Estimating the Parameters of a Distributed Lag Model from Cross-Section Data: The Case of Hospital Admissions and Discharges

CHARLES R. NELSON and G. WILLIAM SCHWERT *

It is natural to think of estimating distributed lag models from time series data since it is in time series contexts that such models arise. This article studies the estimation of rational distributed lags in the context of hospital admissions and discharges where a relevant body of cross-section data exists in the form of lengths-of-stay and appears to be more appropriate for this purpose than do the available time series data. The cross-section approach presumably has applications in many situations in which length-of-stay in the system or its counterpart can be observed.

1. INTRODUCTION

It is natural to think of estimating the parameters of distributed lag or transfer function models from time series data since it is in time series contexts that such models arise. The present article deals with the estimation of a distributed lag model in a situation where a relevant body of cross-section data not only exists but appears to be more appropriate for this purpose than does the available time series data. In particular, we are concerned with investigating the input-output dynamics relating hospital admissions to subsequent discharges. The primary motivation for studying this relationship is the desire on the part of hospitals to schedule admissions so as to stabilize occupancy or "census" near some optimal level and at the same time provide patients and doctors with some lead-time before admission. A model of the input-output dynamics of a hospital would provide a framework in which to compute expected discharges and thus forecast available bed space.

Previous studies of hospital systems have made rather strong assumptions about the statistical properties of the admissions sequence and/or the dynamic properties of the system.[1] Our analysis is based on the assumption that a hospital may be represented by a linear filter which transforms the input sequence of admissions into the output sequence of discharges, that is,

$$D_t = \sum_{i=0}^{\infty} c_i A_{t-i} + \epsilon_t, \qquad (1.1)$$

where D_t is the number of patients discharged on day t, A_t the number admitted, the c_i are fixed constants, and ϵ_t is a disturbance (presumably autocorrelated) which prevents discharges from being predicted perfectly from past admissions. The disturbance process is assumed to be exogenous to the system in the sense that it is not under the control of the admissions scheduler either directly or through adjustment of admissions. The nature of the physical system in question imposes useful restrictions on the impulse response weights; namely, each weight must be positive, since c_i represents the contribution of admissions during day $t - i$ to current discharges, and the weights must sum to unity, since all patients are ultimately discharged.

There is, of course, a substantial literature dealing with the estimation of the c_i in models such as (1.1) from time series data when the input sequence is generated independently of (or is at least uncorrelated with) the disturbance. It is clear, however, that this requirement is *not* likely to be met if the input is being manipulated to control the output as would be the case in a hospital setting. If some attempt was being made to control occupancy during the sample period, then A_t would in general be correlated with past values of the disturbance and, if the disturbance is autocorrelated, with current and future values as well. In view of this it may not be surprising that attempts to estimate the c_i from 700 daily observations were unsuccessful in the sense of leading to negative estimates of some of the c_i, most importantly a large negative weight at lag one day. The results varied

* Charles R. Nelson is associate professor and G. William Schwert is a doctoral candidate, both with the Graduate School of Business, University of Chicago, Chicago, Ill. 60637. This research was supported by the National Science Foundation under Grant 34501, and by the Center for Health Administration Studies, University of Chicago. The authors are grateful to J. Phillip Cooper, Warren H. Hausman, G.S. Madalla, Harry V. Roberts, David Wallace and Arnold Zellner for helpful comments and to Richard Levich who assisted with data collection and computation. They also acknowledge the cooperation of MacNeal Memorial Hospital in making available the data used in this study.

[1] In a recent paper, Wiorkowski and McLeod [1971] assume that the number of patients admitted on successive days is serially independent, a condition which is unlikely to be met, particularly if admissions are being controlled to stabilize occupancy as the authors assume. The assumption of serially independent admissions is employed by Newell [1967] and Pike, Proctor and Wyllie [1963] in analysis of emergency admissions. These authors require, however, that admissions follow a Poisson distribution. Even emergency admissions are likely to display serial dependence at a lag of seven days, essentially a weekly seasonality. This has been confirmed empirically by the authors for daily emergency admissions at MacNeal Memorial Hospital. Bithell [1969] has suggested that the length of stay distribution for hospital patients may be well represented by a Pascal distribution but does not provide empirical evidence to support that contention. The intuitive appeal of the Pascal model is that it may be visualized as a chain of simple Markovian transitions through successive stages of patient care. Unfortunately from the point of view of simplicity, the empirical results presented in this article are not favorable to the Pascal model.

September 1974, Volume 69, Number 347
Applications Section

little over a range of methods of estimation which included "prewhitening" of the input series with accompanying transformation of the output, truncation of the lag structure, flexible rational lag structures, and experimentation with various autocorrelation schemes for the errors suggested by study of the least squares residuals. Estimates of the c_i also proved to be remarkably robust with respect to alternative specifications of the model.[2]

These results should serve as a word of caution for the interpretation of distributed lag models in many economic contexts where "independent" variables may in fact be instruments of control. An obvious example would be the many attempts to relate national income to variables, such as the money supply, which are simultaneously being manipulated to achieve goals of economic policy.

Fortunately, the impulse response weights of the hospital system lend themselves to estimation from cross-section data. This follows from the fact that the c_i may be interpreted as expected frequencies of length of stay. Out of a group of N patients admitted to the hospital on a given day, the fraction leaving on the same day would be on average c_0, after one day c_1, etc. Empirical frequencies of length of stay are therefore estimates of the c_i. Problems of specifying an appropriate lag structure of rational form and estimating the parameters from cross-section length-of-stay data are discussed in Section 2. The remainder of the article reports results obtained from a sample of about 1,300 patients. Both the time series and cross-section data used in this study were obtained from MacNeal Memorial Hospital in Berwyn, Illinois. MacNeal is a general hospital of about 420 beds which had been stable with respect to size and average level of occupancy for several years preceding the data collection.

2. FITTING DISTRIBUTED LAGS TO CROSS-SECTION DATA

Consider a random sample of N patients discharged from the hospital (or N units emerging from any system of interest) and denote the number staying i days by N_i. The individuals appearing in the sample will in general have been discharged at different points in time; collectively, they represent a cross section of length-of-stay experiences. This use of the term "cross section" should be distinguished from a common usage which refers to a set of observations taken at one point in time. Since c_i is the probability that an individual patient chosen at random has stayed i days, the probability of a given sample is

$$p(\text{sample} \mid \mathbf{c}) = \prod_{i=0}^{\infty} c_i^{N_i}. \tag{2.1}$$

The log likelihood, which will be convenient to work with, is given by

$$\ell(\mathbf{c} \mid \text{sample}) = \sum_{i=0}^{\infty} N_i \ln (c_i). \tag{2.2}$$

In practice we need only be concerned with values of the index up to the longest length of stay observed. Maximizing the log likelihood subject to the constraint $\sum_{i=0}^{\infty} c_i = 1$, it is easy to show that the resulting estimates c_i are given by

$$\hat{c}_i = N_i/N \quad i = 1, 2, \cdots, \text{(longest stay in sample)} \tag{2.3}$$

which are just the sample frequencies of length of stay.

Parsimonious representations of the lag structure will generate the impulse response weights from a small number of fundamental parameters. In this article we assume that the lag structure is of rational form so that we may write

$$(c_0 + c_1 L^1 + c_2 L^2 + \cdots) = \frac{(a_0 + a_1 L^1 + \cdots + a_r L^r)}{(1 + b_1 L^1 + \cdots + b_s L^s)}, \tag{2.4}$$

where L is the lag operator. For rational lag structures given by (2.4), the parameters to be estimated are $(a_0, \cdots, a_r)$ and $(b_1, \cdots, b_s)$. The constraint which requires the sum of the impulse response weights to be unity is easily imposed on the distribution by eliminating one free parameter, say a_0, since

$$c(1) = a(1)/b(1) = 1 \tag{2.5}$$

implies

$$a_0 = 1 + \sum_{i=1}^{s} b_i - \sum_{i=1}^{r} a_i. \tag{2.6}$$

Therefore, the likelihood $\ell(\mathbf{a}, \mathbf{b} \mid \text{sample})$ where $\mathbf{a}$ denotes $(a_1, \cdots, a_r)$ and $\mathbf{b}$ denotes $(b_1, \cdots, b_s)$ is readily evaluated for given $\mathbf{a}$ and $\mathbf{b}$ using (2.2) and the set of relations:

$$\begin{aligned} c_0 &= a_0 \\ c_1 &= -b_1 c_0 + a_1 \\ &\vdots \\ c_i &= -b_1 c_{i-1} - \cdots - b_s c_{i-s} + a_i, \quad i = 1, \cdots, r \\ &\vdots \\ c_i &= -b_1 c_{i-1} - \cdots - b_s c_{i-s}, \quad i > r \end{aligned} \tag{2.7}$$

where c_i is understood to be zero for $i < 0$. The inequality constraints $0 \leq c_i \leq 1$ are not readily translated into restrictions on the rational form parameters. It is clear, however, that if those constraints are violated in either direction, evaluation of the log likelihood will fail since the log of at least one negative number would be required.

Direct solution for the maximum likelihood (ML) estimates is readily seen to be infeasible due to the nonlinearity of the relationships linking the rational form parameters to the impulse response weights. Fortunately, several numerical approaches to the maximization problem are available. The approach adopted in this article is based on the Newton-Raphson (N-R) method with a modification suggested by Vandaele and R. Chowdhury [6].

Statistical inference in this context can be based on likelihood ratio tests, making use of the fact that in "large"

[2] Details of the time series results will be supplied on request by the authors.

samples, $2[\ell(\beta \mid \text{sample}) - \ell(\beta_1 \mid \beta_2 = \beta_2^0, \text{sample})]$, where β_1 and β_2 are subvectors of β and β_2^0 is fixed, will have a χ^2 distribution with degrees of freedom equal to the number of constrained parameters. Generally, the elements of β will be the elements of **a** and **b** and the primary hypothesis of interest will be $\beta_2^0 = \mathbf{0}$. However, in testing the constraints implicit in Pascal distributions, it is convenient to reparameterize the model in terms of the zeros of $b(L)$. In addition, large sample standard errors for the parameter estimates can be computed using the fact that the asymptotic variance of $(\hat{\beta})$ is given by

$$V(\hat{\beta}) = -\{E[\mathbf{G}]\}^{-1}, \tag{2.8}$$

where **G** is the matrix of second derivatives of the log likelihood. An estimate of $V(\hat{\beta})$ is therefore given by

$$\hat{V}(\hat{\beta}) = -\mathbf{G}^{-1}, \tag{2.9}$$

where **G** is evaluated at the final parameter estimates.

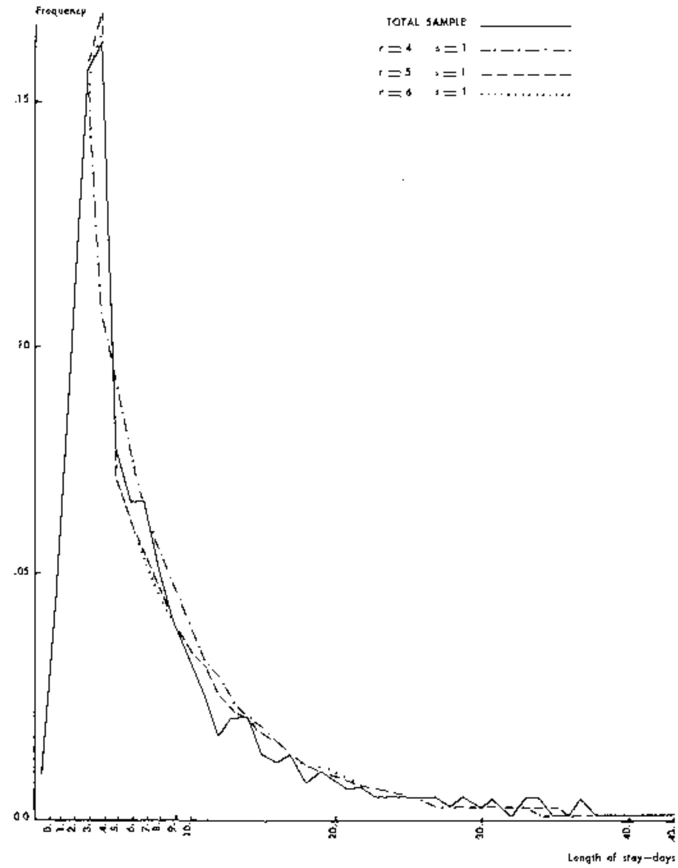
Our experience with standard errors computed from (2.9) was that they were almost always small relative to parameter estimates even when the latter were well within significance bounds as measured by a likelihood ratio test. Study of the actual shape of the log likelihood in the region of its maximum indicated that for our data the departure from a quadratic was substantial, suggesting that the large sample standard errors which are based on a quadratic approximation to the log likelihood may not be very accurate. Certainly from the point of view of operationally meaningful distinctions in the impulse response weights implied by alternative models, the likelihood ratio test proved to be the more useful criterion.

Several practical problems encountered in computation of estimates may be of general relevance. Most important, perhaps, is that the outcome of the N-R iterations was somewhat sensitive to initial guess values for the parameters. Poor guess values often led to infeasible points in the parameter space, i.e., points implying at least one negative impulse response weight. The differential used in numerical computation of derivatives was typically 10^{-3}; however, progress in iteration was sometimes facilitated by increasing the differential to 5×10^{-3} or decreasing it to 10^{-4}.[3]

3. ANALYSIS OF THE LENGTH-OF-STAY DATA

A sample of 1,290 patients was drawn from the discharge records of MacNeal Memorial Hospital for the year 1971.[4] The histogram of lengths of stay through 43 days is displayed in Figure A. It is interesting to note that approximately one percent of those patients stayed "zero" days, i.e., they were discharged on the day of admission. The most common circumstances of zero day stays are expiration and therapeutic treatment which requires only intraday occupancy. The longest stay in the sample is 63 days, but only about .6 percent of the patients in the sample stayed more than 37 days.[5]

A. Histogram of Lengths of Stay—Total Sample



The most important characteristic of the frequency distribution from the viewpoint of modeling is the conspicuous hump at 2, 3 and 4 days and the sharp drop at 5 days after which the frequencies decline at a fairly steady rate. A rational lag structure which is consistent with that shape is

$$c(L) = (a_0 + a_1 L + \cdots + a_5 L^5)/(1 + b_1 L). \tag{3.1}$$

Parameters $a_1, \cdots, a_5$ allow the hump to take form since

$$\begin{aligned} c_0 &= a_0 \text{ (computed using (2.6))} \\ c_1 &= -b_1 c_0 + a_1 \\ &\cdot \quad \cdot \quad \cdot \\ &\cdot \quad \cdot \quad \cdot \\ &\cdot \quad \cdot \quad \cdot \\ c_5 &= -b_1 c_4 + a_5 \end{aligned} \tag{3.2}$$

after which impulse response weights will decline ex-

[3] A listing of the program written for the Hewlett-Packard 2000C is available from the authors.

[4] The data were collected in two subsamples, one taking each twentieth patient in the discharge listing for 1971 and the other taking each fortieth patient. To check on the randomness of the samples, autocorrelations were computed for both ordered sequences of observations. The computed autocorrelations were small relative to the rough standard error $N^{-\frac{1}{2}}$. The first-order autocorrelation for the larger sample was −.06 (with standard error .034), and for the smaller sample, .03 (with standard error .048).

[5] The raw data included one observation at 91 days. Preliminary fitting of the models discussed in this section indicated that the probability of observing a stay that long in a sample of the size actually drawn would be very small, about 1.8 percent at the most. We were unable to verify whether this observation may have been an error in hospital records or an error in transcription of the data. To insure that estimates of the impulse response weights at low lags, which will be of primary importance for forecasting, would not be distorted by inclusion of this observation, it was dropped from the sample.

ponentially according to

$$c_j = -b_1 c_{j-1} \quad j \geq 6. \tag{3.3}$$

We would expect b_1 to be approximately $-.8$, a_1 through a_4 to be positive, and a_5 to be negative to accommodate the sharp drop in frequency at five days.

The estimates obtained for model (3.2)–(3.3) appear in Table 1A. The signs of the coefficients are as expected from study of the histogram. Impulse response weights implied by these estimates are plotted in Figure A where the importance of the numerator parameters in accommodating the hump of the distribution is readily apparent. As a check on the model, versions with $r = 4$ and $r = 6$ were also fitted, and the results are presented in Table 1A and Figure A. Deletion of one numerator parameter reduces the value of the log likelihood by about 27, an amount which is highly significant in the context of a χ^2 test. The addition of a_6, however, increased the log likelihood by only about .26, an amount which is not significant. Further, the estimate of a_6 is very small and estimates of the other parameters change only slightly.

1. Maximum Likelihood Estimates of Rational Lag Parameters

Parameter	*r = 5, s = 1*	*r = 4, s = 1*	*r = 6, s = 1*	*r = 5, s = 2*
	A. From Cross-Section Data, N = 1290			
a_0	.0096[a]	.0100[a]	.0094[a]	.0100[a]
a_1	.0360	.0348	.0337	.0341
a_2	.0597	.0610	.0677	.0557
a_3	.0718	.0740	.0633	.0631
a_4	.0298	− .0304	.0285	.0097
a_5	− .0718	—	− .0662	− .0586
a_6	—	—	− .0029	—
b_1	− .8649	− .8507	− .8666	−1.0190
b_2	—	—	—	.1330
$\ell(\ )$	−3730.99	−3758.02	−3730.73	−3730.74
	B. From Adult Cross-Section Sample, N = 1,124			
a_0	.0105[a]	.0106[a]	.0111[a]	.0110[a]
a_1	.0399	.0412	.0409	.0338
a_2	.0684	.0656	.0657	.0299
a_3	.0306	.0347	.0321	− .0215
a_4	.0117	− .0150	.0106	− .0248
a_5	− .0305	—	− .0296	− .0064
a_6	—	—	.0002	—
b_1	− .8690	− .8628	− .8689	−1.6791
b_2	—	—	—	.7011
$\ell(\ )$	−3372.87	−3377.9	−3372.77	−3368.97

[a] Implied by constraint a(1)/b(1) = 1.

It is interesting to note, however, that the standard error for a_6 implied by $-\{G^{-1}\}$ is only .0002, or about one-tenth the size of the coefficient estimate. If we were to rely on these asymptotic standard errors, we would be obliged to regard as "significant" a parameter which makes no meaningful contribution to the fit of the model as measured by the log likelihood. The source of the inconsistency is probably due, as mentioned earlier, to the fact that the log likelihood is poorly approximated by a quadratic in the neighborhood of the ML estimates.

Bithell [1] has proposed that Pascal distributions offer a promising class of models for hospital length-of-stay distributions. The intuitive appeal of the Pascal distribution derives from the familiar stages-of-servicing interpretation where, in this context, a patient can be thought of as passing through a succession of stages in which the probability of transition to the next stage after i days is given by $(1 - \alpha)\alpha^i$. The number of stages, say s, gives the order of the distribution which can be written

$$c(L) = (1 - \alpha)^s/(1 - \alpha L)^s. \tag{3.4}$$

For values of s greater than one, the Pascal readily generates hump-shaped distributions. It has the important practical advantage of being a one-parameter distribution for given s; therefore, estimation requires nothing more complex than a visual search of the likelihood over values of α. ML estimates of α for $s = 2$ and 3 appear in Table 2. It is clear from the values of $\ell(\ )$ that the larger-order Pascal is not to be preferred to the smaller-order, i.e., $s = 2$ appears to fit the data better. The humps generated by the Pascal distributions were not nearly high enough to account for the observed frequencies at three and four days, and the hump for the third-order model was placed too far to the right. This accounts for the better fit of the second-order model and the fact that both have log likelihoods much smaller than those of the rational models.

2. Maximum Likelihood Estimates of Pascal Lag Parameters from Cross Section Data, N = 1290

Parameter	*Pascal s = 2*	*Pascal s = 3*	*Unconstrained s = 2*
α	.7880	.7124	—
b_1	−1.5759[a]	−2.1372[a]	−1.4999
b_2	.6209[a]	1.5225[a]	.5526
b_3	—	− .3616[a]	—
$\ell(\ )$	−3865.19	−3907.97	−3847.75

[a] Implied by value of α.

In a literal sense, of course, the Pascal models are of rational form with the special constraints that $r = 0$ and $b(L)$ be of the form $(1 - \alpha L)^s$; in other words, the roots of $b(L) = 0$ must be identical. The second constraint can be relaxed in an s parameter model

$$c(L) = b(l)/b(L), \tag{3.5}$$

which we refer to as an unconstrained Pascal model. ML estimates of b_1 and b_2 for the second-order case also appear in Table 2. Plotting of the impulse response weights revealed that the unconstrained model differed relatively little in lag shape from either of the strict Pascals, still failing to follow the sharp hump at three and four days. Relaxation of the constraint in the second-order model can be interpreted as the addition of one parameter, which on the basis of a χ^2 test is clearly significantly different from zero. Thus we reject the constraint implicit in the strict Pascal model. It is apparent that if we wanted to think of patients passing through successive stages of servicing, then the Markovian transition probabilities are different for different stages. In view, however, of the much higher value of $\ell(\quad)$ for the more general rational models of Table 1A, the stages of servicing approach would not seem to be a satisfactory one for the data at hand.[6]

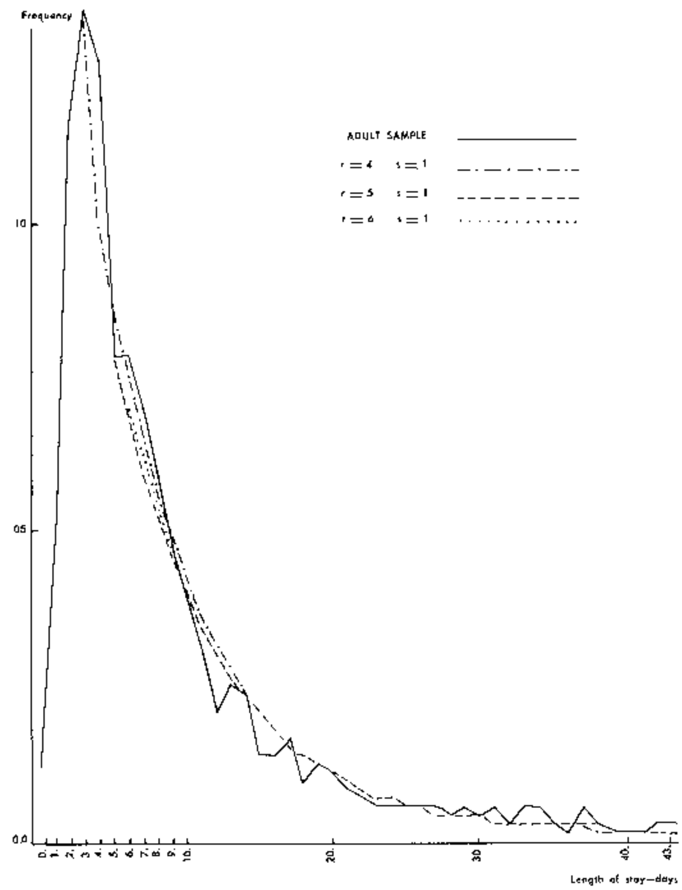
Since the unconstrained Pascal model with $r = 0$ and $s = 2$ and the preferred rational model with $r = 5$ and $s = 1$ are not nested, a direct comparison on the basis of likelihood ratio is not feasible. Both models, however, are nested in a rational model with $r = 5$ and $s = 2$, results for which are reported in the last column of Table 1A. The marginal effect of the five numerator parameters relative to the unconstrained Pascal is an increase in the log likelihood by 17.01, an amount which is highly significant. The marginal effect of the second denominator parameter relative to the rational model with $r = 5$ and $s = 1$, however, is seen to be negligible.[7] The likelihood criterion therefore indicates little advantage to the addition of b_2 to the rational model but considerable advantage from addition of $a_1, \cdots, a_5$ to the unconstrained Pascal model. Clearly, then, numerator parameters are essential for economical fitting of the hump of the frequency distribution.[8]

4. THE EFFECT OF OMITTING INFANTS FROM THE SAMPLE

The ultimate objective of a study of admissions and discharges is to provide a method of forecasting discharges so that future admissions may be scheduled to meet the objectives of hospital management, presumably maintenance of occupancy near some optimal level. The appropriate definition of occupancy might reasonably exclude newborn infants since these patients do not occupy bed space in the usual sense; rather, the number of nursery cribs is readily varied to meet current needs. Since the length-of-stay distribution may be influenced by the presence of nursery patients in the sample, they were deleted to form an "adult" sample of 1,124 patients. Although the "adult" sample includes pediatric cases, the relevant characteristic of this subsample is that it represents occupancy of beds for which there is a definite physical constraint.

The histogram for the adult sample is plotted in Figure B and differs from the total sample histogram primarily in the humped portion of the distribution. In particular, the peak of the distribution is at three rather than four days and the frequency at the peak is .1290 rather than .1612. Since it is in the hump of the total sample distribution that the Pascal models fit poorly, there is some reason to expect them to have more success with the adult sample.

B. Histogram of Lengths of Stay—Adult Sample



Rational models with $r = 4$, 5, and 6 and $s = 1$ yielded the ML estimates displayed in Table 1B. Comparing log likelihoods for the three models, it is evident that a_5 is highly significant while the addition of a_6 represents a negligible gain, as in the results for the total sample.[9]

Differences among the models are evident from the impulse response weights displayed in Figure B. The $r = 4$ model fails to account adequately for the length-of-stay frequency at Day 4, implying instead a considerably lower frequency than that observed in the adult sample. It may appear implausible at first that the model could exhibit that large a discrepancy at Day 4 since a_4 could be chosen to make the discrepancy as small as

[6] The stages of servicing would give results comparable to the rational models of Table 1B if we were willing to expand the order of the $b(L)$ polynomial sufficiently. This follows from the fact that if $a(L)$ and $b(L)$ are the polynomials corresponding to a rational model, then the denominator polynomial $\bar{b}(L)$ for a stage of servicing model is given by equating coefficients of powers of L in the approximation $\bar{b}(L)/\bar{b}(l) \cong b(L)/a(L)$ where the order of $\bar{b}(L)$ is determined by the precision of approximation desired. The rational model, however, clearly provides a much more *economical* representation of the lag in terms of the number of parameters required. For a discussion of the approximation of lag structures by ratios of lag polynomials, see [3, pp. 44–52].

[7] As a final check on the rational models, we fitted a model with $r = 6$, $s = 2$ to allow for the possibility that another numerator parameter might improve the fit significantly. The value of the log likelihood was −3728.96, an increase of 1.78 over $r = 5$, $s = 2$ which is significant at the ten percent level but not at the five percent level. Relative to the simpler $r = 5$, $s = 1$ model, the increase in the log likelihood is 2.03 which is not significant at the 10 percent level. There would seem to be little to be gained, then, from using the more complex model.

[8] See Footnote 6.

[9] In this case, the estimate of a_6 at .0002 is insignificant even by the criterion of its large sample standard error which is also .0002.

desired. If that were done, however, the impulse response weights at longer lags would have exhibited larger discrepancies since they are given by $c_j = -b_1c_{j-1}$ with c_4 as an initial condition. By choosing a seemingly large discrepancy at Day 4, the likelihood criterion has simply selected the best compromise.

While the $r = 5$, $s = 1$ model is the same one preferred for the total sample, differences between the parameter estimates reflect differences in the shape of the length-of-stay distributions. Estimates of a_1 and a_2 are larger for the adult sample while a_3, a_4, and a_5 are smaller. The estimate of b_1 is very little different, reflecting the fact that the tails of the two histograms are very similar.

Pascal models of orders 2 and 3 again yielded impulse response weights which failed to capture adequately the hump in the frequency distribution. The second-order model again had the larger likelihood with the hump in the third-order Pascal being too far to the right relative to that of the histogram. An unconstrained second-order Pascal model produced a hump which was sharper than that of the second-order strict Pascal with a significant increase in the log likelihood. The unconstrained Pascal nevertheless failed to conform adequately to the histogram.

Pursuing the same strategy of model comparison as in the analysis of the total sample, we fitted a model with five numerator parameters and two denominator parameters as shown in the last column of Table 1B. The increase in log likelihood over the second-order unconstrained Pascal is 60.22, which is highly significant as one would expect. The increase in log likelihood over the $r = 5$, $s = 1$ model is 3.90, which is also highly significant. Plotting of the impulse response weights for the $r = 5$ and $s = 1$ and 2 models revealed that the $s = 1$ model fits better in the hump of the distribution while the $s = 2$ model fits better in the area from six to 20 days. These differences are minor from an operational viewpoint, and the slight edge of the $s = 2$ model is probably offset in practice by the simplicity of the $s = 1$ model. Finally, since both the $s = 1$ and $s = 2$ models showed relatively large discrepancies at lags 5 and 6, we fitted an $r = 6$, $s = 2$ model as a final check. The increase in log likelihood was only .49 which is not significant.

5. IMPLICATIONS FOR SCHEDULING ADMISSIONS

The objective of scheduling admissions is to maintain the hospital census near some optimal level. The optimal number of admissions depends therefore on the expected path of future discharges. As of a given day t, the expected number of discharges on day $t + k$ is

$$E_t(D_{t+k}) = \sum_{i=0}^{k-1} c_i A^s_{t+k-i} + \sum_{i=k}^{\infty} c_i A_{t+k-i} + E_t[\epsilon_{t+k}], \quad (5.1)$$

where A^s_{t+k-i} denotes the number of admissions scheduled for day $(t + k - i)$. Parameters c_i may be estimated consistently from cross-section data as we have seen in preceding sections. Clearly, the cross-section data provide no direct information on the serial correlation structure of $\{\epsilon_t\}$ which will be essential for evaluating $E_t[\epsilon_{t+k}]$. This is readily inferred, however, by studying the implied error series $\{\hat{\epsilon}_t\}$ computed from

$$\hat{\epsilon}_t = D_t - (\hat{a}(L)/\hat{b}(L))A_t, \quad (5.2)$$

where $\hat{a}(L)$ and $\hat{b}(L)$ are lag polynomials in the cross-sectional parameter estimates.

Sample autocorrelations of computed errors for the 700 observation time series of admissions and discharges referred to in Section 1 using parameter estimates from the full cross-section sample for model $r = 5$, $s = 1$ suggested a weekly "seasonal" model of the form

$$\hat{\epsilon}_t = \hat{\epsilon}_{t-7} + (1 - \theta_1 L - \theta_2 L^2 - \theta_3 L^3)(1 - \Delta_1 L^7 - \Delta_2 L^{14})u_t, \quad (5.3)$$

where the u_t are i.i.d. disturbances.[10] Least squares estimates of the parameters are:

$$\theta_1 = .1906 \qquad \Delta_1 = .8131$$
$$\theta_2 = .1272 \qquad \Delta_2 = .1601 \quad (5.4)$$
$$\theta_3 = .1309$$

and all are highly significant.

The noise process given by (5.3)–(5.4) is quite different from that exhibited by the residuals from the time series estimates of distributed lag parameters referred to in Section 1. In particular, for a rational lag distribution with $r = 9$ and $s = 1$ we obtained a fitted model for the errors as follows:

$$\hat{\epsilon}_t = .382\hat{\epsilon}_{t-7} + (1 - .187L - .260L^2)u_t. \quad (5.5)$$

The difference should not be surprising in view of the evidence that the time series estimates of the distributed lag parameters themselves are seriously biased due to correlation between the admissions series and the disturbance process. Clearly, admissions scheduling based on the time series estimates would be quite different from that based on the cross-section estimates and possibly seriously suboptimal.

It is important to distinguish between prediction in the sense that it is being used in this discussion and prediction in the sense of least-squares fit. For purposes of control, the relevant prediction is the expected response of discharges to some deliberate change in admissions. The relevant parameters, the c_i, could be estimated consistently from time series data if the admissions series were truly exogenous to the system. This might conceivably be arranged by forcing admissions to follow some fixed pattern regardless of the desirability of the consequences. In the regular operation of the hospital, however, Equation (1.1) is only one in a multiequation system of which the control policy (whether optimal or not) is also an integral part. Thus, both discharges *and* admissions are *endogenous* to the system as a whole and

[10] The identification and estimation of seasonal time series models is developed by Box and Jenkins [2].

the estimation of Equation (1.1) can be thought of as a problem in estimating one equation in a simultaneous system.

As usual in such situations, the least-squares estimates provide by definition the minimum squared error predictions given observed values of the right-hand-side endogenous variable, but nevertheless provide inconsistent estimates of the structural parameters. Of course, it is those structural parameters which are relevant for design of a control scheme, and thus we are obliged to seek consistent estimators. As mentioned in Section 1, the problems posed by the endogeneity of "policy" variables are intrinsic to many economic systems, although in empirical work as well as in textbook discussions such variables are almost always categorized as "exogenous."

6. SUGGESTIONS FOR FURTHER RESEARCH

These results suggest possible extensions in several directions. In the hospital setting, the most obvious direction is probably toward disaggregation of the system into service units which may have materially different dynamic structure, e.g., surgery, maternity, etc. A more extensive data set would allow investigation of the temporal stability of the length-of-stay distribution including the possibility of seasonal variation. Although the admissions scheduler can only adjust the volume of admissions to regulate occupancy in the system, it may be that the response of the system is sensitive to the level of occupancy. Some experimentation with variable-lag models having parameters dependent on the level of occupancy could conceivably be profitable.

In a broader context, the cross-section approach to estimation may be useful in nonhospital settings where length-of-stay measurements can be made. Examples might include the lag distribution between inception and completion of capital investment projects, time to repayment for a class of bank loans, students entering and leaving a Ph.D. program, lapsing of life insurance policies, response from mail-order advertising, time from submission to publication of journal articles, etc. If the input to a particular system is subject to control, the cross-section technique may represent the only feasible approach to consistent estimation. In noncontrol settings, it may offer an alternative to time series estimation in which case the two approaches could serve a corroborative function.

[*Received August 1973. Revised April 1974.*]

REFERENCES

[1] Bithell, J.S., "A Class of Discrete-Time Models for the Study of Hospital Admission Systems," *Operations Research*, 17 No. 1 (January–February 1969), 48–69.

[2] Box, G.E.P. and Jenkins, G.M., *Time Series Analysis*, San Francisco: Holden-Day, 1970.

[3] Dhrymes, P.J., *Distributed Lags, Problems of Estimation and Formulation*, San Francisco: Holden-Day, Inc., 1971.

[4] Newell, D.J., "Immediate Admissions to Hospitals," *Proceedings of the 3rd International Conference on Operations Research* (Oslo, 1963), London: English Universities, 1964, 224–33.

[5] Pike, M.C., Proctor, D.M., and Wyllie, J.M., "Analysis of Admissions to a Casualty Ward," *British Journal of Preventive and Social Medicine*, 17, (1963), 172–76.

[6] Vandaele, W.H., and R. Chowdhury, S., "A Revised Method of Scoring," *Statistica Neerlandica*, 25 (1971), 101–12.

[7] Wiorkowski, J.J., and McLeod, W.R., "The Prediction and Control of the Size of an Input-Output System," *Journal of the American Statistical Association*, 66, No. 336 (December 1971), 712–19.