Volume Title: A Study of Aggregate Consumption Functions

Volume Author/Editor: Robert Ferber

Volume Publisher: UMI

Volume ISBN: 0-87014-453-7

Volume URL: http://www.nber.org/books/ferb53-1

Publication Date: 1953

Chapter Title: Empirical Evaluation of Past Studies

Chapter Author: Robert Ferber

Chapter URL: http://www.nber.org/chapters/c0823

Chapter pages in book: (p. 28 - 63)

Chapter II

EMPIRICAL EVALUATION OF PAST STUDIES

Method of approach

To evaluate the adequacy of the theory underlying a consumption function, it is not sufficient to show that the function fits the data from which it has been derived. It is always possible to find a function that will fit any given set of data perfectly. Hence the fact that a particular theory provides an excellent fit, large values of R^2, even after adjustment for degrees of freedom, and short confidence intervals for the period under observation may give little information as to the validity of the theory. If Theory A provides a better fit than does Theory B for the same period, this does not necessarily mean that A is the better theory (unless one specifically defines the adequacy of the theory as depending on the goodness of fit).

An independent and highly relevant measure of a consumption function theory's validity is its ability to explain fluctuations in consumption in times other than the period for which the theory was formulated. By this criterion, which is the one to be applied here, adequacy is considered to depend inversely on the size of the errors of prediction.[1]

If, during the period for which data are available, some event(s) occurred leading one to suspect on a priori grounds some change in basic conditions, a rational basis exists for evaluating the adequacy of a consumption function fitted to the earlier years by computing the accuracy of its predictions for the years following the event. By this we do not imply that to divide a period of observation arbitrarily into two subperiods, estimate the parameters of the consumption functions from the data for one of the subperiods, and use the resultant functions to predict the values of the dependent variable in the other subperiod is generally superior to the goodness of fit test applied to the entire period. For if there is no reason to suspect a change in basic conditions between the two subperiods, clearly the best estimates of the parameters of the functions are obtained by using all the data available.

In the present study, a basis for a priori, temporal stratification exists between the period 1923-40 and the postwar years 1947-50. The drastic

[1] A somewhat more inclusive definition is the one suggested by Abraham Wald, Jacob Marschak, and others, to the effect that the most adequate theory is the one yielding the highest (net) utility, taking into account the computational and other costs involved in obtaining the final results as well as the (gross) utility achieved by applying the theory to the problems at hand.

changes that have taken place between these two periods provide ample grounds for making such a stratification. In fact, in the light of such drastic changes, it would be highly questionable to estimate the parameters of consumption functions from these two periods combined without conducting beforehand some test similar to that proposed above. In the present case, the test is carried out by considering 1923-40 or some portion thereof as the period of observation and the postwar years as the period to be predicted.

One might object to employing this criterion on the ground that the consumption functions being tested may not have been intended for extrapolation purposes. But a consumption function valid only for the period of observation possesses little value for policy formation or for prediction. Clearly, a function that yields close fits for a number of different periods is of greater general utility than one valid only for a limited period. And in view of the great importance attributed to consumption functions in problems of policy and prediction, it behooves us to determine the extent to which different types of consumption functions may be applicable under more general conditions.

The criterion developed above has been applied to thirteen different forms of aggregate consumption functions. These thirteen forms correspond to fifteen of the functions in Table 1. Six function forms in Table 1 have been omitted for various reasons, generally either because the necessary data were not available at the time the study was undertaken, as in the case of Staehle and Polak, or because past computations have shown that the extra variables used, e.g., T^2, are not likely to be significant. In all instances, savings was used as the variable to be estimated.

All but one of the functions in Table 1 had been fitted on one or another of the old sets of Department of Commerce data. A choice therefore had to be made, either to revise the postwar Commerce data to fit each of the former concepts, or to recompute all of the functions using the revised Commerce figures for 1929 onward and making estimates for the years before 1929. Although the former procedure entails less computation, the second alternative was selected. There are several reasons for the choice.

One reason is the difficulty of revising the new data to the former concepts. A reconciliation of the new and old concepts has been published for 1929-46, in the July 1947 National Income Supplement to *Survey of Current Business* (Department of Commerce), but not for later years. Even if such a reconciliation were available for the entire postwar period, it would be difficult to use because changes in statistical procedures accompanied the change in concepts. To adjust new national income data to the former concepts will not, therefore, reproduce the old estimates. For example, the old estimate for national income in 1946 on the former concept is $165 billion, whereas the statistically revised estimate of the same variable on the same concept is $171.9 billion. Since the difference varies from year to year, and since the aggregate consumption function studies utilized the statistically unrevised estimates, the only way to retain the numerical estimates of the parameters computed in these studies would be by arbitrary estimates of the size of the statistical discrepancy between the revised and unrevised figures on the former concept for each postwar year after 1946. Such a procedure can lead to sizable errors, whose magnitude would be unknown.

Another reason for recomputing the parameters of the consumption functions on the new concepts is that if observation of the relative accuracy of these functions is to be continued, this procedure becomes the simpler of the two. To retain the original numerical functions would mean adjusting the national income estimates for each succeeding year as they become available. A final reason is that in almost all cases the theory underlying the particular function proposed was not framed in terms of any particular definition of the consumption and income variables.[2] For example, none of the theories specified whether consumption expenditures were to be derived as a residual from personal disposable income or by direct estimation, or whether the net imputed rent of owner-occupied dwellings should be included in or excluded from consumption expenditures. The general procedure was to formulate a theory first and then proceed to use the most recent data available for the empirical work. Hence the theories would seem to be equally applicable to the national income data on the new concept.

Accordingly, the parameters of all the aggregate consumption function studies selected for testing were re-estimated from the new Commerce data. In each case the method of fit was the same as that used in the original computations, and the period covered was made as nearly the same as possible. The fact that the new Commerce data only go back to 1929 constituted a limiting factor in the latter regard. For the years before that date, comparable estimates made by Harold Barger were used.

Revised estimates of past consumption functions

Table 2 gives a summary of the results obtained from fitting the functions selected from Table 1 to the revised Commerce data. The same over-all classification is used as in Table 1,[3] the identification number of the function from that table being shown in column 1. The period of observation and method of fit are given in columns 2 and 3, and the actual function in column 4. Column 5 shows the value of R^2 adjusted for sample size and

[2] Stone's was the only study where the consumption and income variables were defined explicitly.
[3] The Modigliani function (1.21) is classified differently. In Table 1, it is classified in its original computational form, but for purposes of comparison with the Duesenberry function (1.23), it was converted into the form shown in Table 2.

Table 2

GOODNESS OF FIT OF AGGREGATE SAVINGS FUNCTIONS AND ACCURACY OF THEIR POSTWAR PREDICTIONS

FUNCTION NUMBER (1)	PERIOD COVERED (2)	METHOD OF FIT[a] (3)	FUNCTION[d] (4)	R^2 ADJUSTED FOR SAMPLE SIZE[e] (5)	AVERAGE ABSOLUTE PERCENTAGE ERROR 1947-49 (6)	1948-49 (7)
			A *Variables in current price aggregates*			
(1.1)	1929-40	l.s.	$S = \mathbf{-7.89} + \mathbf{.152}Y$	.82	195	115
(1.2)	1923-40	l.s.	$S = \mathbf{-6.59} + \mathbf{.133}Y - .014(T - \text{Jan. 1, 1932})^{f}$	.59*	158	87
(1.3)	1929-40	l.s.	$S = \mathbf{-7.02} + \mathbf{.20}Y - .06Y_{-1}$	.86	180	102
(1.6)	1929-40	l.s.	$S = \mathbf{-7.05} + \mathbf{.190}Y - .051Y_{-1} + .013(T - \text{Jan. 1, 1935})^{f}$	.84	186	106
			B *Variables in deflated price aggregates*			
(1.8)	1923-40	l.s.	$S/P = \mathbf{-6.61} + \mathbf{.142}(Y/P)$	.39**	123	61
(1.9)	1923-30, '35-40	l.s.	$S/P = .42 + .0355(Y/P)$	0	39	23
(1.10)	1923-30, '35-40	l.s.	$S/P = \mathbf{-4.12} + .107(Y/P) - .045(T - \text{Jan. 1, 1932})^{f}$	.02	55	12
(1.12)	1923-41	l.s.	$S/P = \mathbf{-10.26} + .195(Y/P) + \mathbf{.016}\frac{Y}{P}(T - 1932) - \mathbf{1.195}(T - 1932)$	.88	409	280
(1.13)	1923-40	l.s.	$S/P = \mathbf{12.50} + \mathbf{.228}(Y/P) - \mathbf{.197}N$	.77	115	51
			C *Variables in deflated price per capita units*			
(1.14)	1929-41	r.f.[b]	$S/NP = -117.6 + .270(Y/NP)$	.70	208	121
(1.15) } (1.16) }	1923-40	l.s.	$S/NP = \mathbf{-94.90} + \mathbf{.227}(Y/NP) - \mathbf{.420}(T - \text{Jan. 1, 1932})^{f}$	.72*	123	59
(1.20)	1929-41	r.f.[c]	$S/NP = -96.35 + .377(Y/NP) - .096(Y/NP)_{-1} - .068(M/NP)_{-1}$	.86	116	76
			D *Savings expressed as ratio to income*			
(1.21)	1923-40	l.s.	$S/Y = \mathbf{.045} + \mathbf{.157}[(Y/NP) - (Y/NP)_0]/(Y/NP)$	.79	19	13
(1.23)	1923-40	l.s.	$S/Y = \mathbf{-.154} + \mathbf{.198}(Y/NP)/(Y/NP)_0$	.74	15	14

For data used and sources, see Appendix Table B1.

NOTES TO TABLE 2

[a] l.s. = least squares; r.f. = reduced forms.

[b] Other equation in model is: $Y/NP = \mathcal{E}_0 + \mathcal{E}_1 X/NP$, X being gross private investment minus corporate savings plus government deficit, assumed exogenous.

[c] Other equations in model are: $GNP = Y - S + I + G$, and $Y + Z = GNP$, where GNP = gross national product, I = private gross investment, G = government expenditures for goods and services, Z = government receipts plus corporate savings plus business reserves minus transfer payments minus inventory profits.

[d] Italicized coefficient differs significantly from zero at the .05 significance level; bold-face coefficients at the .01 level. These significance statements assume that the error terms are normal, independent and homoscedastic. The regression coefficients of the reduced form equations were not tested for significance.

[e] The adjustment of R^2 for the reduced form equations makes an arbitrary (probably too large) allowance for degrees of freedom lost.

[f] One time unit = 6 months.

* Evidence of auto-correlation in the residuals (assuming normality) at the .05 significance level using the von Neumann ratio. This ratio is: $K = \delta^2/\sigma^2$, where $\delta^2 = \sum_{1}^{N\text{-}1} (X_t - X_{t+1})^2/N\text{-}1$, $\sigma^2 = \sum_{1}^{N}(X_t - \bar{X})^2/N$, $\bar{X} = \Sigma X_t/N$. See Hart and von Neumann, "Tabulation of the Probabilities for the Ratio of the Mean Square Successive Difference to the Variance," *Annals of Mathematical Statistics*, Vol. 13 (1942), pp. 207-14.

** Evidence of auto-correlation in the residuals at the .01 significance level using the von Neumann ratio.

column 6 the averages of the absolute relative margins of error in the 1947-49 savings predictions, i.e., the

$$\text{absolute value of} \frac{\text{Actual savings} - \text{Computed savings}}{\text{Actual savings}} \times 100\,.$$

Since the use of 1947 may be open to some objection because of the possible effect of demand backlogs on conditions in that year (particularly in the case of predictions based on functions including lagged variables), the average error, as defined above, for 1948-49 alone is shown in column 7. It should be noted that the postwar predictions in this and the following tables were obtained under the special assumption that income, prices, and population for those years were known. Throughout this paper the terms "forecast" and "prediction" will be used in this sense. The errors might have been considerably larger had the forecasts been based only on information available in the previous year.

A comparison of the values of R^2 of corresponding functions in Tables 1 and 2, where both values are available, reveals that none of the functions is able to explain the fluctuations in savings anywhere near as well as it does the fluctuations in consumption.[4] This is to be expected so long as

[4] This comparison does not allow for the differences between the sets of data used to estimate the parameters of corresponding consumption and savings functions in the original studies, but the effect of such differences on the relative values of R^2 of the consumption and savings functions is not likely to be great.

the marginal propensity to consume estimated by the functions exceeds .5, which it always does. The values of R^2 in Table 2 are essentially monotone functions of those in Table 1. Not only are these values of R^2 lower than before but the absolute spread between them is much larger. In this sense one might say that the discriminatory value of R^2 has been increased by the use of savings as the dependent variable.

Because of the relatively small number of functions involved, little can be said about the characteristics of functions providing good fits to the data as against those that provide poor fits. Nevertheless, a few points are evident. One is that functions attempting to distinguish between so-called secular and cyclical changes by omitting certain "atypical" years, such as 1931-34, do not fit the observed data as well as functions not making this distinction.

Apart from this, restriction of the period of observation to 1929-40 or 1929-41 tends to produce a higher goodness of fit than when functions are fitted to data beginning with 1923. The goodness of fit also seems to be improved when the number of variables included in the function is increased. All three of the four-variable functions in Table 2 yield high values of R^2, in one case even though the function was fitted to 1923-41 data.

Perhaps the most striking fact about Table 2 is the absence of any apparent relationship between the goodness of fit of a function and the accuracy of its postwar predictions. If anything, the relationship is a negative one.[5] The most accurate predictions are obtained from the functions that distinguish secular from cyclical changes, whereas the close-fitting four-variable functions provide some of the most inaccurate predictions. The exclusion of 1931-34 from the period of observation yields zero correlation for all practical purposes, though the accuracy of the postwar predictions of these functions is relatively high; the reason for this phenomenon is considered at a later point. Insofar as the functions in Table 2 are concerned, goodness of fit provides a very misleading indication of the adequacy of a function for prediction purposes. To what extent other criteria of predictive accuracy can be found is a question not easily answered. The results in the table suggest a number of possible criteria — distinction between secular and cyclical changes, the period of observation, adjustment for price changes, adjustment for population shifts — but the data are inadequate to permit generalization. The only means of arriving at the desired generalizations would seem to lie in systematizing the computations so as to obtain more detailed information on each of the points raised above (see below).

[5] The coefficients of correlation between the data in columns 5 and 6 and between the data in columns 5 and 7 are 0.44 and 0.46, respectively. Neither coefficient is significant at the .05 significance level.

Effect of change in data

The functions whose parameters are estimated in Table 2 correspond to certain ones in Table 1, with only minor exceptions as to form and period covered, but with one major exception — the basic data. Thus a comparison of the estimates of the corresponding parameters in Tables 1 and 2 should throw light on the little explored but important question of the stability of empirical estimates in relation to changes in data.[6] Although the fourteen functions in Table 2 are too few and too scattered in scope to permit a detailed investigation, they can nevertheless indicate the approximate magnitudes of error that are involved. This in itself is of major importance.

The comparison of the old and new estimates of the parameters of thirteen of the functions from Table 2 is shown in Table 3; the data used in the original estimates of one of the functions (1.3) differed only slightly from the data used in this study. Nine of the thirteen pairs of functions are perfectly comparable: only algebraic manipulation is needed to convert the consumption function in Table 1 into a savings function corresponding to that in Table 2. The other four pairs differ in varying degrees, as shown by the remarks in the last column of Table 3.

The functions in Table 3 appear in the same order as they do in Table 2. The first three columns of the table provide classifying information for each function. The estimates of the various parameters are shown in the succeeding sets of columns, one set for each parameter. Each set presents, in turn, the original estimate of the particular parameter, the revised or new estimate, and then the percentage deviation of the new estimate from the original estimate, i.e.:

$$\frac{\text{New estimate} - \text{Original estimate}}{\text{Absolute value of original estimate}} \times 100 .$$

Considerable variation in the relative discrepancy between the two estimates for each parameter is apparent in Table 3. About half of the differences are less than 25 per cent, and a third of these are less than 10 per cent. On the other hand, most of the rest of the new estimates are more than 50 per cent away from the original values, four of them differing from the original estimates by more than 100 per cent. In terms of the direction of error, most of the differences are negative; that is, the new estimates tend to be smaller, algebraically, than the old estimates. This is true in 8 out of 13 estimates of the constant term, and in 14 out of 15 estimates of the coefficient of current income. The reason undoubtedly is that in the

[6] Whether such changes are brought about by correction of previous errors or by revised definitions of the main variables is of no import in this analysis. The fact remains that few analysts attempt, or possess the facilities, to adjust one concept of, say, personal disposable income, to another concept which they might prefer.

Table 3

DIFFERENCE IN ESTIMATES OF PARAMETERS OF AGGREGATE SAVINGS FUNCTIONS DUE TO REVISIONS IN BASIC DATA

FUNCTION NUMBER	PERIOD OF OBSERVATION, NEW FUNCTION	METHOD OF FIT[a]	CONSTANT TERM			COEFFICIENT OF Y			COEFFICIENT OF Y_{-1}			COEFFICIENT OF T			OTHER PARAMETERS				DIFFERENCES IN FUNCTIONS OTHER THAN REVISION OF DATA
			ORIG.	NEW	% DIFF.	ORIG.	NEW	% DIFF.	ORIG.	NEW	% DIFF.	ORIG.	NEW	% DIFF.	VARIABLE	ORIG.	NEW	% DIFF.	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
A *Variables in current price aggregates*																			
(1.1)	1929-40	l.s.	−8.62	−7.89	8	.20	.15	−25											
(1.2)	1923-40	l.s.	−5.50	−6.59	−20	.17	.13	−24				−.04	−.028[d]	30					
(1.6)	1929-40	l.s.	−10.26	−7.05	31	.38	.19	−50	−.09	−.05	44	Not comparable							T^2 omitted in new calculations; income variable in original function was national income
B *Variables in deflated price aggregates*																			
(1.8a)	1923-30	l.s.	2.0	5.40	170	.07	−.05	−171											
(1.8b)	1931, ’36-37, ’39-40	l.s.	−1.9	−4.16	−119	.12	.11	−8											
(1.8c)	1932-35, 1938	l.s.	−4.9	2.36	148	.15	.04	−73											
(1.9)	1923-30, ’35-40	l.s.	1.3	.42	−68	.075	.036	−52											
(1.10)	1923-30, ’35-40	l.s.	−9.8	−4.1	58	.25	.11	−56				−.21	−.09[d]	57					
(1.12)	1923-41	l.s.	−9.7	−10.26	−6	.23	.195	−15				−.76	−1.195	−57	$\frac{YT}{P}$	.01	.016	60	Period of observation was 1921-41
(1.13)	1923-40	l.s.	12.6	12.50	−1	.23	.23	0							N	−.17	−.20	−18	Period of observation was 1910-16, ’23-40

C *Variables in deflated price per capita units*

(1.14)	1929-41	r.f.	−95.05	−117.6	−24	.29	.27	−7										
(1.15)	1923-40	l.s.	−76.58	−94.90	−24	.24	.23	−4				−1.15	−.84[d]	27				
(1.20)	1929-41	r.f.	−79.04	−96.35	−22	.42	.38	−10	−.13	−.10	23				$\frac{M}{NP}$	−.06	−.07	−17

D *Savings expressed as ratio to income*[b]

(1.21)	1923-40	l.s.	−2	[c]	.10	.045	−55	$Y\text{-}Y_o$	.12	.16	33	Period of observation was 1921-40
(1.23)	1923-40	l.s.	[c]	[c]	.10	.04	−60	$\frac{Y(Y\text{-}Y_o)}{Y_o}$	.16	.20	25	

[a] l.s. = least squares; r.f. = reduced form. The models for the latter are given in Table 2.

[b] To facilitate comparison with other functions in this table, functions (1.21) and (1.23) are expressed in terms of savings rather than the savings-income ratio. The coefficients of Y are given for the case Y_0 a variable, equal to Y. For (1.21), the marginal propensity to save when $Y < Y_0$, where Y_0 is a constant, is: Original, .227; New, .202. For (1.23), the marginal propensity to save as Y approaches (a constant) Y_0 from below is: Original, .264; New, .242.

[c] No constant term.

[d] Value of this coefficient is double that shown in Table 2 to preserve comparability with corresponding figure in column (13), since unit of T in new function is half of that in the original.

Source: Computed from Tables 1 and 2.

revised Department of Commerce series consumer expenditures are estimated directly rather than as a residual, as was the earlier practice. The direct estimate of consumer expenditures is considerably larger than the residual estimate, with the result that the importance of savings relative to aggregate income is less in the new series than in the old.[7]

Do these data provide a means for making a priori judgments as to the stability of empirical estimates with regard to changes in data? They cannot, to be sure, yield conclusive answers. For one thing, the type of revision is always a major determining consideration; thus, from knowledge of the changed procedure for estimating consumer expenditures in the present case, the direction of change in the coefficient of current income could have been predicted, though probably not the magnitude of the change. Some tentative inferences as to the effect of changes in data on parameter estimates can, however, be drawn from Table 3, with the proviso that they pertain only to data changes similar in some relevant way to the ones that occurred in this case:

1) There is some tendency for the margin of error to be smaller when the period of observation is longer. Of fourteen estimates based on 1923-40 or 1923-41 fits (excluding 1.13 because of the great difference between the periods of observation), seven differences between new and old estimates were not more than 25 per cent and none exceeded 60 per cent.
2) Little correlation seems to exist among relative differences between estimates of the parameters of the same function. In other words, a small percentage deviation in the estimates of one parameter does not necessarily mean that the relative deviations of the other parameters of that function will also be small. A striking illustration of this fact, even allowing for differences in the period of observation, is (1.8b), where the estimates of the income coefficient were within 8 per cent of each other while the estimates of the constant term differed by 119 per cent.
3) Some stability appears in the differences between estimates for the same variable. Thus the margin of error in the five cases where a time coefficient was being estimated (including the estimate of YT/P) ranged from 27 per cent to 60 per cent (though the direction of the differences was not always the same). The margin of error for fourteen out of eighteen estimates of the current income coefficient and of the constant term was 25 per cent or less for continuous periods of observation such as 1929-41, 1929-40, and 1923-40, and between 52 and 68 per cent for the 1923-30, '35-40 period of observation.

[7] See *Survey of Current Business,* National Income Supplement, July 1947, pp. 14 f.

4) No differences in the stability of the estimates as between different functions are apparent. Without a systematic design of estimation, of course, such differences are not easily discernible even if they do exist.

5) No difference was found in the stability of the estimates by method of fit.

All of these inferences are highly tentative, and are useful primarily as points of departure for further study. The one point that emerges clearly is that estimates of parameters are heavily dependent on the particular set of data on which they are based. The difference between significance and nonsignificance of a given coefficient is likely to depend at least as much on the particular set of data employed as on the form of the function or the period of observation. In at least one case, (1.8), the interpretation of the functions would have been altered had the calculations been based on the new data. The original calculations, as the entries for (1.8) in Table 1 show, had led the author to observe that the marginal propensity to consume was smaller in depression than in prosperity; the new calculations (Table 3) throw some doubt on this result. They also imply a marginal propensity greater than unity for 1923-30.

Systematic design of computations

The questions raised earlier about criteria for predictive accuracy can only be answered through further information bearing specifically on those points. Additional computations have been made in order to answer the following questions:

1) How accurate are savings and consumption functions based on prewar data likely to be in their postwar forecasts?

2) What is the effect of the period of observation on the accuracy of prediction of various functions? In particular, does any improvement result from eliminating so-called atypical years from the period of observation?

3) What effect does past income have on the accuracy of prediction? What about income distribution?

4) How effective is goodness of fit as an indicator of predictive accuracy?

5) Does adjustment of the economic variables for price changes tend to improve the predictive value of a function?

6) Does adjustment for population increase the accuracy of prediction?

At least one more question could be raised in such an analysis, namely, the effect of the method of fit on the predictive accuracy of consumption functions. The ramifications of this question are so broad, however, that an adequate treatment did not seem possible with the limited computational facilities available for this study.

Systematic computations were made on seven different types of functions, which will be referred to by letter as follows:

A $S = f(Y)$
B $S = f(Y, Y_{-1})$
C $S = f(Y, N)$
D $S = f(Y, T)$
E $S = f(Y, Y_{-1}, T)$
F $S/Y = f[(Y - Y_0)/Y]$
G $S/Y = f(Y/Y_0)$

These functions include all the forms shown in Table 2 with the exception of (1.12) and (1.20).

Adjustments were made alternately for price changes, for population changes, for both, and for neither, and each specific function was fitted to one or more of three periods: 1929-40; 1923-40; and 1923-40 excluding 1931-34. All in all, sixty functions were fitted in this manner.

Results of systematic computations

A summary of the main results of the systematized computations is given in Table 4. An identification number for each function appears in column 1, with its classification number from Table 2, if any, listed in column 2. The general function form is shown in column 3, adjustment for price and population changes in column 4, and the period of observation for that particular case in column 5. The coefficient of determination for each function adjusted for sample size is shown in column 6. Asterisks opposite these figures indicate evidence of (positive) serial correlation in the residuals, as based on the ratio of the mean-square successive difference to the variance:[8] serial correlation at the .05 significance level is shown by a single, and at the .01 level by a double, asterisk. Columns 7 and 8 show the predictive accuracy of the various functions when applied to the postwar years 1947-49. The average of the absolute relative margins of error of the predictions for 1947-49 is presented in column 7, and for 1948-49 in column 8. The errors of prediction for the first half of 1950 at annual rates are shown separately in column 9.[9] A staggered arrangement of the data in these last four columns has been used to facilitate intertemporal comparisons. Column 10 presents the average error of prediction for

[8] See footnotes to figures in column 5 of Table 2.

[9] Only the first half of 1950 was used because of the possibility that actions brought about by the Korean fighting might have affected the results for the second half. Actually, however, some test computations made after the completion of this study using data for all of 1950 yielded very similar results.

1931-34 for those functions fitted to the period 1923-30, '35-40; the rationale behind these computations is discussed on page 46.

Accuracy of postwar forecasts

A glance at columns 7-9 of Table 4 reveals that, on the whole, the postwar savings forecasts of the functions considered leave much to be desired. Average errors of 100 per cent and more for 1947-49 and 1948-49 occur frequently; and even for 1950 (actually, the first half of 1950 only), few forecasts come closer than within 20 per cent of the true figure. The forecasts for 1947 are markedly less accurate than those for the later years, errors of 200 per cent and more being frequent.[10] Otherwise, however, no noticeable improvement in accuracy is apparent through time — the 1950 forecasts are not significantly more accurate than the 1948 forecasts.[11] With regard to the direction of error, all but one of the functions, (2.6c), overestimated 1947 savings,[12] and most also overestimated 1948-50 savings. Underestimation in 1948-50 occurred generally among those functions fitted to the discontinuous period, and among almost all functions of types F and G, which estimate the savings ratio rather than the level of savings.

Although the margins of error shown in Table 4 are, on the whole, very large, the performance of these functions may be viewed in another way. Because the savings figures are small relative to income, any deviation of a savings forecast from the actual value expressed as a per cent of savings is bound to be much larger than if the same deviation, viewed as an error in a consumption forecast, were expressed as a per cent of aggregate consumption expenditures (the confidence interval of the forecast would be the same whether savings or consumption is the dependent variable). This is illustrated by Table 5, which compares the relative errors of prediction of selected functions when savings and consumption, in turn, are used as the dependent variables. The substitution of consumption for savings reduces the per cent error of the forecast as much as twentyfold. Thus the 114.7 per cent error in the 1948-49 predictions of function (2.1a) reduces to 6.1 per cent when consumption is the variable being estimated. The 33 per cent error of the 1950 prediction of (2.7b) reduces to 2 per cent with consumption as dependent. In this sense, therefore, many of the prediction errors are quite small.

Some idea of the range of the forecasting error is provided by Chart 1. The upper panel of this chart shows the fluctuation in actual per capita consumption at constant prices (the solid line) from 1923 to 1950, together

[10] See Appendix Table B-2.

[11] It is interesting to note that the forecasts from functions of the types A-E tend to be most accurate for 1950, whereas forecasts from functions of types F and G are most accurate for 1949.

[12] Table B-2.

Table 4

GOODNESS OF FIT OF SELECTED AGGREGATE SAVINGS FUNCTIONS AND ACCURACY OF THEIR POSTWAR PREDICTIONS

FUNCTION NUMBER (1)	NUMBER FROM TABLE 2 (2)	FUNCTION TYPE (3)	VARIABLE(S) USED FOR DEFLATION (4)	PERIOD OF OBSERVATION (5)	R^2 ADJUSTED FOR SAMPLE SIZE† (6)	AVERAGE ABSOLUTE PERCENTAGE MARGIN OF ERROR 1947-49 (7)	1948-49 (8)	1950 (9)	1931-34 (10)
(2.1a)	1.1	**A** $S = f(Y)$		a	.82	195	115	79	
(2.1b)				b	.61*	169	95	63	
(2.1c)				c	0	40	26	40	400
(2.2a)			P	a	.81	179	102	74	
(2.2b)	1.8			b	.39**	123	61	37	
(2.2c)	1.9			c	0	39	23	37	375
(2.3a)			N	a	.78	170	95	62	
(2.3b)				b	.64*	140	73	44	
(2.3c)				c	.06	42	10	22	350
(2.4a)			NP	a	.86	161	87	62	
(2.4b)				b	.62**	156	84	58	
(2.4c)				c	0	46	10	19	350
(2.5a)	1.3	**B** $S = f(Y, Y_{-1})$		a	.86	180	102	68	
(2.5b)				b	.70*	146	76	47	
(2.5c)				c	.25	43	7	23	420
(2.6a)			P	a	.80	164	94	69	
(2.6b)				b	.48*	71	35	22	
(2.6c)				c	.23	27	36	41	398
(2.7a)			N	a	.84	158	85	54	
(2.7b)				b	.72*	123	59	33	
(2.7c)				c	.31	50	5	11	319
(2.8a)			NP	a	.85	148	81	58	
(2.8b)				b	.64	119	64	48	
(2.8c)				c	.20	23	25	29	389
(2.9a)		**C** $S = f(Y, N)$		a	.82	216	130	92	
(2.9b)				b	.61*	146	77	47	
(2.9c)				c	0	39	40	52	446
(2.10a)			P	a	.86	139	69	45	
(2.10b)	1.13			b	.77	115	51	29	
(2.10c)				c	.14	64	15	3	176

FUNCTION NUMBER (1)	NUMBER FROM TABLE 2 (2)	FUNCTION TYPE (3)	VARIABLE(S) USED FOR DEFLATION (4)	PERIOD OF OBSERVATION (5)	R^2 ADJUSTED FOR SAMPLE SIZE† (6)	AVERAGE ABSOLUTE PERCENTAGE MARGIN OF ERROR 1947-49 (7)	1948-49 (8)	1950 (9)	1931-34 (10)
(2.11a)		**D** $S = f(Y,T)$		a	.82	206	122	85	
(2.11b)	1.2			b	.59*	158	87	55	
(2.11c)				c	0	39	38	50	463
(2.12a)			P	a	.86	170	95	69	
(2.12b)				b	.70*	141	73	50	
(2.12c)	1.10			c	.02	55	12	8	260
(2.13a)			N	a	.83	205	120	84	
(2.13b)				b	.60*	159	87	56	
(2.14a)			NP	a	.86	149	78	54	
(2.14b)	1.15, 1.16			b	.72*	123	59	40	
(2.14c)	1.17			c	.05	55	11	8	236
(2.15a)		**E** $S = f(Y,Y_{-1},T)$		a	.84	186	106	72	
(2.15b)				b	.69*	131	65	38	
(2.15c)				c	.24	43	20	33	424
(2.16a)			P	a	.86	146	82	60	
(2.16b)				b	.74*	103	54	39	
(2.16c)				c	.24	24	8	16	253
(2.17a)			NP	a	.85	126	65	46	
(2.17b)	1.18			b	.73*	87	40	26	
(2.17c)	1.19			c	.25	24	14	20	398
(2.18b)		**F** $\frac{S}{Y} = f\left[\frac{(Y\text{-}Y_o)}{Y}\right]$		b	.71	45	3	18	
(2.18c)				c	.08	41	9	24	263
(2.19b)			P	b	.82	18	15	19	
(2.19c)				c	.29	23	16	23	134
(2.20b)			N	b	.67*	46	4	19	
(2.20c)				c	.06	41	10	26	292
(2.21b)	1.21		NP	b	.79	19	13	19	
(2.21c)				c	.21	26	14	24	177
(2.22b)	1.23	**G** $\frac{S}{Y} = f\left(\frac{Y}{Y_o}\right)$	NP	b	.74	15	14	19	
(2.22c)				c	.23	26	15	24	216

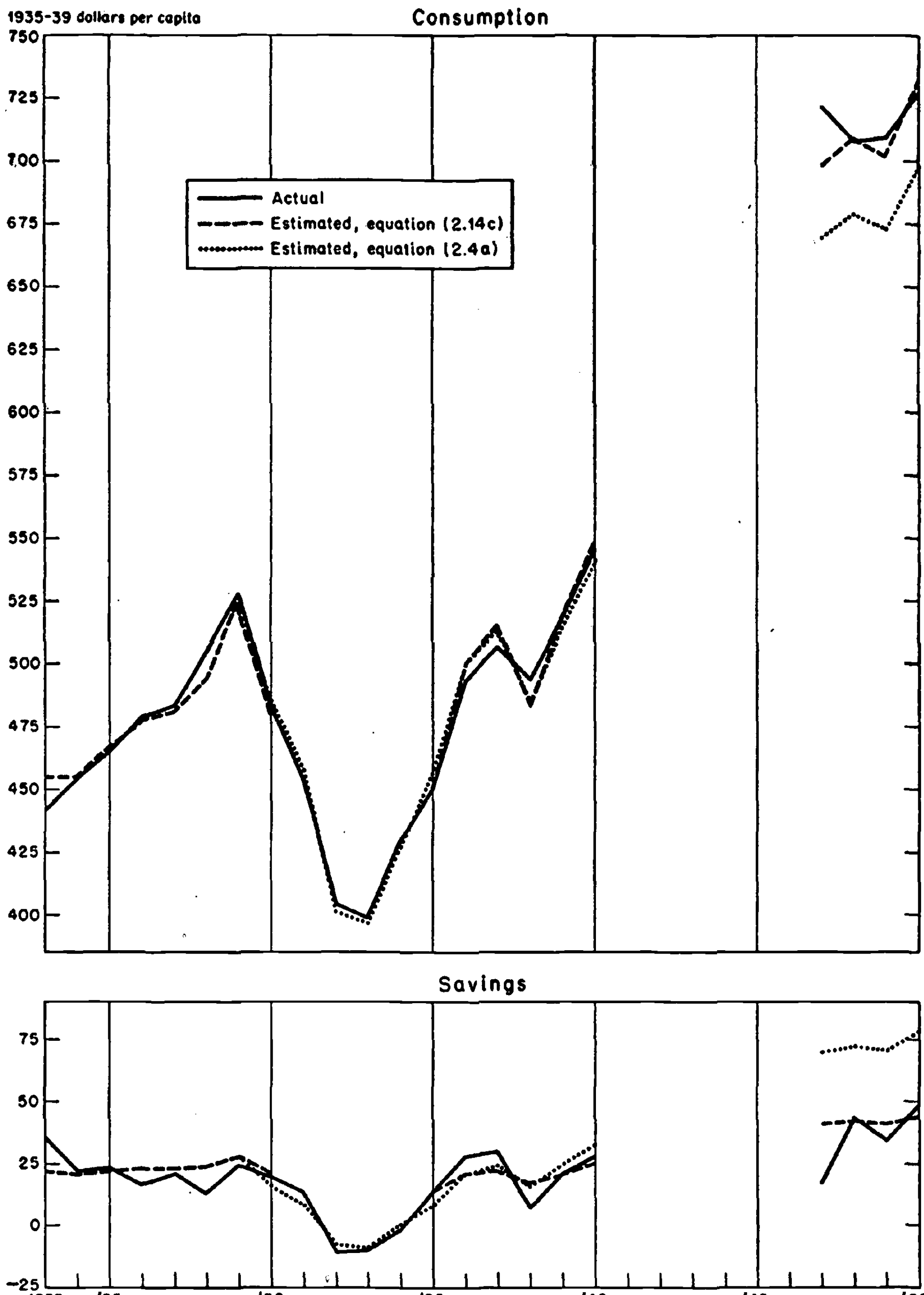
a = 1929-40; b = 1923-40; c = 1923-30, '35-40.

† Single asterisk indicates presence of auto-correlation in the residuals at the .05 level of significance; double asterisk, at the .01 level. All asterisks indicate positive auto-correlation.

Note: Functions (1.17), (1.18), and (1.19) appear in Table 1 but not in Table 2.

Source: Appendix B.

Chart 1
Accuracy of Selected Consumption and Savings Functions



Source: Appendix B.

with the estimates from one of the functions among those yielding the most accurate postwar forecasts, (2.14c), and from one of the poorest predictors, (2.4a). The corresponding series in terms of savings are shown in the lower panel. The difference between the two functions in predictive accuracy is clearly considerable, although the two are almost equally accurate in the period of observation. (The relationship between the accuracies of the function estimates in the two periods is discussed on pages 45 ff.)

Despite the differences in postwar predictive accuracy, the functions, even the most inaccurate among them, tend to reproduce the general time pattern of savings in this period. This is brought out by Chart 2, which compares the postwar estimates from each of six function forms for two

Table 5

RELATIVE ERRORS OF PREDICTION OF SELECTED FUNCTIONS WHEN CONSUMPTION RATHER THAN SAVINGS IS TREATED AS DEPENDENT

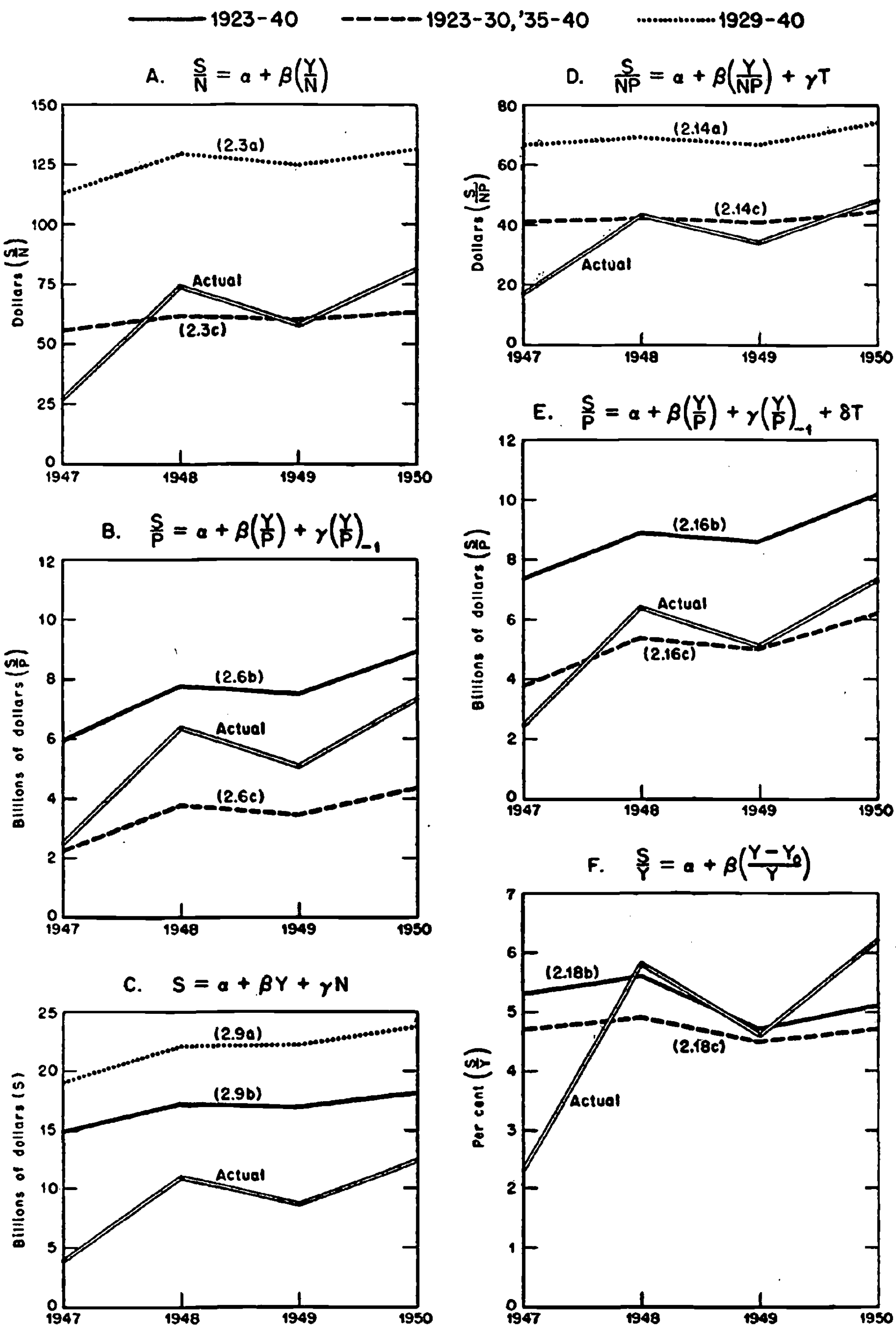
FUNCTION NUMBER	FUNCTION TYPE	PERIOD OF OBSERVATION*	AVERAGE ABSOLUTE PER CENT ERROR IN ESTIMATE OF 1948-49: *Savings*	1948-49: *Consumption*	1950: *Savings*	1950: *Consumption*
(2.1a)	$S = f(Y)$	a	114.7	6.1	78.9	5.3
(2.1b)		b	95.0	5.0	62.6	4.2
(2.1c)		c	26.2	1.5	39.8	2.6
(2.7a)	$\frac{S}{N} = f\left[\frac{Y}{N}, \left(\frac{Y}{N}\right)_{-1}\right]$	a	84.8	4.6	54.3	3.6
(2.7b)		b	59.3	3.2	33.0	2.2
(2.7c)		c	5.3	0.3	11.0	0.7
(2.17a)	$\frac{S}{NP} = f\left[\frac{Y}{NP}, \left(\frac{Y}{NP}\right)_{-1}, T\right]$	a	65.4	3.4	45.6	3.0
(2.17b)		b	40.4	2.2	26.2	1.7
(2.17c)		c	13.9	0.8	19.6	1.3
(2.21b)	$\frac{S}{Y} = f\left[\frac{(Y/NP) - (Y/NP)_0}{Y/NP}\right]$	b	12.8	0.7	19.4	1.3
(2.21c)		c	13.6	0.8	24.2	1.6

* See Table 4.
Source: Appendix B.

of the periods of observation with actual savings for those years. The six function forms shown were selected more or less arbitrarily to represent different types of functions, units of measurement, and pairs of periods of observation, as well as different levels of accuracy in their postwar estimates.

Although many of the functions yielded estimates differing considerably from actual savings (and this was the case especially with functions fitted to 1929-40), almost all of them reproduced, at least faintly, the time pattern of savings from 1947 to 1950. The forecasts rose from 1947 to 1948, declined in 1949, and rose again in 1950, as did actual savings. Only one

Chart 2
Actual Postwar Savings and Savings Estimates from Six Selected Functions



Source: Appendix B.

exception appears to this pattern, and that is in Panel C of the chart, where the estimates of the function fitted to 1929-40 recorded a slight increase from 1948 to 1949.

Another interesting phenomenon brought out by Chart 2 is the damped nature of the estimates. In all cases, actual savings fluctuate more widely than estimated savings. Of the six function forms shown, only the estimates of those in Panels B, E, and F show anything like the sharp movements of actual savings. It may be significant that each of these three functions incorporates past income in some form.[13]

Perhaps the most relevant factor in assessing the value of the forecasts is their absolute accuracies in relation to the forecasts that could have been made without the help of these functions. One yardstick at hand for such an evaluation is the serial correlation in the estimated series. In other words, are the forecasts of the regression functions more accurate than forecasts obtained by projecting the current year's level of savings to the following year? In the case of simple projection forecasts, the average absolute error for total savings would have been 99 per cent for 1947-49 and 45 per cent for 1948-49; for per capita deflated savings, the average errors would have been 115 per cent for 1947-49 and 43 per cent for 1948-49. When these figures are compared with those in columns 7 and 8 of Table 4, many of the functions do not show up very well. None of the forecasts yielded by functions using 1929-40 as the period of observation come as close to actual savings as these simple projections! All the functions fitted to 1923-30, '35-40, however, yield forecasts superior to the simple projections, as do all the functions under types F and G. Some of these functions are far more accurate than the simple projections, so that ability to differentiate these functions from the others would be of great value. At the same time, there are limitations to the use of functions fitted to a period excluding 1931-34, which will be discussed below.

A somewhat different evaluation of the relative accuracy of the forecasts is obtained if the regression functions are used to predict the change in savings from the previous year instead of the absolute level. In other words, the savings function, $S = f(Y)$ could be converted into the difference form, $\Delta S = \Delta f(Y)$, and the estimate of savings in the current year would be obtained as $S_{t-1} + \Delta f(Y_t)$. In the case of (2.1a), for example, the forecast function by this difference method becomes: $S_{t-1} + \beta(Y_t - YS_{t-1})$. In contrast to the results obtained above, the forecasts derived in this manner are generally superior to the estimates of zero change implied by the simple projection of the previous year's savings to

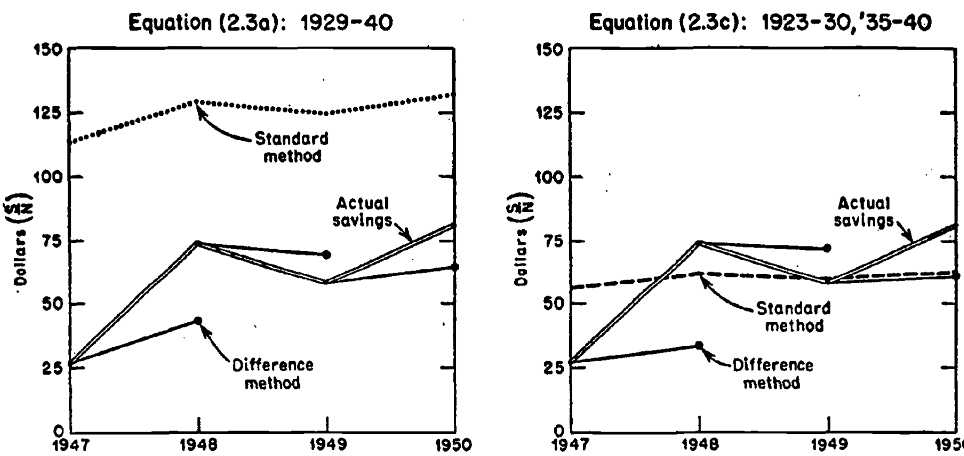
[13] Geoffrey Moore suggests that this understatement of the change in savings — a tendency also present in the period of observation — may be due to the larger amplitude of savings relative to that of the independent variables used. More complex functions, including more variable factors, might explain more fully the fluctuations in savings.

the current year. Needless to say, these forecasts are also usually more accurate than those produced by the more standard method: 8 of the 13 regression functions shown in Charts 1 and 2 are more accurate, on the average, in 1948-50 by the difference method.

The use of this difference method may generally be expected to yield superior forecasts whenever the regression functions trace correctly the direction of change but are consistently too high or too low. This is essentially the case for most of the functions studied for 1948-50, and accounts for the superiority of the method in this case. (Cf. Chart 2, where the only exception is the inaccurate prediction of the 1948-49 movement by function (2.9a). All of the 60 functions fitted forecast the direction of change of saving in 1947-48 and 1949-50 correctly, but in 6 cases incorrectly forecast a rise from 1948 to 1949.) If the function estimates are not consistently too high or too low, or do not trace correctly the direction of change, the use of this difference method is more likely than not to yield less accurate forecasts than the standard procedure. The two functions shown in Panel A of Chart 2 provide a striking example of the difference in accuracy yielded by the two methods. Both functions correctly trace the pattern of savings from 1947 to 1950, but the estimates of (2.3a) deviate consistently and considerably from actual savings whereas those of (2.3c) do not. The result, as shown in Chart 3, is that the difference method proves highly effective with (2.3a) but very ineffective with (2.3c).

Chart 3
Actual Postwar Savings and Savings Estimates from Two Functions by Alternate Methods

$$\frac{S}{N} = \alpha + \beta\left(\frac{Y}{N}\right)$$



Source: Appendix B.

Effect of business fluctuations: the period of observation

It appears in Table 4 that the use of the longer period of observation, 1923-40, invariably improves the accuracy of the estimates as compared with the use of the shorter period, 1929-40. Evidently the longer period tends to provide a more representative base for prediction purposes. The goodness of fit is, however, higher for the shorter period.

Predictive accuracy, moreover, improves markedly when the years 1931-34 are omitted from the period of observation. Most interesting, however, is the lack of uniformity in this improvement. It is with functions of types A-E that the use of the discontinuous period is most effective relative to the use of 1929-40 or 1923-40; in most of these cases the accuracy of prediction is increased sharply. With functions of types F and G, however, the improvement in predictive accuracy, if any, is slighter, and occurs especially when the 1947 estimates are included. These functions, devised by Modigliani and Duesenberry, both incorporate so-called cyclical adjustments; that is, they each contain a variable relating current income to a measure of highest previous income. This suggests one possible explanation for the improved predictive accuracy of the other functions when 1931-34 is excluded from the period of observation: because of the sharp contrast in the stage of business fluctuation characterizing these years and the years for which savings is being predicted, the omission of 1931-34 acts as a sort of cyclical adjustment factor. When explicit account is taken of this factor in the form of the estimating equation, as in function types F and G, the omission of abnormally depressed years becomes superfluous.

If this explanation is correct, then the fact that the omission of years unlike those for which savings is predicted[14] improves the predictive accuracy of a function would seem to point to an inadequacy in the make-up of the function rather than to the inherent desirability of discontinuous periods of observation. In other words, the functions in Table 4 of types other than F and G fail to allow adequately for changes in consumption and savings relationships induced by business fluctuations; variables for current and lagged income and for time trend apparently do not suffice. The identity of the missing element (or elements) remains to be discovered. One answer may be use of the highest previous income variable, Y_0, as in (2.18)-(2.22); but the relatively low values of the coefficients of determination indicate that this is at best only a partial answer.

It should further be noted that the period 1929-40 was weighted heavily by depression, as was 1923-40 to a lesser extent. To omit 1931-34 from either of these periods of observation tends to yield a postwar forecast biased against depression. In view of the characteristics of the early postwar years, this fact undoubtedly improves the 1947-50 predictions of func-

[14] Making the dubious assumption that one would know in advance which years to omit.

tions fitted to periods excluding 1931-34; but the continued use of such functions cannot be justified unless the observer is sure that depressions of such magnitude are out of the question in all future years.

One result stands out in this discussion. It is that the problem of selecting a period of observation — and, to a large extent, the entire problem of improving the accuracy of forecasts — reduces to that of making proper allowance in this period for the various stages of business fluctuations. To exclude one stage of fluctuations is to bias predictions for future years in favor of the other stages. This is the reason for the greater predictive accuracy (for 1947-50) of most of the functions in Table 4 when 1931-34 is omitted from the period of observation. Ideally, we would want the period of observation to include all possible types of business fluctuations. Actually, there is little choice in view of the limited period for which data are available. Given this situation, the best recourse is to seek a function which, although fitted to a limited period of observation, will make the necessary adjustments for business fluctuations. In other words, we seek functions that

a) will supply the most accurate forecasts when fitted to the entire period of observation, and
b) are likely to be as accurate in predicting savings for prosperous years as for depressed years.

Only (2.6) and functions of types F and G satisfy the first criterion. The extent to which the second criterion is met was determined by computing the average absolute percentage of predictive error in "forecasts" for 1931-34 yielded by all the functions in Table 4 fitted to 1923-30, '35-40; the results are shown in column 10 of that table. In no instance is any function as accurate for these years as for the postwar years,[16] but functions of types F and G are again among the most accurate of those tested. Thus, although these results are by no means conclusive, adjustment for cyclical variation by some such method as that employed in function types F and G would seem to be a promising means of obviating the deficiencies of the period of observation.

A by-product resulting from the use of three different periods of observation in these computations is some additional information on how the business cycle affects the marginal propensity to save. Bennion's study had showed a higher propensity to save in depression, but the recomputation of his functions using the revised Department of Commerce data had cast doubt on this conclusion, as noted earlier. Our systematic computations, however, enable us to examine this relationship more thoroughly,

[16] The huge percentage errors are due in part to the fact that savings in 1933 and 1934 were close to zero, which of course tends to magnify relative errors. The absolute residuals, however, far exceeded the residuals obtained for the period of observation and were roughly of the same order of magnitude as the postwar residuals for all the functions.

since these computations are based on, in turn, a period heavily weighted by depression (1929-40), a period less heavily weighted by depression (1923-40), and a period entirely excluding the main depression years (1923-30, '35-40). Some idea of the effect of the cycle can therefore be obtained by comparing the marginal propensities yielded by the same function when fitted to these different periods.

The data required for such a comparison are shown in Table 6, which presents the marginal propensities to save of all the functions for which systematic computations were carried out. Examination of this table leads to the following conclusions:

1) The marginal propensity is less for 1923-40 than for 1929-40 in 10 out of 17 possible comparisons.
2) It is less for 1923-30, '35-40 than for 1929-40 in all 16 possible comparisons.
3) It is also less for 1923-30, '35-40 than for 1923-40 in every one of 21 comparisons.

Table 6

MARGINAL PROPENSITY TO SAVE OF SYSTEMATICALLY-COMPUTED SAVINGS FUNCTIONS

FUNCTION NUMBER	MARGINAL PROPENSITY TO SAVE 1929-40 (a)	1923-40 (b)	1923-30, '35-40 (c)
(2.1)	.1519	.136	.035
(2.2)	.1999	.1421	.0355
(2.3)	.1472	.1295	.0527
(2.4)	.2229	.2082	.0663
(2.5)	.1952	.2030	.1600
(2.6)	.2204	.2391	.1524
(2.7)	.1996	.1998	.1748
(2.8)	.2420	.2635	.1598
(2.9)	.1497	.1314	.0325
(2.10)	.2333	.2278	.1474
(2.11)	.1495	.1334	.0301
(2.12)	.2340	.2256	.1074
(2.13)	.1493	.1357	n.c.
(2.14)	.2333	.2270	.1204
(2.15)	.1902	.2014	.1551
(2.16)	.2700	.2866	.2062
(2.17)	.2649	.2818	.2099
(2.18)*	n.c.	.1297	.0802
(2.19)*	n.c.	.2078	.1653
(2.20)*	n.c.	.1254	.0741
(2.21)*	n.c.	.2015	.1416
(2.22)†	n.c.	.2420	.1526

n.c. = Not computed.

* Marginal propensity for $Y < Y_0$.

† Marginal propensity for $Y \rightarrow Y_0$.

Source: Appendix Table B-2.

Thus the picture which these comparisons provide is of a progressively declining marginal propensity to save as depression years are accorded less weight in the period of observation. It is interesting to note that five of the Modigliani or Duesenberry type functions are included in the last group of 21 comparisons, and that even these functions, which incorporate a "built-in" cyclical adjustment, show a smaller marginal propensity when depression years are omitted.

Past income and income distribution

The insertion of a lagged income variable definitely seems to improve the estimates of consumption functions both in the period of observation and in the extrapolated postwar years. The goodness of fit of function types A and D is increased substantially when Y_{-1} is added, the only exceptions occurring where 1929-40 is the period of observation. The predictive accuracy of the functions is also increased generally, though a number of exceptions occur when the discontinuous period of observation is used, usually where the average error of prediction is already fairly small. For the functions considered, therefore, lagged income provides a noticeable improvement in the relationship.

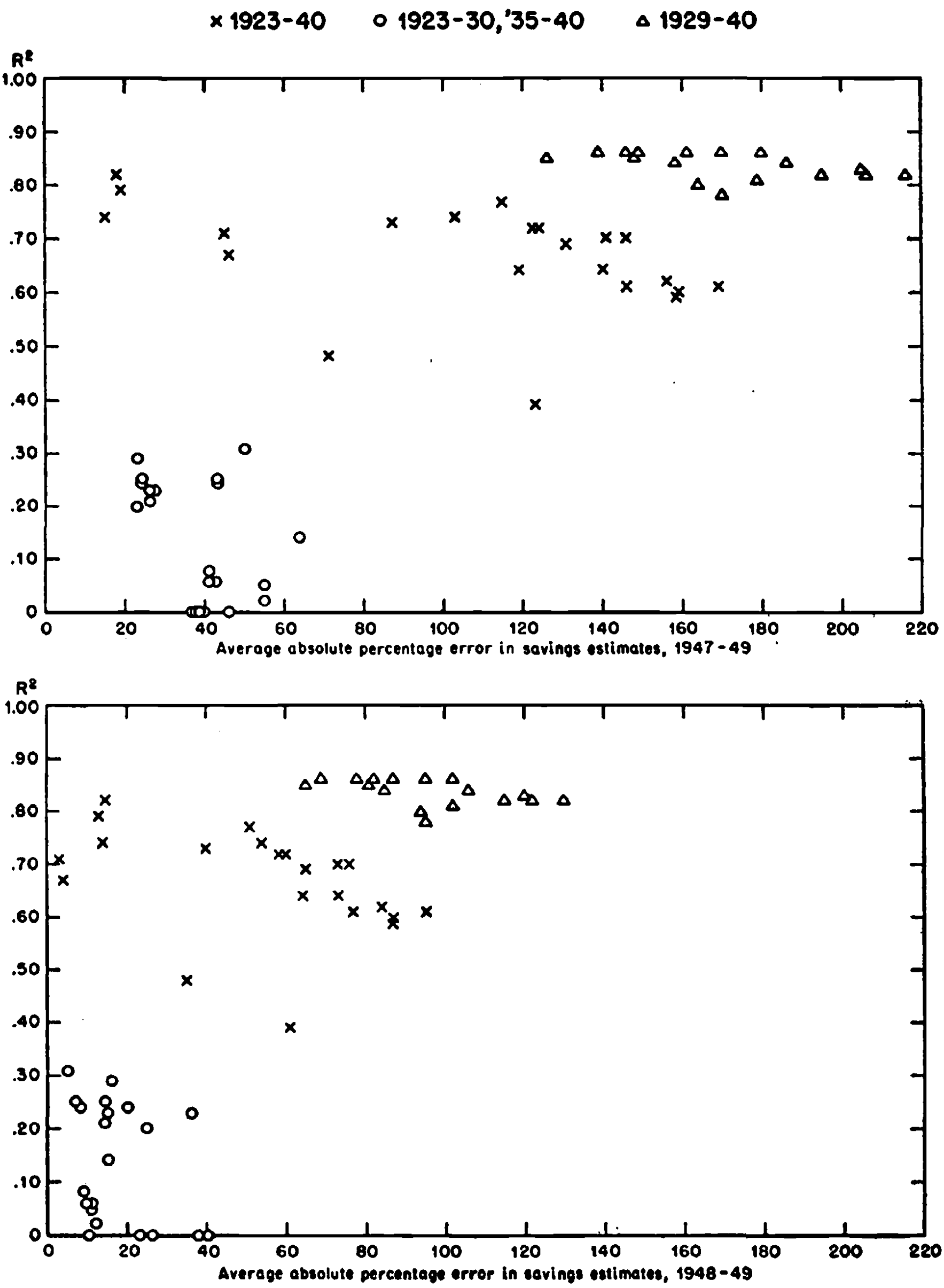
Another important result of using some form of lagged income is that it appears to adjust for the cyclical effect. This follows from the fact that the only functions whose postwar forecasts are at least as accurate when fitted to 1923-40 as when fitted to this period excluding 1931-34 are (2.2), (2.6), (2.9), and (2.18)-(2.22), almost all involving past income in one form or other. It also appears that Y_0 is more effective in this regard than Y_{-1}, at least when the variables are considered singly; but an alternate form reflecting the past income effect, either combining or excluding Y_0 and Y_{-1}, may well be better still.[16]

The relevance of an income distribution variable in an aggregate consumption function was not studied in any great detail. The only test made was to correlate the residuals of certain of the functions in Table 4 with Kuznets' estimates of the relative income shares of the upper income groups.[17] No clear-cut relations were obtained by this procedure, however, and the question of the relevance of income distribution to consumption functions remains a matter for further study.

[16] The use of highest previous consumption, C_0, as recommended by Davis, presents still another possibility. By substituting C_0 for Y_0 in the Modigliani and Duesenberry functions, Davis improved considerably the accuracy of their predictions for 1948-50. For 1947, however, the deviations of the estimates from the actual values were much larger when C_0 was used, and for 1951 the deviations of both types of functions were far above those of such functions as (2.14a) and (2.10a). It may be important to note, however, that Davis' period of observations was 1929-40.

[17] Kuznets, *op. cit.*, p. 67.

Chart 4
Relation between Coefficient of Determination (R^2) and Predictive Accuracy



Source: Table 4. R^2 is adjusted for sample size.

Goodness of fit and predictive accuracy

The relationship between the coefficient of determination and the predictive accuracy of the various functions shown in Table 4 is not clear offhand. However, the two scatter diagrams in Chart 4, where the value of R^2 is plotted against the average absolute per cent error for 1947-49 in the upper panel of the chart, and for 1948-49 in the lower panel, show that the over-all relationship is positive.

Further analysis reveals, however, that this rather fantastic result — a rise in predictive accuracy as goodness of fit declines—is only a superficial one. This is brought out when account is taken of the different periods of observation, as is done in Chart 4 by means of the three types of points used, one for each period of observation. In both panels of the chart, the points for each period of observation are seen to cluster about each other. All the points for the 1929-40 period of observation (the triangles) lie in a narrow horizontal band in the upper right-hand corner of the scatter; the functions fitted to this period have uniformly high coefficients of determination but also uniformly high percentages of error in their postwar predictions. Nevertheless, the relationship among this set of points is, if anything, slightly negative (in fact, $r = -.46$ and $-.49$ for 1947-49 and 1948-49, respectively), thereby leading to the more plausible inference of a positive relationship between goodness of fit and predictive accuracy. Much the same phenomenon is apparent when the relationship between the points corresponding to the 1923-40 fits and to the 1923-30, '35-40 fits is studied separately. The former set covers a wide range of predictive error but a relatively narrow range of R^2, generally from .4 to .8. Since the goodness of fit of these functions is less than that of the 1929-40 group, and their predictive accuracy higher, this set of points lies below and to the left of the points representing functions fitted to 1929-40. The same thing is true with regard to the position of the points for the 1923-30, '35-40 functions in relation to that of those for 1923-40. As a result, the over-all picture is one of positive correlation; but isolation of each of the three sets of points reveals, if anything, the more plausible negative relationship.[18]

In other words, the apparent inverse relationship between goodness of fit and predictive accuracy is attributable largely to differences in the period of fit, which leads us to conclude that when different periods of fit are involved, the use of goodness of fit as a criterion of predictive accuracy can be highly misleading. Chart 5 provides some striking illustrations of this statement. This chart compares the residuals of the same six function forms used in Chart 2 for two of the periods of observation to which they

[18] The coefficients of correlation for the 1923-40 and the 1923-30, '35-40 groups are —.43 and —.45, respectively, using the absolute average of the 1947-49 errors as the postwar variable.

were fitted, and for the postwar years. For both periods of observation, the unadjusted coefficient of determination of the function is noted.

Two points are evident from Chart 5. One is that although the coefficients of determination of the same function fitted to different periods may vary greatly, they provide no indication of the accuracy of these functions in predicting postwar savings. In fact, the relationship is clearly negative; the predictive accuracy of the pairs of functions on five of the six panels of the chart is the opposite of what one would expect on the basis of the relative sizes of the coefficients of determination. This, of course, confirms the over-all relationship shown in Chart 4; but why should such a relationship exist?

The reason becomes clear when we consider the second major point brought out by the chart. It is a striking one: Despite the great differences in the coefficients of determination of the same function fitted to different periods, the residuals of the functions where the periods of observation overlap are roughly of the same order of magnitude. For example, the residuals of the function fitted to 1923-30, '35-40 on Panel F of the chart are almost identical with those for the same years when the period of observation is 1923-40, although the coefficient of determination in the former case is .15 as against .73 for the 1923-40 data. Again, in Panel D the residuals are very similar for the overlapping years, 1929-30, 1935-40, although one coefficient of determination is .88 as against .19 for the other.[19] From Panel C it can be seen that the same phenomenon holds for 1929-40 functions in relation to 1923-40 functions. In fact, the time series pattern of the residuals of all the functions shown is roughly similar — negative in the late twenties, 1933-34 and 1939-40, and positive in 1930-31 and 1935-37.

What this means is that the reason why the coefficient of determination for a function fitted to, say, 1929-40 exceeds that of the same function fitted to 1923-30, '35-40 is *not* that the former yields so much more accurate estimates over the range of observation, but that any function incorporating 1931-34 data is bound to have a high value of R^2 simply because of the wide fluctuation of savings and income in those years. The same reason explains the higher value of R^2, even after adjustment for sample size, for the identical function when 1929-40 is the period of observation rather than 1923-40; that is, the over-all fit is *not* so much better for the

[19] The spread between the coefficients of determination would be substantially increased if adjusted coefficients were used. This is because the adjustment-for-sample-size formula used,

$$R'^2 = 1 - (1 - R^2)\left(\frac{N-1}{N-m}\right)$$

(N being the number of observations, and m being the degrees of freedom lost), "penalizes" low values of R^2 more than high values. For example, if a three-parameter function is fitted to 13 observations, the formula deducts 16 units if R^2 is .20; i.e., $R'^2 = .04$; but it deducts only 4 units if R^2 is .80.

Chart 5
Actual Minus Estimated Savings
in Period of Observation and in Postwar Period

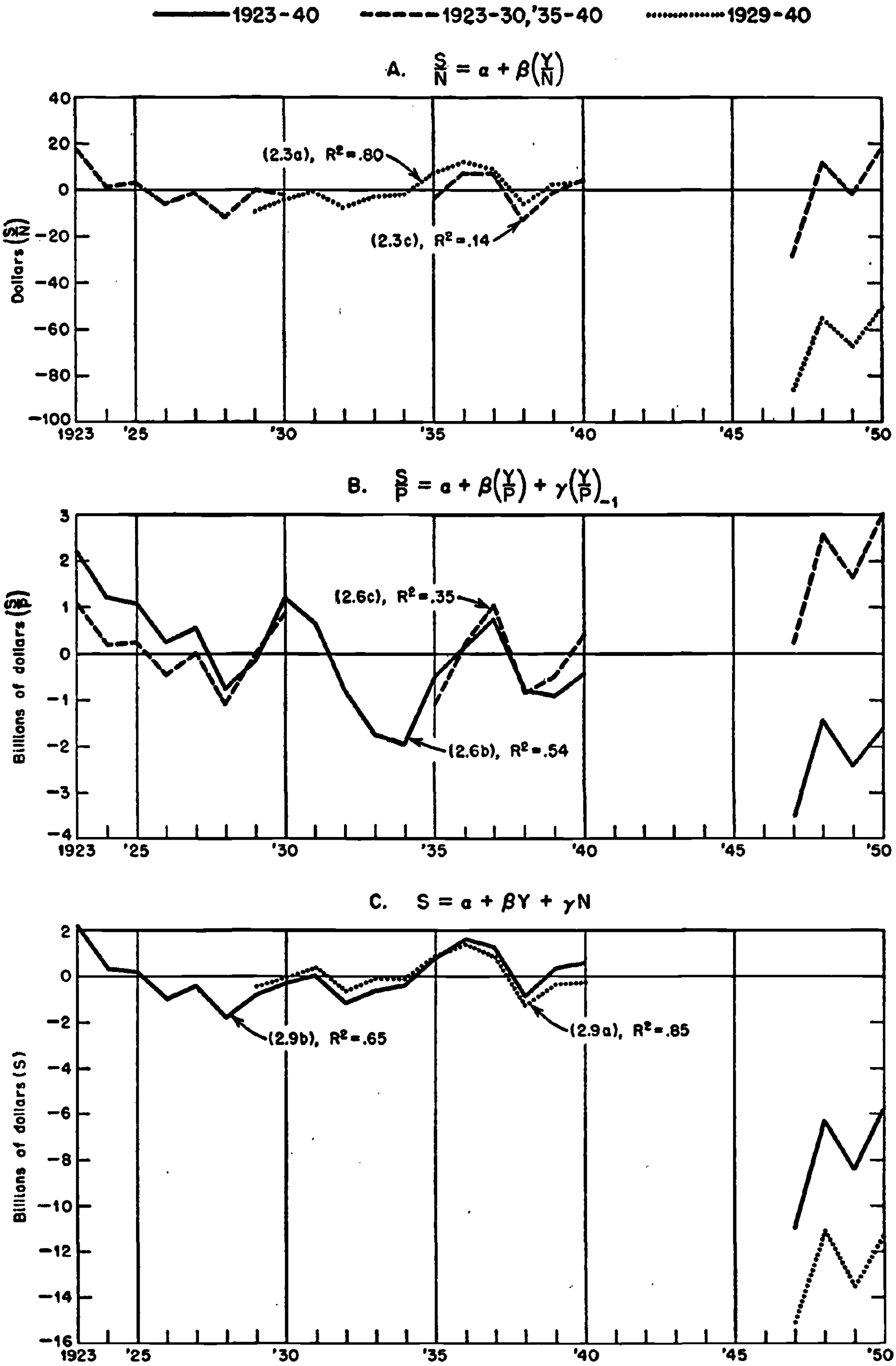
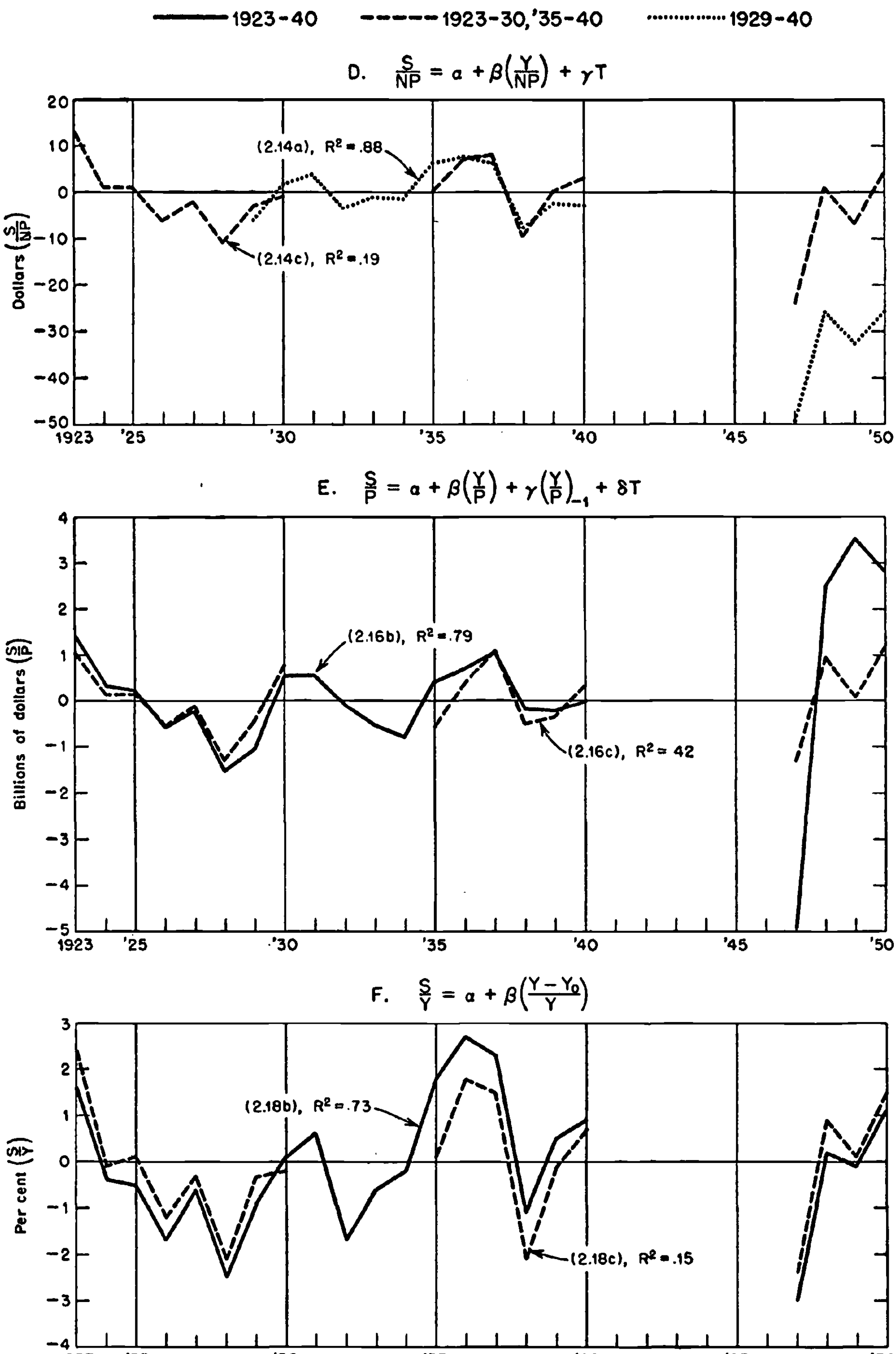


Chart 5 (concl.)

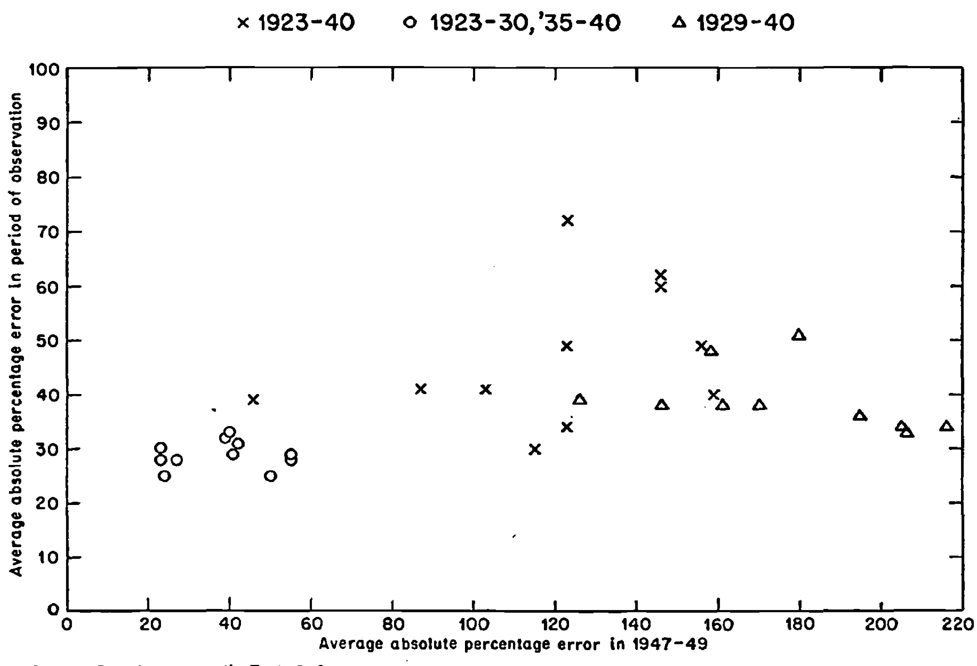
Actual Minus Estimated Savings in Period of Observation and in Postwar Period



Source: Appendix Table B-2.

former, but rather the fewer number of years covered increases the relative importance of the 1931-34 observations. The inclusion of 1931-34 increases the variance of savings, and so, in this case, raises the values of R^2.[20] But 1931-34 was a depression period, and as we have seen, the inclusion of depression years in most of the functions used to predict postwar prosperity levels of savings tends to reduce the accuracy of the forecasts. Therefore, the inclusion of 1931-34 in the period of observation tends to increase the goodness of fit and at the same time to lower the accuracy of the predictions. In the case of functions that seem to make adequate allowance for cyclical effects, as in Panel F of Chart 5, the inclusion of 1931-34 in the period of observation has no effect on predictive accuracy, though the value of R^2 is boosted tremendously. Thus, not only can goodness of fit be a highly misleading criterion of predictive accuracy as between two different periods of observation, but wherever the variances of the dependent variable(s) differ markedly, goodness of fit is meaningful in a relative

Chart 6
Relation between Average Errors in Savings Estimates in Period of Observation and in Postwar Period



Source: Based on Appendix Table B-2.

[20] For total undeflated savings, the variances are 1929-40, 3.39; 1923-40, 2.88; 1923-30, '35-40, 0.90. The corresponding variances of the residuals when computed from equation (2.1) are 0.54, 1.04, and 0.86. Although the variance of the residuals is, in general, smallest for 1929-40 and largest for 1923-40 for each function, the differences between periods either reinforce or are not large enough to counterbalance the effect of the differences in the actual variances on the relative values of R^2.

sense only and may provide no indication of the absolute accuracy of the functions in the years where the periods of observation overlap.

One question that might be raised is whether the results may have been affected by use of two different measures of goodness of fit, R^2 for the period of observation and the average absolute error for the period of prediction. In other words, might a closer relationship be obtained between the goodness of fit in the period of observation and that in the prediction period if the same measure were used throughout? For examination of this possibility, the average absolute errors for the period of observation were computed for a 50 per cent sample[21] of the functions in Table 4. The relationship between these values and the corresponding error figures for 1947-49 is shown in Chart 6.

In Chart 6, as in Chart 4, the over-all relationship is not a very close one; but contrary to the earlier case, it now is in the direction one would expect. Being positive, it indicates that functions whose estimates have a high average margin of error in the period of observation also tend to have a high margin of error for prediction purposes. The tendency is not strong, however, the coefficient of correlation between the two measures of error being .45. When the relationship is examined by period of observation, a distinction which is of major importance in this analysis, the tendency does not become any stronger, as evidenced by the coefficients of correlation for the three periods shown below:

Period of observation		*Coefficient of correlation*
a	1929-40	−.41
b	1923-40	.36
c	1923-30, '35-40	.07

None of the three "period" coefficients is statistically significant at the .05 probability level. Thus the relationship existing between these measures of goodness of fit in the period of observation and in the period of prediction, whatever it may be, appears to be slight and of little aid in practice.[22]

[21] The sample was selected rather haphazardly, but in such a way that: (a) each period of observation was represented by one-third of the functions; (b) all function types were represented; and (c) the number of times a function form fitted to one period coincided with the same form fitted to another period was roughly equal for any two periods of observation.

[22] Another striking phenomenon concerning the periods of observation to which these functions were fitted is that significant evidence of auto-correlation — in all cases, positive — appears only for the 1923-40 period (though all but three of the nonsignificant values of K also differ from the expected values in the direction of positive auto-correlation). The reasons for this are: 1) the denominator of K, the variance of the residuals, tends to increase with length of period fitted, whereas, 2) the numerator, the mean-square successive difference of the residuals, is smallest for 1929-40 and largest for 1923-30, '35-40 for each function. Positive correlation of the residuals is particularly in evidence in 1932-34, the deep depression years in which savings were consistently overestimated, and in 1925-30, when savings were relatively stable and the estimates of most functions were either all above or all below the actual figures. K, the quotient of these two terms, is therefore commonly lowest for 1923-40, larger for 1929-40, and largest for 1923-30, '35-40.

Adjustment for price changes

Deflation for price changes generally increases the predictive accuracy of the functions, judging by a comparison of the deflated and undeflated forms of the functions in Table 4. Differences in the effect of price deflation are apparent, however, among the various periods of observation and for certain function types, as follows:

a) In 24 comparisons of postwar predictive accuracy (based on columns 7-9 in Table 4) for functions of types A to E fitted to 1929-40 data, price deflation improved accuracy in 21 cases, made no change in one case, and decreased accuracy in two cases.

b) In 30 comparisons for functions fitted to 1923-40 data, price deflation improved accuracy in 20 cases, made no change in one case, and decreased it in 9. Three of the cases where accuracy decreased were functions of type F.

c) In 27 comparisons for functions fitted to 1923-30, '35-40 data, price deflation helped in 17 cases, made no change in one case, and decreased accuracy in 9 instances. For functions of type F, price deflation helped in 4 out of 6 comparisons. In 4 out of 9 instances where price deflation was used in addition to population deflation, less accurate predictions resulted.

While price deflation improves predictive accuracy more often than it does not, its effect would seem to depend on the accuracy of the undeflated function. The functions which in 1947-50 are more accurate — primarily those fitted to 1923-30, '35-40 data and those deflated by population — are less likely to be improved by price deflation. Since the validity of generalizations based upon results obtained from functions fitted to 1923-30, '35-40 is subject to some suspicion, however, the hypothesis that the effect of price deflation depends on the accuracy of the undeflated function is at best highly tentative.

It is interesting to note that reliance upon an examination of coefficients of determination would provide misleading information in this respect, as the coefficients of determination of the functions using undeflated variables frequently exceed those of the corresponding functions based on deflated variables.

Adjustment for population changes

On the question of adjustment for population the results are also uniform. For the four types of functions where a comparison was made between variables expressed in aggregates and variables expressed in per capita units (A, B, D, and F), the accuracy of the forecasts was improved, or at least not decreased, by population deflation in almost every instance. The evidence therefore points to the predictive usefulness of expressing the economic variables in per capita units. The goodness of fit of the functions is not affected appreciably by population deflation.

Given the desirability of allowing for population changes, is it preferable to make this adjustment by deflation or by including population as a separate variable in the function? The results in Table 4 are not conclusive on this point. The accuracies of prediction of (2.9) are less than those of the corresponding deflated function (2.3), but the reverse is true when the economic variables are deflated for price changes. In terms of absolute accuracy, the difference between (2.4) and (2.10) is not great. Essentially, function type C amounts to using $S = f(Y,T)$ (type D), and this is substantiated by the similarity of the results, although the former function is the more accurate of the two when price deflation is used.

Reliability of results

A pertinent question in interpreting the results concerning predictive accuracy is their consistency. In other words, do particular functions tend to yield consistently high or consistently low forecasts in all years, or do the averages shown in columns 7 and 8 of Table 4 conceal substantial variations in accuracy? In such matters one cannot, of course, generalize for all future years; but on the basis of the data at hand, a high order of consistency in the results would seem to be present. The functions yielding the most accurate forecasts for 1947 are almost invariably the most accurate for 1948 and for 1949 as well, though where price deflation is used, the consistency may be due at least in part to the fact that 1947-50 were years of almost constant real income. The exceptions to the pattern of consistency are concentrated almost exclusively among functions fitted to 1923-30, '35-40 data, and here the number of inconsistencies between the 1947 and 1948 forecasts is large. With this exception, however, the consistency among forecasts from the same function for different postwar years is quite high.

A further test of consistency derives from the use of the functions in Table 4 to estimate savings for the first half of 1950. The resultant estimates, shown in column 9 of Table 4, were compared in each case with the estimates furnished for 1949 (Appendix Table B-2). Comparison revealed that the general accuracy of the predictions was considerably increased in 1950. Functions that had yielded highly inaccurate predictions in 1948-49, such as (2.1), (2.5a), (2.11a), and (2.12a), showed the greatest improvement. But those functions that had been most accurate in 1948-49 — (2.3c), (2.4c), (2.5c), (2.7c), and (2.18b) — were among the most accurate for 1950, and those which were least accurate in 1948-49 also tended to be least accurate in 1950. It would seem, therefore, that the results are consistent in this sense.

Another question that might be raised is the effect of inaccuracies in the 1922-28 data on the validity of the results. Because of the relatively small magnitudes involved, large errors in the savings estimates for those years are quite probable. This is particularly so in view of the fact that savings was estimated as a residual, as income minus consumption. In some years, an error of as little as one billion in estimated consumption or income — only about 1.5 per cent of these aggregates — would spell the difference between a rise or a decline in savings. Such errors are unavoidable with the data at hand. What does their possible existence imply as to possible modification of the results of this empirical analysis? Without fitting the functions in Table 4 to alternative sets of savings-income estimates for 1922-28, it is difficult to tell for sure; yet short of such a tremendous operation, a number of inferences are possible. One is that wherever the functions fitted in part to the early years yield the same results as those fitted to 1929-40, there would seem to be little doubt as to the validity of the results. This is the case for the results bearing on the value of deflation by price and population, on the relative superiority of function type B over types A, C, and D, and on the consistency of the forecasts. But conclusions necessarily involving the 1922-28 estimates, such as the greater accuracy of the postwar forecasts when a longer period of observation is used, and the superiority of function types E and F, may be suspect in this regard, though the first of these two would seem plausible on an a priori basis. The improvement in predictive accuracy due to the omission of 1931-34 from the period of observation is another result that might fall in this suspect category, were it not for an additional test that was made. The test consisted of fitting (2.1), (2.2), and (2.3) to data for the period 1929-30, '35-40 (that is, omitting 1923-28 as well as 1931-34) and computing the relative error in the postwar residuals. The results confirmed the previous findings with reference to the exclusion of 1931-34 from the period of observation for this function type (A) in that the margin of predictive error for 1947-49 was considerably below that for the corresponding functions fitted to 1929-40 data.

Another consideration supporting the validity of the results is the fact that much the same results were obtained when savings and income estimates for 1922-28 made independently by the writer were used. These estimates had been made before Barger's estimates were provided and were constructed by means of regressions for the post-1929 period and extrapolation backward, rather than by building up components as Barger did. Although these estimates differed from Barger's estimates both as to the level of savings and, at times, as to the direction of change, the general nature of the results obtained by using these data to fit the seven function types in Table 4 did not differ in any major respect from those obtained here. On balance, therefore, it would seem that most of the results of the empirical analysis are largely unimpaired by possible errors in the estimates for 1922-28.

Summary of findings

1) Predictions of savings in the postwar years, 1947-50, obtained by inserting actual values of income, prices, etc. into seven different types of consumption functions derived from prewar data, are on the whole not very accurate. (See page 36 for the list of types of functions used and pages 5-6 for the definitions of variables.) Each function was fitted to one or more of three different time periods — 1923-40; 1929-40; 1923-30, '35-40 — and adjustments were made alternately for price changes, for population changes, for both and for neither. The largest discrepancies in the postwar predictions occur in 1947, where errors of 200 per cent and more are common; but even for the years more accurately forecasted, 1948 and the first half of 1950, the forecasts infrequently come within 20 per cent of the true figure. In terms of consumption, the errors of prediction are very much smaller, ranging generally between one and six per cent. A practical judgment of the value of these postwar forecasts, derived by comparing their accuracy with what would have been obtained had the forecast been simply the preceding year's value, reveals that none of the functions fitted to 1929-40 data was as accurate as these simple projections. Many of the 1923-40 functions, however, proved superior to the simple projections, as did all the functions fitted to 1923-30, '35-40 data, and the functions that took the highest previous income into account. In addition, many of the functions tended to reproduce well the time pattern of actual savings during 1947-50, although the amplitude of variation in the estimates was generally much smaller than the actual variation. Considerable improvement in accuracy was obtained, however, when the functions were converted to first differences and used to predict the change in savings rather than the absolute level.

2) The heart of the prediction problem seems to lie in securing some adjustment for changes in the savings-income relationship during the business cycle. Equations fitted to data for the longer period of observation, 1923-40, invariably yielded more accurate predictions than those derived from the shorter period, 1929-40; the latter is, of course, heavily weighted by depression years. Most interesting, however, was the finding that use of a discontinuous period (1923-40 with 1931-34 omitted) yielded greater accuracy than the use of 1923-40 in the case of functions of the type, $S = f(Y)$, $S = f(Y,N)$, $S = f(Y,T)$, and $S = f(Y,Y_{-1},T)$. It failed to do so for one form of $S = f(Y,Y_{-1})$ and for all forms taking the highest previous income into account; these were also among the most accurate of the functions tested.

Since the exclusion of 1931-34 from the period of observation does not seem justifiable on logical grounds, it would appear that the most desirable functions of the ones tested are those including a variable to allow for income in a previous peak year, functions whose accuracy did not improve when the years 1931-34 were excluded. Some such variable would therefore seem to be a desirable attribute of an aggregate consumption function, though it remains to be seen whether such functions continue to perform better in later years.

3) For the functions studied, lagged income provides a noticeable improvement both in the goodness of fit and in the accuracy of postwar forecasts in addition to its value for cyclical adjustment purposes, as noted above.

4) Price deflation generally improves predictive accuracy, but its effect seems to depend on the accuracy of the undeflated function. Those functions that were most accurate — especially those fitted to the discontinuous period and those deflated by population — were less often improved by price deflation.

5) Population deflation invariably increases predictive accuracy. However, the use of population as a separate variable does not seem to produce better results than are obtained with a function of the type $S = f(Y)$ deflated by population.

6) The coefficient of determination is not a reliable indicator of predictive accuracy. When two different periods of observation are involved, the coefficients of determination can be highly misleading as to which function is likely to be the more accurate predictor. For the periods under study the relationship was in fact the reverse of what might be expected, because inclusion of the depression years 1931-34 in a period of observation tended to increase greatly the coefficient of determination, and at the same time to reduce sharply the accuracy of the predictions for the prosperous postwar period. It also appeared that when, say, two functions are fitted to data for different but overlapping periods, the differences between actual and estimated savings may be almost identical for both functions during the overlapping period, even though the coefficients of determination differ substantially.

If the period of observation and the variable being estimated are held constant, a more plausible relationship is obtained between the coefficient of determination and predictive accuracy. However, even in this case the relationship did not turn out to be a very close one.

7) Another result of some interest is the heavy dependence of the empirical estimates of the parameters on the particular set of data employed. Differences of 25 per cent in the estimated parameters of the functions are common as a result of the substitution of the revised (1947) Department of Commerce national income series for earlier data, and differences of even 100 per cent and over occur frequently.