8) and uses distributions of  $\lambda$  assuming that the parameters take values equal to those estimated by maximum likelihood from the data. This approach is justified in general by the asymptotic sufficiency of the maximum likelihood estimators.

For each model in turn, therefore, the question is asked 'what is the distribution of  $\lambda$  if the model is assumed true with parameters as estimated by maximum likelihood from the data?' These distributions are denoted by  $\Lambda_r$  and  $\Lambda_p$ .

Cox developed asymptotic expressions for the expectation and variance of  $\log \lambda$  and established that  $\log \lambda$  is asymptotically normal. His results, however, assume that the likelihood function is differentiable with respect to all of the parameters and are therefore not valid for our problem. In place of his results I have used simulation to indicate the ranges of the distribution of  $\lambda$ .

Let  $\hat{\theta}$  and  $\hat{\psi}$  be the ML estimates from the observations of the regression parameters under the two models. Let  $\hat{\sigma}_r^2$  and  $\hat{\sigma}_p^2$  be the estimates of residual variance obtained by dividing the residual sums of squares by  $n - 6$  and  $n - 3$ , respectively.

On the assumption that the segmented model is true a set of simulated enzyme concentrations is obtained by calculating  $f(\hat{\theta}, t_i) + \epsilon_i$ , where the  $\epsilon_i$  are a set of pseudo-random normal deviates with variance  $\hat{\sigma}_f^2$ . To these simulated concentrations both regression models are fitted and the ratio  $\lambda$  of the residual sums of squares calculated. This is repeated 10 times using different sets of random normal deviates to obtain 10 observations, which are denoted by  $\lambda_{fk}$  for  $k = 1, 2, \dots, 10$ , on the distribution  $\Lambda_f$ .

Similarly, 10 observations  $\lambda_{gk}$  on the distribution  $\Lambda_g$  are obtained by fitting both regression models to simulated concentrations  $g(\hat{\psi}, t_i) + \epsilon_i$ , where the  $\epsilon_i$  are further sets of pseudo-random normal deviates with variance  $\hat{\sigma}_g^2$ .

Thus 10 observations are made on each of the two distributions  $\Lambda_f$  and  $\Lambda_g$ . The mean of  $\Lambda_f$  is less than the mean of  $\Lambda_g$ . The further observation  $\lambda_0$ , the value of  $\lambda$  obtained by fitting both regressions to the data, is to be allocated to one of the two distributions.

Let  $m_f, m_g, s_f$ , and  $s_g$  be the means and standard deviations of the  $\lambda_{fk}$  and the  $\lambda_{gk}$ , respectively. Let  $d_f = \max\{m_f + 2s_f, \max \lambda_{fk}\}$  and  $d_g = \min\{m_g - 2s_g, \min \lambda_{gk}\}$ .

Then I regard  $\lambda_0$  as a possible observation from  $\Lambda_f$  if  $\lambda_0 < d_f$ , and as a possible observation from  $\Lambda_g$  if  $\lambda_0 > d_g$ . This leads naturally to 4 possible conclusions:

- if  $\lambda_0 < d_f$ ,  $\lambda_0 < d_g$  the true model is segmented,  
 $\lambda_0 > d_f$ ,  $\lambda_0 > d_g$  the true model is smooth,  
 $\lambda_0 > d_f$ ,  $\lambda_0 < d_g$  neither model is true,  
 $\lambda_0 < d_f$ ,  $\lambda_0 > d_g$  no discrimination between the two models is possible.

No justification of this rule in terms of misclassification probabilities is claimed, for the form of the distributions  $\Lambda_f, \Lambda_g$  is not known.

## 6. NUMERICAL EXAMPLE

The above analysis is illustrated using as an example the concentration of sucrose from a synchronous culture. A graph of the concentrations is given in Figure 1.

The parameters of the fitted segmented model were  $\hat{T}_1 = 22.6$ ,  $\hat{T}_2 = 52.7$ ,  $\hat{\alpha} = 74.6$ ,  $\hat{\beta}_1 = 2.55$ ,  $\hat{\beta}_2 = 4.43$ ,  $\hat{\beta}_3 = 10.50$ , and  $\hat{\sigma}_f = 5.33$ . The residual sum of squares was 1901.

The parameters of the fitted smooth model were  $\hat{a} = 16.65$ ,  $\hat{b} = 63.38$ ,  $\hat{c} = 0.02711$ , and  $\hat{\sigma}_g = 7.14$ . The residual sum of squares was 3571.

The ratio of residual sums of squares was  $\lambda_0 = 0.532$ .

10 sets of concentrations were simulated assuming the segmented model true by adding random normal deviates with standard deviation  $\hat{\sigma}_f = 5.33$  to the fitted expected concentrations. The 10  $\lambda_{fk}$  values were:

0.549 0.426 0.437 0.344 0.508 0.551 0.461 0.490 0.423 0.536.

Similarly, by adding sets of random normal deviates with s.d.  $\hat{\sigma}_g = 7.14$

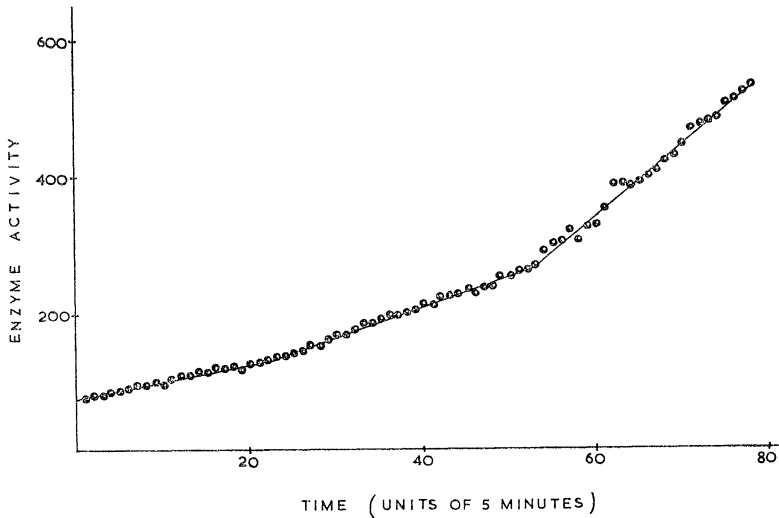


FIGURE 1

THE ACTIVITIES OF SUCRASE (TO WHICH ENZYME CONCENTRATIONS ARE PROPORTIONAL) IN SAMPLES FROM A SYNCHRONOUS CULTURE. THE FITTED SEGMENTED LINEAR REGRESSION IS SHOWN.

to fitted expectations  $16.65 + 63.38 \exp(0.02711t_i)$  and fitting both regression models to these simulated concentrations the following 10 values of  $\lambda_{\sigma k}$  were obtained:

1.213 1.227 1.269 1.183 1.264 1.000 0.998 1.044 0.951 1.031.

The value of  $\lambda$  calculated from the data, 0.532, lies within the range of the  $\lambda_{fk}$  but well outside the range of the  $\lambda_{\sigma k}$  and I conclude that the true model is segmented. In fact, in the majority of the experiments the conclusion was as obvious as in this example and clearly valid irrespective of the form of the distributions  $\Lambda_f$  and  $\Lambda_\sigma$ .

#### 7. USE OF ASYNCHRONOUS CONTROLS AND A SUMMARY OF THE EXPERIMENTS

Initially the results from only 3 or 4 experiments on each of the 3 enzymes were available and the observations were found to favour the segmented model in most but not all cases. It was clearly necessary to obtain further sets of observations on synchronous cultures and also, for two reasons, to obtain similar observations on asynchronous cultures which would act as controls. Firstly, the segmented form of the concentrations was possibly caused not by the synchrony of the cells but by some feature of the experiment such as the process of centrifuging. Secondly, the observations were possibly better fitted by the segmented model because the smooth model had less freedom-of-



fit. The slopes of the fitted segmented model were not constrained to be in the ratio 1 : 2 : 4 because slight changes may have occurred in the culture medium during growth which could alter these ratios. The exponential curve does not allow for such changes and can be at best only an approximation to the form of smooth increase in enzyme concentration. If enzyme concentration increase in the synchronous cultures was smooth but not exactly exponential, then the segmented model, because of its greater freedom, could have given the better fit. Asynchronous cultures were formed by remixing all of the cells in a tube after centrifuging and enzyme concentrations were determined and analysed exactly as for the synchronous cultures.

For two of the enzymes, alkaline phosphatase and acid phosphatase, the observations from asynchronous cultures favoured the smooth model in only a minority of the experiments. Study of the graphs of the enzyme concentration suggested that this was caused by the failure of the exponential to describe the smooth curve, rather than by true segmentation, and that a more flexible form of smooth curve was needed. A polynomial of degree 5 was chosen because this gives an easily fitted family of smooth curves which, from consideration of the magnitude of the parameter  $c$  and the power series expansion of  $g(\psi, t)$  over the range of observations, approximately includes the exponential as a special case. The exponential was therefore replaced by a polynomial in  $t$  of degree 5 and the experiments on these 2 enzymes from both synchronous and asynchronous cultures were completely reanalysed, the analysis being exactly as before except for the change in the form of  $g(\psi, t)$ . The polynomial will be an unsuitable model for enzyme concentration increase unless the range of observations lies between a minimum and the adjacent point of inflexion. No constraints were imposed on the fitted parameters of the polynomial but it was confirmed by inspection that the first and second differences of the fitted values were all positive.

This change in  $g(\psi, t)$  had the desired effect of altering the conclusions about the asynchronous cultures in favour of the smooth model, while not affecting the conclusions about the synchronous cultures. The sucrase experiments were not reanalysed.

The final results of the analyses of all the experiments is given in Table 1. From these results I conclude that there is strong evidence that the concentrations of all 3 enzymes follow the segmented model in synchronous cultures and, therefore, that there is a characteristic point in the cell cycle at which the rate of synthesis of enzyme doubles.

#### 8. EFFECT OF HETEROSCEDASTIC AND NON-INDEPENDENT ERRORS

The sources of variation about the regression models were discussed in section 3; heteroscedastic and non-independent errors were anticipated. This is confirmed by an examination of the residuals. In most experiments there is evidence that variation about the regression model increases with expected concentration. In several experiments also there is evidence for some serial correlation between the residuals, the first autocorrelation being usually positive but rarely greater than 0.3. The simulations, which have assumed

TABLE 1  
SUMMARY OF THE 39 EXPERIMENTS ANALYSED. THE TABLE GIVES THE  
NUMBER OF EXPERIMENTS IN EACH CATEGORY.

Enzyme	Culture	Model favoured by the observations			
		Segmented	Smooth	Neither	No discrimination
Sucrase	Synchronous	6	1	2	0
	Asynchronous	1	3	1	0
Alkaline Phosphatase	Synchronous	5	1	0	1
	Asynchronous	0	6	0	1
Acid Phosphatase	Synchronous	5	0	0	0
	Asynchronous	1	4	0	1
Total	Synchronous	16	2	2	1
	Asynchronous	2	13	1	2

that errors are independently distributed with constant variance, will be irrelevant if the distributions  $\Lambda_r$  and  $\Lambda_p$  are critically dependent on these assumptions.

A limited investigation of this, for  $\Lambda_r$  only, has been made. A representative set of parameters for  $\theta$  was chosen and, using a set of 70 pseudo-random normal deviates with unit variance, which we denote by  $x_i$ , 70 values of  $f(\theta, i) + \epsilon_i$  were constructed in 3 ways:

- (a) control  $\epsilon_i = K_1 x_i$ ,
- (b) heteroscedastic  $\epsilon_i = K_2 f(\theta, i) x_i$ ,
- (c) correlated  $\epsilon_i = 0.333 \epsilon_{i-1} + 0.943 K_3 x_i$  for  $i > 1$ ,  $\epsilon_1 = K_3 x_1$ .

$K_1$  was taken as 6, and  $K_2$ ,  $K_3$  chosen so that the expected residual sum of squares to the fitted segmented model was the same for all 3 cases. To each of the 3 sets of simulated observations were fitted the segmented, exponential, and polynomial regressions, and values of  $\lambda$  for both types of smooth model were calculated. This was repeated 40 times using different sets of random deviates.

The distributions  $\Lambda_r$ , as represented by the 40 sets of observations, were very similar for all 3 types of error. The means and standard deviations (Table 2) suggest that the conclusions made in discriminating between regression models were valid even though the assumptions of independent and homoscedastic errors were not.

### 9. THE EFFECT OF IMPERFECT SYNCHRONY

The synchronous cultures used in these experiments were not perfectly synchronous. Observations on cell numbers indicated that the cell populations took about 1 hour to double in size at the first division. The question arises as to whether the discrimination should still have favoured the segmented model when the synchrony was as imperfect as this. If not, the

TABLE 2  
 MEAN AND STANDARD DEVIATION OF 40 OBSERVATIONS FROM  $\Lambda_f$  UNDER  
 DIFFERENT ASSUMPTIONS ON THE DISTRIBUTION OF ERRORS  
 ABOUT THE SEGMENTED MODEL

Errors	$g(t)$	Exponential	$g(t)$	Polynomial
	Mean	S. D.	Mean	S. D.
Homoscedastic, independent	0.629	0.076	0.693	0.084
Heteroscedastic, independent	0.623	0.090	0.692	0.093
Homoscedastic, correlated	0.613	0.095	0.692	0.105

conclusions may lead to the conjecture that the rate doubling times are more synchronous than the cell division times.

The effect of this asynchrony is to cause a rounding-off of the intersections of the segments. Visually, a segmented curve with corners rounded off to a degree corresponding to a variation in rate doubling times over 1 hour looks very similar to a strictly segmented curve, particularly when random variation is added. Simulations confirmed that the distribution  $\Lambda_f$  is altered very little by rounding off the corners of the true segmented model to this degree.

Although the synchrony is imperfect the method of discrimination would still therefore favour the segmented model.

DISCRIMINATION ENTRE MODELES DE REGRESSION POUR DETERMINER  
 LA DESCRIPTION DE LA SYNTHESE ENZYMATIQUE DANS DES  
 CULTURES SYNCHRONES DE CELLULES

RESUME

Des théories alternatives pour la synthèse enzymatique pendant le cycle cellulaire mènent à deux modèles de régression rendant compte de l'augmentation de la concentration de l'enzyme dans des cultures synchrones de cellules de levure. L'un des modèles est une régression linéaire en trois segments de droite; l'autre est une exponentielle bien régulière. L'analyse des observations à partir d'une série d'expériences montrées pour discriminer entre l'un et l'autre modèles est discutée. On décrit des méthodes pour estimer la régression linéaire segmentée et pour déterminer sous les deux hypothèses par simulation les distributions du critère de discrimination (rapport des vraisemblances maximisées). On a examiné les conséquences de la non validité de quelques unes des hypothèses faites pour bâtir les modèles de régression.

REFERENCES

- Cox, D. R. [1961]. Tests of separate families of hypotheses. *Proc. Fourth Berkeley Symp.* 1, 105-23.  
 Hudson, D. J. [1966]. Fitting segmented curves whose join points have to be estimated. *J. Amer. Statist. Ass.* 61, 1097-1129.  
 Mitchison, J. M. and Creanor, J. [1969]. Linear synthesis of sucrose and phosphatases during the cell cycle of *Schizosaccharomyces pombe*. *J. Cell Sci.* 5, 373-91.

Received June 1969