CHAPTER VIII

PREDICTION TESTS

PURPOSES OF AVAILABLE TESTS

Tests like those that have thus far been applied to various fitted equations have been described by Christ[1] as tests of internal consistency. The Durbin-Watson and the Rubin-Anderson tests check the extent to which the sample data are consistent with some of the assumptions on which estimates of parameters are based. When we check the plausibility of coefficients, we are checking implications of the data and a priori assumptions used in estimation against a priori information that was not employed in the estimation process.

The ultimate purpose of deriving economic relations is useful prediction, and success in prediction is the final test of any structure. Structures or equations that are internally consistent do not necessarily lead to useful predictions. On the other hand, it is conceivable that a prediction scheme, initially unrelated to any consistent theory, could meet with considerable success in prediction. In the latter case, we would naturally try to incorporate a rationalization for the success of the scheme in our theories of the phenomena predicted. This would be done partly to improve our theories, possibly enabling us to extend successful prediction to related areas, and partly to see whether the revised theory indicated circumstances in which the scheme might be expected to fail. If the latter were true, we would be cautious in applying the scheme in these circumstances, until a reasonable number of observations were available. These would either tend to confirm our revised theory or lead to new revisions.

Though a successful prediction formula would be valuable, regardless of its source, the fact that we cannot investigate every possible formula means that some discretion concerning sources must be exercised. In the simultaneous-equations approach, an effort is made to use existing theory, knowledge of institutional arrangements, and historical data as sources, and to arrive at promising methods of prediction through formal

[1] Carl Christ, *A Test of an Econometric Model for the United States, 1921-47,* Conference on Business Cycles, National Bureau of Economic Research, New York, 1951.

statistical procedures. Unfortunately limitations imposed by both limited sources of data and available statistical procedures typically force an investigator to include some dubious information in his formal procedure and to exclude some potentially useful information. Tests of internal consistency are efforts to get clues about the validity of the doubtful assumptions, on which formal procedures have been partly based. They thus offer some opportunity for selecting, and possibly improving, structures before or in connection with more direct tests of predictive usefulness.

The more direct tests of predictive usefulness presented here involve the use of data for the year 1950. These data became available while the computations already reported were in process. The tests applied follow much the same procedures as those used by Christ and Marshall.[2] They involve calculating residuals from fitted equations, using observations for 1950 and parameters estimated from observations for the period 1920–49. The residuals are used in a formal test of significance and are also compared with residuals from certain "naïve" relations.[3]

An Acceptance Region for a Calculated Residual

The test of significance is straightforward when applied to an equation estimated by single-equation least-squares methods. Certain complications, to be discussed below, arise if one makes the assumptions appropriate to simultaneous-equations methods. The test is directed toward substantially the same purpose as the Christ-Marshall tolerance interval test,[4] namely, to check whether the estimated relation fits data outside the sample period as well as should be expected from its fit during the sample period.[5] Let the equation in question be written

[2] Christ, *op. cit.*

Andrew W. Marshall, A Test of Klein's Model III for Changes of Structure, unpublished, M. A. Thesis, University of Chicago, 1949.

[3] Marshall, *op. cit.*, p. 21–25.

Christ, *op. cit.*, pp. 56–59.

[4] There are some conceptual difficulties with the way in which tolerance intervals were obtained by Christ and Marshall. The tables they used were constructed to obtain tolerance intervals from successive independent observations from a stable normal population. Successive calculated residuals are not independent.

[5] Of course, the fit during the outside period should not be expected to be so good as the fit during the sample period. In the single-equation case, and in the notation developed in the text,

$$E(\hat{u}_t^2) = \sigma^2(1 - x_t M_{xx}^{-1} x_t')$$

if t is part of the sample period, and

$$E(\hat{u}_t^2) = \sigma^2(1 + x_t M_{xx}^{-1} x_t')$$

$$y_t - \alpha x_t' = u_t \tag{8.1}$$

where y_t is the value of the dependent variable at time t, x_t' is a column vector of the values of independent variables at time t, α is a row vector of constant coefficients. u_t is the value of an unobserved random disturbance at time t; it is assumed to be normally distributed, nonautocorrelated, and independent of x_t. Ordinarily the equation contains a constant term, which we allow for by letting one of the components of x_t take the value unity at each time point. The constant term is then the coefficient of this particular independent variable.

The least squares (also maximum-likelihood under appropriate assumptions) estimates of the coefficients based on observations of y_t, x_t for $t = 1 \cdots T$ are given by

$$\tilde{\alpha} = M_{yx} M_{xx}^{-1} \tag{8.2}$$

where M_{yx} is a vector of sums of cross products of the dependent variable and each of the independent variables in turn. M_{xx} is a matrix whose elements are sums of squares and cross products of the independent variables. The residual for time t is given by

$$\tilde{u}_t = y_t - \tilde{\alpha} x_t' \tag{8.3}$$

Let the subscript r denote an interval of time outside the sample period. The variance of the residual for the rth period is indicated by[6]

$$E(\tilde{u}_r^2) = \sigma^2(1 + x_r M_{xx}^{-1} x_r') \tag{8.4}$$

where σ^2 is the variance of u_t and is unknown. However, let

$$s^2 = \sum_{t=1}^{T} \tilde{u}_t^2 \tag{8.5}$$

Then, if k is the number of elements in x_t,

$$t = \frac{\tilde{u}_r(T - k)}{\sqrt{s^2(1 + x_r M_{xx}^{-1} x_r')}} \tag{8.6}$$

has the t distribution, with $T - k$ degrees of freedom, and contains no unknown parameters. This relation can be used to construct confidence intervals for predictions or tests of significance for calculated residuals. The latter are more suitable for our present purposes.

The null hypothesis under which the quantity given in (8.6) has the t distribution consists of two parts:

---

if t is outside the sample period. $\tilde{u}_t$ is the calculated residual for time t, and σ^2 is the variance of the true disturbance.

[6] Mood *op cit.*, p. 304.

(*a*) The assumption of the appropriateness of the statistical specification for the sample period.

(*b*) The assumption of no structural change between the sample period and the test period.

An acceptance region of desired size for t can be found from standard tables and converted to an acceptance region for $\tilde{u}_r$ by (8.6). The occurrence of a value of $\tilde{u}_r$ outside the acceptance region (in the critical region) can be explained in one or a combination of three ways:

(*a*) As an indication that an inappropriate specification was used during the sample period.

(*b*) As an indication of structural change.

(*c*) As a statistical accident.

An investigator can never know for sure which explanation is true in a particular case. He can exercise some control over the probability that *c* is the sole factor, by his choice of the size of the acceptance region. He can usually exercise some judgment between *a* and *b*. Important structural changes are likely to have manifold effects and to leave various kinds of evidence behind. Indications of inappropriate specification may be supported by the results of tests of internal consistency or by the results of experiments with alternative specifications. To the extent that the investigator bases his initial specification on peculiarities of the data for the sample period, he increases the chance that the test will reject the null hypothesis for the reason of inappropriate specification.

TESTS BASED ON 1950 DATA

Intervals of 95% acceptance for the first five relations of the original statistical model of Chapter IV and for the linear versions of demand for feed grains and demand for livestock products are given in the fourth column of Table XXII. The linear version of demand for feed grains 6.1 was included because it was rejected by none of the tests of internal consistency of Chapter VI and should thus be regarded as a possibly useful equation. Demand for livestock products in linear form 7.21 was included because the logarithmic version 4.5 fits the 1950 data so badly. If the linear version had fit the 1950 and subsequent data well, there would have been an incentive to re-examine the grounds on which some of the implications of this equation were labeled implausible in the preceding chapter. However, it can be seen from the last row of the table that this is not the case.

No attempt was made to compute an acceptance interval under the simultaneous-equations assumptions as no satisfactory procedure has been developed that could be applied to our case. Rubin[7] has developed an approximation formula for the variance of a residual, but it involves all of the predetermined variables of a system and thus does not apply to an incomplete model. We conjecture that, if a reasonable procedure were found for our case, it would lead to larger intervals than our least-squares intervals.[8] If this is true and if we regard the occurrence of a limited-information residual outside the least-squares interval as grounds for questioning the limited-information equation, we shall be applying a more stringent test to the limited-information equations than to the

[7] Herman Rubin, The Approximate Distribution of Calculated Disturbances, *Cowles Commission Discussion Paper, Statistics 318*, Oct. 1947 (hectographed). Rubin's formula is given in Christ, *op. cit.*, p. 55.

[8] Let $\tilde{u}_r$ be the calculated residual for time r (outside the sample period) under least-squares procedures and $\hat{u}_r$ the calculated residual from a limited-information equation.

$$\text{(i)} \qquad \hat{u}_r = y_r + \hat{\alpha}x'_r = (\hat{\alpha} - \alpha^*)x'_r + u^*_r$$

where y_r is the variable whose coefficient has been normalized in limited information. It is assumed that this variable was treated as dependent in least squares. $\hat{\alpha} = (\hat{\beta}\,\hat{\gamma})$ is the vector of estimated limited-information coefficients. α^* is the vector of true coefficients, u_r is the true disturbance. $x_r = (w_r\, z_r)$ is the vector of other variables in the equation; w_r is a vector of other current endogenous variables. Let $w'_r = \pi s'_r + v'_r$ be the reduced form equations for w_r, with s_r being all the predetermined variables and v_r the reduced-form disturbances. We compare $E(\tilde{u}^2_r)$ and $E(\hat{u}^2_r)$. The latter depends on two sets of random variables, the $\hat{\alpha}$ and the v_r, which are independent. We can write

$$\text{(ii)} \qquad E(\hat{u}^2_r) = E_v[E(\hat{u}^2_r \mid v_r]$$

where the expectation sign outside the square brackets refers to expected value over the distribution of v_r.

$$\text{(iii)} \qquad E(\hat{u}^2_r \mid v_r) = x_r\psi x'_r + 2\hat{a}x'_r u_r + u^{*2}_r$$

where $\psi = E(\hat{\alpha} - \alpha^*)'(\hat{\alpha} - \alpha^*)$, and $\hat{a} = E(\hat{\alpha}) - \alpha$ is a vector of biases in the limited information estimates.

$$\text{(iv)} \qquad E(\hat{u}^2_r) = E_v(x_r\psi x'_r) + 2\hat{b}E_v(v'_r u^*_r) + \sigma^{*2}$$

where σ^{*2} is the variance of u^* and $\hat{b} = E(\hat{\beta}) - \beta$. We compare this with

$$\text{(v)} \qquad E(\tilde{u}^2_r) = x_r\theta x'_r + \sigma^2 \quad \text{where} \quad \theta = E(\hat{\alpha} - \alpha)'\,(\hat{\alpha} - \alpha)$$

Estimates of diagonal elements of ψ and θ were obtained when coefficients were estimated, and the estimates of diagonal elements of ψ were uniformly larger than estimates of corresponding elements of θ. Estimates of σ^{*2} obtained in the calculations were uniformly larger than estimates of σ^2. The sign of the middle term of (iv) is not known, but its absolute value becomes arbitrarily small as the sample size increases.

TABLE XXII
TESTS BASED ON OBSERVATIONS FOR 1950

Equation No.	Variable of Normalization	1950 Value of Variable of Normalization	95% Acceptance Interval for Least-Squares Residual	Least-Squares Residual	Limited-Information Residual	Residual I	Residual II
4.1	Y_1	9.978	±0.023	−0.007	−0.005	0.007	−0.018
4.2	Y_2	1.500	±0.108	0.001	−0.010	0.050	0.225
6.1	y_2	31.63	±4.53	−2.04	−1.82	3.41	17.42
4.3	Y_3	1.766	±0.056	0.017	0.007	−0.011	0.057
4.4	Y_4	9.970	±0.035	0.028	0.021	0.006	−0.021
4.5	Y_5	0.239	±0.057	−0.078	−0.078	0.013	0.087
7.21	y_5	1.736	±0.147	−0.270	−0.271	0.051	0.365

IDENTIFICATION OF EQUATIONS AND VARIABLES OF NORMALIZATION. 4.1—production relation, Y_1 : log of index of livestock production. 4.2—demand for feed grains, Y_2 : log of price of feed grains. 6.1—linear version of demand for feed grains, y_2 : price of feed grains. 4.3—demand for protein feeds, Y_3 : log of price of protein feeds. 4.5—demand for livestock products, Y_5 : log of price of livestock products. 7.21—linear version of demand for livestock products, y_5 : price of livestock products.

least squares. In the absence of an appropriate limited-information test, we elected to follow this procedure. This is a crude makeshift which we hope can be improved on in later studies through investigations of the small-sample properties of limited-information procedures.

The variables of normalization in the second column of Table XXII are the variables whose coefficients were set equal to unity in the limited-information estimation of the indicated equations and were treated as dependent variables in the least-squares procedures.

Residual I in the next to last column of the table was obtained by subtracting the 1949 value of the variable of normalization from the 1950 value. It is the residual obtained by applying naïve model I,[9] which arbitrarily postulates that this year's value of any variable is equal to last year's value plus a random element. Residual II was obtained by adding to the 1949 value the algebraic difference obtained by subtracting the 1948 value from the 1949 value. The result was then subtracted from the observed 1950 value. This corresponds to the application of naïve model II, which arbitrarily assumes that the change in any variable from last year to this will be, except for a random disturbance, the same as the

[9] Christ, *op. cit.*, pp. 56–59, 69.
Marshall, *op. cit.*, pp. 21–22.

change in that variable from year before last to last year. When applied to logs of observed variables, this model projects the most recently observed proportionate change; when applied directly to observe variables, it projects the latest absolute change.

The use of naïve model tests is largely intuitive. As yet there is no formal theory of just how the tests are to be interpreted or when they may significantly reject an equation.[10] Nevertheless, their intuitive appeal is strong, and they may easily detect difficulties against which other tests are relatively weak. Of course, we obtain only a few fragments of evidence from a single year's comparisons in any case. With more data it may often happen in practice that the implications of the tests will be rather apparent and convincing, even in the absence of a formal theory.

The most striking features of Table XXII are the very large residuals from the fitted equations for demand for livestock products. They are uniformly large whether one looks at the linear or logarithmic version,

[10] For special structures, it could happen that accurate structural estimates could be obtained; yet the calculated residual from a given structural equation could be consistently larger (in absolute value) than the residuals from one or the other of the naïve models. As an example consider the following.

Let

$$\beta y_t' + \Gamma z_t' = u_t \qquad \text{(i)}$$

be a structural system, and let

$$y_t' = -\beta^{-1}\Gamma z_t' + \beta^{-1}u_t = \pi z_t' + v_t' \qquad \text{(ii)}$$

be the reduced form. Let

$$y_{kt} + bw_{kt}' + gz_t' = u_{kt} \qquad \text{(iii)}$$

represent the kth structural equation normalized on y_{kt}. w_{kt} is a vector of other current endogenous variables; u_{kt} is the random disturbance. b and g are vectors of coefficients. The kth equation of the reduced form might be written

$$y_{kt} = \pi^{(k)}z_t' + v_{kt} \qquad \text{(iv)}$$

Now suppose that, for each predetermined variable z_{it}, one of the two following statements holds: (a) The year to year variation in z_{it} is small, or (b) the coefficient $\pi_i^{(k)}$ of z_{it} in (iv) is small. If, in addition, the variance of v_{kt} is small, then year-to-year variation in y_{kt} will be small, and residuals from applying naïve model I will be small. However, the variance of u_{kt} might be large, relative to that of v_{kt}, and the calculated residuals from the kth structural equation might thus be larger than the calculated residuals from naïve model I. Nothing in the above assumptions prevents the obtaining of accurate estimates of b and g. The illustration is admittedly highly special, but it does suggest the need for further examination of the interpretations of naïve model tests of structural relations. This possibility would not arise in naïve model tests of reduced-form relations. Such tests were not undertaken in this study because of the incompleteness of the original model.

the least-squares or limited-information residuals, or whether one compares them with the acceptance interval or with residual I. Possible interpretations of these residuals are offered below. Other aspects of the tabulated results perhaps deserve passing mention. From 1948 to 1949 the changes in all of the listed variables were quite large. From 1949 to 1950 they were either modest in size or opposite in direction to the 1948–49 changes. Thus residuals II are uniformly large. The slight tendency for limited-information residuals to be less in absolute value than the least-squares residuals is interesting, but attempts to interpret this might well be deferred until evidence for other years is available.

Residuals for equation 4.4, supply of livestock products, are rather large. It will be recalled from Chapter VI that anticipations of producers enter critically into this relation and that our search for indicators of anticipations had been unrewarding. Also, the Rubin-Anderson and Durbin-Watson tests tended to reject the limited-information version of this relation, and the latter test was inconclusive when applied to the least-squares calculations. Thus the various scraps of evidence relating to this equation tend to suggest that it may not be very useful in its present form. The authors believe strongly that better handling of anticipations is necessary here. We think that the prospects for this will be much better when particular types of livestock are studied individually. There will then be better opportunities for relating such things as surveys of intentions and outlook information for particular crops to producers' plans and behavior.

Postwar Developments in Demand for Livestock

To return to the livestock demand relation, the price calculated from the limited-information estimate of equation 4.5 is compared with the observed price for each year 1920–42 and 1947–50 in Figure 2. The comparison would be negligibly altered if the least-squares parameters had been used instead. The badness of fit of the relation in 1950 is also apparent in 1947 and 1949.

It was suggested earlier that a tested residual could fall in the critical region because of a statistical accident, inappropriate specification, or structural change. The probability that the 1950 result was due to statistical accident was 0.05 in the formal test, and the similar discrepancy for 1949 makes this explanation even less acceptable. The generally good fit in the prewar period suggests structural change as a strong possibility. In this instance, there are a number of reasons for believing that a structural change has taken place in the demand for some livestock products. Though there is no reason to believe that other factors may not also be at work, it is possible to account rather well for the postwar

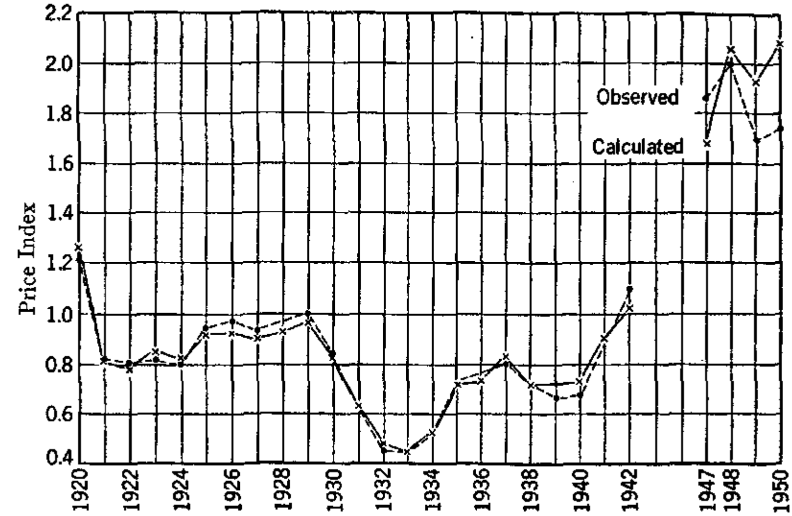

FIG. 2. Observed and calculated price of livestock products.

behavior of our calculated and observed livestock prices with the hypotheses that a significant change in the substitutability of other materials for animal fats and oils took place during World War II.

Oleomargarine was widely substituted for butter during the war. Many consumers became acquainted with the product for the first time, and, in the postwar period, restrictions on the sale of oleomargarine were gradually removed. Detergents were developed which compete with soap in many uses, and various vegetable oils came into wider use for cooking. If the above factors account for a significant part of the apparent decline in demand for livestock products, then prices of hogs and milk should be most affected since animal fats account for a larger fraction of the total value of these products than of the others. Farm prices and per-capita consumption of individual and aggregate livestock products are shown as ratios of their levels in four postwar years to 1920–49 averages in Table XXIII. In 1949 and 1950 prices of both hogs and milk are below the calculated price for all livestock products while per capita consumption of each of these products is relatively low.

Prices of poultry products are also low, but this can be fairly adequately explained by increased production.[11] The data on price and consumption of individual products for 1949 and 1950 are thus roughly

[11] The supply curve for poultry seems to have moved rather steadily to the right during much of our period. New breeds were developed, eggs per hen increased substantially; the broiler industry developed along with new techniques of specialized poultry production, pushing poultry off many general farms.

TABLE XXIII

POSTWAR PRICE AND CONSUMPTION OF LIVESTOCK PRODUCTS

Figures in Table Are Ratios to 1920–49 Averages

Commodity	1947		1948		1949		1950	
	Price	Per-capita consumption	Price	Per-capita consumption	Price	Per-capita consumption	Price	Per-capita consumption
Aggregate livestock								
(*a*) Observed	1.84	1.14	2.00	1.08	1.69	1.10	1.74	1.12
(*b*) Calculated	1.67	....	2.06	....	1.92	....	2.08	....
Milk	1.67	1.02	1.88	1.00	1.54	0.99	1.50	0.94
Cattle	2.09	1.24	2.51	1.08	2.24	1.07	2.60	1.05
Calves	1.92	1.35	2.30	1.19	2.14	1.07	2.49	0.96
Hogs	2.41	1.02	2.31	0.97	1.81	1.03	1.73	1.03
Sheep & lambs	1.80	0.93	2.01	0.85	2.01	0.69	2.31	0.59
Eggs	1.62	1.19	1.68	1.17	1.61	1.17	1.30	1.23
Chicken & broilers	1.40	1.16	1.61	1.09	1.32	1.25	1.26	1.33
Turkeys	1.43	1.45	1.83	1.32	1.38	1.75	1.29	1.82

TABLE XXIV

POSTWAR DEVELOPMENTS IN FATS AND OILS

Quantity Figures Are in Millions of Pounds; Price Indices, 1935–39 = 100

Year	Domestic Disappearance of Oleomargarine	Use of Fats and Oils in Soap	Total Vegetable Fats and Oils				Total Animal Fats and Oils		
			Domestic production	Domestic disappearance	Index of wholesale prices		Domestic production	Domestic disappearance	Index of wholesale prices
					Domestic	Foreign			
1947	719	2347	3520	4658	318	295	6431	5856	244
1948	892	2122	4163	4901	303	278	6093	5580	248
1949	854	1833	4820	4825	177	224	6695	5512	183
1950	930	1873	4844	5358	200	200	6935	5960	188

Data from *Agricultural Statistics, 1951.*

consistent with the structural change hypothesis. The fact that observed price of the aggregate was well above calculated price in 1947 is as yet unexplained.

In 1947 liquid assets held by individuals were abnormally high, consumers' durable goods were scarce and informally rationed, and rationing and price control of livestock products had just been abandoned in the fall of 1946. These are all factors that were not taken into account in our equation. Therefore they did not influence calculated price but undoubtedly did influence observed price. In addition, many of the fats and oils that compete with animal fats were in short supply in 1947 so that the increased postwar substitutability may have been largely ineffective. Regulations affecting sale of oleomargarine were gradually liberalized during the observed postwar period, and detergents became better known. Evidence of these developments is shown in Table XXIV.

There thus exists substantial evidence of a structural change in the postwar demand for various animal fats. The fact that this and other changes (such as technical developments in poultry) affect the various products differently furnishes an additional incentive to construct fairly complete models for individual products in future studies.