72-17,177

SMITH, Gary Nance, 1945-
ESTIMATING A GENERAL DISEQUILIBRIUM MODEL OF THE FINANCIAL SECTOR.

Yale University, Ph.D., 1971
Economics, finance

University Microfilms, A XEROX Company, Ann Arbor, Michigan

ESTIMATING A GENERAL DISEQUILIBRIUM MODEL OF THE FINANCIAL SECTOR

by

Gary Smith

A Dissertation Presented to the Faculty of the
Graduate School of Yale University
in Candidacy for the Degree of
Doctor of Philosophy

1971

PLEASE NOTE:

Some pages may have

indistinct print.

Filmed as received.

University Microfilms, A Xerox Education Company

SUMMARY

The primary task of this thesis is the estimation of a quarterly econometric model of the United States financial sector. A general portfolio approach is discussed and a specific model constructed. Some anticipated econometric problems are then considered, with particular attention paid to the effective dimensionality of the explanatory variables.

The model is estimated in ten ways using postwar flow of funds data. In five cases, extensive a priori restricitions are imposed on the model in response to collinearity patterns. After an examination of the in-sample and out-of-sample forecasts, a final model is estimated, using all of the available data, by single equation ordinary least squares with a moderate number of a priori constraints and residual-minimizing first order autoregressive parameters.

TABLE OF CONTENTS

Chapter Page

I. Preliminary Remarks

1. General Equilibrium Approaches

In comparative statics analysis, the pitfalls of a partial equilibrium approach are well known, and usually avoided. In intermediate macro theory, for instance, we know that in a world of commodities, money and bonds, it is inadequate to look at only one of these markets but sufficient to look at any two, and thus that, if used properly, Hicksian IS-LM and Patinkin Commodity-Bond analyses are equivalent alternative techniques.

However, the realization that we should look at all but one of the markets, with conditions in the last market implicit in the conditions in the other markets, should not mislead us into forgetting about that last market. Rather, the very interdependence which allows us to look at only two markets, also means that anything we say about any one market is also saying something about the other two markets, considered together. And anything we say about two markets is also a very specific statement about the last market.

If, for example we hypothesize that interest rates do not affect the demand or supply of commodities or money, then we've also proposed that interest rates do not affect the bond market. Similarly, if we assume that expected price changes matter in the commodity market but not in the money market, then such expectations must work in the bond market so as to exactly offset the net effect in the commodity market. (i.e., if an increased anticipation does not affect money holdings but results in $\dot{x}$ dollars of aggregate dissaving, then all $\dot{x}$ of these dollars must come out of bond holdings.)

In a dynamic macro model, similar rules apply. For example, in an IS-LM curve framework, one might be tempted to theorize that in a disequilibrium situation interest rates will rise or fall as there is an excess demand or supply of money, and prices will rise or fall as there is an excess demand or supply of goods.[1] However, an excess demand or supply cannot exist in one market without at least one other market also being in disequilibrium, and we would normally expect it to make a difference here which of the other markets were in disequilibrium and to what extent. For example, an excess demand for money can coexist with either an excess demand for or supply of bonds (depending on conditions in the commodity market), and we would usually think that which was the case would make a difference at least for the movement of interest rates.

Exactly analagous cautions should govern the construc-

1. Patinkin [11] wrote that, "For the most part, this assumption is uncritically taken for granted in the contemporary theory of economic dynamics," admitted its "inappropriateness", and then went on to make the strange assumption that the interrelationships "are never strong enough to change the directions of [adjustment implied by the inappropriate analysis]."

tion of a static or dynamic model of the financial sector. In considering any group's behavior, we should look at all but one market, but not forget the residual market, and should remember that funds going into (or leaving) one market must have come from (or be going into) at least one other market. If we instead were to look at only a few selected sectors or markets, then we would be likely to inadvertently lump together in the residual sector or market items which don't belong together, and to thereby assign to these residual items implicit behavior that is objectionable and would never have been explicitly postulated.

In their "pitfalls" paper [13], professors Brainard and Tobin advocate general equilibrium and disequilibrium approaches to financial model building. The heart of their proposal requires little more than an explicit recognition of the interrelatedness of markets and an alertness to avoiding theoretical restrictions which are unintended or inconsistent. Among the choices still to be made by each model builder are the degree of aggregation, the list of explanatory variables, and the functional form of the behavioral equations.

This has also been called a "general portfolio approach", and seems to have become the dominant view of the financial sector. For example,[2] Mann [10] wrote that

> There seems now to be little quarrel that monetary policy works through the supply schedules and prices of financial assets, thereby influencing the real sector of the economy. While there are various interpretations of the transmission process, most are concerned with the effects of portfolio adjustments in response to disequilibria induced by monetary policy.

2. Another summary article with this conclusion is W. Smith [12].

2. Some Econometric Problems

Brainard and Tobin's general disequilibrium approach and accompanying advice to model builders is not especially controversial but has not very often been followed. A glance at the more widely known econometric models reveals that the prevalent financial sectors consist of only a few mildly related quasi-reduced form equations.

There are some models which are more than this characterization, but mostly in that they contain a large number of financial equations. In the FRB-MIT model [5], for example, there is more reliance on an interest rate structure than on asset demand equations; those asset demand functions which do appear represent partial adjustment mechanisms, assume many different functional forms (linear, log linear, quadratic), and display a great variation in arguments, with the result that each residual asset is a combination of scarcely related items which implicitly depends on a long list of variables in a very awkward functional way. Most of the rate equations are "simply empirical relationships", instead of theoretically inspired descriptions of non-market determined rates or the bringing of perfect substitutability to the market place.

The Brookings model [4] is closer to a general portfolio approach, especially in its theoretical justification. De Leeuw here sticks to a general functional form, and uses twelve demand relations and only three rate equations; further, two of the rate equations are explicitly (and reasonably) justified as being not market determined, and the third rate has an accompanying derivation which makes clear the assumptions that it relies on. On the other hand, the residual asset categories are too large and heterogeneous; the explanatory variables vary too much across equations; lagged rather than current wealth is used; and partial adjustment mechanisms are used.

The Goldfeld model [6] has the virtue that balance sheet identities are clear, with residual assets easily identified and sensical. The explanatory variables again vary too much with assets, however, as illustrated by the fact that the banking sector's residual asset (noncommercial loans and misc.) implicitly depends on 45 explanatory variables, many of which have implied coefficients that are at least unusual. The adjustment mechanism is assumed to be a partial equilibrium one, although it is stated in one footnote that lagged stocks of all portfolio items should have been included but were left out for econometric reasons. Goldfeld also uses an interest rate structure, but admits that this is because a general equilibrium approach is beyond the scope of his study.

Among the justifications given for the use of these and much more simplified financial models are: the financial sector doesn't matter much; monetary events are influenced too much by non-quantifiables; the financial markets are too protean; and monetary variables move so closely together that we cannot break down their separate influences and/or we need not look at more than a few monetary variables. Some of these reasons are clearly theoretical propositions which can only be properly tested by the construction of a large financial model, since theorists are splintered and simple correlations, quasi-reduced forms, and not very rigorous data mining are not sufficient evidence.

Standing in the way of such a test however is the fear that the financial markets are too complicated and the many associated parameters impossible to estimate because of the small effective dimensionality of the data. Faced with this problem, economic model builders are sorely tempted to limit the complexity of their models to suit the apparently small amount of information contained in the data. This is an acceptable practice if carefully done,[3] but it unfortunately seldom apparent that a simple model has been arrived at by means of a systematic, theoretically justified pruning. More often, it seems that those financial variables which are found useful in explaining expenditure items are

3. Harman [8] wrote that, "It is a commonly accepted scientific principle that a theoretical law should be simpler than the observed data upon which it is based."

simply made to depend on enough variables to be satisfactorily explained themselves. In other cases, explanatory variables have been included on theoretical grounds and then discarded because their explanatory power was redundant. One major task of this thesis is therefore to consider methods of measuring the significant degrees of freedom of a data set, and rigorously set forth some possible ways of reducing the number of parameters to be estimated in a consistent explicit way, keeping in mind neglected assets, implicit assumptions, and a priori information.

For the much-discussed "autocorrelation problem", there are a goodly number of well-defined remedies at hand but it is normally unclear which remedy should be employed. Similarly, we are offered instrumental variables as a means of handling simultaneity, but the choice of instruments is uncertain; and while instrumental variables (in contrast to ordinary least squares) yield consistent estimates, the variances may be larger.

Chapter III considers these various problems and estimators, in anticipation of applying a selected group of them to a quarterly econometric model of the U.S. financial sector constructed in Chapter II along the lines advocated by Brainard and Tobin.[4]

4. The appropriate data has recently been made available by the Federal Reserve Board's construction [2] of a consistent set of quarterly flow of funds accounts for the period from 1952 to 1970.

3. Forecasting Tests

3.1 The Case for Forecasting Tests

A comparison of estimation techniques or models based on sample period fits is likely to be inconclusive and may be a misleading guide to relative forecasting powers. Since forecasting (either in actual or postulated situations) is normally the major goal of model building, the out-of-sample fit is usually a more objective and interesting test of a model's success.

Many people have been aware of the divergence between in-sample and out-of-sample performance,[5] but only very recently have rigorous out-of-sample comparisons become popular. Since many of these recent comparisons have been primarily interested in choosing the best model or have otherwise glossed over the point, I have stressed below in a few examples the fact that to date in-sample fit has been a poor guide to out-of-sample powers.

One conclusion to be drawn is that we cannot rely solely on sample period statistics to judge models if we're interested in forecasts or if we think that substantial differences between in-period and out-of-period

---

5. In his presidential address to the Royal Economic Society [14], Alec Cairncross said that, "Anyone who has watched econometrics applied to short-term forecasting knows that a wide selection of equations can be generated with highly impressive R^2's from the same set of past observations, and that there tends to be a distressing scatter in the predictions yielded by these same equations when they are projected into the future."

performance suggest that we may have been misled by the estimation techniques and data. In Chapter IV of this thesis, forecasting results are reported for the various estimators of the model of Chapter II.

3.2 Tests of the Wharton Model

In a paper presented at the November 1969 Conference on Research in Income and Wealth, Evans, Haitovsky, and Treyz include calculations of ex post in-sample and out-of-sample forecasts with the Wharton model.

Results are reported for three different mechanical rules for adjusting the constant term in each period in each predictive equation as a means of offsetting the effects of first order autocorrelated residuals. These rules can be summarized as follows: let $\hat{\rho}_k$ be the estimate of ρ_k in

$$r_{kt} = \rho_k r_{kt-1} + U_{kT}$$

where r_{kT} = K^{Th} equation residual in period t. Let A_{it} = adjustment of i^{th} equation for period t forecast. Then the three rules are

1. $A_{it} = 0$
2. $A_{i(t+T)} = \hat{\rho}_i^T e_{it}$
3. $A_{i(t+T)} = \hat{\rho}_i^T \left(\frac{e_{it} + \hat{\rho}_i e_{it-1}}{2}\right)$

The sample period forecasts spanned the 48 quarters 1953I-1964IV. The out-of-sample period forecasts covered the 16 quarters 1965I-1968IV. Forecasts were made for one through six quarters ahead, whenever possible; for example, using 1964I data, in-sample forecasts were made for one, two, and three quarters ahead. Of the various resultant statistics reported one of the most useful is the Theil U coefficient,

$$U = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_i - O_i)^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} O_i^2}}$$

where P_i and O_i are the predicted and observed changes. The U coefficient is the root mean square forecast error divided by the root mean square of the observed changes - which is the root mean square error of a naive "no change" forecasting technique. This coefficient thus incorporates a normalization which makes comparisons of forecasts of different variables more satisfactory, and also implies a rough standard: $U > 1$ implies a technique worse than a very naive method. However, this standard is very rough indeed and the intended normalization is more appropriate to relatively random variables (such as inventory investment and unemployment) than to trend dominated variables. For these latter variables the "constant changes" naive model would be a better normalizer.

In any case, results are reported for five variables and I've reproduced below the U coefficients for one quarter forecasts, in order to give some flavor of the results.

Variable	Adjustment Method	U Coefficient In-Sample Period (U_s)	Out-of-Sample Period (U_f)
GNP	1	.83	1.37
	2	.84	.51
	3	.49	.73
C	1	.87	1.52
	2	.92	.59
	3	.37	.84
I	1	.92	2.28
	2	.78	.78
	3	.33	1.36
Real GNP	1	.98	1.90
	2	1.06	.75
	3	.17	1.01
Unemploy.	1	2.77	6.36
	2	2.29	5.69
	3	2.71	4.85

A glance at the data suggests that the sample period U-coefficients (U_s) were a poor guide to the out-of-sample U coefficients (U_f); the latter seem generally larger than the former, and the correlation between values seems slight.

The regression of U_f on U_s is an obvious test and pseudo-quantification of these observations:

$$U_f = -.319 + \underset{(8.29)}{2.16U_s} \ , \ \bar{R}^2 = .841$$

which would seem to contradict our casual empiricism, in that the usual measure of association - the correlation coefficient - is impressively large. A slightly deeper examination reveals however that the problem is more with the test than with our original feelings.

A plot of the data (Figure 1) shows that:

(1) The data does lie preponderantly above the line of $U_f = U_s$, which is reinforced by the supplementary calculation

FIGURE 1

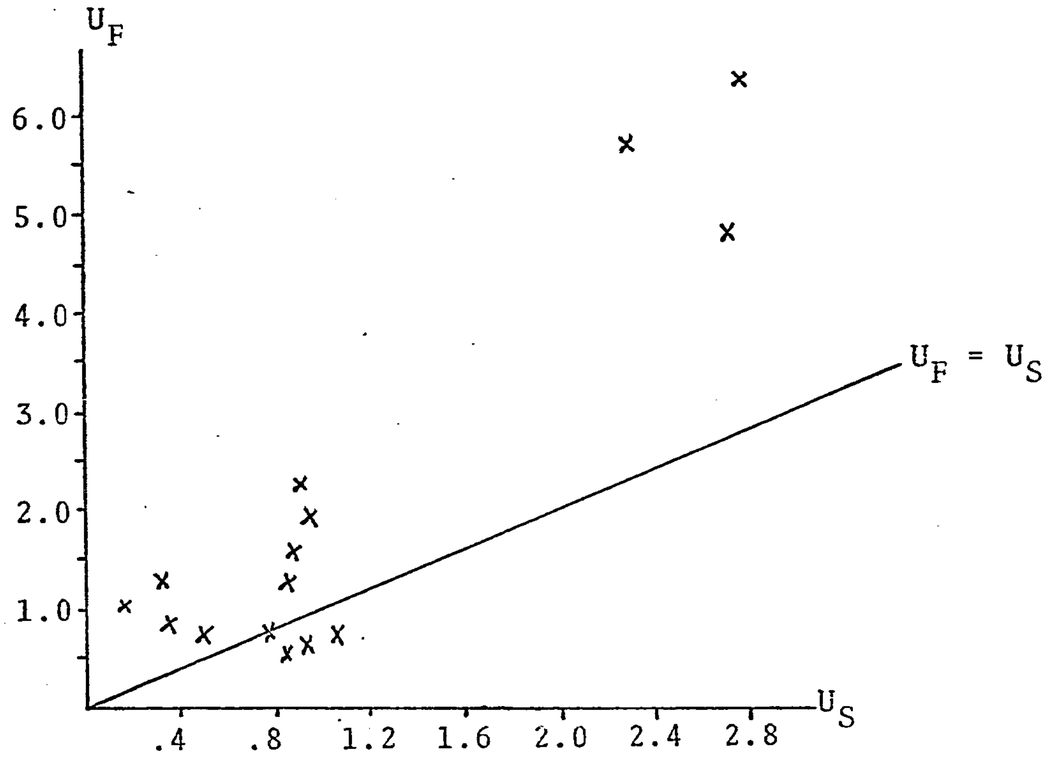
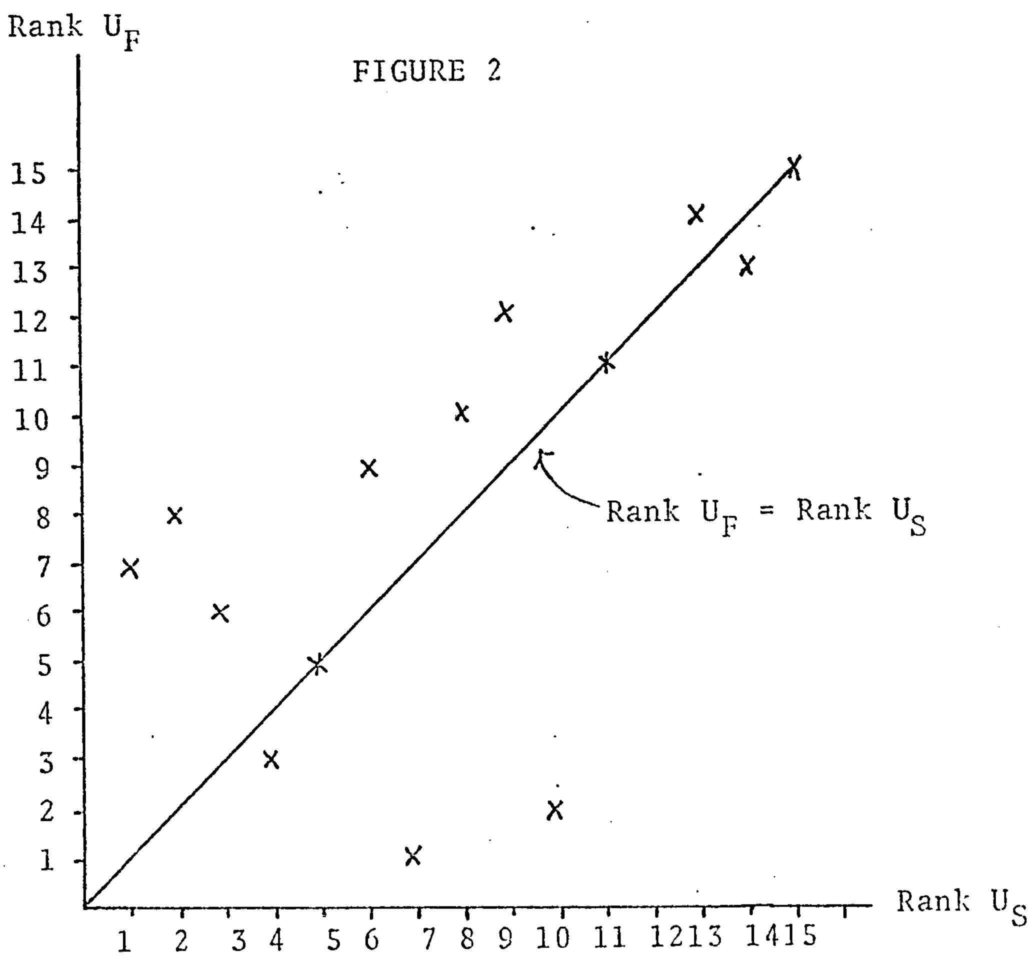


FIGURE 2



that $\bar{U}_f = 2.036$ while $\bar{U}_s = 1.089$;

(2) The fit is indeed sparse. For the 12 observations with $U_s < 1.1$, no correlation is apparent. For the seven observations with $U_f < 1.1$, there appears to be a fairly good perverse correlation: A larger U_s is associated with a lower U_f; and

(3) The size of $\bar{R}^2$ was probably substantially enhanced by the boost to the dispersion of U_f around $\bar{U}_f$ resulting from the three $U_F > 4.5$.

Points (2) and (3) are partly confirmed by the elimination from the data set of the three very large observations associated with predicting unemployment: without these observations, $\bar{R}^2$ falls to .039.

Further confirmation comes from the consideration of the ordering powers of U_s. For each of the five economic variables there are three possible pairwise comparisons: method A versus B, A versus C, and B versus C. Of the fifteen pairs in all, the method that has the lower U_s has the lower U_f too only seven times and the higher U_f eight times. Stated differently, if with each variable we were to choose between two prediction methods on the basis of sample period performance, we would be misled eight times out of fifteen.

This suggests repeating our original regressions of U_f on U_s, but using ranks instead of absolute size. This should eliminate the disproportionate weight of the three very large observations, and give a more realistic $\bar{R}^2$. I, consequently calculated the Spear and Kendall coefficients of rank correlation; the former is the R that would emerge from the suggested regression, but the distribution of the second is more precisely known.[6] Spearman's ρ here is .532 so that the unadjusted R^2 is .282. This is much smaller than when absolute values were compared, but is still significant at the 5% level, though not at the 1% level. Kendall's τ is .391, which is just significant at the 5% level.

A glance at the scatter diagram of ordered values (fig. 2) reveals the fit to still appear very random, with the exception of the three ranks (13, 14, 15) associated with unemployment. Since we will hopefully not have to consider choosing a model with such high values of U_s and U_f,[7] the more appropriate comparison is probably between the 12 observations which are associated with values of U_s close to or less than 1. For these observations, Spearman's ρ is .091 and Kendall's τ is .061, neither

6. A good, simple treatment of these statistics is still M.G. Kendall [9].

7. Put differently, we would not bother to choose between models which do so poorly in relation to a naive model.

of which is significant.[8]

Besides cautioning against an indiscriminate use of R^2, the apparent conclusions regarding this data are

(1) On average, U_f was twice as large as U_s;

(2) The three very large values of U_s were the forerunners of even larger values of U_f, but for those values of U_s in which we might be interested, U_s was both quantitatively and by rank of minimal assistance in predicting U_f.

3.3. Bischoff's Tests of Investment Equations

Using the in-sample period 1953-68 and the out-of-sample period 1969-70, Bischoff [1] calculated these rankings for seven investment models:

Models	Equipment		Construction	
	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample
FMP	1	4	1	5
Cash flow	2	6	2	6
Wharton	3	2	5	7
Accelerator	4	3	7	2
Market Value	5	1	4	1
Neoclassical	6	7	6	3
DHL	7	5	3	4

8. Part of the reason for these low values may be that we are splitting hairs in asking to differentiate between values of U_s and U_f which differ only on the order of 10^{-2}. However, the reduction of U_s and U_f to one significant place and the permitting of ties has only a slight effect on the coefficients: $\rho_{15}=.593$, $\rho_{12}=.200$, $\tau_{15}=.400$, $\tau_{12}=.049$ where the subscript indicates the number of observations. To deny the accuracy of U_s to even fewer significant places is to deny it any power in the crucial range $0 \leq U_s \leq 1$.

The rank coefficients for the equipment equations are $\rho=.14$ and $\tau=.24$; for the construction equations $\rho=-.43$ and $\tau=-.34$. None of these are significant at the 5% level.

3.4. Cooper's Tests of Quarterly Models

Cooper [3] tested 7 quarterly models on the in-sample period 1949-I to 1960-IV and the out-of-sample period 1961-I to 1965-IV. His rankings of some of the more important nonfinancial variables are reproduced below, along with the rank correlation coefficients.

	Real GNP		Current GNP		Employ-ment		Un-employ-ment		Gross Output		Real Plant And Equip. Expenditure		Current Pla And Equip. Exp.	
	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out
Naive	2	2	3	1	1	1	2	1	1	1	3	5	3	2
Friend-Taubman	7	7	5	7							5	4	5	4
Fromm	1	1	1	3			1	2			2	1	2	1
Liu	3	4	2	6							1	2	1	3
Klein	5	6	8	2	2	3	4	5	4	2	6	6	6	6
OBE	4	5	4	5	4	4	5	4	2	4	4	3	4	5
Wharton-EFU	6	8	7	8	3	2	3	3	3	3	7	7	7	7
Goldfeld	8	3	6	4										
ρ	.619		.165		.800		.800		.200		.857		.857	
τ	.572		.142		.667		.600		.000		.714		.714	

The correlation coefficients are insignificant except for plant and equipment expenditures and τ for real GNP (all of which are significant at the 5% level though not at the 1% level).

Because most of the models tested had primitive financial sectors, there were only three financial variables which were endogenous to more than one model. For these three variables, the rankings were:

	4-6 Month Prime Commercial Paper Rate		Moody's AAA Corporate Bond Yield		Current Total Consumer Liquid Assets	
	In	Out	In	Out	In	Out
Naive	4	1	2	4	1	1
Liu	1	4	1	3	2	2
Wharton-EFU	3	3	3	1		
Klein	{2	2}	{4	2}	3	4
OBE					4	3
ρ	-.800		-.600		.800	
τ	-.667		-.333		.333	

None of the correlation coefficients are significant at the 5% level.

References

1. Bischoff, C., "Investment Demand in 1971 and 1972: A Comparison Alternative Models," (forthcoming Brookings Paper).

2. Board of Governors of the Federal Reserve System, Flow of Funds Accounts 1945-1968, Washington: Board of Governors, 1970.

3. Cooper, R., "The Predictive Performance of Quarterly Econometric Models of the United States," Working Paper No. 113, University of California, Berkeley, November 1969.

4. DeLeeuw, F., "A Model of Financial Behavior," in The Brookings Quarterly Econometric Model of the United States, J. Duesenberry, G. Fromm, L. Klein, and E. Kuh, eds., Chicago: Rand McNally, 1965, 465-532.

5. DeLeeuw, F., and E. Gramlich, "The Federal Reserve-MIT Econometric Model," Federal Reserve Bulletin, LIV (January 1968), 11-40.

6. Goldfeld, S., Commercial Bank Behavior and Economic Activity, Amsterdam: North-Holland, 1966.

7. Gramlich, E., and J. Kalchbrenner, "A Constrained Estimation Approach to the Demand for Liquid Assets," Federal Reserve System Committee on Financial Analysis, (mimeographed) 1969.

8. Harman, H., Modern Factor Analysis, 2nd edition, Chicago: University of Chicago Press, 1967.

9. Kendall, M., Rank Correlation Methods, 2nd edition, New York: Hafner, 1955.

10. Mann, M., "How Does Monetary Policy Affect the Economy?," Journal of Money, Credit, and Banking, I (August 1969), 538-48.

11. Patinkin, D., Money, Interest, and Prices, 2nd edition, New York: Harper and Row, 1965.

12. Smith, W., "On Some Current Issues in Monetary Economics: An Interpretation," The Journal of Economic Literature, VIII (September 1970), 767-82.

13. Brainard, W., and J. Tobin, "Pitfalls in Financial Model Building," American Economic Review, Papers and Proceedings, LVIII (May 1968), 99-122.

14. Cairncross, A., "Economic Forecasting," The Economic Journal, LXXIX (December 1969), 797-818.

II. THE MODEL

A complete economic model would simultaneously explain all of the economic transactions made by every sector of the economy. This is but a model of financial flows which relies on the very useful simplifying assumption that we are able in each sector to concentrate on a group of financial decisions which are made separately from other decisions. While therefore neither complete nor thoroughly interdependent, this model is less myopic than the usual examination of a few selected assets and rates.

Section 1 describes the general behavioral structure that is assumed, and Section 2 presents a specific model which is to be used to explain and forecast quarterly U.S. financial data.

1. A General Financial Model

1.1. Sector Demand Equations

Each sector's financial assets and liabilities (including net worth) can be grouped into s items, which are subject to the balance sheet constraint

$$\sum_{i=1}^{s} X_i = 0$$

These X_i can then be divided into two groups, on the basis of whether or not the sector exercises direct control over its holdings of each item. I've labeled the directly adjusted items "assets," the remaining items "liabilities," and the sum of either "wealth".

Thus when the X_i are appropriately rearranged,

$$[1] \qquad W = \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} X_i = \sum_{i=n+1}^{s} - X_i = \sum_{i=1}^{m} \ell_i = W$$

where a_i = i^{th} asset

ℓ_i = i^{th} liability

W = sectoral wealth

Behaviorally, it is assumed that a sector's desired holdings of each asset are determined by the m sectoral liabilities, the rates of return on all n assets, and q other variables such as sectoral income or time. That is,

$$a_i^* = f_i(\ell, r, s)$$

where a_i^* = desired holding of a_i

ℓ = vector of m sectoral liabilities

r = vector of nonconstant rates of return on $n-z$ assets

s = vector of q other explanatory variables

It is also assumed that these functions can be written as

$$a_i^* = \sum_{j=1}^{m} f_i^j(r, s)\ell_j$$

or

[2a] $$\frac{a_i^*}{W} = \sum_{j=1}^{m} f_i^j(r, s) \frac{\ell_j}{W}$$

so that a rise in sectoral wealth with liability shares constant will induce a proportionate rise in all desired asset holdings.

In some sectors, it is assumed that liability shares do not have differential effects on desired asset holdings, or there may be only one liability. In these cases, we have

[2b] $$\frac{a_i^*}{W} = f_i(r, s) \frac{\Sigma\ell_j}{W} = f_i(r, s)$$

In order to proceed further, some assumptions must be made about the functional form of the f_i^j or f_i. As is often done,[1] it is assumed here that these

---

[1]Motley [4] has a similar model of the household sector, but assumes that desired holdings are loglinear functions. He does not use the assumption $\Sigma a_i^* = W$, possibly because there are no nonzero parameters which will enforce this constraint on loglinear functions. He does assume linear adjustment equations (similar to my equation [7]) which permit the constraint $\Sigma a_i = W$, but he does not impose this constraint either.

functions are linear (have constant partial derivatives) over the range of variation in the independent variables that will be encountered. Equations [2a] and [2b] can then be written as

$$[3a] \qquad \frac{a_i^*}{W} = \sum_{j=1}^{m} \left(\sum_{h=1}^{n-z} \alpha_{ih}^{j} r_h + \sum_{h=1}^{q} \beta_{ih}^{j} s_h \right) \frac{\ell_j}{W}$$

and

$$[3b] \qquad \frac{a_i^*}{W} = \sum_{h=1}^{n-z} \alpha_{ih} r_h + \sum_{h=1}^{q} \beta_{ih} s_h$$

which can be condensed by using matrix notation:

$$[4a] \qquad \underset{n\times 1}{a^*/W} = \sum_{j=1}^{m} \left(\underset{n\times n-z}{A_j} \cdot \underset{n-z\times 1}{r} + \underset{n\times q}{B_j} \cdot \underset{q\times 1}{s} \right) \frac{\ell_j}{W}$$

$$[4b] \qquad \underset{n\times 1}{a^*/W} = \underset{n\times n-z}{A} \cdot \underset{n-z\times 1}{r} + \underset{n\times q}{B} \quad \underset{q\times 1}{s}$$

a^* , r , and s are now column vectors; the A_j and B_j are coefficient matrices (with the α_{ih}^{j} and β_{ih}^{j} as elements); and W and ℓ_j are scalars.

[4b] is of course still a special case of [4a], but I have kept separate track of it since it is an attractive alternative to the cumbersome and many-parametered [4a].

For both of these relationships, it is assumed that

[5] $$\sum_{i=1}^{n} a_i^* = W ,$$

which imposes a variety of restrictions on the parameters.

In particular, it can be shown[2] that

$$\sum_{i=1}^{n} \alpha_{ih}^{j} = 0 \qquad \forall h, j$$

$$\sum_{i=1}^{n} \beta_{ih}^{j} = 0 \qquad h \neq 1, \forall j$$

$$\sum_{i=1}^{n} \beta_{i1}^{j} = 1 \qquad \forall j$$

where β_{i1}^{j} are the constant terms in each equation.[3] For [3b], these reduce to

$$\sum_{i=1}^{n} \alpha_{ih} = 0 \qquad \forall h$$

$$\sum_{i=1}^{n} \beta_{ih} = 0 \qquad h \neq 1$$

$$\sum_{i=1}^{n} \beta_{i1} = 1$$

---

[2] See Appendix II-A.

[3] And there must be constant terms in each equation; see Appendix II-A.

Verbally, these restrictions are that the constant terms (coefficients of $s_1 = 1$) must sum across equations to one, while the coefficients of all other variables must sum across equations to zero. In the matrices A_j and B_j , the elements in the first column of each B_j must sum to 1, while the elements in any other column must sum to zero.

This is because if we are careful to put all changes in wealth (including capital gains) into ΔW , then a change in an independent variable can only induce a shuffling of asset proportions. That is, if it is desired to hold a larger proportion of one's wealth in the form of some asset, then the net demand for all other asset proportions must fall by an equal amount. This also means that knowing n-1 of the asset demands will give us the demand for the n^{th} asset:

$$\frac{a_n^*}{W} = 1 - \sum_{i=1}^{n-1} \frac{a_i^*}{W}$$

The individual parameters of B_j will of course depend on what the variables in s are. If the asset proportions are gross substitutes, then the matrix $\sum_j A_j \ell_j / W$ (or the matrix A in the case of [4b]) will have positive diagonal elements and elsewhere negative elements.

Gramlich and Kalchbrenner [3] use a quasi-quadratic utility function to derive linear demand functions that display the property of equal cross elasticities

$$\frac{\delta a_i}{\delta r_j} = \frac{\delta a_j}{\delta r_i}$$

(which they find econometrically to be very useful). They attribute this to the inclusion of W having eliminated all income effects. However, it can be easily shown that there are still income effects in their model, but that their utility function has constrained them to be equal--which is not the case with a genuine quadratic utility function.

The simpler system [4b] can be depicted in a Euclidean n-space in which each axis records the desired asset proportion holdings of one of the n assets. The wealth constraint would appear as an n-1 dimensional hyperplane (described by $1 = \sum_{i=1}^{n} a_i^*/W$) with corners at 1 unit out each axis.

The vector $(\beta_{11}, \beta_{21}, \ldots, \beta_{n1})$ of the constants specifies the relative holdings of all assets when all independent variables are zero. Since the endpoint of this vector must lie in the wealth hyperplane, we have again

$$1 = \sum_{i=1}^{n} \beta_{i1} .$$

The effect of a change in an independent variable, say r_k , will be given by a move dr_k units along a vector

$(\alpha_{1k}, \alpha_{2k}, \ldots, \alpha_{nk})$ --i.e. the vector $(\alpha_{1k}, \alpha_{2k}, \ldots, \alpha_{nk})$ dr_k is added to $(\beta_{11}, \beta_{21}, \ldots, \beta_{n1})$. For this move to leave one on the wealth plane, any vector perpendicular to the vector $V_k = (\alpha_{1k}, \alpha_{2k}, \ldots, \alpha_{nk})$ must also be perpendicular to the wealth plane. The unit vector perpendicular to the wealth plane is $U = (1, 1, \ldots, 1)$ so that if θ is the angle between U and V_k , then from vector theory

$$U \cdot V_i = |U||V_k|\cos\theta = |U||V_k|\cos 90^\circ = 0$$

therefore

$$0 = U \cdot V_k = 1 \cdot \alpha_{1k} + 1 \cdot \alpha_{2k} + \ldots + 1 \cdot \alpha_{nk} = \sum_{i=1}^{n} \alpha_{ik}$$

which is the balance sheet constraint we presented earlier and algebraically derived in Appendix II-A.

A sector is not always at its desired asset position, but is presumably always adjusting toward its desired portfolio. The adjustment in the holdings of any asset is assumed to depend linearly on all of the disequilibria in the portfolio and also on all changes in liabilities, to reflect the fact that the origin of a change in wealth is likely to initially influence the form in which it is held. For example, an equities capital gain will probably be initially held almost entirely as increased equities, and then be gradually distributed into other assets according to the discrepancies between desired and actual holdings that are created by the capital gain and that arise from other sources.

One complication which arises here is that the sum of the discrepancies will always equal the sum of the changes in liabilities:

$$[6] \qquad \sum_{j=1}^{n} [a_j^* - a_j(-1)] = W - W(-1) = dW = \sum_{j=1}^{m} d\ell_m$$

so that one of these $m+n$ proposed explanatory variables is unnecessary (and cannot be specified independently of the other variables without, in general, leading to an inconsistency).

The choice of which variable to delete will not affect the model or the econometric results, but will alter the interpretation of the model's coefficients.

In particular, if we write

$$da_i = \sum_{j=1}^{n-1} \varepsilon_{ij}[a_j^* - a_j(-1)] + \sum_{j=1}^{m} f_{ij}[d\ell_j]$$

with the n^{th} discrepancy omitted, then a dollar increase in $a_j^* - a_j(-1)$ with other explanatory variables constant must be accompanied by a dollar decrease in $a_n^* - a_n(-1)$, in order to maintain the identity [6]. ε_{ij} should therefore be interpreted as the partial effect on holdings of asset i for desiring to hold \$1 more of asset j and \$1 less of asset n. Since the overall effect on the portfolio of such a disequilibrium can only be a shuffling of asset holdings, the ε_{ij} summed across equations should be zero. Similarly, f_{ij} should be interpreted as the partial effect on holdings of asset i of a $1 increase in liability j accompanied by a $1 increase in desired holdings of asset n. The net effect of such a change across all assets $(\sum_{i=1}^{n} f_{ij})$ will clearly be $1. In Appendix II-B; it is explicitly shown that

$$\sum_{i=1}^{n} \varepsilon_{ij} = 0 \qquad \forall j$$

$$\sum_{i=1}^{n} f_{ij} = 1 \qquad \forall j$$

If we instead delete the m^{th} liability:

$$da_i = \sum_{j=1}^{n} \varepsilon_{ij}[a_j^* - a_j(-1)] + \sum_{j=1}^{m-1} f_{ij}[d\ell_j]$$

then ε_{ij} should be viewed as the partial effect on holdings of asset i of a $1 increase in $[a_j^* - a_j(-1)]$ accompanied by an equal rise in liability m, and f_{ij} should be interpreted as the marginal effect of a $1 gain in liability j offset by a $1 fall in liability m. Appendix II-C explicitly demonstrates the intuitively justifiable:

$$\sum_{i=1}^{n} \varepsilon_{ij} = 1 \qquad \forall j$$

$$\sum_{i=1}^{n} f_{ij} = 0 \qquad \forall j$$

I've arbitrarily decided to consistently delete one of the liability changes since this allows me to retain my notation a_j^* , a_{-1} , A , and B and because I think that it's slightly easier to interpret the coefficients in this framework.

Writing the adjustment equations in matrix form:

$$[7] \qquad \underset{n\times 1}{\Delta a} = \underset{n\times n}{E} \cdot (\underset{n\times 1}{a^*} - a_{-1}) + \underset{n\times m-1}{F} \cdot \underset{m-1\times 1}{\Delta\ \ell} ,$$

and substituting [4a] or [4b]

$$[8a] \quad \frac{\Delta a}{W} = [E\cdot A_1 r + E\cdot B_1 s]\ \frac{\ell_1}{W} + [E\cdot A_2 r + E\cdot B_2 s]\ \frac{\ell_2}{W}$$

$$+ \ldots + [E\cdot A_m r + E\cdot B_m s]\ \frac{\ell_m}{W} - \frac{Ea_{-1}}{W} + \frac{F\cdot\Delta\ell}{W}$$

$$[8b] \quad \underset{n\times 1}{\frac{\Delta a}{W}} = \underset{n\times n-z}{E\cdot A} \cdot \underset{n-z\times 1}{r} + \underset{n\times q}{E} \cdot B \cdot \underset{q\times 1}{s} - \underset{n\times n}{E}\ \underset{n\times 1}{\frac{a_{-1}}{W}} + \underset{n\times m-1 \quad m-1\times 1}{\frac{F\cdot\Delta\ell}{W}}$$

where the scalar W has been divided through to eliminate secular growth in the variables, which is often accompanied by heteroscedastic error terms.

These are the basic sector demand equations of the model, which clearly could have been reached by other theoretical routes. For example, [8b] can be quickly derived with a Koyck transformation from the hypothesis that holdings of an asset are a distributed lag function of $w\cdot r$, $w\cdot s$, a_{-1} , and ℓ with identical geometrically declining weights.

1.2. Real Versus Nominal Units

To this point, we have not explicitly distinguished between nominal and real holdings or rates. For desired holdings

$$a^*/w = \sum_{j=1}^{m} (A_j r + B_j s)\ \ell_j/w$$

it seems appropriate that all variables be in real terms; we can therefore let the existing symbols stand for real amounts. In the adjustment equations, however, a sector should make nominal transactions when nominal desired holdings differ from nominal actual holdings:

$$pa - p_{-1}a_{-1} = E(pa^* - p_{-1}a_{-1}) + F(p\ell - p_{-1}\ell_{-1})$$

where p is whatever price scalar a sector uses to compute its real holdings. Equivalently, a sector will adjust when its current real desired holdings differ from the current real value of actual holdings:

$$a - \frac{p_{-1}a_{-1}}{p} = E\left(a^* - \frac{p_{-1}a_{-1}}{p}\right) + F\left(\ell - \frac{p_{-1}\ell_{-1}}{p}\right)$$

It can be seen that it is not necessary that p be the correctly perceived current price level; but only that the sector evaluate its prevailing portfolio in the same units that it calculates its desired portfolio and its new actual portfolio.

Combining equations, we obtain

$$\frac{pa - p_{-1}a_{-1}}{pw} = \sum_j (EA_j r + EB_j s)\frac{p_j \ell_j}{pw} - \frac{Ep_{-1}a_{-1}}{pw}$$

$$+ F(p\ell - p_{-1}\ell_{-1})/pw$$

so that it is acceptable to use [8a] or [8b] with real rates of return and all holdings in nominal units.

If an asset yields a nominal return r', then its real rate of return is

$$r = r' - \dot{P}$$

where $\dot{P}$ is the rate of inflation. But while the nominal rate of return is usually known exactly, the rate of inflation is known only imprecisely and/or with some lag. This suggests replacing $\dot{P}$ with $\dot{P}^E$, the expected rate of inflation during the time period the interest rate applies to. If we use

$$\dot{P}^E = \sum_{i=0}^{n} \lambda_i \dot{P}_{-i} \quad , \quad \sum_{i=0}^{n} \lambda_i = 1$$

then $\dot{P}^E$ will be affected by the current rate of inflation but not be identical to it.

The rate part of a typical demand equation is

$$\sum_i \alpha_i r_i = \sum_i \alpha_i (r_i' - \dot{P}^E)$$

$$= -(\sum_i \alpha_i)\dot{P}^E + \sum_{i \neq 1} \alpha_i r_i'$$

where r_1' is zero, the nominal return on currency. If we substitute in

$$\dot{P}^E = \dot{P} - \sum_{i=1}^{n} \lambda_i (\dot{P} - \dot{P}_{-i})$$

then we obtain

$$-(\sum_i \alpha_i)\dot{P} + \sum_{i=1}^{n} \lambda_i (\sum_i \alpha_i)(\dot{P} - \dot{P}_{-i}) + \sum_{i \neq 1} \alpha_i r_i'$$

from which it is clear that unique estimates of all α_i and λ_i can be obtained.

1.3. Noncompetitive Markets

So far we have considered cases in which rates adjust to clear markets.[4] In some markets, it may be thought more realistic to have rates which are, for institutional or other reasons, set at non-market clearing levels. By definition, in such markets at least one participant must accept more (or less) of the asset than he would like. Thus, in this type of market we need descriptions of how the rate is set, how much excess demand is created, and how that excess is divided among the sectors.

[4]The rates adjust to temporarily equilibrate the markets. These will not be dynamic equilibria (in the sense that, ceteris paribus, the endogenous variables will not move as time passes), since at any moment in time there will generally be some sector which is not at its long run desired position.

A very simple model could consist of a rate setting mechanism, one sector in which the asset is an exogenous "liability", and normal asset demand functions for all other sectors. A more complicated scenario would allow all sectors to include the asset in their calculation of desired portfolios, and would allow all sectors to move toward their desired portfolios with some sectors accepting the surplus created by such moves. These possibilities are illustrated in the following model.

Consider a 2-sector, 2-asset world. Sector 1's holdings of assets a and b are described by

$$\Delta \begin{pmatrix} a_1 \\ b_1 \end{pmatrix} = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} \\ \varepsilon_{21} & \varepsilon_{22} \end{pmatrix} \begin{pmatrix} a_1^* - a_1(-1) \\ b_1^* - b_1(-1) \end{pmatrix} + \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} S_b$$

sector 2's holdings are

$$\Delta \begin{pmatrix} a_2 \\ b_2 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} a_2^* - a_2(-1) \\ b_2^* - b_2(-1) \end{pmatrix} + \begin{pmatrix} \emptyset_1 \\ \emptyset_2 \end{pmatrix} S_b$$

The new variable, S_b , is the surplus in market b that arises from the sectors' adjustments toward their desired holdings of asset b at prevailing interest rates. That is,

$$S_b = [\varepsilon_{21}(a_1^* - a_1(-1)) + \varepsilon_{22}(b_1^* - b_1(-1))]$$

$$+ [h_{21}(a_2^* - a_2(-1)) + h_{22}(b_2^* - b_2(-1))]$$

or

$$S_b = (\Delta b_1 - \theta_2 S_b) + (\Delta b_2 - \phi_2 S_b)$$

$$= -(\theta_2 + \phi_2) S_b$$

since <u>ex post</u> $\Delta b_1 + \Delta b_2 = 0$. Therefore

$$\theta_2 + \phi_2 = -1$$

which is to say that the surplus must be absorbed by b_1 and b_2 . Also, it can easily be shown that

$$\theta_1 + \theta_2 = \phi_1 + \phi_2 = 0 \ ,$$

which says that for a given change in wealth, an absorption of more asset b must be at the expense of asset a .

In the case where sector 1 absorbs all of the market surplus,

$$\phi_2 = 0 \Longrightarrow \begin{array}{l} \phi_1 = 0 \\ \theta_1 = 1 \\ \theta_2 = -1 \end{array}$$

Further, since sector 2's holdings are

$$\Delta \begin{pmatrix} a_2 \\ b_2 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} a_2^* - a_2(-1) \\ b_2^* - b_2(-1) \end{pmatrix} \quad ,$$

we have

$$S_b = [\varepsilon_{21}(a_1^* - a_1(-1)) + \varepsilon_{22}(b_1^* - b_1(-1))] + \Delta b_2$$

which, substituted into sector 1's equations, yields

$$\Delta \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} \varepsilon_{11} + \theta_1\varepsilon_{21} & \varepsilon_{12} + \theta_1\varepsilon_{22} \\ \varepsilon_{21} + \theta_2\varepsilon_{21} & \varepsilon_{22} + \theta_2\varepsilon_{22} \end{bmatrix} \begin{bmatrix} a_1^* - a_1(-1) \\ b_1^* - b_1(-1) \end{bmatrix} + \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \Delta b_2$$

$$= \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_1^* - a_1(-1) \\ b_1^* - b_1(-1) \end{bmatrix} + \begin{pmatrix} 1 \\ -1 \end{pmatrix} \Delta b_2$$

2. A Specific Financial Model

This model has five sectors, which are defined below in terms of the sector classifications used in the Federal Reserve Board's flow of funds accounts:

Public	=	households + noncorporate, nonfinancial business + private nonbank finance not included in Savings Institutions + rest of world
Corporations	=	corporate nonfinancial business
Banks	=	commercial banks
Savings Institutions	=	Savings and Loan Associations + Mutual Savings Banks
Government	=	U.S. government + monetary authorities + federally sponsored credit agencies

The financial assets assumed to be held by each sector are displayed in the following capital accounts balance sheet:

	Public (1)	Corporations (2)	Banks (3)	Savings Institutions (4)	Government (5)	Net Holdings (6)
Currency	C1	C2	RR+NFR	C4	-H	0
Demand Deposits	DD1	DD2	-DD3	DD4	DD5	0
Time Deposits	TD1		-TD3			0
CD's		CD2	-CD3			0
Savings Accounts	SA1			-SA4		0
Short Term Bonds	Short1	Short2	Short3		-Short5	0
Long Term Bonds	Long1	-Long2	Long3	Long4	-Long5	0
Public Loans	-PL1		-PL3			0
Corporate Loans		-CL2	CL3			0
Mortgages	-Mort1	-Mort2	Mort3	Mort4	Mort5	0
Equities	E1					E1
Other	Oth1	Oth2	Oth3	Oth4	Oth5	0
Net Worth	W1+Oth1	W2+Oth2	0	0	-DT	E1

All of the public's financial holdings are assumed to be part of its portfolio of directly adjustable assets. This sector's demand equations are thus

$$\Delta\begin{bmatrix} C1+DD1 \\ TD1 \\ SA1 \\ Short1 \\ Long1 \\ -PL1 \\ -Mort1 \end{bmatrix}\frac{1}{W1} = E_1A_1\begin{bmatrix} \dot{p}^E \\ r_{TD} \\ r_{SA} \\ r_{short} \\ r_{long} \\ r_{PL} \\ r_{Mort} \\ r_E \end{bmatrix} + E_1B_1\begin{bmatrix} 1.0 \\ \frac{Y_d}{W1} \end{bmatrix} - E_1\begin{bmatrix} C1+DD1 \\ TD1 \\ SA1 \\ Short1 \\ Long1 \\ -PL1 \\ -Mort1 \\ E1 \end{bmatrix}_{-1}\frac{1}{W1}$$

$$+ F_1\frac{\Delta E1}{W1} + U_1$$

$$\Delta W1 = S_p + \Delta E1$$

The meaning of each subscripted rate variable should be apparent. The other new variables are

Y_d = disposable income = $Y - T - S_b$

Y = net national product

T = net government taxes

S_b = corporate saving

S_p = household saving

U = vector of disturbance terms

$\Delta W1$ has been disaggregated into household saving and the capital gain on equity, with the former being deleted as a redundant explanatory variable. No attempt is made to distinguish between holdings of currency and demand deposits, as the flow of funds data does not separate the two assets. When we get to commercial banks, however, we will have to explain the economy's total holdings of demand deposits.

The corporate sector is also assumed to directly control its entire financial portfolio, and thus has demand equations

$$\Delta \begin{bmatrix} C2+DD2 \\ CD2 \\ Short2 \\ -Long2 \\ -CL2 \end{bmatrix} \frac{1}{W2} = E_2 A_2 \begin{bmatrix} \dot{p}^E \\ r_{CD} \\ r_{short} \\ r_{long} \\ r_{CL} \\ r_{Mort} \end{bmatrix} + E_2 B_2 \begin{bmatrix} 1.0 \\ q \\ \frac{Y-G}{W2} \end{bmatrix} - E_2 \begin{bmatrix} C2+DD2 \\ CD2 \\ Short2 \\ -Long2 \\ -CL2 \\ -Mort2 \end{bmatrix}_{-1} \frac{1}{W2}$$

$$+ F_2 \frac{\Delta ATL}{W2} + U_2$$

$$-Mort2 = W2 - (C2 + DD2) - CD2 - Short2 + Long2 + CL2$$

$$\Delta W2 = G - T - \Delta E1 - \Delta W1 - \Delta Oth2 = G - T - S_p - \Delta Oth2$$

$$= S_b - I_g - \Delta Oth2$$

where G = government expenditures

I_g = gross nonfinancial investment

ΔATL is the change in accrued tax liability, and has been separated from the rest of the nonfinancial flow to allow for the corporate practice of funding against tax liabilities with short term assets. This remainder is approximately equal [1, p. 30] to retained earnings + depreciation expense + Δ net accounts payable - inventory change (book value) - change in gross noncurrent assets, and is the deleted variable in the demand equations.

The variable q is the market value of \$1 (replacement cost) of capital, and is included because profitability should influence corporate willingness to take on financial debt. We can define the market value of the existing capital stock as

$$V_K = E1 - W2 - Oth2$$

and let

$$q = \frac{V_K}{P_K K} = \frac{E1 - W2 - Oth2}{P_K K}$$

where K is the capital stock and P_K is the replacement cost of a unit of capital.

Commercial banks are assumed to passively accept demand deposits, time deposits, certificates of deposit, and the associated required reserves; these are treated then as "liabilities." Its remaining holdings (net free reserves, short and long term bonds, public and corporate loans, and mortgages) constitute its portfolio of adjustable assets. The markets for public and corporate loans are however assumed to be imperfect, with the banking sector absorbing whatever surplus arises. The demand equations are

$$\Delta\begin{bmatrix}NFR\\ Short3\\ Long3\end{bmatrix}\frac{1}{W3} = \sum_{i=1}^{2}\{E_3A_{3i}\begin{bmatrix}\dot{p}^E\\ r_d\\ r_{short}\\ r_{long}\\ r_{PL}\\ r_{CL}\\ r_{Mort}\end{bmatrix}h_i + E_3B_{3i}h_i\} - E_3\begin{bmatrix}NFR\\ Short3\\ Long3\\ Mort3\\ PL3\\ CL3\end{bmatrix}_{-1}\frac{1}{W3}$$

$$+ f_3\frac{\Delta\ell_1}{W3} + \theta_{PL}\frac{\Delta PL3}{W3} + \theta_{CL}\frac{\Delta CL3}{W3} + U_3$$

$PL3 = PL1$

$CL3 = CL2$

$Mort3 = W3 - NFR - Short3 - Long3 - PL3 - CL3$

$W3 = DD3 + TD3 + CD3 - RR - Oth3$

$DD3 = \alpha(C1 + DD1 + C2 + DD2 + C4 + DD4) + U_{DD}$

$TD3 = TD1$

$CD3 = CD2$

$RR = \lambda_1 k_{DD} DD3 + \lambda_2 k_{TD}(TD3 + CD3)$

$h_1 = (1 - \lambda_1 k_{DD})DD3/W3 = \ell_1/W3$

$h_2 = ((1 - \lambda_2 k_{TD})(TD3 + CD3) - Oth3)/W3 = 1 - h_1$

The k's are reserve requirements and the λ's are member bank proportions of total deposits. Four seasonal dummies were originally included and then dropped, because the moment matrix was too large (31 x 31 as compared to 25 x 25) and too nearly singular for its inverse to be accurately calculated. Aggregate demand deposits are assumed to be a constant proportion of total currency plus demand deposits.

Banks are assumed to adjust towards desired holdings of public and corporate loans by means of rate rather than acceptance changes. The rate adjustment equations are

$$\Delta r_{PL} = \emptyset_{PL} \frac{(PL^* - PL_{-1})}{W3} = \emptyset_{PL} \sum_{i=1}^{2} (A_i^{PL} r + B_i^{PL}) \frac{h_i}{W3} - \emptyset_{PL} \frac{PL3_{-1}}{W3} + U_{PL}$$

$$\Delta r_{CL} = \emptyset_{CL} \frac{CL^* - CL_{-1}}{W3} = \emptyset_{CL} \sum_{i=1}^{2} (A_i^{CL} r + B_i^{CL}) \frac{h_i}{W3} - \emptyset_{CL} \frac{CL3_{-1}}{W3} + U_{CL}$$

where r is the vector of rates included in the other demand equations, and the superscripts on the A_i and B_i indicate the appropriate rows. r_{TD} and r_{CD} are assumed exogenous since during the forecast period they were very close to their legal ceilings.

Savings Institutions are assumed to accept all savings accounts, and directly allocate the remainder of their holdings:

$$\Delta \begin{bmatrix} C4+DD4 \\ Long4 \end{bmatrix} \frac{1}{W4} - E_4 A_4 \begin{bmatrix} \dot{p}^E \\ r_{long} \\ r_{Mort} \end{bmatrix} + E_4 B_4 \begin{bmatrix} S1 \\ S2 \\ S3 \\ S4 \end{bmatrix} - E_4 \begin{bmatrix} C4+DD4 \\ Long4 \\ Mort4 \end{bmatrix}_{-1} \frac{1}{W4} + F_4 \frac{\Delta SA4}{W4} + U_4$$

$$Mort4 = W4 - (C4 - DD4) - Long4$$

$$W4 = SA4 - Oth4$$

$$SA4 = SA1$$

$$Si = \text{seasonal dummy for } i^{th} \text{ quarter}$$

ΔOth4 is the deleted variable in the demand equations. In the forecast period, r_{SA} was generally at the ceiling level set by the Federal Home Loan Bank Board. Since they have only had the power to do this since September 21, 1966, we should consider r_{SA} to be endogenous in our regressions with earlier data. To make the pre-1966 forecasts comparable with the post-1966 predictions, we can assume r_{SA} to always be forecast perfectly.

The remaining balance equations are

$$(1-\alpha)(C1 + DD1 + C2 + DD2 + C4 + DD4) + NFR + RR - H = 0$$

$$Short1 + Short2 + Short3 - Short5 = 0$$

$$Long1 - Long2 + Long3 + Long4 - Long5 = 0$$

$$-Mort1 - Mort2 + Mort3 + Mort4 + Mort5 = 0$$

$$G - T = \Delta Short5 + \Delta Long5 + \Delta H - \Delta Mort5 - \Delta DD5 - \Delta Oth5$$

Overall, there are 20 stochastic relations and 23 identities and market clearing equations which determine 43 endogenous variables. There are 12 policy variables (k_{DD} , k_{TD} , r_{TD} , r_{CD} , r_{SA} , H , DD5 , Short5 , Long5 , Mort5 , Oth5 , T , G) and 13 other exogenous variables (Oth1 , Oth2 , Oth3 , Oth4 , λ_1 , λ_2 ,

Y , S_p , I_g , p_K , K , ΔATL , ATP) .

The sum of the income identity

$$S_b + S_p + T - G = I_G$$

and the two equations

$$\Delta W1 = E1 + S_p$$

$$\Delta W2 = S_b - I_G - \Delta Oth2$$

yields

$$\Delta W1 + \Delta W2 + \Delta Oth2 = \Delta DT + \Delta E1$$

which is also implicit in the market clearing equations and the equations which sum the 5 columns of the capital accounts balance sheet. To eliminate this redundancy, I dropped the first column-summing equation

$$W1 = C1 + DD1 + TD1 + \ldots + E1$$

On the other hand, we are missing an equation which will serve as a supply equation for equities. To meet this need I postulated the following model.

Assume that corporate after tax profits are expected in the short run to be similar to current levels, but in the long run to return to a trend path. Thus, if

ATP_τ = actual corporate after tax profits in year τ

ATP^*_τ = expected corporate after tax profits in year τ

$ATPN_\tau$ = normal corporate after tax profits in year τ

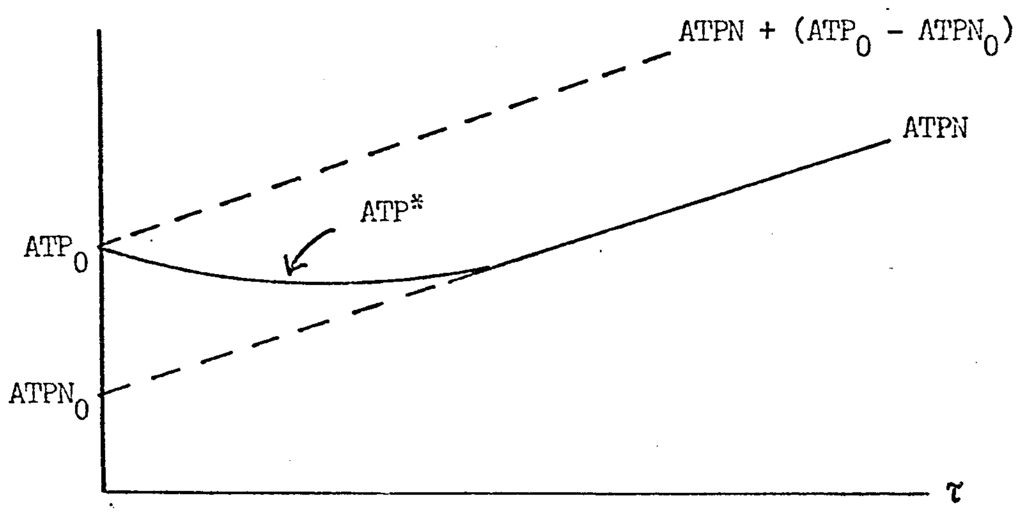
$= ATPN_0(1+g)^\tau$

then we might use

$$ATP^*_\tau = ATPN_\tau + \alpha^\tau(ATP_0 - ATPN_0) \;, \quad 0 \le \alpha \le 1$$

$$= (1+g)^\tau ATPN_0 + \alpha^\tau(ATP_0 - ATPN_0)$$

Graphically, a smoothed representation would be



If El_0 is the present cost of the title to these profits, then the implicit rate of return on that title is the r_E which solves

$$El_0 = \sum_{\tau} \frac{ATP^*_\tau}{(1+r_E)^\tau} = ATPN_0 \sum_{\tau} \left(\frac{1+g}{1+r_E}\right)^\tau + (ATP_0 - ATPN_0) \sum_{\tau} \left(\frac{\alpha}{1+r_E}\right)^\tau$$

For $g < r$ and $\alpha < 1+r$, this reduces to

$$El_0 = ATPN_0 \left(\frac{1+g}{r_E - g}\right) + (ATP_0 - ATPN_0) \left(\frac{\alpha}{1+r_E-\alpha}\right)$$

If $ATP_0 = ATPN_0$, then

$$r_E = \frac{ATPN_0}{El_0}(1+g) + g$$

The following table gives some annual nominal rates of return implied by various annual nominal growth rates and overall equity price-earnings ratios $[P/E = El_0/ATP_0 = El_0/ATPN_0]$, when ATP_0 equals ATPN.

P/E \ g	.04	.05	.06
10	.144	.155	.166
15	.109	.120	.131
20	.092	.103	.113

Since the logarithm of

$$ATPN_\tau = ATPN_0(1+g)^\tau$$

is

$$\log \mathrm{ATPN}_\tau = \log \mathrm{ATPN}_0 + \tau \log(1+g) \ ,$$

I regressed log ATP on τ, and obtained

$$\mathrm{ATPN}_0 = \mathrm{ATPN}_{1952 \cdot I} = 18.3$$

$$g = .054$$

In this model I used these estimates and the assumption that $\alpha = .7$

APPENDIX II-A

We have

$$1 = \sum_{i=1}^{n} \frac{a_i^*}{W} = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\sum_{h=1}^{n-z} \alpha_{ih}^{j} r_h + \sum_{h=1}^{q} \beta_{ih}^{j} s_h \right) \frac{\ell_j}{W}$$

$$= \sum_{j=1}^{m} \left[\sum_{h-1}^{n-z} \left(\sum_{i=1}^{n} \alpha_{ih}^{j} \right) r_h + \sum_{i=1}^{n} \beta_{i1}^{j} + \sum_{h=2}^{q} \left(\sum_{i=1}^{n} \beta_{ih}^{j} \right) s_h \right] \frac{\ell_i}{W}$$

where $s_1 = 1$ is the constant 1. Notice that there must be a constant, or this equation could not hold for all values of the independent variables--in particular, $r_h = s_h = 0$ would give $1 = 0$.

Now, substituting $\ell_m = W - \sum_{j=1}^{m-1} \ell_j$ yields

$$1 = \sum_{j=1}^{m-1} \left[\sum_{h=1}^{n-z} \left(\sum_{i=1}^{n} \alpha_{ih}^{j} \right) r_h + \sum_{i=1}^{n} \beta_{i1}^{j} + \sum_{h=2}^{q} \left(\sum_{i=1}^{n} \beta_{ih}^{j} \right) s_h \right] \frac{\ell_i}{W}$$

$$+ \left[\sum_{h=1}^{n-z} \left(\sum_{i=1}^{n} \alpha_{ih}^{m} \right) r_h + \sum_{i=1}^{m} \beta_{i1}^{m} + \sum_{h=2}^{q} \left(\sum_{i=1}^{n} \beta_{ih}^{m} \right) s_h \right] \left(1 - \frac{\sum_{j=1}^{m-1} \ell_j}{W} \right)$$

$$= \sum_{j=1}^{m-1} \left[\sum_{h=1}^{n-z} \left(\sum_{i=1}^{n} \alpha_{ih}^{j} - \sum_{i=1}^{n} \beta_{ih}^{m} \right) r_h + \left(\sum_{i=1}^{n} \beta_{i1}^{j} - \Sigma \beta_{i1}^{m} \right) \right.$$

$$\left. + \sum_{h=2}^{q} \left(\sum_{i=1}^{n} \alpha_{ih}^{j} - \sum_{i=1}^{n} \beta_{ih}^{m} \right) s_h \right] \frac{\ell_i}{W}$$

$$+ \sum_{h=1}^{n-z} \left(\sum_{i=1}^{n} \alpha_{ih}^{m} \right) r_h + \sum_{i=1}^{n} \beta_{i1}^{m} + \sum_{h=2}^{q} \left(\sum_{i=1}^{n} \beta_{ih}^{m} \right) s_h$$

For this to hold for all values of the r_h , s_h , and ℓ_i , it must be true that the constant

$$\sum_{i=1}^{n} \beta_{i1}^{m} = 1$$

while the coefficients of all variables are zero:

$$\left(\sum_{i=1}^{n} \alpha_{ih}^{j} - \sum_{i=1}^{n} \beta_{ih}^{m}\right) = 0 \qquad j \neq m , \quad \forall h$$

$$\sum_{i=1}^{n} \alpha_{i1}^{j} - \sum_{i=1}^{n} \beta_{i1}^{m} = 0 \qquad j \neq m$$

$$\sum_{i=1}^{n} \alpha_{ih}^{j} - \sum_{i=1}^{n} \beta_{ih}^{m} = 0 \qquad j \neq m , \quad h \neq 1$$

$$\sum_{i=1}^{n} \alpha_{ih}^{m} = 0 \qquad \forall h$$

$$\sum_{i=1}^{n} \beta_{ih}^{m} = 0 \qquad h \neq 1$$

These reduce to

$$\sum_{i=1}^{n} \beta_{i1}^{j} = 1 \qquad \forall j$$

$$\sum_{i=1}^{n} \beta_{ih}^{j} = 1 \qquad h \neq 1 , \quad \forall j$$

$$\sum_{i=1}^{n} \alpha_{ih}^{j} = 0 \qquad \forall h , \quad \forall j$$

APPENDIX II-B

For the system

$$da_i = \sum_{j=1}^{n-1} \alpha_{ij}[a_j^* - a_j(-1)] + \sum_{j=1}^{m} \beta_{ij} d\ell_j ,$$

we have

$$dW = \sum_{i=1}^{n} da_i = \sum_{i=1}^{n} [\sum_{j=1}^{n-1} \alpha_{ij}(a_j^* - a_j(-1)) + \sum_{j=1}^{m} \beta_{ij} d\ell_j]$$

$$= \sum_{i=1}^{n} [\sum_{j=1}^{n-1} \alpha_{ij}(a_j^* - a_j(-1)) + \sum_{j=1}^{m-1} \beta_{ij} d\ell_j + \beta_{im}(dW - \sum_{j=1}^{m-1} d\ell_j)]$$

$$= \sum_{j=1}^{n-1} [(\sum_{i=1}^{n} \alpha_{ij})(a_j^* - a_j(-1))] + (\sum_{i=1}^{n} \beta_{im}) dW + \sum_{j=1}^{m-1} [(\sum_{i=1}^{n} \beta_{im})] d\ell_j$$

which can hold for all possible $(a_j^* - a_j(-1))$, dW , and $d\ell_j$ if and only if

$$\sum_{i=1}^{n} \alpha_{ij} = 0 \qquad \forall j$$

$$\sum_{i=1}^{n} \beta_{im} = 1$$

$$\sum_{i=1}^{n} \beta_{ij} - \sum_{i=1}^{n} \beta_{im} = 0 \qquad j \neq m$$

Thus

$$\sum_{i=1}^{n} \alpha_{ij} = 0 \qquad \forall j$$

$$\sum_{i=1}^{n} \beta_{ij} = 1 \qquad \forall j$$

APPENDIX II-C

For the system

$$da_i = \sum_{j=1}^{n} \alpha_{ij}(a_j^* - a_j(-1)) + \sum_{j=1}^{m-1} \beta_{ij} d\ell_j$$

we have

$$dW = \sum_{i=1}^{n} da_i = \sum_{i=1}^{n} [\sum_{j=1}^{n} \alpha_{ij}(a_j^* - a_j(-1)) + \sum_{j=1}^{m-1} \beta_{ij} d\ell_j]$$

$$= \sum_{i=1}^{n} [\sum_{j=1}^{n-1} \alpha_{ij}(a_j^* - a_j(-1)) + \beta_{in}[dW - \sum_{j=1}^{n-1}(a_j^* - a_j(-1))]$$

$$+ \sum_{j=1}^{m-1} \beta_{ij} d\ell_j]$$

$$= \sum_{j=1}^{n-1} (\sum_{i=1}^{n} \alpha_{ij} - \sum_{i=1}^{n} \beta_{in})(a_j^* - a_j(-1)) + (\sum_{i=1}^{n} \beta_{in}) dW + \sum_{j=1}^{m-1} (\sum_{i=1}^{n} \beta_{ij}) d\ell_j$$

which can hold for all $(a_j^* - a_j(-1))$, dW , and $d\ell_j$ if and only if

$$\sum_{i=1}^{n} \alpha_{ij} - \sum_{i=1}^{n} \beta_{in} = 0 \qquad j \neq n$$

$$\sum_{i=1}^{n} \alpha_{in} = 1$$

$$\sum_{i=1}^{n} \beta_{ij} = 0 \qquad \forall j$$

Thus

$$\sum_{i=1}^{n} \alpha_{ij} = 1 \qquad \forall j$$

$$\sum_{i=1}^{n} \beta_{ij} = 0 \qquad \forall j$$

REFERENCES

[1] Anderson, W., Corporate Finance and Fixed Investment: An Econometric Study, Boston: Harvard University Graduate School of Business Administration, 1964.

[2] Brainard, W., and J. Tobin, "Pitfalls in Financial Model Building," American Economic Review, Papers and Proceedings, LVIII (2) (May 1968), 99-122.

[3] Gramlich, E., and J. Kalchbrenner, "A Constrained Estimation Approach to the Demand for Liquid Assets," Federal Reserve System Committee on Financial Analysis (mimeographed) 1969.

[4] Motley, B., "Household Demand for Assets: A Model of Short-Run Adjustments," Review of Economics and Statistics, LII (August 1970), 236-41.

[5] Royama, S., and K. Hamada, "Substitution and Complementarity in the Choice of Risky Assets," in Risk Aversion and Portfolio Choice, edited by D. Hester and J. Tobin, New York: John Wiley and Sons, 1967, 27-40.

[6] Tobin, J., "A General Equilibrium Approach to Financial Model Building," Journal of Money, Credit and Banking, I (February 1969), 15-29.

[7] Turnovsky, S., "Some Empirical Evidence on the Formation of Price Expectations," University of Pennsylvania Discussion Paper No. 143 (December 1969).

III. SOME ECONOMETRIC CONSIDERATIONS

The general demand equations [8a] of Chapter II can be estimated directly by single equation ordinary least squares. This will provide unique estimates of all elements in E, A_i, B_i, and F:[1]

$$\hat{A}_i = (\hat{E})^{-1} (\hat{EA}_i) \ ; \ \hat{B}i = (\hat{E})^{-1} (\hat{EB}_i)$$

The balance sheet restrictions will also be obeyed: The elements in any column of $\hat{A}_i$ or $\hat{B}_i$ will sum to zero (except for the constant terms, which will sum to one); the elements in each column of $\hat{E}$ will sum to one; and the elements in each column of $\hat{F}$ will sum to zero. This is explicitly proven in Appendix III-A.

In the case of noncompetitive markets, the estimated parameters will obey all adding-up restrictions; but unique estimates of all structural parameters will not in general result unless in each noncompetitive market there is one sector that absorbs the entire surplus. The occurrence of unique estimates in the latter case is proven in Appendix III-A.

---

1. If one of the (a^*-a_{-1}) had been deleted (instead of one of the $\Delta\ell_i$), then E would not have an inverse. In this case, if we delete the last row of $\hat{E}$ and postmultiply the inverse of this by $(\hat{EA}_i)$ with the last row deleted, then the result will be a unique estimate of A_i, with the last row deleted. The last row can then be calculated from the balance sheet restriction that the elements in any column of A_i must sum to zero.

Part D of this chapter explains why system estimators have been neglected. Parts A, B, and C are concerned with some common econometric problems that are expected to be encountered.

One likely problem is that there will be very few "true" degrees of freedom. For our general (competitive markets) demand equation ([8a] of chapter II), there are $T - [m(n-z+q)+n+m-1]$ conventionally measured degrees of freedom. A sector with 8 assets, 7 rates, 4 other explanatory variables, and 3 liabilities will require 44 observations just for the degrees of freedom to be positive so that estimates can be obtained. With 4 liabilities, this minimal criterion rises to 56 observations; with 5 liabilities, 68 observations are needed. Since my data consists of 72 observations -- of which I intend to save a substantial part for out-of-sample tests -- the potential squeeze is obviously worrisome.

Beyond these minimal requirements, we also have to contend with the fact that the collinear autocorrelated nature of the data series will make the effective dimensionality of the data set much less than its simply measured rank. Since we will not in general be able to accurately sort out the separate influences of more variables than we have significant dimensions, we must be wary of the common temptation to limit the complexity of our model to agree with the limited information contained in the data.

This would be an acceptable practice if it were done carefully,[2] but simple models are seldom arrived at by means of a systematic, well-designed, theoretically justified pruning. More typically, a financial variable which is found to be useful in predicting expenditure items is simply made a function of enough variables to be satisfactorily predicted itself. In other cases, explanatory variables are included on theoretical grounds and then quickly discarded because their explanatory power is redundant.

In part A of this chapter, I summarize some methods for quantifying the collinearity problem, and indicate some possible techniques for reducing the number of parameters to be estimated in a consistent explicit way, keeping in mind residual assets, implicit assumptions, and a priori information.

Parts B and C are concerned with simultaneity and autocorrelation -- which are familiar problems, with many proposed solutions, These various solutions are described and their merits briefly considered, in anticipation of narrowing down the list of estimation techniques to be applied to the model of Chapter II.

---

2. Harman [30] wrote that, "It is a commonly accepted scientific principle that a theoretical law should be simpler than the observed data upon which it is based."

A. MULTICOLLINEARITY

1. TESTING ITS SEVERITY

This section is concerned with measuring the severity of collinearity in the sample data. My discussion here follows Farrar and Glauber [17] closely and is somewhat sketchy. A more detailed development of the arguments, and proofs which are easier to follow than those in Farrar and Glauber, are available from this author.

Our basic demand equations for any sector can be written as a system of η equations,

$$\underset{t\times 1}{Y_i} = \underset{t\times p}{X}\ \underset{p\times 1}{\beta_i} + \underset{t\times 1}{U_i}$$

each with the same p explanatory variables. The variance-covariance matrix of the single equation OLS parameter estimates is

$$\text{COV}\ (\hat{\beta}) = \nabla^2_{u_i}\ (X'X)^{-1}$$

with the variance of the k^{th} element of $\hat{\beta}_i$ given by $\nabla_{u_i}\ a_{kk}$, where a_{kk} is the k^{th} diagonal element of $(X'X)^{-1}$, and

$$\frac{1}{\sum\limits_t X_k^2} \leq a_{kk} \leq \infty$$

with a_{kk} equal to one over the sample moment of X_k if X_k is orthogonal to the other X_i, and with a_{kk} infinite if X_k is linearly dependent on the other X_i.

The variances of the η estimated coefficients of X.

are

$$\mathrm{Var}\begin{bmatrix}\hat{\beta}_1^R\\ \beta_1^R\\ \cdot\\ \cdot\\ \cdot\\ \beta_\eta^R\end{bmatrix} = a_{kk}\begin{bmatrix}\nabla^2_{u_1}\\ \nabla^2_{u_1}\\ \cdot\\ \cdot\\ \cdot\\ \nabla^2_{u_\eta}\end{bmatrix}$$

and will all expand together the more collinear X_k is with the rest of the explanatory variables. And if the $\nabla^2_{u_i}$ are of roughly the same order of magnitude, then the imprecision in the estimates of the coefficients of a particular explanatory variable across equations will be roughly comparable.

The imprecision created by such collinearity will affect the normal t-statistics reported for each parameter estimate

$$\hat{\beta}_i^R \Big/ \sqrt{\nabla^2_{u_i} \cdot a_{kk}} = \tau_{T-P}$$

but cannot be accurately gauged from such a statistic, because of the presence of $\hat{\beta}_i^R$ (which is not the same for all variables, and may be very small) and $\hat{\nabla}^2_{u_i}$ (which is not constant across equation equations and is likely to underestimate the true $\hat{\nabla}^2_{u_i}$ if u_i is autocorrelated [3]).

3. This will occur if the observations fall predominantly above or below the true relation, which is likely with small samples from data and disturbances which are serially correlated. Note also that $\sqrt{\nabla^2_{u_i}} \cdot a_{RR}$ is not the correct expression with autocorrelated disturbances; see, for example, Johnston [34, p. 188].

Stated more plainly, multicollinearity will lower the calculated t-statistics (reflecting uncertain and unstable parameter estimates), but we cannot infer from the size of such a t-statistic how serious the multicollinearity problem is.

An overall indication of the extent of the collinearity among the explanatory variables can be gained by using principal component analysis to determine the effective dimensionality of the data set. In general, we will not be able to reliably estimate more parameters than there are significant dimensions, and such estimates will be subject to the caveat that the collinearity patterns among the explanatory variables not change, unless we impose at least as many a priori constraints as there are insignificant dimensions.

Principal components are discussed in some detail in section 2.4. below; this particular use of component analysis is discussed in section 2.4.6.

Farrar and Glauber advocated a cruder test based on Bartlett's finding that, with the normalized data set z multivariated normal and drawn from an orthogonal population,

$$- [T-1-1/6\ (2p+3)] \log |z'z| \sim \chi^2(1/2(p-1)(p-2))$$

As Haitovsky [29] has argued, Bartlett's test is conservative in that strict orthogonality is not necessary to obtain good parameter estimates.

He proposes that the null hypothesis be that Z'Z is singular, on the thin grounds that this is what some people mean by multicollinearity. There is obviously no test of this hypothesis comparable to Bartlett's test since the data will either be singular or nonsingular, and a population with less than p dimensions could not generate p-dimensional data. Nevertheless, Haitovsky proposes the heuristic test statistic

$$- [T-1-1/6(2p+3)] \text{ Log } (1-|Z'Z|) \sim \chi^2(V)$$

which does indeed give more liberal results than Bartlett's statistic. When there are two explanatory variables, for example, the tests are

$$\text{Bartlett:} \quad -(T-2\ 1/2) \text{ Log } (1-r^2) \sim \chi^2(1)$$

$$\text{Haitovsky:} \quad -(T-2\ 1/2) \text{ Log } (r^2) \sim \chi^2(1)$$

where r is the correlation between the two variables. At the 1% level, with T=50, Bartlett's test would signal trouble with $r^2 > .13$, while Haitovsky would require $r^2 > .87$.

This is the usual statistical dilemma of which event should receive the presumptive weight of being the null or alternative hypothesis, and here too there is unfortunately no good answer. This problem is also encountered below in the more general form of choosing and testing the effective dimensionality of a data set. Luckily, the models being considered in this thesis are of such large dimensionality that the roughness of the tests should not swamp the alternatives being tested.

Given an approximate idea of how many significant dimensions there are, it is crucial to determine where a priori restrictions are most needed. We can first determine which X_i

are the most troublesome by calculating the standard multiple correlation coefficients and associated F-statistics between each explanatory variable and the remaining explanatory variables. Then we can locate the causes of such troublesome correlations by calculating the partial correlation coefficients (and standard associated t-statistics) between each pair of the explanatory variables, holding constant the correlations with the remaining explanatory variables.

2. CORRECTING FOR ITS EFFECTS

2.1. The possibilities

Farrar and Glauber [17, p. 92] wrote that, "Economists are coming more and more to agree that...[correction] requires the generation of additional information." They suggest "...additional primary data collection, the use of extraneous parameter estimates from secondary data sources, or the application of subjective information." The fruits of the first two possibilities are easily enough employed, but difficult to obtain. The third possibility is much more practical, though not as mechanical.

Section 2.4. below is concerned with principal components and factor analysis, and supplements the previous section's discussion of where and when to apply subjective information. I also consider here some proposed direct applications of component and factor analyses to alleviate collinearity problems.

Sections 2.2. and 2.3. are concerned with some of the ways in which a priori information can be used to reduce a collinearity problem. In these sections, I consider the system

$$Y_1 = \sum_{i=1}^{p} \beta_{1i} X_i$$
$$\vdots$$
$$Y_n = \sum_{i=1}^{p} \beta_{ni} X_i$$

where $x_1=1$ and x_p is a problem variable in that it is severely collinear with the remaining x_i. This analysis can be expanded to fit the needs of most collinearity problems. If, for example, X_p, X_{p-1}, X_{p-2}, X_{p-3} are all pairwise highly collinear, but negligibly correlated with the remaining variables, then we should treat any three of these as problem variables and handle them in the ways discussed in 2.2. and 2.3. If, on the other hand, X_p is highly correlated with X_{p-1} and X_{p-2} is correlated with X_{p-3}, then we should treat X_p or X_{p-1} and X_{p-2} or X_{p-3} as problem variables.

Sections 2.2. and 2.3. consider two basic ways of using a priori information[4] so as to eliminate X_p as an independent variable:

(1) Replacing the terms $\beta_{jp}X_p$ with known linear functions of the other variables appearing in each equation, or

4. This a priori information is assumed to be certain, or else the uncertainties are assumed to not be well enough understood to be quantifiable. Methods which incorporate subjectively known probability distributions of the a priori information (for example, Baysian techniques (discussed and referenced in Christ [11, pp. 256, N. 18 and 537]) or Theil-Goldberger "Mixed Estimation" (discussed and referenced in Christ [11, pp. 534-7]) are not considered in this paper.

(2) Restricting each β_{jp} to be linearly related to the coefficients of variables in the other equations.

I have referred to these two methods rather loosely in the titles of sections 2.2. and 2.3. as "horizontal" and "vertical" applications of information, although elsewhere I use these phrases in more restricted senses: for example, in the system

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} \Sigma_1 \\ \Sigma_2 \end{pmatrix} = BX+\Sigma$$

I use the terms "horizontal" or "row" restrictions on B to mean

$$\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} = \begin{pmatrix} \lambda_1 \bar{\beta}_{11} & \lambda_1 \bar{\beta}_{12} \\ \lambda_2 \bar{\beta}_{21} & \lambda_2 \bar{\beta}_{22} \end{pmatrix}$$

where the $\bar{\beta}_{ij}$ are known constants. This means that we know

$$\beta_{i1} / \beta_{i2} = \bar{\beta}_{i1} / \bar{\beta}_{i2}$$

which is the magnitude of each element relative to any other element in its row. This is also equal to

$$\left(\partial Y_i / \partial X_1 \right) \Big/ \left(\partial Y_i / \partial X_1 \right)$$

The relative response of Y_i to unit changes in X_1 and X_2.

By "vertical" or "column" restrictions on B, I mean

$$\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \bar{\beta}_{11} & \lambda_2 & \bar{\beta}_{12} \\ \lambda_1 & \bar{\beta}_{21} & \lambda_2 & \bar{\beta}_{22} \end{pmatrix}$$

which implies that we know $\beta_{1i}/\beta_{2i} = \bar{\beta}_{1i}/\bar{\beta}_{2i}$, the magnitude of each element relative to other elements in its column -- which is also $(\partial Y_1/\partial X_i)/(\partial Y_2/\partial X_i)$, the relative responses of Y_1 and Y_2 to a change in a single X_i.

2.2. Horizontal Application of a priori Information

2.2.1. A general single equation substitution

Written in its most general form, we can substitute for the p^{th} term in the h^{th} equation

$$\beta_{hp} X_p = \sum_{i-1}^{p} \phi_i \beta_{hi} + \sum_{i=1}^{p} \theta_i X_i + \sum_i \sum_j \psi_{ij} \beta_{hi} X_j$$

where ϕ_i, θ_i, and ψ_{ij} are known constants.

If we momentarily ignore the ψ_{ij}, and rearrange the terms other than x_1 and x_p so that

$$\theta_i = 0 \qquad a < i \leq p-1$$

then the h^{th} equation can be rewritten after substitution as

$$Y_h - \theta_p X_p = (\beta_{h1} + \theta_1 + \sum_{i=1}^{a} \phi_i \beta_{hi} + \phi_p \beta_{hp}) 1 + \sum_{i=2}^{a} (\beta_{hi} + \theta_i) X_i + \sum_{i=a+1}^{p-1} \beta_{hi} (X_i + \phi_i)$$

or as

$$Y_h - \theta_p X_p = (\beta_{h1} + \theta_1 + \sum_{i=1}^{p} \phi_i \beta_{hi}) 1 + \sum_{i=2}^{a} (\beta_{hi} + \theta_i) X_i + \sum_{i=a+1}^{p-1} \beta_{hi} X_i$$

In either case, the presence of θ_1 and the ϕ_i does not affect the final equations, the forcasts, or the implicit estimates of any of the structural parameters other than β_{h1}.

Turning to $\sum_i \sum_j \psi_{ij} d_{hi} X_j$

there is no point to using $x_j = x_1 = 1$, since the terms $\sum_i \psi_{i1} d_{hi} 1$ duplicate the ϕ_i. To go further, we must notice that in the final equation that will be estimated, each of the original $\beta_{hj}x_j$ will appear either as β_{hj} multiplied by a linear combination of x_i's or as x_j multiplied by a linear combination of parameters. It will have to appear one of these ways, and to appear both ways would in general result in more parameter estimates than parameters, with one of the overidentified parameters being β_{hj}.

Since our choice of which θ_i and ψ_{ij} will be nonzero will determine how terms will be collected, we must be wary of specifications which necessitate conflicting methods of collection. For example, each non-zero θ_j requires that the associated $\beta_{hj}x_j$ appear with x_j factored out, and thus that β_{hj} not be used to collect terms. Consider then

$$\sum_i \psi_{ip}\beta_{hi}X_p$$

Since we are seeking to eliminate x_p as a separate independent variable, we will have to collect these terms by β_{hi}, which was just shown to be impossible for any i such that $\theta_i \neq o$. That is

$$\theta_i \neq 0 \Rightarrow \psi_{ip} = 0$$

And conversely the nonzero ψ_{ip} which do appear will fix the associated β_{hi} as factors.

Continued verbal dissection of the implicit restrictions on the $\sum_{ij}\sum \psi_{ij}\beta_{hi}X_j$ is difficult, and it will be easiest (for both the author and the reader) to fall back on an explicit layout of the terms involved. For p=6, we can display

$\sum_{i}\sum_{j}\psi_{ij}\beta_{hi}X_j$ as

$$\begin{array}{llll}
\text{A} & & \text{B} & \\
\psi_{11}\beta_{h1}1 & + \psi_{12}\beta_{h1}X_2+\psi_{13}\beta_{h1}X_3 & + \psi_{14}\beta_{h1}X_4+\psi_{15}\beta_{h1}X_5+\psi_{16}\beta_{h1}X_6 & \\
\psi_{21}\beta_{h2}1 & + \psi_{22}\beta_{h2}X_2+\psi_{23}\beta_{h2}X_3 & + \psi_{24}\beta_{h2}X_4+\psi_{25}\beta_{h2}X_5+\psi_{26}\beta_{h2}X_6 & \\
\psi_{31}\beta_{h3}1 & + \psi_{32}\beta_{h3}X_2+\psi_{33}\beta_{h3}X_3 & + \psi_{34}\beta_{h3}X_4+\psi_{35}\beta_{h3}X_5+\psi_{36}\beta_{h3}X_6 & \text{C} \\
\psi_{41}\beta_{h4}1 & + \psi_{42}\beta_{h4}X_2+\psi_{43}\beta_{h4}X_3 & + \psi_{44}\beta_{h4}X_4+\psi_{45}\beta_{h4}X_5+\psi_{46}\beta_{h4}X_6 & \\
\psi_{51}\beta_{h5}1 & + \psi_{52}\beta_{h5}X_2+\psi_{53}\beta_{h5}X_3 & + \psi_{54}\beta_{h5}X_4+\psi_{55}\beta_{h5}X_5+\psi_{56}\beta_{h5}X_6 & \\
\psi_{61}\beta_{h6}1 & + \psi_{62}\beta_{h6}X_2+\psi_{63}\beta_{h6}X_3 & + \psi_{64}\beta_{h6}X_4+\psi_{65}\beta_{h6}X_5+\psi_{66}\beta_{h6}X_6 &
\end{array}$$

where 3 sets have been displayed in order to facilitate analysis. $X_1 = 1$ and $X_6 = X_p$ have been fixed, while X_2, X_3, X_4 and X_5 have been arranged according to the way terms will be collected.

We've already seen that all elements of set A should be zero, since the ψ_{i1} repeat the θ_i. Set B was drawn to consist of elements grouped by x_i (i=2,3), and set C was drawn to consist of elements which will be grouped by β_{hi} (i=4,5). Presumably $\theta_4 = \theta_5 = 0$.

Given the items included in set B, set C could only consist of the indicated elements. This is because our equation already has p-1 (five) terms:

$$Y_h-\theta_pX_p=(\beta_{h1}+\theta_1+\sum_{i=1}^{p}\phi_i\beta_{hi})1+(\beta_{h2}+\theta_2+\sum_{i=1}^{6}\psi_{i2}\beta_{hi})X_2$$
$$+(\beta_{h3}+\theta_3+\sum_{i=1}^{6}\psi_{13}\beta_{hi})X_3+\beta_{h4}X_4+\beta_{h5}X_5$$

where some of the θ_i, ϕ_i, ψ_{i2}, and ψ_{i3} may be zero.

If we are not going to group terms by X_4 and X_5 (put them in set B), then we must group them by β_{h4} and β_{h5}. Further, we cannot group terms by β_{h1}, β_{h2}, β_{h3}, or β_{h6} without going to six or more explanatory variables. From the Gauss-Markov Theorem, we know that with six explanatory variables, we have not affected the structural estimates or the collinearity problem. More than six explanatory variables cannot be linearly independent, and their partial effects cannot be estimated.

On the other hand, if β_{h6} appears in the final equation, (either ϕ_6 or one of the ψ_{6i} is non-zero), then less than six explanatory variables will mean that not all of the structural parameters can be estimated.[5]

To repeat, the ψ_{ij} outside of AUBUC must be zero if the collinearity problem is to be reduced, and there is no point to having the elements of A be non-zero. If any of the coefficients of β_{h6} are non-zero, then we will not obtain unique estimates of all the β_{hi}, unless we go back to six explanatory variables and our original parameter estimates. Finally, note that the elements of B∩C can be collected either by X_2 and X_3 or by β_{h4} and β_{h5}.

From this diagram and these remarks, we can infer that the non-zero elements of $\sum_{ij}\sum \psi_{ij} d_{hi} X_j$ can generally be written as

$$\sum_{j=z}^{a} \left(\sum_{i=1}^{a} \psi_{ij}\beta_{hi} + \psi_{pj}\beta_{hp}\right) X_j + \sum_{i=a+1}^{p-1} \beta_{hi} \left(\sum_{j=a+1}^{p} \psi_{ij} X_j\right)$$
$$+ \sum_{j=2}^{a} \sum_{i=a+1}^{p-1} \psi_{ij}\beta_{hi} X_j$$

5. This assumes that the substitution equation for $\beta_{h6}X_6$ does not imply an equation in which β_{h6} is a linear function of the other β_{hi}. If the contrary were true, then we could have substituted the implicit equation, and β_{h6} would not appear in the final equation to be estimated.

where the third term (analogous to $B \cap C$) can be incorporated with either the first term or the second term. When incorporated with the first term, the substitution equation will be

$$d_{hp}X_p = \theta_p + (\sum_{i=1}^{p} \theta_i \beta_{hi} + \theta_1) + \sum_{i=2}^{a} \theta_i X_i + \sum_{j=2}^{a} (\sum_{i=1}^{p} \psi_{ij}\beta_{hi}) X_j$$

$$+ \sum_{i=a+1}^{p-1} \beta_{hi} \; (\sum_{j=a+1}^{p} \psi_{ij} X_j)$$

and the equation to be estimated will be after substitution,

$$Y_h - \theta_p X_p = (\beta_{h1} + \theta_1 + \sum_{i=1}^{p} \phi_i \beta_{hi}) 1 + \sum_{j=2}^{a} (\beta_{hj} + \theta_j + \sum_{i=1}^{p} \psi_{ij}\beta_{hi}) X_j$$

$$+ \sum_{i=a+1}^{p-1} \beta_{hi} (X_i + \sum_{j=a+1}^{p} \psi_{ij} X_j)$$

where it is preferable to have $\phi_p = \psi_{pj} = 0 \; \forall_j$

2.2.2. Substitution into all n equations

Since the same variables appear in each equation in the system, the multicollinearity problem will be the same in each equation, so that if we want to substitute something for $d_{hp}X_p$ in one equation then we should do so in all equations.

We should therefore introduce superscripts into our notation in order to identify the equation into which each substitution is made

$$\beta_{hp}X_p = \sum_{i=1}^{p} \phi_i^h \beta_{hi} + \sum_{i=1}^{p} \theta_i^h X_i + \sum_{ij}\sum \psi_{ij}^h \beta_{hi} Xj \qquad 1 \leq h \leq n$$

and we should consider what restrictions the balance sheet identities impose on these substitutions.

We can write the h^{th} equation as

$$Y_h - \theta_p^h X_p = \left(\beta_{h1} + \theta_1^h + \sum_{i=1}^{p} \phi_i^h \beta_{hi}\right) 1 + \sum_{j=2}^{a} \left(\beta_{hj} + \theta_j^h + \sum_{i=1}^{p} \psi_{ij}^h \beta_{hi}\right) X_j$$

$$+ \sum_{i=a+1}^{p-1} \beta_{hi} \left(X_i + \sum_{j=a+1}^{p} \psi_{ij}^h X_j\right)$$

and we know from the Gauss-Markov Theorem that analysis with the terms arranged this way will be sufficient.[6]

With single equation ordinary last squares, we will obtain

$$\sum_{h=1}^{n} \left(\widehat{\beta_{hj} + \theta_1^h + \sum_{i=1}^{p} \phi_i^h \beta_{hi}}\right) = 1$$

$$\sum_{h=1}^{n} \left(\widehat{\beta_{hj} + \theta_j^h + \sum_{i=1}^{p} \psi_{ij}^h \beta_{hi}}\right) = 0 \qquad 2 \le j \le a$$

$$\sum_{h=1}^{n} \hat{\beta}_{hi} = 0 \qquad a+1 \le i \le p-1$$

in accordance with Appendix III-A if

$$\psi_{ij}^h = \psi_{ij} \qquad 1 \le h \le n \; ; \; a+1 \le j \le p, \; a+1 \le i \le p-1$$

(so that each equation has the same set of explanatory variables) and if $1 = \sum_{h=1}^{n} (Y_h - \theta_p^h X_p) = 1 - X_p \sum_{h=1}^{n} \theta_p^h \Rightarrow \sum_{h=1}^{n} \theta_p^h = 0$

---

6. See for example Graybill [26, p.116, Thm 6.3] for an explicit statement of the relevant corollary.

Thus, the fact that we want the structural estimates to be such that

$$\sum_{h=1}^{n} \hat{\beta}_{h1} = 1 \quad ; \quad \sum_{h=1}^{n} \hat{\beta}_{hi} = 0 \qquad 1<i\leq p$$

Requires that

$$[1a] \qquad \sum_{h=1}^{n} (\theta_1^h + \sum_{i=1}^{p} \phi_i^h \hat{\beta}_{hi}) = 0$$

$$[1b] \qquad \sum_{h=1}^{n} (\theta_j^h + \sum_{i=1}^{p} \psi_{ij}^h \hat{\beta}_{hi}) = 0 \qquad 2\leq j\leq a$$

$$[1c] \qquad X_p \sum_{h=1}^{n} \hat{\beta}_{hp} = \sum_{h=1}^{n} \hat{\beta}_{hp} X_p = 0$$

This third condition [1c] will be met if the first two conditions hold:

$$\sum_{h=1}^{n} \hat{\beta}_{hp} X_p = \sum_{h=1}^{n} \Big[\theta_p^h + \theta_1^h + \sum_{i=1}^{p} \phi_i^h \hat{\beta}_{hi} + \sum_{j=2}^{a} (\theta_j^h + \sum_{i=1}^{p} \psi_{ij}^h \hat{\beta}_{hi}) X_j$$

$$+ \sum_{i=a+1}^{p-1} \sum_{j=a+1}^{p} \hat{\beta}_{hi} \psi_{ij} X_j \Big]$$

$$= \sum_{h=1}^{n} [\theta_1^h + \sum_{i=1}^{p} \phi_i^h \hat{\beta}_{hi}] + \sum_{j=2}^{a} (\sum_{h=1}^{n} \theta_j^h + \sum_{i=1}^{p} \psi_{ij}^h \hat{\beta}_{hi}) X_j$$

$$= \sum_{i=a+1}^{p-1} \sum_{j=a+1}^{p} \psi_{ij} X_j (\sum_{h=1}^{m} \hat{\beta}_{hi}) = 0$$

since the first two terms are zero from conditions [1a] and [1b], respectively, and $\sum_{h=1}^{n} \hat{\beta}_{hi} = 0 \quad a+1\leq i\leq p-1$ directly from the estimation.

Thus, we need only look at conditions [1a] and [1b], although these admittedly are not very strict or illuminating. A theoretically plausible restriction is to let β_{hp} depend on the parameters in each particular equation, but to replace X_p with variables which do not depend on the equation that is going to accept the substitution. That is, if we divide both sides of our substitution equation by $d_{hp}-\theta_p$, we obtain

$$X_p = \frac{1}{\beta_{hp}-\theta_p^h}\left(\theta_1^h + \sum_{i=1}^{p} \phi_i^h \beta_{hi}\right) + \sum_{j=2}^{a} \frac{1}{\beta_{hp}-\theta_p^h}\left(\theta_j^h + \sum_{i=1}^{p} \psi_{ij}^h \beta_{hi}\right) X_j$$

$$+ \sum_{j=a+1}^{p} \left(\sum_{i=a+1}^{p-1} \frac{\beta_{hi}}{\beta_{hp}-\theta_p^h} \psi_{ij}\right) X_j$$

$$= \sum_{j=1}^{p} \Pi_j X_j$$

where we want to make the Π_j independent of h:

$$\frac{1}{\beta_{hp}-\theta_p^h}\left(\theta_1^h + \sum_{i=1}^{p} \phi_i^h \beta_{hi}\right) = \Pi_1$$

$$[2] \qquad \frac{1}{\beta_{hp}-\theta_p^h}\left(\theta_j^h + \sum_{i=1}^{p} \psi_{ij}^h \beta_{hi}\right) = \Pi_j \qquad 2 \le j \le a$$

$$\frac{1}{\beta_{hp}-\theta_p^h} \sum_{i=1+a}^{p-1} \psi_{ij}^h \beta_{hi} = \Pi_j \qquad a+1 \le j \le p$$

If these rules are followed, then conditions [1a] and [1b] will hold:

$$[1a] \quad \sum_{h=1}^{n} (\theta_1^h + \sum_{i=1}^{p} \phi_i^h \hat{\beta}_{hi}) = \sum_{h=1}^{n} \Pi_1 (\hat{\beta}_{hp} - \theta_p^h) = 0$$

$$[1b] \quad \sum_{h=1}^{n} (\theta_j^h + \sum_{i=1}^{p} \psi_{ij}^h \hat{\beta}_{hi}) = \sum_{h=1}^{n} \Pi_j (\hat{\beta}_{hp} - \theta_p^h) = 0$$

$$2 \le j \le a$$

2.2.3 Special Cases

a. Known Parameter

The simplest case is where the parameter is known:

$$\beta_{hp} = \theta_p^h \Rightarrow \beta_{hp} X_p = \theta_p^h X_p$$

The equation to be estimated then becomes

$$Y_h - \theta_p^h X_p = \sum_{i=1}^{p-1} \beta_{hi} X_i$$

Condition [1a] is met and condition [1b] implies that we must choose

$$\sum_{h=1}^{n} \theta_p^h = 0$$

b. β_{hp} Known Function of Other β_{hi}

We might know (or want to restrict)

$$\beta_{hp} = \theta_p^h + \sum_{i=2}^{p-1} \psi_{ip} \beta_{hi} \Rightarrow \beta_{hp} X_p = \theta_p^h X_p + \sum_{i=2}^{p-1} \psi_{ip} \beta_{hi} X_p$$

as for example with

$$\beta_{hp} = -\beta_{hk} \Rightarrow \beta_{hp} X_p = -\beta_{hk} X_p$$

$$\text{or} \quad \sum_{i=2}^{p} \beta_{hi} = \theta_p^h \Rightarrow \beta_{hp} X_p = (\theta_p^h - \sum_{i=2}^{p-1} \beta_{hi}) X_p$$

The first example corresponds to two explanatory variables having equal but opposite effects on Y_h; in the second case a unit rise in all explanatory variables (other than the constant) has a known effect θ_p^h on Y_h.

Our general equation to be estimated is

$$Y_h = \sum_{i=1}^{p-1} \beta_{hi} X_i + \theta_p^h X_p + \sum_{i=2}^{p-1} \psi_{ip} \beta_{hi} X_p$$

$$Y_h - \theta_p^h X_p = \beta_{hi} \cdot 1 + \sum_{i=2}^{p-1} \beta_{hi} (X_i + \psi_{1p} X_p)$$

and the appropriate column restrictions on the β_{hi} will hold if

$$\sum_{h=1}^{n} \theta_p^h = 0$$

c. <u>X_p Known Function of Other X_i</u>

We might know that

$$X_p = \phi_p + \sum_{j=2}^{p-1} \psi_{pj} X_j \Rightarrow \beta_{hp} X_p = \phi_p \beta_{hp} + \sum_{j=2}^{p-1} \psi_{pj} \beta_{hp} X_j$$

where superscripts have been omitted, indicating that this relation among the explanatory variables is independent of the dependent variables. Conditions [2] are thus satisfied, and we know directly that the estimated parameters will sum across equations properly.

Unfortunately, an estimate of β_{hp} does not occur, and hence a unique estimate of any other β_{hi} will not be obtainable unless the appropriate ϕ_p or ψ_{pj} is zero. This is so because with p-1 explanatory variables we only obtain p-1 parameter estimates, and our substitution equation contains no information about (or restriction on) the parameter values.

We might note that our regression equation is identical to that which would have been obtained if we had just deleted X_p.[7] One advantage of making the explicit substitution is that we are reminded that the estimated coefficients are not just the partial responses of Y_h to a unit change in X_j, but do include the change in X_p associated with a unit change in X_j and the effect of that change in X_p on Y_h:

$$(\beta_{hj} \hat{+} \beta_{hp} \psi_{pj}) = \frac{\partial Y_h}{\partial X_j} + \frac{\partial Y_h}{\partial X_p} \frac{\partial X_p}{\partial X_j}$$

Stated less sternly, the coefficient of X_j is the change in Y_h that can be expected when X_j changes by one unit, taking into account X_j's effect on Y_h through X_p. If this latter relationship is stable, we should be able to forecast fairly well; if it is unstable, then we will need an a priori estimate of β_{hp}. In either case, we will see in 2.4. that an approximately correct specification of β_{hp} can be beneficial and is more likely to be beneficial the more severe is the collinearity of X_p with the remaining explanatory variables.

7. This is not an uncommon practice, and is sometimes accompanied by the estimation of an equation $X_p = \emptyset_p + \sum_{i=1}^{p-1} \psi_{pj} X_j$ so that X_p can be kept an endogenous variable.

2.3 Vertical Application of A Priori Information

This section considers the possibility of restricting each β_{hp} to being linearly related to the coefficients of variables in the other equations. In the first part, I illustrate how a "stacking" procedure can accomodate strategically chosen a priori restrictions which relate the β_{hp} to coefficients of variables other than X_p and which cross equations. In the second and third parts, I show two ways of alleviating multicollinearity by assuming that we know the relative magnitudes of the coefficients associated (across equations) with a particular problem variable.

2.3.1. A General Approach Illustrated

The most general technique is, for X_p a problem variable, to just make each β_{hp} (h=1,...,n) a linear function of some β_{hi} somewhere else in the system,[8] and then regress the whole system as one equation with these functions, as well as the balance sheet restrictions, explicitly observed. Since the possibilities here are obviously very numerous, I am going to be content to give what I think is a reasonable example.

8. Since each equation contains an identical list of explanatory variables, there is no point to having $\beta_{hp}X_p$ depend on X_i which appear in other equations.

In the system

$$Y_1 = \beta_{10} + \beta_{11} X_1 + \beta_{12} X_2 + \beta_{13} X_3$$

$$Y_2 = \beta_{20} + \beta_{21} X_1 + \beta_{22} X_2 + \beta_{23} X_3$$

$$Y_3 = \beta_{30} + \beta_{31} X_1 + \beta_{32} X_2 + \beta_{33} X_3$$

where all of the X_1 are pairwise highly correlated, we may believe[9] that

$$\beta_{ij} = \beta_{ji} \quad , \quad i \neq j$$

After substitution, we obtain

$$Y_1 = \beta_{10} - (\beta_{21} + \beta_{31}) X_1 + \beta_{21} X_2 + \beta_{31} X_3$$

$$Y_2 = \beta_{20} + \beta_{21} X_1 - (\beta_{21} + \beta_{32}) X_2 + \beta_{32} X_3$$

$$Y_3 + \beta_{30} + \beta_{31} X_1 + \beta_{32} X_2 - (\beta_{31} + \beta_{32}) X_3$$

which we can rearrange by collecting terms by coefficients:

$$Y_1 = \beta_{10} + \beta_{21} (X_2 - X_1) + \beta_{31} (X_3 - X_1)$$

$$Y_2 = \beta_{20} + \beta_{21} (X_1 - X_2) + \beta_{32} (X_3 - X_2)$$

$$Y_3 = 1 - \beta_{10} - \beta_{20} + \beta_{31} (X_1 - X_3) + \beta_{32} (X_2 - X_3)$$

With five coefficients to be estimated, we can use all our information and guarantee unique parameter estimates by regressing this as a single equation:

9. This is actually done by Gramlich and Kalchbrenner [25, pp. 6-7, 16] in a model where the X_i are interest rates and the Y_i are various liquid asset proportions. Unfortunately, these symmetry conditions are more convenient than realistic, as I mentioned in Chapter II.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3-1 \end{bmatrix} = \beta_{10}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + \beta_{20}\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} + \beta_{21}\begin{bmatrix} X_2-X_1 \\ X_1-X_2 \\ 0 \end{bmatrix} + \beta_{31}\begin{bmatrix} X_3-X_1 \\ 0 \\ X_1-X_3 \end{bmatrix} + \beta_{32}\begin{bmatrix} 0 \\ X_3-X_2 \\ X_2-X_3 \end{bmatrix}$$

This will be an asymptotically efficient technique if the disturbances are contemporaneously independent, but have equal variances.

If only X_1 and X_3 (or X_1 and X_2) were correlated, we might use restrictions such as

$$\beta_{11} = \theta\ \beta_{22}$$
$$\beta_{21} = \beta_{12}$$

which would yield

$$\begin{bmatrix} Y_1 \\ Y_2 \\ 1-Y_3 \end{bmatrix} = \beta_{10}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + \beta_{20}\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} + \beta_{12}\begin{bmatrix} X_2 \\ X_1 \\ -X_1-X_2 \end{bmatrix} + \beta_{22}\begin{bmatrix} \theta X_1 \\ X_2 \\ -\theta X_1-X_2 \end{bmatrix} + \beta_{13}\begin{bmatrix} X_3 & & 0 \\ 0 & +\ \beta_{23} & X_3 \\ -X_3 & & -X_3 \end{bmatrix}$$

This approach can obviously be extended and modified to accomodate different ways of making all but one of any set of β_{hp} (h = 1,...n) be linear functions of other β_{hi}:

This approach can also be used to constrain coefficients to be zero without violating the balance sheet constraints, which is not true when the system is regressed as n separate equations.

2.3.2. Regression With Different LHS Variables.

In the n equation system $\underset{n\times1}{Z} = \underset{n\times p}{B}\ \underset{p\times1}{X} + \underset{n\times1}{V}$

we can replace the Z_i with different LHS variables. In particular, there are n-1 linearly independent equations

$$\underset{}{0} = -I_{n-1} \underset{n-1x1}{Y} + \underset{n-1xp}{D} \underset{px1}{X} + \underset{n-1x1}{U}$$

which can be solved for any n-1 variables, and in particular for the n-1 variables such that the remaining p variables are the least collinear of all possible sets of RHS variables.[10]

For simplicity, assume that only members of X are used as LHS variables. In this case, we have

$$X^{(1)} = D_1^{-1} Y - D_1^{-1} D_2 X^{(2)} - D_1^{-1} U$$

which will yield unique estimates of all parameters:

$$\hat{D}_1 = (\hat{D}_1^{-1})^{-1} \qquad \hat{D}_2 = \hat{D}_1 \ (\widehat{D_1^{-1} D_2})$$

We can reduce the collinearity of the explanatory variables still further by building in column restrictions on D_1 which have the effect of fixing the relative responses of the Y_i to changes in a single element of $X^{(1)}$, but leave to be estimated the differential changes in any single Y_i to be associated with changes in different elements of $X^{(1)}$. This can be accomplished by substituting

$$\underset{n-1xn-1}{D_1} = \underset{n-1xn-1}{\overline{D_1}} \ \underset{n-1xn-1}{C}$$

where C is a diagonal matrix, and $\overline{D_1}$ is a matrix of a priori

10. These will be the p variables which (normalized to unit variance) have the largest determinant of their correlation matrix.

coefficients for which the relevant consideration is that in any column the relative magnitudes of the elements be reasonable accurate. The relative magnitudes of the elements in any one column with respect to the elements in any other column are not consequential since these will ultimately be fixed by our estimates of the γ_{ii}.

Now, substituting

$$D_1^{-1} = C^{-1} \bar{D}_1^{-1}$$

into our system, we obtain

$$X^{(1)} = C^{-1} [\bar{D}_1^{-1} Y] - D_1^{-1} D_2 X^{(2)} - D_1^{-1} U$$

And since

$$C^{-1} = \begin{pmatrix} 1/\gamma_{11} & & 0 \\ & \ddots & \\ 0 & & 1/\gamma_{n-1,n-1} \end{pmatrix},$$

we have only to regress each element of $X^{(1)}$ on one linear combination[11] of Y_i and the p-n+1 elements of $X^{(2)}$.

One cost of these restrictions will be the errors in the variable $(\bar{D}_1)^{-1} Y$ that are created by an incorrect specification of $(\bar{D}_1)^{-1}$. Looking at the K^{th} equation, and the term $\sum_{i=1}^{n-1} \bar{\psi}_{Ki} Y_i$ (where the ψ_{ij} are the elements of $\bar{D}_1^{-1}$), we

11. If the elements of $(\bar{D}_1)^{-1}$ are $\bar{\psi}_{ij}$, then in the K^{th} equation the linear combination will be $\frac{1}{\gamma_{KK}} [\sum_{i=1}^{n-1} \bar{\psi}_{Ki} Y_i]$.

can make some headway by assuming that each $\bar{\psi}_{ij}$ differs from the true ψ_{ij} by a term

$$\varepsilon_{\bar{\psi}_{ij}} \sim N\,(0,\ \nabla_{ij}^2) \quad \text{with} \quad E\,(\varepsilon_{\psi_{ij}}\varepsilon_{\psi_{iK}}) = 0$$

With these assumptions, we have introduced into the K^{th} equation an error

$$\gamma_{kk}\varepsilon_k \sim N(0,\ \gamma_{KK}^2 \sum_{i=1}^{n-1} (Y_i)^2\, \nabla_{Ki}^2\,)$$

If the Y_i are autocorrelated, then this error will be also, and the correlation between the Y_i and the other explanatory variables will induce biases. These freshly created biases may of course offset previously introduced ones.

The size of these errors may be relatively small. If, for example, we substitute $Y_i = \Delta a_{i/W}$ and assume that the asset proportions are equal (and constant) and that $\nabla_{Ki}^2 = \nabla_K^2$, $\forall i$, then

$$\nabla_{\varepsilon_K}^2 = \sum_{i=1}^{n-1} (\frac{\Delta a_i}{W})^2\, \nabla_{Ki}^2 = (n-1)\,(\frac{\Delta a}{W})^2\, \nabla_K^2 = (n-1)\,(\frac{a}{W})^2 (\frac{\Delta a}{a})^2\, \nabla_K^2$$

$$= (n-1)\ (\frac{1}{n})^2\ (\frac{\Delta a}{a})^2\ \nabla_K^2$$

More specifically, if $n=8$ and $\Delta a/a = .10$, then

$$\nabla_{\varepsilon_K}^2 = \frac{7}{8} \cdot \frac{1}{8}\,(.1)^2\, \nabla_K^2 < \frac{1}{800}\ \nabla_k^2$$

Another cost of this procedure is that we have new disturbance terms which (ignoring errors in $(\bar{D}_1)^{-1}Y$) are linear combinations of the original disturbances, and the new disturbances may be more or less well behaved then the old ones. Similarly, the new RHS variables may cause more or less simultaneity problems, depending on their feedbacks and the available instruments.

These problems are of course true of all single equation methods, since only full information maximum likelihood estimates are invariant to the choice of LHS variables. In this situation, the best procedure [19] is probably to choose the LHS variables on the basis of a causal interpretation of the model, so that they are the variables which have the least direct and indirect feedback on the remaining variables. This criterion would demand regression of the equations in the form they were originally justified (with changes in asset holdings on the left), without regard for collinearity goals.

2.3.3. A Complication Resolved

From Chapter II, our basic equation is

$$[8a]\quad \underset{n\times 1}{\Delta a/w} = \underset{n\times n-z}{(E\cdot A)}\ \underset{n-z\times 1}{r} + \underset{n\times q}{(E\cdot B)}\ \underset{q\times 1}{s} - \underset{n\times n}{E}\ \underset{n\times 1}{a_{-1}/w} + \underset{n\times m-1}{F}\ \underset{m-1\times 1}{\Delta \ell/w}$$

(I've used this equation rather than [8b] because the notation is simpler; but the results can easily enough be extended).

For our purposes here it will be useful to combine terms so that the system can be written as

$$\Delta a/w = EHX - Ea_{-1}/w + F\Delta\ell/w$$

Now, it may be that some of the X_i are problem variables, but that we have little a priori knowledge about the elements in EH, though we do know something about E and/or H, considered separately. Unfortunately, only very restricted types of knowledge about E and H is useful in constraining their product EH.

If E is specified completely, then specifying the coefficients of X_i in the "desired relations" will fix X_i's coefficients in the final equations, since

$$EHX = E[H_1 \vdots H_2 \vdots \cdots \vdots H_{n-z+q}] \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{n-z+q} \end{bmatrix}$$

$$= EH_1X_1 + EH_2X_2 + \cdots EH_{n-z+q} X_{n-z+q}$$

If, on the other hand, E is not specified completely but is subject to vertical or horizontal constraints, then the problem is more difficult.

For example, consider the simple case where

$$EH = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} \\ \varepsilon_{21} & \varepsilon_{22} \end{pmatrix} \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} = \begin{pmatrix} \varepsilon_{11} h_{11} + \varepsilon_{12} h_{21} & \varepsilon_{11} h_{12} + \varepsilon_{12} h_{22} \\ \varepsilon_{21} h_{11} + \varepsilon_{22} h_{21} & \varepsilon_{21} h_{12} + \varepsilon_{22} h_{22} \end{pmatrix}$$

If we impose the vertical restriction on H

$$\begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} = \begin{pmatrix} \gamma_1 \bar{h}_{11} & \gamma_2 \bar{h}_{12} \\ \gamma_1 \bar{h}_{21} & \gamma_2 \bar{h}_{22} \end{pmatrix}$$

Then

$$EHX = \begin{bmatrix} \gamma_1(\varepsilon_{11}\bar{h}_{11} + \varepsilon_{12}\bar{h}_{21})X_1 + \gamma_2(\varepsilon_{11}\bar{h}_{12} + \varepsilon_{12}\bar{h}_{22})X_2 \\ \gamma_1(\varepsilon_{21}\bar{h}_{11} + \varepsilon_{22}\bar{h}_{21})X_1 + \gamma_2(\varepsilon_{21}\bar{h}_{12} + \varepsilon_{22}\bar{h}_{22})X_2 \end{bmatrix}$$

But now neither simple vertical nor horizontal restrictions on E will collapse EHX into two terms,[12] which is exactly the number of parameter estimates that we must obtain from EHX if we are to have unique estimates of all parameters. Again, this is because Ea-1/w will give us unique estimates of all ε_{ij}, and our two vertical restrictions on H have reduced the number of necessary parameter estimates from 4 to 2.

On the other hand, we can generally apply the following two-step procedure, which makes restrictions on E unnecessary for the estimation of EH -- although either vertical or horizontal restrictions could be imposed on E a-1/w if some of the $a_i(-1)$ are problem variables.

We can rearrange the X_i and the columns of H so that $X^{(1)}$ contains the "problem" X_i. Then

$$\frac{\Delta a}{w} = E[H^{(1)} \vdots H^{(2)}] \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix} - E\ a\text{-}1/w + F\ \Delta\ell/w$$

$$= EH^{(1)}X^{(1)} + EH^{(2)}X^{(2)} - E\ a\text{-}1/w + F\ \Delta\ell/w$$

If we regress the system in this form (with possible restrictions on E and F), then we will presumably obtain good

---

12. For example, horizontal restrictions on E

$$\begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} \\ \varepsilon_{21} & \varepsilon_{22} \end{pmatrix} = \begin{pmatrix} \lambda_1\ \bar{\varepsilon}_{11} & \lambda_1\ \bar{\varepsilon}_{12} \\ \lambda_2\ \bar{\varepsilon}_{21} & \lambda_2\ \bar{\varepsilon}_{22} \end{pmatrix}$$

yield

$$\begin{bmatrix} \gamma_1\lambda_1(\bar{\varepsilon}_{11}\bar{h}_{11}+\bar{\varepsilon}_{12}\bar{h}_{21})X_1 & + & \gamma_2\lambda_1(\bar{\varepsilon}_{11}\bar{h}_{12}+\bar{\varepsilon}_{12}\bar{h}_{22})X_2 \\ \gamma_1\lambda_2(\bar{\varepsilon}_{21}\bar{h}_{11}+\bar{\varepsilon}_{21}\bar{h}_{21})X_1 & + & \gamma_2\lambda_2(\bar{\varepsilon}_{21}\bar{h}_{12}+\bar{\varepsilon}_{22}\bar{h}_{22})X_2 \end{bmatrix}$$

estimates[13] of E, F, and $H^{(2)}$, but unreliable estimates of $(\hat{EH}^{(1)})$.

If we now impose horizontal restrictions on $H^{(1)}$:

$$H^{(1)} = C\bar{H}^{(1)} = \begin{bmatrix} \gamma_1 & & 0 \\ 0 & \gamma_2 & \\ & & \ddots \end{bmatrix} \begin{bmatrix} \bar{h}_{11} & \bar{h}_{12} & \cdots \\ \bar{h}_{21} & & \\ \vdots & & \end{bmatrix} = \begin{bmatrix} \gamma_1\bar{h}_{11} & \gamma_1\bar{h}_{12} & \cdots \\ \gamma_2\bar{h}_{21} & & \\ \vdots & & \end{bmatrix}$$

Then we can rewrite

$$EH^{(1)}X^{(1)} = EC\bar{H}^{(1)}X^{(1)} = \begin{bmatrix} \sum_i \varepsilon_{1i}\gamma_i\bar{h}_{i1} & \sum_i \varepsilon_{1i}\gamma_i h_{i2} & \cdots & X_1 \\ \sum_i \varepsilon_{2i}\gamma_i\bar{h}_{i1} & & & X_2 \\ \vdots & & & \vdots \end{bmatrix}$$

$$= \begin{bmatrix} \sum_j \sum_i \varepsilon_{1i}\gamma_i\bar{h}_{ij}X_j \\ \sum_j \sum_i \varepsilon_{2i}\gamma_i\bar{h}_{ij}X_j \\ \vdots \end{bmatrix} = \begin{bmatrix} \sum_i \gamma_i\varepsilon_{1i} \sum_j \bar{h}_{ij}X_j \\ \sum_i \gamma_i\varepsilon_{2i} \sum_j \bar{h}_{ij}X_j \\ \vdots \end{bmatrix}$$

$$= \gamma_1 \begin{bmatrix} \varepsilon_{11}\sum_j \bar{h}_{ij}X_j \\ \varepsilon_{21}\sum_j \bar{h}_{ij}X_j \\ \vdots \end{bmatrix} + \gamma_2 \begin{bmatrix} \varepsilon_{12}\sum_j \bar{h}_{2j}X_j \\ \varepsilon_{12}\sum_j \bar{h}_{2j}X_j \\ \vdots \end{bmatrix} + \cdots$$

If we now use $\hat{E}$, $\hat{F}$, and $\hat{H}^{(2)}$, then we can regress

$$\frac{\Delta a}{w} - (E\hat{H}^{(2)})\ X^{(2)} + \hat{E}\ \frac{a-1}{w} - \hat{F}\ \frac{\Delta \ell}{w} = \gamma_1 \begin{bmatrix} \hat{\varepsilon}_{11}\sum_j \bar{h}_{1j}X_j \\ \hat{\varepsilon}_{21}\sum_j \bar{h}_{1j}X_j \end{bmatrix} + \begin{bmatrix} \hat{\varepsilon}_{12}\sum_j \bar{h}_{2j}X_j \\ \hat{\varepsilon}_{22}\sum_j \bar{h}_{2j}X_j \end{bmatrix} + \cdots$$

as one equation, thereby obtaining unique estimates of the γ_i and hence of $H^{(1)}$.

---

13. These estimates will be inefficient to the extent that the restrictions imposed in the second step are correct.

Alternatively, using a similar procedure, it is now profitable to employ vertical restrictions on $H^{(1)}$:

$$H^{(1)} = \bar{H}^{(1)}C = \begin{bmatrix} \bar{h}_{11} & \bar{h}_{12} & \cdots \\ \bar{h}_{21} & & \\ \vdots & & \end{bmatrix} \begin{bmatrix} \gamma_1 & & 0 \\ & \gamma_2 & \\ 0 & & \ddots \end{bmatrix} = \begin{bmatrix} \gamma_1\bar{h}_{11} & \gamma_2\bar{h}_{12} & \cdots \\ \gamma_1\bar{h}_{21} & & \\ \vdots & & \end{bmatrix}$$

We can rewrite

$$EH^{(1)}X^{(1)} = E\bar{H}^{(1)}CX^{(1)} = \begin{bmatrix} \sum_i \varepsilon_{1i}\bar{h}_{i1} & \Sigma\varepsilon_{1i}\bar{h}_{i2} & \cdots \\ \sum_i \varepsilon_{2i}\bar{h}_{i1} & & \\ \vdots & & \end{bmatrix} \begin{bmatrix} \gamma_1 X_1 \\ \gamma_2 X_2 \\ \vdots \end{bmatrix}$$

$$= \begin{bmatrix} \gamma_1(\sum_i \varepsilon_{1i}\bar{h}_{i1})\, X_1 & \gamma_2(\sum_i \varepsilon_{1i}\bar{h}_{i2})\, X_2 & \cdots \\ \gamma_1(\Sigma\varepsilon_{2i}\bar{h}_{i1})\, X_1 & & \\ \vdots & & \end{bmatrix}$$

$$= \gamma_1 \begin{bmatrix} (\Sigma\varepsilon_{1i}\bar{h}_{i1})\, X_1 \\ (\Sigma\varepsilon_{2i}\bar{h}_{i1})\, X_1 \\ \vdots \end{bmatrix} + \gamma_2 \begin{bmatrix} (\Sigma\varepsilon_{1i}\bar{h}_{i2})\, X_2 \\ (\Sigma\varepsilon_{2i}\bar{h}_{i2})\, X_2 \\ \vdots \end{bmatrix} + \cdots$$

Using $\hat{E}$, $\hat{F}$ and $\hat{H}^{(2)}$, we can regress

$$\frac{\Delta a}{w} - (E\hat{H}^{(2)})X^{(2)} + \hat{E}\frac{a-1}{w} - \hat{F}\frac{\Delta \ell}{w} = \gamma_1 \begin{bmatrix} (\Sigma\hat{\varepsilon}_{1i}\bar{h}_{i1})X_1 \\ (\Sigma\hat{\varepsilon}_{2i}\bar{h}_{i1})X_1 \\ \vdots \end{bmatrix} + \gamma_2 \begin{bmatrix} (\Sigma\hat{\varepsilon}_{1i}\bar{h}_{i2})X_2 \\ (\Sigma\hat{\varepsilon}_{2i}\bar{h}_{i2})X_2 \\ \vdots \end{bmatrix} + \cdots$$

as one equation to obtain unique estimates of the γ_i.

2.4. Principal Components and Factor Analysis

2.4.1. Introduction

Principal Components and factor analysis have had long histories, but their practical applications have remained primarily limited to psychology. In 1957, Hotelling (who had made many significant contributions to component analysis) concluded that although these methods might have heuristic value, "some, but not all, of the things people have undertaken to do by (such analyses) can be done better in other ways" [32, pp. 69-70], such as by regression techniques.

Despite the authority of this pronouncement, principal components (and, less often, factor analysis) have been intermittently advocated as being possible sources of improvement in regression analysis -- particularly when severe multicollinearity problems are present. In this section I consider these suggestions, with my primary emphasis being on principal components since this is the more frequently suggested remedy.

Since the literature on component and factor analyses is very large and often poorly written and obscure,[14] I have included brief elementary descriptions of both techniques. I have used similar notations in both discussions since this is often done, although I do not (as some writers have done) consider component analysis to be just a special case of factor analysis -- as my introductory remarks to section 2.4.7. make clear.

---

14. Some better than average introductory discussions are in Kendall [37], Tintner [67, pp. 102-14], and Dhrymes [14, ch.2].

2.4.2. The Nature of Principal Component Analysis

Algebraically, in the equation

$$\underset{T\times 1}{Y} = \underset{T\times P}{X}\ \underset{P\times 1}{\beta} + \underset{T\times 1}{\varepsilon}$$

we want to find a basis P defined in terms of the p column vectors X_i

$$\underset{T\times p}{P} = \underset{T\times p}{X}\ \underset{p\times p}{M}$$

whose p elements (the column vectors of P) are uncorrelated with one another. Therefore, we want

$$D = P'P = M'X'XM$$

to be a diagonal matrix. But then the columns M_i of M are an orthonormal set of eigenvectors of X'X, and the diagonal elements of D are the eigenvalues, λ_i, (or characteristic roots or latent roots) of X'X.

If we order the eigenvalues and the associated eigenvectors so that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p,$$

then it can be shown that each of the P_i in turn extracts the maximum amount of the summed variances of the X_i, consistent with being orthogonal to the preceding P_i.

Stated differently, if we want to use K linear functions of the p X_i to estimate the X_i, and if ∇_i^2 is the residual sample variance associated with the estimation of each X_i by the best linear estimator based on the K functions, then $\sum_{i-1}^{p} \nabla_i^2$ is a measure of the overall success of the K linear functions and $\sum_{i=1}^{p} \nabla_i^2$ is a minimum when these functions are the first K principal components of $X'X$.[15]

2.4.3. The Application of Component Analysis to Singular Moment Matrices

If the rank of $X = m < p$, then $X'X$ will be singular, of rank m , and have m non-zero eigenvalues, and the m associated eigen vectors will capture all the variation in the X_i .

OLS cannot be applied to the untransformed data since $(X'X)^{-1}$ does not exist, and it is suggested[16] that principal components be utilized. To do this, we can define

$$F_i = p_{1i}\lambda_i^{-1/2} = XM_i\lambda_i^{-1/2} \;, \quad \lambda_i \neq 0$$

so that $F_i'F_i = P_i'P_i\lambda^{-1} = \lambda\lambda^{-1} = 1$, and thus $F_1, \ldots, F_m$ is an orthonormal basis of X .

---

15. Formally, if

$$\underset{Txp}{\hat{X}} = \underset{TxK}{(XB)}\underset{Kxp}{\hat{A}} = (XB)[(XB)'(XB)]^{-1}(XB)'X = XB[B'(X'X)B]^{-1}B'X'X \;,$$

Then

$$\begin{aligned}\sum_{i=1}^{p} \nabla_i^2 &= \sum_{i=1}^{p} \sum_{t=1}^{T} (X_i(t) - \hat{X}_i(t))^2 = \text{trace}[(X-\hat{X})'(X-\hat{X})] \\ &= \text{trace}\,(X'X - \hat{X}'X - X'\hat{X} + \hat{X}'\hat{X}) \\ &= \text{trace}\,[X'X - (X'X)B(B'(X'X)B)^{-1}B'(X'X)] \\ &= \text{trace}\,(X'X - \hat{X}'\hat{X})\end{aligned}$$

can be shown to be minimized when $XB = XM = \underset{TxK}{p}$ where M contains the K eigenvectors with the highest eigenvalues.

A stronger condition can also be shown: That $XB = P$ minimizes the Euclidean norm $||X'X - X'X|| = [\sum_{ij}\sum_{ij}\xi^2]^2$, where ξ_{ij} is an element of $(X'X - \hat{X}'\hat{X})$, which means that $XB = P$ will yield the $\hat{X}'\hat{X}$ which is "closest" to $X'X$.

16. e.g., Massy, [54], pp. 239-240.

We can substitute $X = FA$:

$$\underset{T\times 1}{Y} = \underset{T\times p}{X}\ \underset{p\times 1}{\beta} + \underset{T\times 1}{\varepsilon} = \underset{T\times m}{F}\ \underset{m\times p}{A}\ \underset{p\times 1}{\beta} + \underset{T\times 1}{\varepsilon} = \underset{T\times m}{F}\ \underset{M\times 1}{\gamma} + \underset{T\times 1}{\varepsilon}$$

from which we can obtain estimates of m parameters

$$\hat{\gamma} = (A\hat{\beta})$$

We cannot however solve these m equations for the p β_i.

More generally, there are an infinite number of possible m-dimensional bases for X, and for any such basis B

$$\underset{T\times p}{X} = \underset{T\times m}{B} \cdot \underset{M\times p}{C} = \underset{T\times p}{(X}\ \underset{p\times m}{D)}\ \underset{m\times p}{C}$$

so that we can substitute

$$Y = X\beta + \varepsilon = B \cdot C\,\beta + \varepsilon = B\,\pi + \varepsilon$$

and obtain estimates of m parameters, $\hat{\pi} = (C\hat{\beta})$, which are the original parameters transformed so that they represent the partial effect on Y of a change in the basis variables rather than a change in the observed X_i.

Each of these π_i must be interpreted very carefully, inasmuch as they assume that the other basis variables are constant and that the previous linear dependencies continue to hold, which will in general mean that some if not all of the X_i are changing in a prescribed way. With different bases, these prescribed movements will be different, and thus if one basis element, b_k, appears in two bases, (as will generally be possible) then the associated π_k will depend on the basis.[17] Thus, principal components are of no particular use here, excepting that they do detail a method for finding a basis, and the possibility that it is of interest to have the m parameters obtained correspond to orthogonal basis variables.

The first of these justifications is of little significance since there will always be m of the X_i which are linearly independent,[18] and such a basis is easily enough located. This will yield estimates of the m partial effects on Y of a change in one of the basis X_i, assuming that the other p-m X_i continue to show the same linear dependencies. If we happen to know p-m of the β_i corresponding to X_i such that the remaining m columns of X are linearly independent, then estimates of all the β_i are available.

The second possible justification also turns out to be unimportant, due to the fact that the coefficients assoc-

---

17. For example, in the model $Y=X_1\beta_1+X_2\beta_2+X_3\beta_3+\varepsilon$ with $X_1=X_2+X_3$, two possible bases are $[X_2X_3]$ and $[X_1X_3]$, with associated equations $Y=X_2(\beta_2+\beta_1) + X_3(\beta_3+\beta_1) + \varepsilon$

$$Y=X_1(\beta_2+\beta_1) + X_3(\beta_3-\beta_2) + \varepsilon$$

and different coefficients for X_3. The "reason" for this is that the first coefficient assumes $\Delta X_2=0$, which implies $\Delta X_1+\Delta X_3$, so that $\Delta Y|\overline{X}_2=\beta_3\Delta X_3+\beta_1\Delta X_1 = (\beta_3+\beta_1)\Delta X_3$; but the second coefficient assumes $\Delta X_1=0 \Rightarrow \Delta X_3=-\Delta X_2 \Rightarrow \Delta Y|\overline{X}_1=\beta_3\Delta X_3-\beta_2\Delta X_2= (\beta_3-\beta_2)\Delta X_3$.

18. If there were not m linearly independent X_i, then there would not be m linearly independent combinations of the X_i.

iated with any particular basis are easily located linear functions of the parameters associated with any other basis. This can be shown as follows:

Consider two bases, B_1 and B_2, described by

$$\underset{T\times p}{X} = \underset{T\times m}{B_1}\ \underset{m\times p}{C_1} = \underset{T\times p}{(X}\ \underset{p\times m}{D_1)}\ \underset{m\times p}{C_1}$$

$$X = B_2\ C_2 = (X\ \ D_2)\ \ C_2$$

By definition, there is some matrix G such that

$$\underset{T\times m}{B_1} = \underset{T\times m}{B_2}\ \underset{m\times m}{G}$$

and we know that G is a one-to-one transformation.[19]

Substituting,

$$Y=X\beta+\varepsilon=B_1C_1\beta+\varepsilon=B_1\gamma_1+\varepsilon=B_2G\gamma_1+\varepsilon$$

But also

$$Y=X\beta+\varepsilon=B_2C_2\beta=B_2\gamma_2+\varepsilon$$

so that

$$B_2G\gamma_1=B_2\gamma_2$$

$$B_2'B_2G\gamma_1=B_2'B_2\gamma_2$$

$$G\gamma_1=\gamma_2 \quad \text{and} \quad \gamma_1=G^{-1}\gamma_2$$

As for the estimated parameters,

$$\hat{\gamma}_1 = (B_1'B_1)^{-1}B_1'Y = (G'B_2'B_2'G)^{-1}G'B_2'Y$$

$$= G^{-1}(B_2'B_2)^{-1}(G')^{-1}G'B_2'Y = G^{-1}(B_2'B_2)^{-1}B_2'Y$$

$$= G^{-1}\hat{\gamma}_2$$

---

19. e.g., James [33], Thm. 12-5.2, p.737.

Finally, the implicit forcases of Y are independent of the basis:

$$\hat{Y}^{(1)} = B_1\hat{\gamma}_1 = B_1G^{-1}\hat{\gamma}_2 = B_2\hat{\gamma}_2 = \hat{Y}^{(2)}$$

These results demonstrate the unimportance of the particular m-dimensional basis chosen for the regression. For example, one might think that it would be advantageous to choose a basis that was orthogonal, in order to avoid multicollinearity problems. However, we've shown that the implicit estimates of the parameters associated with any basis are independent of the basis we use to estimate those parameters. That is, although basis B_k may consist of orthogonal variables, the use of any other basis (orthogonal or not) will yield identical estimates of the parameters associated with B_k or any other basis.

Although unimportant for the regression, the principal components may be of interest as variables per se. This is unlikely, though, since the forcasts are independent of the basis, and since each component is just a linear combination of X_i and probably qualitatively incomprehensible.

In the long history of component analysis, there have only occasionally appeared principal components which were effectively identified as variables in their own right, and cases where more than a few components were so identified have been very rare.[20]

2.4.4. The Application of Component Analysis to Nonsingular Moment Matrices

In the case where X is of full rank p, we can normalize the principal components matrix P, so that the new basis vectors F_i are of unit length simply by postmultiplying by $D^{-1/2}$.

$$\underset{Txp}{F} = \underset{Txp}{P}\ \underset{pxp}{D^{-1/2}} = \underset{Txp}{X}\ \underset{pxp}{M}\ \underset{pxp}{D^{-1/2}}$$

Now, $MD^{-1/2}$ is the inverse of A, the set of "normalized factor loadings" (or weights) appearing in the factor analysis equation X=FA and can be expressed as

$$A = (MD^{-1/2})^{-1} = D^{1/2}\ M^{-1} = D^{1/2}\ M'$$

With X of full rank, none of the eigenvalues (λ_i) are zero and consequently all of the factors are required to totally reproduce the variance of the X_i. Some of the factors are, however, usually dropped since one of the aims of factor analysis is a "considerable parsimony" in analysis. In particular, if the amount of variance accounted for by the

20. One such rarity is the oft-cited Stone [64], where three components captured 97.45% of the variation in seventeen series (variable such as employees' compensation, consumer expenditures, construction, net change in inventories, and foreign trade balance for the years 1922-38) and where these three components exhibited simple correlation coefficients of .995, .948 and -.836 with national income, change in national income and time, respectively. Kendall [37, pp. 26-7] in advocating translation to principal components, wrote that, " The remarkable feature of Stone's work, however, is that he was able to interpret his components." Dhrymes [14, pp. 64-5], another advocate, also called this identification "rather remarkable", and wrote that, "in general one does not or cannot interpret principal components in an intuitively meaningful way."

p-m components associated with the smallest eigenvalues is sufficiently small,[21] then these factors are often dropped. The closer the last p-m eigenvalues are to zero, the smaller the variance contained in those p-m dimensions, the closer X'X is to being of rank m, and the greater is the "multicollinearity problem."

One problem with this technique is that there is no reason why components which are powerful in explaining the variation in X will be useful in explaining Y (see for example Hotelling [32, pp. 74-5]), and thus the components that we drop because of their small eigenvalues might have been of great help in forcasting Y.

Another possibility is therefore to drop the components which add insignificantly to the explanation of Y in the regression of Y on all of the components. Unfortunately, this criteria may lead us to discard much of the variation in X, and thereby aggravate rather than alleviate the collinearity problem. This point is illustrated in section 2.4.5. below.

If all p of the components are retained, then the implicit estimates β^* of the structural parameters will be identical to those estimates, $\hat{\beta}$, which would have resulted

21. There are approximate significance tests for the null hypothesis that $p-m_o$ components are insignificant (see for example Lawley & Maxwell [47, pp. 22-7]), but for the different problem of selecting a value for p-m, we do not know which procedure and criterion are most likely to lead to a correct choice (see Anderson, [5, pp. 137-8]).

from a direct estimation using the original X_i. This is an immediate corollary of the Gauss-Markov theorem, but can be directly shown quite easily. Our model is

$$\underset{Tx1}{Y} = \underset{Txp}{X}\ \underset{px1}{\beta} + \underset{Tx1}{\varepsilon} = \underset{Txp}{F}\ \underset{pxp}{A}\ \underset{px1}{\beta} + \varepsilon = \underset{Txp}{F}\ \underset{px1}{\gamma} + \varepsilon$$

where $F = XMD^{-1/2} \Rightarrow Y = XMD^{-1/2}\gamma + \varepsilon$ so that

$$X\beta = XMD^{-1/2}\gamma$$

$$X'X\beta = X'XMD^{-1/2}\gamma$$

$$\beta = MD^{-1/2}\gamma$$

The implicit estimate of β is therefore

$$\beta^* = MD^{-1/2}\hat{\gamma} = MD^{-1/2}\ (F'F)^{-1}\ F'Y = MD^{-1/2}I_p^{-1}\ (XMD^{-1/2})'Y$$

$$= MD^{-1/2}\ D^{-1/2}\ M'X'Y = MD^{-1}M'X'Y = (X'X)^{-1}\ X'Y = \hat{\beta}$$

If p-m of the components are dropped, then our model is

$$\underset{Txp}{Y} = \underset{Txp}{X}\ \underset{px1}{\beta} + \underset{Tx1}{\varepsilon} = [\underset{Txm}{F^{(1)}} \vdots \underset{Txp-m}{F^{(2)}}] \begin{bmatrix} \underset{mxp}{A^{(1)}} \\ \underset{p-mxp}{A^{(2)}} \end{bmatrix} \beta + \varepsilon$$

$$= \underset{Txm}{F^{(1)}}\ \underset{mxp}{A^{(1)}}\beta + \underset{Txp-m}{F^{(2)}}\ \underset{p-mxp}{A^{(2)}}\ \beta + \varepsilon$$

$$= F^{(1)}\ \underset{px1}{\gamma} + (U\beta + \varepsilon)$$

$$= \underset{Txp}{X}\ \underset{pxm}{M^{(1)}}\ \underset{mxm}{D^{(1)-1/2}}\ \underset{mx1}{\gamma} + (U\beta + \varepsilon)$$

Kendall [37, pp. 70-5] proposes regressing Y on $F^{(1)}$ to obtain $\hat{\gamma}$ and then calculating

$$\beta^* = M^{(1)}\ D^{(1)-1/2}\ \hat{\gamma}$$

But this seems to worsen matters, since β is not equal to $M^{(1)}\ D^{(1)-1/2}\ \gamma = M^{(1)}M^{(1)}\beta'$, but rather

$$\beta = M^{(1)}D^{(1)-1/2}\gamma + M^{(2)}D^{(2)-1/2}A^{(2)}\beta = M^{(1)}M^{(1)'}\beta$$
$$+ M^{(2)}M^{(2)'}\beta$$

Further, in our regression we've discarded possibly useful information in $U\beta$, not because of insignificance in explaining Y, but rather because of $F^{(2)}$'s weakness in explaining X. Or, alternatively, we will discard significant variation in X which can lead to even more uncertain parameter estimates than if we had used OLS.

The bias in our estimates will be[22]

$$\beta - E(\beta^*) = M^{(2)} M^{(2)\prime} \beta$$

with variance-covariance matrix.

$$COV(\beta^*) = \nabla^2 \; M^{(1)} D^{(1)-1} M^{(1)\prime}$$

as compared to this, the OLS estimates will be unbiased and have $COV(\hat{\beta}) = \nabla^2 \; (X^1X)^{-1} = \nabla^2 M D^{-1} M'$

$$= \nabla^2 M^{(1)} D^{(1)-1} M^{(1)\prime} + \nabla^2 \; M^{(2)} D^{(2)-1} M^{(2)},$$

So that

$$COV(\hat{\beta}) - COV(\beta^*) = \nabla^2 M^{(2)} D^{(2)-1} M^{(2)\prime}$$

which is positive semi-definite. One implication of this is that the diagonal elements (which are $VAR\;(\hat{\beta}) - VAR\;(\beta^*)$) are all nonnegative.

These smaller variances may offset the bias, so that the mean squared errors (MSE) of β^* are smaller than those of $\hat{\beta}$. It can be shown that $MSE\;(\beta^*) - MSE(\hat{\beta})$ = the diagonal elements of $M^{(2)} M^{(2)\prime} [\beta\beta' - \nabla^2 (X'X)^{-1}] \; M^{(2)} M^{(2)\prime}$. If $\beta\beta' - \nabla^2 (X'X)^{-1}$ were positive or negative semi-definite, then

22. See Appendix III-B for the derivation of the bias, var-cov matrix, and mean squared error.

$M^{(2)}M^{(2)'}$ $[\beta\beta' - \nabla^2(X'X)^{-1}]$ $M^{(2)}M^{(2)'}$ would be also. But $\beta\beta' - \nabla^2(X'X)^{-1}$ is not semi-definite.

For our purposes, it is only important to show that MSE (β^*) can be less than MSE $(\hat{\beta})$; this is in fact true since $(X'X)^{-1}$ can explode while $\beta\beta'$ remains finite. However, without knowing $\beta\beta'$ and ∇^2 we cannot tell which term will actually be smaller, nor can we tell which components to use in order to get small MSE's.

Another point to bear in mind is that our transformation to a smaller dimensioned space and back again necessitated p-m linear restrictions on the β_i. Our definition

$$\underset{m\times 1}{\gamma} = \underset{m\times p}{(D^{(1)1/2}M^{(1)})}\underset{p\times 1}{\beta}$$

maps the p-dimensional vector β onto the m-dimensional vector γ. But knowing γ (or an estimate of γ) cannot yield a unique β (or $\hat{\beta}$) since there are an infinite number of β which will be mapped to each particular γ.[22]

To overcome this, we can use Kendall's "incorrect" transformation

$$\underset{p\times 1}{\beta} = \underset{p\times m}{(M^{(1)}\ D^{(1)-1/2})}\underset{m\times 1}{\gamma}$$

which "solves" our problem by putting p-m linear restrictions on the β_i.

22. Equivalently, we have defined m equations in the p unknown β_i, and these cannot be solved uniquely.

We can explicitly see that there are these restrictions by rewriting the transformation as

$$\underset{p\times 1}{\beta} = \underset{p\times m}{H}\ \underset{m\times 1}{\gamma}$$

H has at most m linearly independent rows. If we rearrange terms so that the first m rows are independent, then we can partition

$$\begin{bmatrix} \underset{m\times 1}{\beta^{(1)}} \\ \underset{p-m\times 1}{\beta^{(2)}} \end{bmatrix} = \begin{bmatrix} \underset{m\times m}{H^{(1)}} \\ \underset{p-m\times m}{H^{(2)}} \end{bmatrix} \underset{m\times 1}{\gamma}$$

and solve for

$$\underset{p-m\times 1}{\beta^{(2)}} = H^{(2)}\gamma = \underset{p-m\times m}{(H^{(2)}\ H^{(1)-1})}\ \underset{m\times 1}{\beta^{(2)}}$$

which are p-m restrictions on the β_i.

In summary, the main advantage of applying principal components here is that the variances can be kept small and that in some cases the MSE's are less than would be the case with OLS. But in a non-exploratory situation, it doesn't make much sense to throw away valuable information in order to preserve small variances. Nor does it seem very wise to apply non-theoretical restrictions to the parameters by means of an incorrect relationship. Instead, it would seem more reasonable to apply a priori information since this will reduce variances, without necessitating a discarding of information.

In an exploratory situation, it is true that we will not have a priori information to apply -- but then we will have little idea of what the matrix $\beta\beta' - \nabla^2 (X'X)^{-1}$ is like and therefore no knowledge about which technique will lead to smaller MSE's. In addition, we will not be able to tell how seriously wrong the p-m restrictions on the β_i are.

These comments are illustrated in the following example.

2.4.5. A Two-Component Example

Let

$$Y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon = [X_1 \ \ X_2] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \varepsilon = X\beta + \varepsilon$$

with all variables standardized to zero mean and unit length. The correlation matrix is

$$(X'X) = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} \qquad [r = X_1'X_2]$$

which has the characteristic roots and vectors

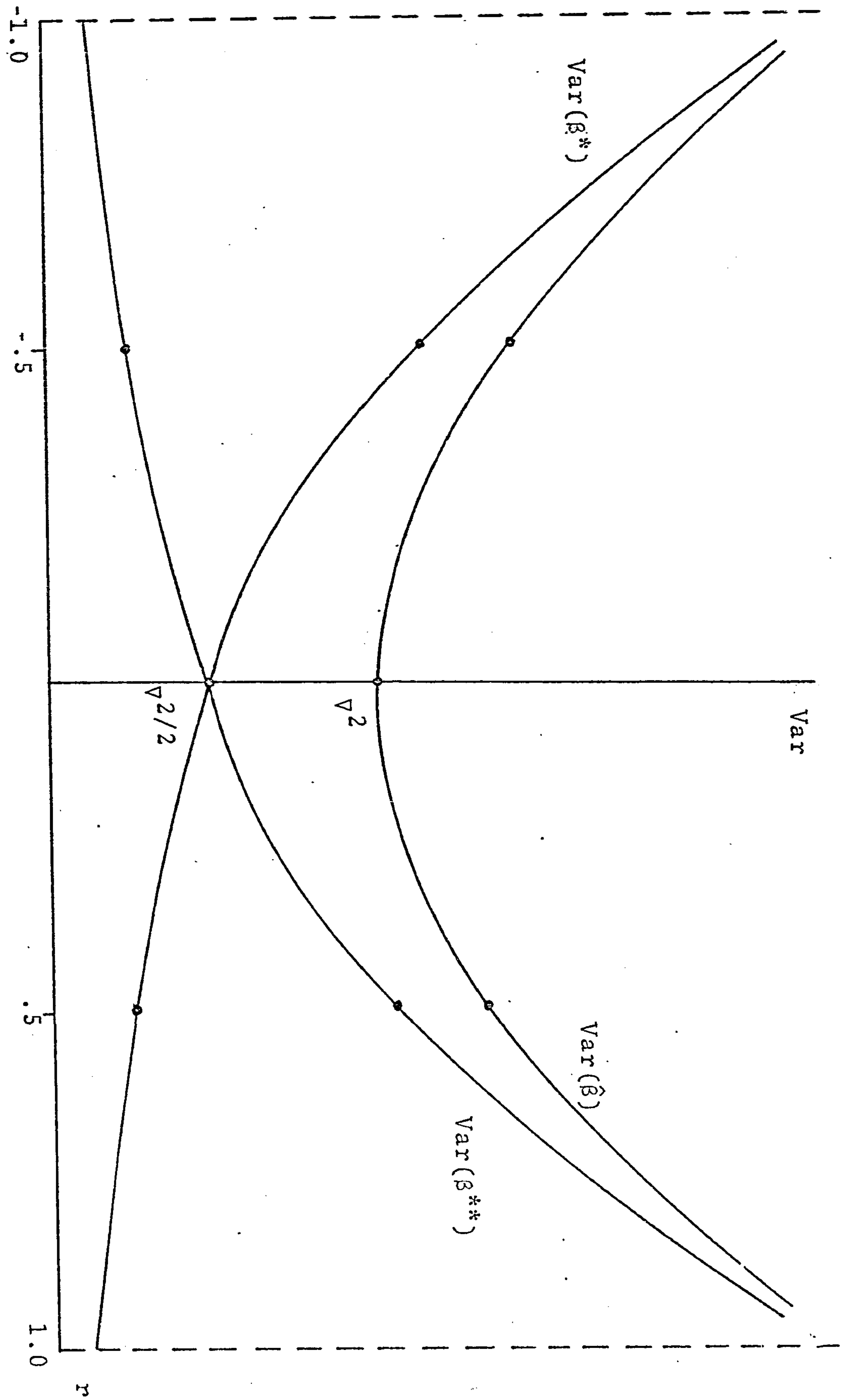
$$\lambda_1 = 1 + r \qquad M^{(1)} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \qquad M^{(2)} = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

$$\lambda_2 = 1 - r$$

With a properly behaved disturbance term, our three possible sets of estimates can be summarized by

	Data Used		
	X	$F^{(1)}=XM^{(1)}\sqrt{1/\lambda_1}$	$F^{(2)}=XM^{(2)}\sqrt{1/\lambda_2}$
Symbol used to identify method	$\hat{\beta}$	β^*	β^{**}
Estimate of β_1	$(X_1'Y-rX_2'Y)/(1-r^2)$	$\frac{1}{2}\,\frac{1}{1+r}(X_1'Y+X_2'Y)$	$\frac{1}{2}\,\frac{1}{1-r}(X_1'Y-X_2'Y)$
Estimate of β_2	$(X'Y-rX'Y)/(1-r^2)$	β_1^*	$-\beta_1^{**}$
Bias (β_1)	None	$\frac{1}{2}\ (\beta_1 - \beta_2)$	$\frac{1}{2}\ (\beta_1 + \beta_2)$
Bias (β_2)	None	$\frac{1}{2}\ (\beta_2 - \beta_1)$	$\frac{1}{2}\ (\beta_1 + \beta_2)$
Variance (β_1)	$\nabla^2/(1-r^2)$	$(\nabla^2/2)/(1+r)$	$(\nabla^2/2)/(1-r)$
Variance (β_2)	Same	Same	Same
MSE (β_1)	$\nabla^2/(1-r^2)$	$\frac{(\beta_1-\beta_2)^2}{4}+\frac{(\nabla^2/2)}{1+r}$	$\frac{(\beta_1+\beta_2)^2}{4}+\frac{\nabla^2/2}{1-r}$
MSE (β_2)	Same	Same	Same

The variances of the three estimates are graphed on the following page.



-1.0
-.5
.5
1.0
r
Var
Var(β*)
Var(β̂)
Var(β**)

The MSE of $\hat{\beta}$ will be the same as var $(\hat{\beta})$, while MSE (β^*) and MSE (β^{**}) will be equal to the variances shifted upward by $(\beta_1-\beta_2)^2/4$ and $(\beta_1+\beta_2)^2/4$, respectively.

If we use the component which best explains Y, then we could easily have larger MSE's in our parameter estimates than would result from OLS.

If, on the other hand, we use the component which has the largest characteristic root, then we will choose according to the sign of r:

$$0<r\leq 1 \quad \Rightarrow \quad \lambda_1>\lambda_2$$

$$r=0 \quad \Rightarrow \quad \lambda_1=\lambda_2$$

$$-1\leq r<0 \quad \Rightarrow \quad \lambda_1<\lambda_2$$

This means that we will always use the principal component estimator which has the smallest variance, though not necessarily the smallest MSE. On the other hand, the power of these estimates in explaining Y may drop dramatically.[23]

A primary determinant of how much larger than the variance the MSE will be is the size of β_1 and β_2. The closer β_1 is to β_2, the smaller the bias of β^* will be and the closer MSE(β^*) will be to var (β^*). On the other hand, the closer β_1 is to $-\beta_2$, the smaller will be the bias of β^{**} and the closer MSE (β^{**}) will be to var (β^{**}). This is intuitively appealing because using the first component $(F^{(1)})$ forces

23. In fact, may disappear entirely; see Hotelling [32, p. 75]

$\beta_1{}^* = \beta_2{}^*$, while using the second component forces $\beta_1{}^{**} = -\beta_2{}^{**}$.

This suggests the rule: for r close to one and β_1 known a priori to be close to β_2, use β^*; for r close to -1 and β_1 known to be close to $-\beta_2$, use β^{**}. But there is another clue. If we use OLS with the constraint $\hat{\beta}_1 = \hat{\beta}_2$

$$Y = (X_1 + X_2) \quad \beta_1 + \varepsilon$$

Then we will obtain

$$\hat{\beta}_i = \frac{X_1'Y + X_2'Y}{2\ (1+r)} = \beta_i^*$$

And if we use OLS with the constraint $\hat{\beta}_1 = -\hat{\beta}_2$: $Y = (X_1 - X_2)\beta_1 + \varepsilon$
Then we will obtain

$$\hat{\beta}_i = \frac{X_1'Y - X_2'Y}{2\ (1+r)} = \beta_i^{**} \quad \text{and} \quad \hat{\beta}_2 = \hat{\beta}_2^{**}$$

This suggests a more reasonable rule: somewhat uncertain a priori information can be very beneficial in collinearity situations, and is more flexible than component analysis which may result in more uncertain parameter estimates and/or poorly fitting equations, and will result in possibly unintended parameter restrictions.

2.4.6. Other Uses of Component Analysis

Although I have been pessimistic about the potential of component analysis in alleviating collinearity problems, I do think that there are several areas in which principal components can be interesting and useful.

(i) The translation of a data set into its principal components can provide an easily understood description of the extent of the collinearity in the sample data. If almost all of the variation in the p explanatory variables is contained in k dimensions, then we should not expect reliable estimates of more than k parameters, and can expect to have to resort to at least p-k theoretically inspired restrictions on the date in order to obtain good estimates of all p β_i.

It is true that we do not know how to best estimate the truly significant number of components (see note 21, p. 96 above). But rather than an exact representation of the underlying population, all we seek here is an approximate appraisal of the data set. Our goal will not often be defeated if we choose some significance level α and then find the smallest value of k for which the remaining p-k components are insignificant.[24]

It is also true that we need to know much more than the effective dimensionality of the data. In particular, the tests of section 1. can lead us to the places most in need of subjective information -- since the indiscriminate application of a priori information may have little effect on a collinearity problem.[25]

---

24. For more elegant though time-consuming tests, see Jöreskog [35, Ch. IV].

25. For example, if X_1 and X_2 are very collinear, then the independent variables in the equation

$$Y=X_1\beta_1+X_2\beta_2+X_3\beta_3+\varepsilon$$

will have about 2 significant dimensions. Putting in a subjective value for β_3 will yield

$$Y-\overline{\beta}_3X_3=X_1\beta_1+X_2\beta_2+\varepsilon$$

which has 2 parameters to estimate and about one significant dimension.

(ii) If the number of explanatory variables, p, is greater than T, the number of observations, then $(\underset{p\times p}{X'X})^{-1}$ does not exist (since its rank will be no greater than T<p), and OLS estimates cannot be calculated. We have seen in Chapter II that this is a very real possibility for this model.

If our model does have this problem, one solution is to aggregate some groups of assets and use the first principal component of the associated rates as an index rate for each group.

An extreme example would occur if we attempted to list each equity that a sector could purchase. We obviously do not have enough quarterly observations to estimate the partial effects of a change in each equity price (or rate of return). But we could aggregate all equities and use one index or, if this was unsatisfactory, group the stocks into a small number of subsets, each with its own index.[26]

Another potential source of large savings in the number of parameters would occur if we tried to include a long list of bonds of different maturities or from different sources. While not as spectacular, we can also use this technique to

26. Feeney & Hester's [18] component analysis of the 30 stocks included in the Dow Jones Industrial Average found that the first components captured 76% of the variance in stock prices but only 40% of the variance in rates of return, and that these components were highly correlated (R=.992 and .931, respectively) with the Dow Jones Price Average and a constructed Dow Jones return average.

group small numbers of similar assets.

In general, we clearly want to form groups of assets whose rates move closely together, since these will be cases in which one component can accurately reflect a good deal of the variation in the rates of return available on member assets.

This brings to mind the rate structure approach which uses one rate in place of several because the other relevant rates are so strongly correlated with this rate. This technique must however be inferior to using the first component of these rates since we've seen that this is the linear combination of the rates which these rates are (weighted equally) most closely related to. More specifically, if we have K rates, r_i, which we want to summarize by one linear combination of the rates

$$I = \sum_{i=1}^{K} b_i r_i$$

and if V_i^2 is the variance in r_i which is not explained by its correlation with I, then $\sum_{i=1}^{K} V_i^2$ is minimized when I is the first principal component.

(iii) We cannot calculate K-class estimators using all of the predetermined variables in the system if there are more than T such variables -- since the moment matrix for the predetermined variables will be singular. One way around this problem is to use just the first K (<T) principal components of the predetermined variables to provide estimates of the RHS endogonous variables.

It is unfortunately unclear how we should choose K. Identification requires that K be at least as large as the number, m, of RHS endogenous variables. And we know that asymptotically it won't matter where K is between m and T. But for small samples, it will matter -- with a smaller number being less able to explain the endogonous variables and a larger number using up degrees of freedom (causing the estimates of the disturbances to become unreliable and increasing the bias).

Dhrymes [14] suggests $K \approx \sqrt{T}$ while Kloek and Mennes [43] suggest $\ell + K \leq T/3$, where ℓ is the number of included predetermined variables.[27] Alternatively, after the characteristic roots have actually been extracted, it may become clear that some moderately sized number of components captures almost all of the variance in the predetermined variables. One would probably be satisfied with this K unless it grossly violated the Dhrymes or Kloek-Mennes rules.

27. Klein [41] reports for one of his models, with T=31 and ℓ never greater than 3, that 4 components did better than 8 and that experiments with 6, 10 and 12 components showed "no improvement" over 4. Dhrymes' rule would have suggested $K \approx 5$ or 6, and the Kloek-Mennes rule implied $K \leq 7$.

2.4.7. Factor Analysis

John Scott [62, 63] has advocated the use of factor analysis regression "whenever there are errors in variables and when high intercorrelation causes least squares procedures to break down."

The factor analysis model can be written as

$$\underset{Txp}{Z} = \underset{Txm}{F}\ \underset{mxp}{A} + \underset{Txp}{U}$$

where

Z = matrix of observed variables

F = matrix of unobserved factors

A = factor loading matrix (constants)

U = matrix of error terms

and

$E(F) = E(U) = 0 \quad F'F = I \quad U'U = \Sigma$, a diagonal matrix.

There is a strong resemblance here to the principal components model, and some writers[28] have treated the latter as just a special case of the former, despite the presence of significant conceptual differences.[29]

---

28. For example, Scott [62, pp. 553-4]. Interestingly, Anderson [1, Ch. 11] does this, but then elsewhere [2, p.8] makes the appropriate distinction and writes that the term "factor analysis" is often used in a very loose sense, particularly by "statisticians not well acquainted with statistics as used in psychology."

29. Such differences have been stressed for example by Bartlett [8, pp. 38-41], Harman [30], and Kendall [37, p.37].

The computation of the principal components of a given data set is a purely mathematical manipulation which is intended to reveal interdependencies which are not very well understood but regarding which one eventually hopes to construct a model. In factor analysis, on the other hand, one is trying to test a postulated model in which underlying unobservable traits or phenomena generate observable variates.[30] Reflecting this difference, the factor analysis model has a specific or unique term (which is not just p-m discarded dimensions), and the factors are chosen to maximally reproduce the correlations between the observed variables rather than their variances.

For the regression equation

$$Y = X\beta + \varepsilon \ ,$$

Scott advocates setting up a factor analysis model in which the matrix of observed variates includes both Y and X:

$$\underset{T\times 1}{(Y} \quad \underset{T\times p-1}{X)} = Z = F\,A + U = \underset{T\times m}{F} \; (\underset{m\times 1}{A^{(1)}} \; \underset{m\times p-1}{A^{(2)}}) + (\underset{T\times 1}{U^{(1)}} \; \underset{T\times p-1}{U^{(2)}})$$

We are not given any advice on how to choose m (the number of factors),[31] but we are advised to estimate A by whichever of six recounted methods (including principal components) appeals to us.

30. In psychology, test scores might reflect such postulated traits as general intelligence, mathematical aptitude, and verbal aptitude. In economics, plant and equipment expenditures might depend on such unobserved factors as sales and price expectations, technological opportunities, and credit availability.

31. See note 21 above.

An estimate of F could be obtained from

$$F = (Z-U)\,A'(AA')^{-1}$$

if we had an estimate of U; but we cannot estimate U without knowing F. One possibility is to ignore U and use

$$F = Z\,\hat{A}'\,(\hat{A}\hat{A}')^{-1}.$$

Scott recommends using instead the procedure first proposed by Thomson:[32]

$$\hat{F} = Z\,(Z'Z)^{-1}\,\hat{A}'$$

expanding this expression to

$$\hat{F} = (Y\ X)\begin{bmatrix}\hat{G}^{(1)}\\ \hat{G}^{(2)}\end{bmatrix} = Y\hat{G}^{(1)} + X\hat{G}^{(2)}$$

and substituting into $Y = FA^{(1)} + U^{(1)}$ yields

$$Y = Y\hat{G}^{(1)}\hat{A}^{(1)} + X\hat{G}^{(2)}\hat{A}^{(1)} + U^{(1)}$$

$$= X\hat{G}^{(2)}\hat{A}^{(1)}(I-\hat{G}^{(1)}\hat{A}^{(1)})^{-1} + U^{(1)}(I-G^{(1)}A^{(1)})^{-1}$$

Scott therefore proposes

$$\beta^{**} = \hat{G}^{(2)}\hat{A}^{(1)}(I-\hat{G}^{(1)}\hat{A}^{(1)})^{-1}$$

which can be shown to be equivalent to

32. If we postmultiply the right hand side of this equality by $[I+\hat{A}(\hat{U}'\hat{U})^{-1}\hat{A}'][I+\hat{A}(\hat{U}'\hat{U})^{-1}\hat{A}']^{-1}$ and substitute $\hat{A}'\hat{A}=Z'Z-\hat{U}'\hat{U}$, it will reduce to

$$\hat{F}=Z(\hat{U}'\hat{U})^{-1}\hat{A}'[I+\hat{A}(\hat{U}'\hat{U})^{-1}\hat{A}']^{-1}$$

King [39] is very suspicious of factor analysis regression, but says that Scott would have done better to use Bartlett's

$$\hat{F}=Z(\hat{U}'\hat{U})^{-1}\hat{A}'[\hat{A}(\hat{U}'\hat{U})^{-1}\hat{A}']^{-1}.$$

Lawley and Maxwell [47, Ch. 7] references both of these procedures and discusses their derivation. See also Bartlett [8, pp. 42-4].

$$\beta^{**}=\hat{\beta}-(X'X)^{-1}\hat{U}^{(2)}{}'\hat{U}^{(1)}(\hat{U}^{(1)}{}'\hat{U}^{(1)}-Y'X'(X'X)^{-1}\hat{U}^{(2)}{}'\hat{U}^{(1)})^{-1}$$

$$(Y'Y-Y'X(X'X)^{-1}X'Y)$$

where $\hat{\beta}$ is $(X'X)^{-1}X'Y$, the OLS estimate of β.

Asymptotically, $\hat{U}^{(2)}{}'\hat{U}^{(1)}$ should approach zero, and β^{**} will consequently converge to $\hat{\beta}$. In small samples, however, the way in which this estimate differs from $\hat{\beta}$ is very complicated and heavily dependent on the errors, $U^{(1)}$ and $U^{(2)}$, in the factor analysis equation.

These error terms depend of course on the method chosen to estimate A, which should disturb at least those followers of Scott who aren't certain about which of the six methods to use to estimate A. More distressing, though, is the nature of these error terms as compared to their ultimate use here. First, the transformation of the data to a smaller dimensioned space of artificial factors is a discarding of information, which should aggrevate collinearity problems. Second, this information is discarded not on the basis of its inability to explain Y, but according to its inability to explain the correlation between all Z_i.[33] Third, the use of a factor analysis model ignores any a priori notions of causation which we may possess, and treats all variables as if they were generated by underlying factors. This is an especially dubious procedure when some of the X_i are in fact exogeneous, as in such extreme cases as with time and dummy variables.

33. Hotelling [32] discusses and Massy [54, p. 241] stresses the fact that these are very different criteria. See also my 2.4.3. above.

There is, however, another justification for a factor analysis model which does modify some of these difficulties. It has been known for some time [50], that the factor analysis model is mathematically equivalent to a situation in which p-m variables (observed with error) are generated by the other m variables (observed with error).

For example, if

$$\underset{Txp}{Z} = \underset{Txm}{F}\ \underset{mxp}{A} + U,$$

then we can find m linearly independent columns of A and partition accordingly:

$$(\underset{Txm}{Z^{(1)}}\ \underset{Txp-m}{Z^{(2)}}) = \underset{Txm}{F}\ (\underset{mxm}{A^{(1)}}\ \underset{mxp-m}{A^{(2)}}) + (\underset{Txm}{U^{(1)}}\ \underset{Txp-m}{U^{(2)}})$$

thus

$$F = (Z^{(1)}-U^{(1)})\ (A^{(1)})^{-1}$$

and

$$Z^{(2)}-U^{(2)} = FA^{(2)} = (Z^{(1)}-U^{(1)})\ (A^{(1)})^{-1}\ A^{(2)}$$

or

$$Z^{(2)*} = Z^{(1)*}\ (A^{(1)})^{-1}\ A^{(2)}$$

King [34] made the partly correct argument that if this is our justification for using factor analysis, then

34. King is however overly concerned about the fact that we can choose different partitionings of A and hence different errors in variables models with associated different coefficients to be interpreted as marginal effects of one variable on another. This is exactly the same problem as I discussed earlier and involves nothing more than the fact that the partial derivatives of different equations assume that different variables are being held constant.

we should be ultimately interested in an estimate of $(A^{(1)})^{-1}A^{(2)}$. This is true if our interest is in knowing the structural parameters, but is not what we want if our goal is prediction, either of $Z^{(2)}$ or of $Z^{(2)*}$. This is so because $(A^{(1)})^{-1}A^{(2)}$ describes the marginal effects of changes in $Z^{(1)*}$, and if we use these coefficients to predict the marginal effects of changes in $Z^{(1)}$ then we will introduce an error equal to $-U^{(1)}(A^{(1)})^{-1}A^{(2)}$. This error will vanish--so that $E(\hat{Z}^{(2)}) = E(Z^{(2)})$ -- only for those forcast periods in which $Z^{(1)}=E(Z^{(1)*})$[35]

If we do believe in this error in variables justification, but are concerned most about prediction, then we should set up our system as p-m structural equations

$$Z^{(2)*} = Z^{(1)*} (A^{(1)})^{-1} A^{(2)}$$

and estimate these relations with $Z^{(2)}$ and $Z^{(1)}$ substituted for $Z^{(2)*}$ and $Z^{(1)*}$.

Returning to our original model, if we believe that Y is a function of X, and that all variables are observed with error, then we should use[36]

$$(Y - U^{(2)}) = (X - U^{(1)}) \beta$$

which has the corresponding factor analysis model

35. See Johnston [34, pp. 162-4] for the details of the two variable case.

36. This unrealistically ignores ε, but is typically done.

$$\underset{T\times p-1}{(X} \vdots \underset{T\times 1}{Y)} = \underset{T\times p-1}{F} \; (\underset{p-1\times p-1}{A^{(1)}} \vdots \underset{p-1\times 1}{A^{(2)})} + (\underset{T\times p-1}{U^{(1)}} \vdots \underset{T\times 1}{U^{(2)})}$$

so that m = p-1. This reduces, as before, to

$$Y-U^{(2)} = (X-U^{(1)}) \; (A^{(1)})^{-1} \; A^{(2)}$$

so that we should be estimating this $(A^{(1)})^{-1}A^{(2)}$ if we are interested in structural parameters, and estimating

$$Y = X\,\beta + \varepsilon$$

if we are more concerned with forcasting.

If we use principal components to estimate A, then we've implicitly imposed the restriction that E(UU') = $\nabla^2 I$ [14, p. 80]; there is no need to estimate F by Scott's or Bartlett's formulas since F will be simply the first m columns of XA^{-1} [see 2.4.4. above]; and our approach is equivalent to the classical errors in variable approach when E(UU') is known [50].

If we use an alternative estimator of A and Bartlett's formula for F (note 32, p.112), then it can be shown that using the rest of Scott's procedure will result in

$$\hat{\beta} = (A^{(1)})^{-1} \; A^{(2)}$$

If we instead use Scott's formula for F, then we should get worse structural estimates but may get better predictive estimates. The trouble is that we have no reason to expect this, since we're still discarding information in the dropping of one factor, and doing so according to a dubious criterion. Another remaining problem is the forcing of such things as time and dummy variables, which are observed without error, to depend inexactly on artificial factors. One solution is to constrain the associated U_i to be zero; alternatively,[37] Anderson [2, pp. 11, 15-8] suggests regressing all variables observed with error on all variables observed without error, and then applying a factor analysis to the residuals.

37. These approaches are equivalent if the maximum likelihood estimate of Λ is used [2, p. 17].

B. SIMULTANEITY

Very briefly, the OLS regression of an equation which has endogenous[38] variables on the right hand side will yield biased and inconsistent parameter estimates. If the equation is just identified, then translation to its reduced form will yield unbiased, consistent, maximum likelihood (assuming normality) estimates of the reduced form parameters and biased but consistent, maximum likelihood estimates of the structural parameters. If the structural equation is underidentified, then its parameters cannot be estimated without the introduction of a priori information (such as parameter restrictions) which turns the equation into an identified one. With an overidentified equation, we cannot get from the reduced form back to the structural parameters, and it is at least aesthetically preferable to estimate the structural equation directly with a K-class estimator such as OLS, Two-Stage Least Squares (2SLS), or Limited Information (LI).[39]

Since I have interpreted the primary objective here to be accurate forcasting (in either actual or postulated

38. Strictly speaking, we are only concerned with variables which are correlated with the disturbance term. In recursive systems, for example, variables endogenous to the system will be independent of many of the disturbances.

39. The term LI here includes the three equivalent methods: LI/maximum likelihood, LI/least variance ratio, LI/ least generalized variance.

situations, it might be thought that we should only be interested in good estimates of the reduced form parameters. However, a minimal requirement should be that we do not contradict our postulated model -- since it doesn't make much sense to forcast with reduced form parameters which are inconsistent with the theory that was used to justify those reduced forms. Further, if we are able to calculate the structural parameter estimates and find that some of them are nonsensical -- but we really believe our model -- then we can constrain those parameters to be acceptable and thereby presumably avoid being misled by the particular data available. If, for example, some highly collinear data yields very uncertain parameter estimates, then we can help fill information gaps in the data by imposing restrictions on the structural parameters. Also, in cases where we could anticipate structural changes,[40] knowledge of the structure is obviously valuable.[41]

Finally, there is good reason to believe that the application of OLS to the reduced form (LSRF) in an overidentified situation is not even a good technique for prediction purposes. Klein [40] has shown that LSRF is asymptotically no more efficient and possibly less efficient than

40. This will most likely occur with policy simulations, such as the imposing or dropping of interest rate ceilings.

41. A very good illustrative discussion is Marschak (53).

K-class techniques, because LSRF does not utilize the a priori information that we presumably have about which parameters belong in the structural equation. This implies that the relative forcasting performance of LSRF is going to depend heavily on how correctly the structural equation is specified--and indeed, in Monte Carlo studies[42] LSRF has been very weak in correctly specified models and done better in misspecified models.

This suggests that how well LSRF performs should be a good measure of how correct we are in excluding some predetermined variables from the structural equation,[43] with the limit being the case in which the equation is really just identified -- since here LSRF will be the maximum likelihood, best linear unbiased estimator of the reduced form parameters -- and will be identical with 2SLS or LI applied to the correctly specified structural equation.

Turning to a comparison of the K-class estimators, 2SLS and LI -- unlike OLS -- are both consistent estimators, and in fact are asymptotically equivalent. Further, it can be shown [9] that they are the "best linear consistent" estimators. Small sample properties have not been firmly

---

42. In Johnston [34, Ch. 10], see the oft-quoted Summers' study and Klein's calculations with Ladd's data.

43. Thus, because Liu [48] believes that most economic relationships are truly underidentified but are misspecified to be overidentified, he concludes that LSRF's are likely to be superior forcasting equations.

established analytically,[44] but Monte Carlo studies have given us some apparent patterns.[45]

LI seems to have not often given the best results, and then usually to have done so only marginally. On the other hand, LI can be very unstable, especially in the face of collinearity problems. 2SLS estimates also seem to be somewhat unstable, but noticeably less so than LI.[46] OLS estimates are generally the most biased, but seldom have large variances.[47] In cases where there are very few true degrees of freedom (i.e., in relatively small and/or collinear samples), OLS's smaller, finite variance may offset its larger bias so that its mean squared error is the smallest of the three. These apparent relationships have prompted the Quandt-Tukey

---

44. See Dhrymes [14, p. 358] for references on some exact results that have recently been obtained. Good summaries of several studies can be found in Christ [11, pp. 473-8], Johnston [34, Ch. 10], and Wonnacott & Wonnacott [70, p. 399]. For a well-designed Monte Carlo study that had ambiguous results, see Cragg [12].

45. In particular, Klein and Nakamura [42] argue on the basis of heuristic reasoning and long experience that 2SLS is more sensitive than OLS to multicollinearity, and that LI is the most sensitive of the three.

46. The use of instrumental variables can also go very far awry if the instruments are not very well correlated with the variables to be replaced. In this simple 2 variable equation, for example,

$$Y_1 = \alpha + \beta Y_2 + \varepsilon$$

the use of X as an instrument for Y_2 yields $\hat{\beta} = S_{XY_1}/S_{XY_2}$, where S_{XY_i} is the sample covariance of X and Y_i. Clearly the smaller S_{XY_2} is, the more unstable the estimate becomes.

47. For some theoretical support for this point, see Theil [66, p. 553].

[58, p. 100] decision rule: Calculate the OLS and 2SLS estimates, and accept the 2SLS estimate unless the difference between them is as great as the size of the OLS estimate, in which case accept the OLS estimate. This would seem to be a reasonable procedure in cases where we are very uncertain about the value of a parameter, but would be dominated by a well chosen a priori restriction. Also, in our particular model, we cannot replace one or a few parameter estimates without violating the adding-up restrictions. We would therefore have to replace the coefficients of a particular variable in every equation or reestimate the system subject to explicit constraints. The latter approach would take time and be inferior to a priori constraints. The former method would require a more elaborate rule so that we would know how many violations of the Quandt-Tukey Rule justified replacing a parameter in every equation.

Once again, a relevant point is that in forcasting we're ultimately interested in the reduced forms rather than the structural equations, and a ranking of various K-class estimators according to their structural results may be very different from a ranking based on the success of their implicit reduced form estimates.

In particular, OLS estimation of a structural equation will minimize the sample period residuals of that equation with the endogenous RHS variables assumed to be known exactly, whereas in forcasting we want to minimize the residuals with the RHS endogenous variables given by their simultaneously predicted values. Ball [7] gives a good example of where both 2SLS and LI yielded better implicit sample period fits for the reduced forms than did OLS, presumably because the instrumental approaches take into account the fact that the RHS endogenous variables have to be predicted from predetermined variables.

If we do use an instrumental variables approach, then the question arises as to which variables to use as instruments. We know that we desire those elusive instruments which are exogenous but highly correlated with the variables to be replaced. And we know that we must use less than T instruments but more instruments than the number of RHS endogenous variables to be replaced, and that in between these extremes there are costs to using both a small and a large number of instruments.[48]

Fisher [20,22] has given us some authoritative preference ordering rules, but these are vitiated here by two considerations: (i) consistency requires that we do not use different instruments for variables in the same equation; and (ii) Appendix A of this chapter tells us that our balance sheet restrictions will be imposed by single equation regressions if we use the same instruments across equations.

Since the treatment of autocorrelation also has implications for our choice of instruments, I will postpone further comment to section 5. below.

48. This is discussed more specifically in section 2.4.6., in the third application that I list for principal components.

C. AUTOCORRELATION

3. The Effects of Autocorrelation

Briefly, the presence of lagged endogenous variables and serially uncorrelated disturbances will lead to biased, but consistent and asymptotically efficient estimates. Autocorrelated disturbances in the presence of only exogenous RHS variables will lead to inefficient forecasts and may lead to inefficient parameter estimates.

The presence of autocorrelated disturbances and RHS endogenous and/or lagged endogenous variables will lead to biased and probably inconsistent parameter estimates. The usual measures of the variances of the parameter estimates and of the disturbances are inappropriate, and if used are likely to underestimate the true variances.

4. Detection of Autocorrelation

The most well known test for first order autocorrelation and the one almost invariably reported is the Durbin-Watson (DW) statistic.[49] With lagged dependent variables present, however, this statistic is not valid and if used gives the wrong probability of Type I error. Nerlove and Wallis [55], among others, have criticized its use in this situation on the grounds that it has little power, being strongly biased toward 2.

---

49. The most well known alternative is the BLUS Statistic. Koerts and Abrahamse [44,45] describe this statistic and compare its power relative to the DW statistic.

Durbin [15] recently proposed an alternative large sample statistic which has the same asymptotic distribution as the likelihood ratio statistic under both the null and alternative hypothesis. And a small sample Monte Carlo study by Maddala [51] found that this statistic did well for the model

$$y = \alpha y_{-1} + \beta x + (\rho u_{-1} + \varepsilon)$$

when α was large.

But as Durbin [15] stressed, the point of these and other tests is to save one the trouble of searching or iterating to find the maximum likelihood estimates. Since maximum likelihood estimates will be calculated here in Chapter IV, there is no real need for detection.

5. SOLUTIONS

5.0. The Assumed Problem

Each of the asset demand equations in our model is of the type

$$y_i = Y_i A_i + X_i B_i + U_i$$

where

y_i = LHS endogenous variable

Y_i = RHS endogenous variables

X_i = predetermined variables

We will assume that

$$U_i = \rho_i U_i(-1) + \varepsilon_i$$

where $|\rho_i| \leq 1$ and the ε_i are serially independent.

5.1. The Balance Sheet Restrictions

The balance sheet identities will impose certain restrictions on any autoregressive structure that we assume. It is worthwhile to consider here the general first-order[50] autoregressive structure

$$U_i = \sum_{j=1}^{n} \rho_{ij} U_j(-1) + \varepsilon_i \ , \quad i = 1, \ldots, n \ .$$

It can easily be shown that our balance sheet identities cannot be maintained for all possible values of the independent variables unless $\sum_{i=1}^{n} U_i = 0 \ , \ \forall T$. From this we have

$$0 = \sum_{i=1}^{n} U_i = \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij} U_j(-1) + \sum_{i=1}^{n} \varepsilon_i$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n-1} (\rho_{ij} - \rho_{in}) U_j(-1) + \sum_{i=1}^{n} \varepsilon_i$$

$$= \sum_{j=1}^{n-1} U_j(-1) \sum_{i=1}^{n} (\rho_{ij} - \rho_{in}) + \sum_{i=1}^{n} \varepsilon_i$$

which can always be true only if

$$\sum_{i=1}^{n} (\rho_{ij} - \rho_{in}) = 0 \ , \quad j = 1, \ldots, n-1$$

$$\sum_{i=1}^{n} \varepsilon_i = 0 \ .$$

---

[50] As will be clear from the derivation, the introduction of higher order terms of the form $\sum_{j=1}^{n} \lambda_{ij} U_j(-\tau)$ will necessitate restrictions on the λ_{ij} identical to those on the ρ_{ij} .

This conclusion is exactly analogous to our balance sheet restrictions of Appendix III-A, as can be seen in the following 3-equation example:

$$U_1 = \rho_{11}U_1(-1) + \rho_{12}U_2(-1) + \rho_{13}U_3(-1) + \varepsilon_1$$

$$U_2 = \rho_{21}U_1(-1) + \rho_{22}U_2(-1) + \rho_{23}U_3(-1) + \varepsilon_2$$

$$U_3 = \rho_{31}U_1(-1) + \rho_{32}U_2(-1) + \rho_{33}U_3(-1) + \varepsilon_3$$

The independent variables are linearly dependent since $U_1(-1) + U_2(-1) + U_3(-1) = 0$; so we can eliminate $U_3(-1)$ by a substitution, yielding

$$U_1 = (\rho_{11} - \rho_{13})U_1(-1) + (\rho_{12} - \rho_{13})U_2(-1) + \varepsilon_1$$

$$U_2 = (\rho_{21} - \rho_{23})U_1(-1) + (\rho_{22} - \rho_{23})U_2(-1) + \varepsilon_2$$

$$U_3 = (\rho_{31} - \rho_{33})U_1(-1) + (\rho_{32} - \rho_{33})U_2(-1) + \varepsilon_3$$

Because $U_1 + U_2 + U_3 = 0$, the coefficients of any variable must sum across equations to zero--just as in Appendix III-A.

Now with the assumption we are making $\rho_{ij} = 0$, $i \neq j$, the restrictions on ρ collapse to

$$\rho_{jj} - \rho_{nn} = 0 , \quad j = 1, \ldots, n-1$$

or

$$\rho_{jj} = \rho \qquad \forall j$$

5.2. Maximum Likelihood Estimators

The equation to be estimated can also be written as

$$y_i = \rho_i y_i(-1) + Y_i A_i - \rho_i Y_i(-1) A_i + X_i B_i - \rho_i X_i(-1) B_i + \varepsilon_i \ .$$

The correlation of ε_i with Y_i implies that we want to use instruments to get $\hat{Y}_i$ and then regress

$$y_i = \rho_i y_i(-1) + \hat{Y}_i A_i - \rho_i Y_i(-1) A_i + X_i B_i - \rho_i X_i(-1) B_i$$
$$+ [\varepsilon_i - (Y_i - \hat{Y}_i) A_i] \ .$$

We can guarantee that the residual in this equation $(\varepsilon_i - (Y_i - \hat{Y}_i)A_i)$ will be uncorrelated with all RHS variables by using all RHS variables (other than Y_i) as instruments in the first stage.

If we than find (by search or iteration) the $\hat{\rho}_i$, $\hat{A}_i$, and $\hat{B}_i$ which minimize the sum of squared residuals, then we will have estimates (S2SLS) which are consistent, maximum likelihood and asymptotically efficient among single equation estimators [6, 60].

If in each sector we constrain the ρ_i to be constant across equations, then each such equation will have the same RHS variables and the left hand side variables $(y_i - \rho y_i(-1))$ will sum across equations to $1-\rho$. We can therefore infer from Appendix III-A that each coefficient in A_i and B_i will sum across equations to zero, with the exception of the constant term $(1-\rho)$ in

$X_i - \rho X_i(-1)$, whose coefficients will sum to one.

Sargan proposed using as instruments the current and lagged values of all predetermined variables and the lagged values of all endogenous variables (which aren't also counted as predetermined), since all of these will be in the reduced form for Y_i . But the moment matrix for the instruments can be singluar (if there are more than T instruments) or may be so large and nearly singular that its inverse will be very difficult to calculate accurately. In addition, Fair [16] has argued heuristically that adding "unnecessary" instruments[51] does not affect the asymptotic efficiency but does use up degrees of freedom and increase the small sample bias.

This prompted Fair to suggest an iterative procedure which eliminates one instrument for every exogenous variable that appears in only one equation; this is clearly of no use in our model. Alternatively, Fair proposed a procedure in which all autoregressive parameters are assumed to be approximately equal to some known constant; this would seem to be of uncertain value.

Amemiya [6] had previously proposed dropping the lagged endogenous variables from the list of instruments,

---

[51] Those instruments which (in the limit) "add nothing to the explanation of the endogenous variables in the reduced form and which are uncorrelated with the reduced form error term."

which also preserves consistency in the case of the specification error created by the disturbances really being of the form

$$U_i = \sum_j \rho_{ij} U_j(-1) + \varepsilon_i$$

as compared to our assumption that $\rho_{ij} = 0$, $i \neq j$. If there is no specification error, then this method (MS2SLS) is less efficient than S2SLS [61]. Additionally, Amemiya assumed that there were no lagged endogenous variables present in the equation; his method can be modified by treating any lagged endogenous variables as endogenous, but Fair suspects that this will result in a considerable efficiency loss.

Amemiya's MS2SLS is also consistent (while S2SLS is not) if all of the predetermined variables are exogenous and the true disturbances follow a general order autoregressive scheme. But Wickens [69] has shown that 2SLS estimates are also consistent in this situation as are what he calls Theil- and Madansky-type estimators which use the residuals from 2SLS estimated equations,[52] and he shows that the Theil-type estimator is at least as efficient as (and generally more efficient than) any of these consistent estimators. The problem with this

[52] 2SLS and the Theil- and Madansky-type estimators are in the class of instrumental variables estimators, which are discussed in the following section.

again is that our equations contain a large number of lagged endogenous variables, and doing nothing about them will take away the hard-earned consistency while treating them as endogenous may be costly by Fair's argument.

One final question of the maximum likelihood technique is the following: Since I have placed so little faith in in-sample statistics and so much reliance on out-of-sample prediction, it might be thought that the proper criterion for choosing $\hat{\rho}$ is to find the estimate which forecasts best. One problem with this argument is that the suggested criterion is identical to fitting the model to the out-of-sample data, while ignoring the in-sample data. The purpose of the forecasting tests is to avoid being misled by how well a model can be made to fit a given set of data, so that we can ultimately discriminate between techniques in order to estimate a final version of the model using all available data. A rule which says "fit ρ to the out-of-sample data" is clearly of no interest; the rule that we should test is the maximum likelihood rule "choose the ρ which minimizes the in-sample residuals."

5.3. Instrumental Variables Estimators

We can consistently estimate the equation

$$y_i = Y_i A_i + X_i B_i + U_i$$

by using instruments for the Y_i and for those members of X_i which are lagged endogenous variables. The great advantage of this approach is that we do not have to make any assumptions about the structure of the disturbances. On the other hand, to the extent that our assumption that the disturbances follow a first order autoregressive pattern is accurate, not using that information will be inefficient. Another disadvantage here is the introduction of the usual problem of finding instruments which are close to exogenous but well correlated with y_i .[53]

Malinvaud [52] reports some Monte Carlo studies in which instrumental variables did worse than OLS. In a paper discussed in the following section, Grether and Maddala find that when the autoregressive parameter is not large, instrument variables are not far from asymptotic efficiency--unfortunately, we have come to expect economic models to have large autoregressive parameters.

5.4. Two Step Procedures

Grether and Maddala [27] have identified and discussed two types of two-step procedures commonly used with serially correlated disturbances. The use of instruments for Y_i and the lagged endogenous members of X_i in the OLS regression of

[53]Liviatan [49] suggested using the lagged dependent values of the exogenous X_i as long as they add explanatory power.

$$y_i = Y_i A_i + X_i B_i + U_i$$

will yield consistent estimates of A_i and B_i, and hence consistent estimates of the disturbances.

Method (1) then uses these estimated disturbances in the OLS regression of the equation

$$y_i = Y_i A_i + X_i B_i + \rho_i \hat{U}_i(-1) + \varepsilon_i$$

Method (2), on the other hand, uses the estimated residuals to consistently estimate the autoregressive parameter ρ and thereby the covariance matrix of the disturbances. Prais-Winston or Cochrane-Orcutt transformations can then be made and A_i and B_i can be consistently estimated.

Taylor and Wilson's three-pass least squares [65] is in the spirit of method (1); the alternative advocated by Wallis [68] in response is an example of method (2).

Grether and Maddala give the asymptotic variances of various estimators for the model

$$y = \alpha y_{-1} + \beta x + u$$

$$u = \rho u_{-1} + \varepsilon$$

where ε and x are not serially correlated. These variances reveal that the 2-step procedures are asymptotically inefficient not only when compared with a maximum likelihood estimator, but in some cases when compared with the first step of the two-step procedures. (This first step is the instrumental variables method listed in 5.3.) They also found that, for the high values of ρ usually encountered in economic work, the inefficiency of the two-step procedures tended to become very large. Neither 2-step procedure was superior to the other although method (2) did seem more sensitive to values of ρ close to one.

D. SYSTEMS ESTIMATION

6. Aitken System Estimators

6.1. The Efficiency of Aitken Estimation of a System of Equations

In this section I will discuss a general class[54] of estimators and give examples of some members of that class. Consider a system of m equations, each of the type

$$\underset{T\times 1}{y_i} = \underset{T\times p_i}{X_i} \; \underset{p_i\times 1}{B_i} + \underset{T\times 1}{u_i}$$

The system as a whole can be written as

[54] This class of estimators has been associated very closely with Zellner, with [71] being a seminal paper. Many things said here are proven in that paper, as well as in Dhrymes [14, pp. 153-67].

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} X_1 & & 0 \\ & \ddots & \\ 0 & & X_m \end{bmatrix} \begin{bmatrix} \beta_1 \\ \\ \beta_m \end{bmatrix} + \begin{bmatrix} U_1 \\ \\ U_m \end{bmatrix}$$

$$TM\times 1 \qquad TM\times\Sigma p_i \qquad \Sigma p_i\times 1 \qquad TM\times 1$$

or

$$Y = X\beta + U$$

Now, if $E(U) = 0$, $E(X'U) = 0$, and $E(UU') = \Phi$, then the best linear unbiased estimator of β is the Aitken (generalized least squares) estimator

$$\tilde{\beta}_{GLS} = (X'\Phi^{-1}X)^{-1}X'\Phi^{-1}Y$$

It can also be shown [72] that, in the class of linear unbiased predictors, the forecasts $Y = X\tilde{\beta}_{GLS}$ have the minimal generalized mean-squared forecast error and also minimize every positive definite quadratic form in the forecast errors.

Since Φ is normally unknown, it must be estimated--which will necessitate some a priori assumptions.[55] If the U_i are serially uncorrelated, then we can write

---

[55] Φ is of rank TM, while the actual observations are of rank $\Sigma p_i < TM$.

$$\underset{TM\times TM}{\Phi} = \underset{M\times M}{\Sigma} \otimes I_\tau$$

where $\otimes$ denotes the Kronecker product and Σ consists of the contemporaneous covariances, σ_{ij}, between the disturbances. The σ_{ij} can be consistently estimated from the residuals of single equation OLS regressions;[56] from these we can construct the consistent estimator

$$\hat{\Phi}^{-1} = \hat{\Sigma}^{-1} \otimes I_\tau$$

and from this form the "feasible Aitken estimator"

$$\hat{\beta}_{GLS} = (X'\hat{\Phi}^{-1}X)^{-1}X'\hat{\Phi}^{-1}Y$$

which will be consistent and asymptotically equivalent to $\tilde{\beta}_{GLS}$.

It can be easily shown that two cases in which single equation OLS yields parameter estimates identical to those resulting from the feasible Aitken estimator are when (i) the residuals are not contemporaneously correlated (i.e., $\hat{\sigma}_{ij} = 0$, $i \neq j$); or (ii) each equation contains the same RHS variables. These estimators can also be shown[57] to be equivalent when the equations can be grouped into subsets within which all equations contain the same RHS variables, and between subsets the residuals are not contemporaneously correlated.

---

[56] $\hat{U}_i'\hat{U}_j/T$ is the maximum likelihood estimator of σ_{ij}, while $\hat{U}_i'\hat{U}_j/\sqrt{(T-p_i)(T-p_j)}$ is an unbiased estimator.

[57] See Appendix III-C.

These general results can cover a wide variety of situations, since we will usually make assumptions about a system which make it possible to transform the system into equations which meet the assumptions made here.

If, for example, some of the RHS variables are correlated with the disturbances terms, then we might use an instrumental variables approach, which will be equivalent to OLS regressions of the equations

$$y_i = \hat{X}_i^{(1)}\beta_i^{(1)} + X_i^{(2)}\beta_i^{(2)} + U_i$$

(where the $X_i^{(1)}$ have been replaced by their estimated values, based on their regression on a set of instruments). From our previous result we know that the additional step of using the estimated residuals from these equations to form a feasible Aitken system estimator will yield estimates which are consistent and asymptotically efficient, though not any different from the single equation estimates if there are no contemporaneous correlations or if the same RHS variables appear in all equations and the same instruments are used across equations.

The second condition will not occur if the reason for the correlation with the disturbances is that some of the RHS variables are currently endogenous, but could hold if for example the problem is that the $X_i^{(1)}$ are observed with error. The second condition could also be of interest if we are contemplating applying an Aitken estimator to a subsystem rather than to the entire set of equations.

When all predetermined variables are used as instruments, the single equation method is 2SLS, and the Aitken extension is 3SLS. It is interesting that 2SLS and 3SLS also coincide when all equations are just identified. This is because in the just-identified situation 2SLS is equivalent to OLS applied to the reduced form (which will have all the properties assumed above) and all predetermined variables will appear in each reduced form.

Another situation in which the original result can be applied is where the disturbances follow first order autoregressive schemes:

$$U_i = \rho_i U_i(-1) + \varepsilon_i \ , \ |\rho_i| \leq 1$$

where the ε_i are not serially correlated but may be contemporaneously correlated.

From consistent estimates of the U_i,[58] we can consistently estimate the ρ_i

[58] If all the RHS variables are exogenous, then OLS regressions of the original equations will suffice; otherwise instruments can be employed.

$$\hat{\rho}_i = \frac{\sum_{\tau=2}^{T} U_i(\tau)U_i(\tau-1)}{\sum_{\tau=2}^{T} U_i^2(\tau)}$$

and apply either the transformation associated with Cochrane-Orcutt:

$$y_i - \hat{\rho}_i y_i(-1) = [X_i - \hat{\rho}_i X_i(-1)]\beta_i + \varepsilon_i$$

$$y_i^* = X_i^* \beta_i + \varepsilon_i$$

or[59] we can use the transformation associated with Prais-Winston:

$$H_i^{-1} Y_i = H_i^{-1} X_i \beta_i + H_i^{-1} U_i$$

$$Y_i^* = X_i^* \beta_i + U_i^*$$

where

$$\hat{H}_i^{-1} \begin{bmatrix} (1-\hat{\rho}_i^2)^{1/2} & & & & 0 \\ -\hat{\rho}_i & & & & 0 \\ 0 & -\hat{\rho}_i & 1 & & \vdots \\ \vdots & & \ddots & \ddots & \\ 0 & \cdots & & -\hat{\rho}_i & 1 \end{bmatrix}$$

[59] The Cochrane-Orcutt estimator costs us one observation but has the offsetting advantage of preserving the assumed contemporaneous covariance structure of the disturbance terms, which the Prais-Winston procedure does not do. The latter method can be modified to be strictly correct, but

so that $E(U^*U^{*\prime})$ can be shown to be of the form $\Sigma \otimes I$.

In either case, the transformed equations are of the type initially discussed,[60] and we thereby know that an Aitken system estimator will be consistent and asymptotically efficient, but not an improvement if the disturbances are not correlated across equations or if the X_i^* are invariant across equations. The latter condition will be true of both the Cochrane-Orcutt and Prais-Winston transformations if the X_i and ρ_i are constant across equations.

6.2. Potential for the Model of Chapter II

We can logically consider first the possible step from single equation methods to an Aitken system estimate of the equations associated with each sector. To do this we will have to delete an equation from each sector and estimate the residual parameters from the balance sheet restrictions; Powell [57] has shown that the resultant parameter estimates do not depend upon which equation is deleted.

---

then the first period's residuals will have to be discarded and the Aitken system estimation will be slightly different than the class of estimators we are considering here. Parks [56] proposed the Aitken estimation of this type of system and discusses this point.

[60]If some of the X_i^* are endogenous, then they can be replaced by their estimates based on instruments.

However, from 6.1 we know that the system estimates obtained in this way will be identical to the single equation estimates on which they are based if the final step in the single equation method is the OLS regression of equations of the type

$$Y_i = X\beta_i + U_i$$

where the RHS variables do not depend on the equation. For each sector, we did in fact generally impose the constraint that the (possibly transformed) equations contain the same RHS variables since we knew that single equation OLS applied to such equations would impose adding-up properties on the parameters.[61] Thus, in general the application of Aitken system estimates to sectors will not make any difference.

The next question is whether anything can be gained from an Aitken estimation of the complete financial sector; i.e., the asset demand equations of all sectors. Here we've seen (Appendix III-C) that such an approach will be inconsequential if the explanatory variables are invariant within sectors and the disturbances are not correlated for equations in different sectors. We've said that the first condition will generally hold; the second

---

[61]For example, we required that instruments and autoregressive parameters be constant across equations, which we saw in 6.1 do make the Aitken system estimators impotent.

condition may occur if sectors hold different assets or if the disturbances affect supply and demand differently or if the disturbances reflect omitted factors unique to each sector. If on the other hand the different sectors have explanatory variables which are similar (or very highly correlated), then the Aitken estimates will again be very close to the single equation estimates. This could occur if: (i) sectors hold the same assets or if the rates corresponding to the assets which are different are highly correlated; and (ii) the different wealth scalers used by different sectors are highly correlated.

Since it seems reasonable to expect that across sectors the explanatory variables will be highly correlated while the disturbances will not be very correlated, I would not expect the Aitken system estimates to be much different from the single equation estimates. Against this small expected return, we have to weigh the computational time and difficulties that may be encountered.

In particular, because of its size and the collinearity of the variables, $X\Phi^{-1}X$ may be very close to singularity and very difficult to accurately compute.[62] We also have the problem that some of our assumptions may be inappropriate, and Aitken system estimates will transfuse such

---

[62] See Zellner and Thornber [73] for some 3SLS estimates of a 22 equation Klein model that were several times calculated inaccurately.

errors throughout the entire model and affect all parameter estimates. This may exaggerate whatever problem exists and will make it very difficult to locate the difficulty.

Another possibility is to apply an Aitken system estimator to the entire model. The problem with this is that I do not have as much faith in the specification of the nonfinancial equations as I have in the financial sector. This type of estimator is known to be very sensitive to specification error, and the presence of such an error in any equation will again affect every parameter estimate. This could make the model function worse than before, and it will again be difficult to discover which variables and/or equations are responsible.

Because of the considerations discussed here; it was decided that Aitken system estimates using the results of each single equation method would probably not be worth the effort. If eventually a few single equation techniques do seem to stand out from the rest, then it may be profitable to apply Aitken estimators to these few results. The profitability of such an approach will depend on the validity of the propositions that single equation estimators which forecast well will also estimate the disturbances well, and that Aitken system estimators will be useful and not get one into trouble when the disturbances are well estimated.

7. Full Information Maximum Likelihood (FIML)

The essential difference between FIML (or the equivalent Full Information Least Generalized Variance) and the Aitken system estimators discussed in 6. is that the former simultaneously finds the parameters and the covariance matrix of the disturbances which maximize the likelihood function, while the latter method maximizes the likelihood function conditionally on Σ and thereafter uses a consistent estimator of the covariance matrix; asymptotically, they can be shown to be equivalent [14, pp. 367-72].

On the other hand, the simultaneous estimation process can be very difficult and unstable. For example, whereas in a single equation instrumental variables approach or an Aitken generalization of such methods we could use a judiciously chosen subset of the set of all predetermined variables, in FIML we must invert the matrix of the moments of all predetermined variables in the system, and then use this inverse in the calculation of every parameter estimate. This undoubtedly renders FIML the most sensitive to multicollinearity of all the estimation techniques considered here.[63] In addition, FIML techniques are particularly sensitive to specification errors of almost any kind [31].

[63]Klein and Nakamura [42] discuss the sensitivity of FIML relative to other methods.

A final problem which alone is sufficient to warrant a postponement of FIML methods is that the model is linear in parameters and can therefore be easily estimated by single equation methods, but is nonlinear in variables which makes FIML estimation very complicated.[64]

[64]See Klein [41] for a discussion of his difficulties in computing FIML estimates of a Wharton model.

APPENDIX III-A

(i) The Adding-Up Restrictions

Consider n equations of the type

$$\underset{Tx1}{y_i} = \beta_i^1 + \underset{Txp-1}{X} \; \underset{p-1x1}{\beta_i} + \underset{Tx1}{\varepsilon_i}$$

where $\sum_{i=1}^{n} y_i = \iota\lambda$, $\underset{Tx1}{\iota} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$. It is well known

that the parameter estimates will be unaffected if we regress the equations in deviations-from-means form:

$$\underset{Tx1}{y_i - \iota\bar{y}_i} = \underset{Txp-1}{(X - \iota\bar{X})} \; \underset{p-|x|}{\beta_i} + \underset{Tx1}{\varepsilon_i}$$

A column vector whose elements are the sums across equations of the estimated coefficients is

$$\sum_{i=1}^{n} \hat{\beta}_i = \sum_{i=1}^{n} [(X - \iota\bar{X})'(X - \iota\bar{X})]^{-1}(X - \iota\bar{X})^{-1}(y_i - \iota\bar{y}_i)$$

$$= [(X - \iota\bar{X})'(X - \iota\bar{X})]^{-1}(X - \iota\bar{X})' \sum_{i=1}^{n} (y_i - \iota\bar{y}_i)$$

$$= \underset{p-1x1}{0}$$

since

$$\sum_{i=1}^{n}(y_i - \iota\bar{y}_i) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\iota\sum_{\tau=1}^{T}\iota\frac{1}{T}\, y_i(\tau)$$

$$= \iota\lambda - \iota\frac{1}{T}\sum_{\tau=1}^{T}\sum_{i=1}^{n} y_i(\tau) = \iota\lambda - \iota\frac{1}{T}\sum_{\tau=1}^{T}\lambda = \iota\cdot 0$$

The intercept term is

$$\hat{\beta}_i^1 = \bar{y}_i - \bar{X}\hat{\beta}_i$$

The sum across equations of the intercepts is

$$\sum_{i=1}^{n}\hat{\beta}_i^1 = \sum_{i=1}^{n}\bar{y}_i - \bar{X}\sum_{i=1}^{n}\hat{\beta}_i = \lambda - \bar{X}\cdot 0 = \lambda$$

<u>(ii) Competitive Markets</u>

Our specific demand equations([8a] of Chapter II)

$$\frac{\Delta a}{\omega} = \sum_{i=1}^{m}(EA_i r + EB_i s)\,\frac{\ell_i}{\omega} - E\,\frac{a-1}{\omega} + F\,\frac{\Delta\ell}{\omega}$$

are equivalent to the equations

$$\frac{a}{\omega} = EA_m r + EB_b s + \sum_{i\neq m}[E(A_i - A_m)r + E(B_i - B_m)s]\,\frac{\ell_i}{\omega}$$

$$+ (I-E)\,\frac{a-1}{\omega} + F\,\frac{\Delta\ell}{\omega}$$

If B_i^1 is the column of B_i corresponding to the constant term in s, then from (i) we know[1] that the elements of $(\widehat{EB_m^1})$ will sum to 1, and that the elements in any other column will sum to zero. The elements of F are therefore constrained as desired. It is also immediately clear that the elements of

$$\hat{E} = (\widehat{I-E}) + I$$

will sum to zero as desired.

Now for each matrix of the form

$$(\widehat{EX}) = \begin{pmatrix} \hat{q}_{11} & \hat{q}_{12} & \cdots \\ \hat{q}_{21} & & \\ \vdots & & \vdots \end{pmatrix}$$

we can define

$$(\hat{E}^{-1}) = (\hat{E})^{-1} = \frac{1}{|\hat{E}|} \begin{pmatrix} C_{11} & C_{21} & \cdots & C_{2n} \\ C_{12} & & & \vdots \\ \vdots & & & \\ C_{1n} & \cdots & & C_{nn} \end{pmatrix}$$

where C_{ij} is the cofactor the element $\hat{\varepsilon}_{ij}$ of $\hat{E}$.

1

$$EB_m s = E[B_m^1 \quad B_m^2 \quad \cdots \quad] \begin{pmatrix} s_1 \\ s_2 \\ \vdots \end{pmatrix} = EB_m^1 s_1 + EB_m^2 s_2 + \cdots$$

Then

$$\hat{X} = (\hat{E}^{-1})(\widehat{EX}) = \frac{1}{|\hat{E}|}\begin{bmatrix} \sum_{i=1}^{n} C_{i1}\hat{q}_{i1} & \sum_{i=1}^{n} C_{i1}\hat{q}_{i2} & \cdots \\ \sum_{i=1}^{n} C_{i2}q_{i1} & & \\ \vdots & & \end{bmatrix}$$

and the sum of the elements in the k^{th} column of $\hat{X}$ will be

$$k_X = \frac{1}{|\hat{E}|} \sum_{j=1}^{n} \sum_{i=1}^{n} C_{ij}\hat{q}_{ik}$$

$$= \frac{1}{|\hat{E}|} \sum_{i=1}^{n} \hat{q}_{ik} \sum_{j=1}^{n} C_{ij}$$

Thus,

$$k_X = \frac{1}{|\hat{E}|} \sum_{i=1}^{n} q_{ik} |\hat{E}| = \sum_{i=1}^{n} q_{ik}$$

the sum of the elements in the k^{th} column of $(\widehat{EX})$.

Therefore, the elements in each column of $\hat{A}_n$ and $(A_i \hat{-} A_n)$ sum to zero, which implies that the elements in any column of each $\hat{A}_i$ sum to zero. A similar result holds by the same reasoning for all columns of $\hat{B}_i$ other than $\hat{B}_i^1$. In this case, the elements of $\hat{B}_n^1$ sum to one because the elements of $(\widehat{EB_n^1})$ sum to one. Since the elements of $E(\widehat{B_i^1 - B_n^1})$ sum to zero, the elements of $(\widehat{B_i^1 - B_n^1})$ must sum to zero, and hence the elements of each $\hat{B}_i^1$ must sum to one.

(iii) Noncompetitive Markets

For notational simplicity, consider the desired relations

$$\underset{n\times 1}{a^*/w} = \underset{n\times p}{A}\ \underset{p\times 1}{r}$$

where the elements of s are now included in r. We can now partition the sector's holdings into q markets in which it adjusts toward its desired asset position, and $n-q$ markets in which is accepts the holdings implied by the other sectors' net demands:

$$\Delta \begin{bmatrix} \underset{q\times 1}{a^1} \\ \underset{n-q\times 1}{a^2} \end{bmatrix} \frac{1}{w} = \begin{bmatrix} \underset{q\times n}{E} \\ \underset{n-q\times n}{0} \end{bmatrix} \underset{n\times 1}{(a^* - a_{-1})} \frac{1}{w} + \begin{bmatrix} \underset{q\times n-q}{\Phi} \\ I_{n-q} \end{bmatrix} \Delta a^2/w$$

$$= \begin{bmatrix} EA \\ 0 \end{bmatrix} r - \begin{bmatrix} E \\ 0 \end{bmatrix} \frac{a_{-1}}{w} + \begin{bmatrix} \Phi \\ I \end{bmatrix} \frac{\Delta a^2}{w}$$

From our earlier result, we know that the estimated coefficients will obey the adding-up restrictions, and it is obvious that the estimates will be zero and I_{n-q} where it is indicated above that they should be.

If for each of the $n-p$ a_i in a^2 we regress

$$\Delta r_i = \emptyset_i \frac{a_i^* - a_i(-1)}{w} = \emptyset_i A_i r - \emptyset_i \frac{a_i(-1)}{w}$$

(where A_i is the i^{th} row of A), then we will obtain unique estimates of the $\emptyset_i$ and these $n-q$ A_i , which we can group together as $\underset{n-qxp}{A^{(2)}}$.

Now, since

$$EA = (\underset{qxq}{E^1} \quad \underset{qxn-q}{E^2}) \begin{bmatrix} \underset{qxp}{A^1} \\ \\ \underset{n-qxp}{A^2} \end{bmatrix} = E^1A^1 + E^2A^2$$

we can uniquely estimate A^1 :

$$\hat{A}^1 = (\hat{E}^1)^{-1}(\widehat{E^1A^1}) = (\hat{E}^1)^{-1}(\widehat{EA} - \hat{E}^2\hat{A}^2)$$

APPENDIX III-B

We want to examine the expected value, variance, and mean squared error of

$$\beta^* = M^{(1)} D^{(1)^{-1/2}} \hat{Y}$$

where

$$\hat{Y} = [F^{(1)'} F^{(1)}]^{-1} F^{(1)'} Y$$

$$= I_m^{-1} (XM^{(1)} D^{(1)^{-1/2}})' Y$$

$$= D^{(1)^{-1/2}} M^{(1)'} X'Y$$

substituting,

$$\beta^* = M^{(1)} D^{(1)^{-1}} M^{(1)'} X'Y = M^{(1)} D^{(1)^{-1}} M^{(1)'} X'(X\beta + \varepsilon)$$

$$= M^{(1)} D^{(1)^{-1}} M^{(1)'} MDM'\beta + M^{(1)} D^{(1)^{-1}} M^{(1)'} X'\varepsilon$$

Now since

$$M^{(1)'}MDM' = \underset{m\mathrm{x}p}{M^{(1)'}} (\underset{p\mathrm{x}m}{M^{(1)}} \vdots \underset{p\mathrm{x}p-m}{M^{(2)}}) \begin{pmatrix} \underset{m\mathrm{x}m}{D^{(1)}} & \vdots & \underset{m\mathrm{x}p-m}{0} \\ \cdots & & \cdots \\ \underset{p-m\mathrm{x}m}{0} & \vdots & \underset{p-m\mathrm{x}p-m}{D^{(2)}} \end{pmatrix} \begin{pmatrix} \underset{m\mathrm{x}p}{M^{(1)'}} \\ \\ \underset{p-m\mathrm{x}p}{M^{(2)'}} \end{pmatrix}$$

$$= (I_m \vdots \underset{m\mathrm{x}p-m}{0}) \begin{pmatrix} D^{(1)} & M^{(1)'} \\ D^{(2)} & M^{(2)'} \end{pmatrix} = D^{(1)} M^{(1)'}$$

we have

$$\beta^* = M^{(1)}D^{(1)^{-1}}D^{(1)}M^{(1)'}\beta + M^{(1)}D^{(1)^{-1}}M^{(1)'}X'\varepsilon$$

$$= M^{(1)}M^{(1)'}\beta + M^{(1)}D^{(1)^{-1}}M^{(1)'}X'\varepsilon$$

Thus,

$$E(\beta^*) = M^{(1)}M^{(1)'}\beta$$

We know that

$$I_p = MM' = (\underset{p\times m}{M^{(1)'}} \vdots \underset{p\times p-m}{M^{(2)}}) \begin{pmatrix} M^{(1)'} \\ M^{(2)'} \end{pmatrix} = M^{(1)}M^{(1)'} + M^{(2)}M^{(2)'}$$

$$= \sum_{i=1}^{m} M_i M_i' + \sum_{i=m+1}^{p} M_i M_i'$$

which implies that the squares of the k^{th} elements in each column sum to one, so that every element of M must be smaller than one in absolute value.

Looking at

$$E(\beta_k^*) = k\sum_{i=1}^{m} M_{ki}^2 + \sum_{j\neq k} j\sum_{i=1}^{m} M_{ki}M_{ji} ,$$

if $\sum_{i=1}^{m} M_{ki}^2 = 1$, then $\sum_{i=m+1}^{p} M_{ki}^2 = 0$ and if this is true for all k , then the last $p-m$ eigenvectors could consist only of zeros.

Therefore, β^* is biased. The amount of bias will be

$$\beta - E(\beta^*) = \beta - M^{(1)}M^{(1)'}\beta = (I - M^{(1)}M^{(1)'})\beta = M^{(2)}M^{(2)'}\beta$$

Because $\beta^* - E(\beta^*) = M^{(1)}D^{(1)^{-1}}M^{(1)'}X'\varepsilon$, the var-cov matrix will be

$$\begin{aligned}\text{Cov } \beta^* &= E[\beta^* - E(\beta^*)][\beta^* - E(\beta^*)]' \\ &= E[M^{(1)}D^{(1)^{-1}}M^{(1)'}X'\varepsilon\varepsilon'XM^{(1)}D^{(1)^{-1}}M^{(1)'}] \\ &= \sigma^2 M^{(1)}D^{(1)^{-1}}M^{(1)'}X'XM^{(1)}D^{(1)^{-1}}M^{(1)'} \\ &= \sigma^2 M^{(1)}D^{(1)^{-1}}M^{(1)'}\end{aligned}$$

since

$$D^{(1)^{-1}}M^{(1)'}X'XM^{(1)} = D^{(1)^{-1}}M^{(1)'}MDM'M^{(1)} = D^{(1)^{-1}}D^{(1)}M^{(1)'}M^{(1)} = I$$

On the other hand, the OLS estimates will be unbiased with

$$\text{Cov}(\hat{\beta}) = \sigma^2(X'X)^{-1} = \sigma^2 M^{(1)}D^{(1)^{-1}}M^{(1)'} + \sigma^2 M^{(2)}D^{(2)^{-1}}M^{(2)'}$$

since

$$(X'X)^{-1} = MD^{-1}M' = (M^{(1)}\ M^{(2)})\begin{pmatrix} D^{(1)^{-1}} & 0 \\ 0 & D^{(2)^{-1}} \end{pmatrix}\begin{pmatrix} M^{(1)'} \\ M^{(2)'} \end{pmatrix}$$

$\mathrm{MSE}(\hat{\beta}) = \mathrm{Var}(\hat{\beta})$ = diagonal elements of

$$\sigma^2(X'X)^{-1} = \sigma^2 M^{(1)} D^{(1)^{-1}} M^{(1)'} + {}^2 M^{(2)} D^{(2)^{-1}} M^{(2)'}$$

Thus, $\mathrm{MSE}(\beta^*) - \mathrm{MSE}(\hat{\beta})$ = diagonal elements of

$$M^{(2)} M^{(2)'} \beta\beta' M^{(2)} M^{(2)'} - \sigma^2 M^{(2)} D^{(2)^{-1}} M^{(2)'}$$

$$= M^{(2)} M^{(2)'} [\beta\beta' - \sigma^2 (X'X)^{-1}] M^{(2)} M^{(2)'}$$

since it can be shown that

$$D^{(2)^{-1}} = M^{(2)'} (X'X)^{-1} M^{(2)}$$

APPENDIX III-C

Consider the system

$$\underset{Tmx1}{Y} = \underset{Tmx\sum_i p_i}{X} \; \underset{\sum_i p_i x1}{B} + \underset{Tmx1}{U_i}$$

discussed in 7.1. where it is now assumed that the equations can be grouped into q sectors, within which each of η_i equations has the same p_i regressors:

$$\underset{T\Sigma\eta_i x\Sigma p_i}{X} = \begin{bmatrix} I_{\eta_1} \otimes \underset{Txp_1}{X_1} & & & 0 \\ & I_{\eta_2} \otimes \underset{Txp_2}{X_2} & & \\ & & \ddots & \\ 0 & & & I_{\eta_q} \otimes \underset{Txp_q}{X_q} \end{bmatrix}$$

where $\otimes$ denotes the Kronecker product.

If the disturbance terms for equations in different groups are not correlated, then we can represent the covariance matrix of disturbances as

$$= (M^{(1)}D^{(1)^{-1}} \quad M^{(2)}D^{(2)^{-1}}) \begin{pmatrix} M^{(1)'} \\ M^{(2)'} \end{pmatrix} = M^{(1)}D^{(1)^{-1}}M^{(1)'}$$

$$+ M^{(2)}D^{(2)^{-1}}M^{(2)'}$$

This means that

$$\mathrm{Cov}(\hat{\beta}) - \mathrm{Cov}(\hat{\beta}^*) = \sigma^2 M^{(2)}D^{(2)^{-1}}M^{(2)'}$$

$$= \sigma^2 M^{(2)}D^{(2)^{-1/2}}D^{(2)^{-1/2}} =$$

$$= \sigma^2 (M^{(2)}D^{(2)^{-1/2}})_{p-mxp} (M^{(2)}D^{(2)^{-1/2}})'$$

which is sufficient to show that this difference is a positive semi-definite matrix.

Now, the squared bias terms are the diagonal elements of

$$[\beta - E(\beta^*)][\beta - E(\beta^*)]' = M^{(2)}M^{(2)'}\beta\beta' M^{(2)}M^{(2)'}$$

so that the mean squared errors (MSE) are (bias^2 + variance =) the diagonal elements of

$$M^{(2)}M^{(2)'}\beta\beta' M^{(2)}M^{(2)'} + \sigma^2 M^{(1)}D^{(1)^{-1}}M^{(1)'}$$

as compared to

$$\underset{T\Sigma\eta_i x T\Sigma\eta_i}{\Phi} = \begin{bmatrix} \underset{\eta_1 x \eta_1}{\Sigma_1} \otimes I_\tau & & & 0 \\ & \underset{\eta_2 \otimes \eta_2}{\Sigma_2} x\ I_i & & \\ & & \ddots & \\ 0 & & & \underset{\eta_q \otimes \eta_q}{\Sigma} \end{bmatrix}$$

$$(X'\Phi^{-1}X) = \begin{bmatrix} I_{\eta_1} \otimes X_1' & & 0 \\ & I_{\eta_2} \otimes X_2' & \\ 0 & & \ddots \end{bmatrix} \begin{bmatrix} \Sigma_1^{-1} \otimes I_\tau & & 0 \\ & \Sigma_2^{-1} \otimes I_\tau & \\ 0 & & \ddots \end{bmatrix} \begin{bmatrix} I_{\eta_1} \otimes X_1 & & 0 \\ & I_{\eta_2} \otimes X_2' & \\ 0 & & \ddots \end{bmatrix}$$

$$= \begin{bmatrix} (I_{\eta_1} \otimes X_1')(\Sigma_1^{-1} \otimes I_\tau)(I_{\eta_1} \otimes X_1) & & 0 \\ & (I_{\eta_2} \otimes X_2')(\Sigma_2^{-1} \otimes I_\tau)(I_{\eta_2} \otimes X_2) & \\ 0 & & \ddots \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_1^{-1} \otimes X_1' X_1 & & 0 \\ & \Sigma_2^{-1} \otimes X_2' X_2 & \\ 0 & & \ddots \end{bmatrix}$$

since $(A \otimes B)(C \otimes D) = AC \otimes BD$.

Now,

$$\hat{\beta}_{GLS} = (X'\Phi^{-1}X)^{-1}X'\Phi^{-1}y$$

$$= \begin{bmatrix} \Sigma_1 \otimes (X_1'X_1)^{-1} & & 0 \\ & \Sigma_2 \otimes (X_2'X_2)^{-1} & \\ 0 & & \ddots \end{bmatrix} \begin{bmatrix} \Sigma_1^{-1} \otimes X_1 & & 0 \\ & \Sigma_2^{-1} \otimes X_2' & \\ 0 & & \ddots \end{bmatrix} Y$$

$$= \begin{bmatrix} I_{n_1} \otimes (X_1'X_1)^{-1}X_1' & & 0 \\ & I_{n_2} \otimes (X_2'X_2)^{-1}X_2' & \\ 0 & & \ddots \end{bmatrix} Y = \hat{\beta}_{OLS}$$

REFERENCES

1. Anderson, T.W., An Introduction to Multivariate Statistical Analysis, New York: John Wiley & Sons, 1958.

2. ———, "The Use of Factor Analysis in the Statistical Analysis of Multiple Time Series,: Psychometrika, XXVIII (1963), 1-25.

3. Anderson, T.W., and H. Rubin, "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," Annals of Mathematical Statistics, XX (March 1949), 46-63.

4. ———, "The Asymptotic Properties of Estimates of the Parameters in a Complete System of Stochastic Equations," Annals of Mathematical Statistics, XXI (December 1950), 570-82.

5. ———, "Statistical Inference in Factor Analysis," in Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability, J. Neyman ed., V, Berkeley: University of California Press, 1956, 111-50.

6. Amemiya, T., "Specification Analysis in the Estimation of the Parameters of a Simultaneous Equation Model with Auto-correlated Residuals," Econometrica, XXXIV (April 1966), 283-306.

7. Ball, R., "The Significance of Simultaneous Methods of Parameter Estimation in Econometric Models," Applied Statistics, XII (March 1963), 14-25.

8. Bartlett, M., "Factor Analysis in Psychology as a Statistician Sees It," in Essays on Probability and Statistics, New York: John Wiley & Sons, 1962, 37-48.

9. Basmann, R., "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation," Econometrica, XXV (January 1957), 77-83.

10. ———, "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics," Journal of the American Statistical Association, LV (December 1960), 650-9.

11. Christ, C., Econometric Models and Methods, New York: John Wiley & Sons, 1966.

12. Cragg, J., "On the Relative Small Sample Properties of Several Structural Equation Estimators," Econometrica, XXXV (January 1967), 89-110.

13. David, F., and J. Neyman, "Extension of the Markoff Theorem on Least Squares," in Statistical Research Memoirs, II, F. David and J. Neyman, eds., London: University of London, Department of Statistics, 1938, 105-116.

14. Dhrymes, P., Econometrics, New York: Harper & Row, 1970.

15. Durbin, J., "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables," Econometrica, XXXVIII (May 1970), 410-2.

16. Fair, R., "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors," Econometrica, XXXVIII (May 1970), 507-17.

17. Farrar, D., and R. Glauber, "Multicollinearity in Regression Analysis: The Problem Revisited," The Review of Economics and Statistics, XLIX (February 1967), 92-107.

18. Feeney, G., and D. Hester, "Stock Market Indices: A Principal Components Analysis," in Risk Aversion and Portfolio Choice, D. Hester and J. Tobin, eds., New York: John Wiley & Sons, 1967, 110-38.

19. Fisher, F., "On the Cost of Approximate Specifications in Simultaneous Equation Estimation," Econometrica, XXIX (April 1961), 139-170.

20. Fisher, F., "The Choice of Instrumental Variables in the Estimation of Economy-Wide Econometric Models," International Economic Review, VI (September 1965), 245-73.

21. Fisher, F., "Dynamic Structure and Estimation in Economy-Wide Econometric Models," in The Brookings Quarterly Econometric Model of the United States, J. Duesenberry, G. Fromm, L. Klein, and E. Kuh, eds., Chicago: Rand McNally, 1965, 589-636.

22. Fisher, F., and B. Mitchell, "The Choice of Instrumental Variables in the Estimation of Economy-Wide Econometric Models: Some Further Thoughts," International Economic Review, XI (June 1970), 226-34.

23. Freund, R., "A Warning of Roundoff Errors in Regression," The American Statistician, XVII (December 1963), 13-5.

24. Goldberger, A., Econometric Theory, New York: John Wiley & Sons, 1964.

25. Gramlich, E., and J. Kalchbrenner, "A Constrained Estimation Approach to the Demand for Liquid Assets," Federal Reserve System Committee on Financial Analysis, 1969 (mimeographed).

26. Graybill, F., An Introduction to Linear Statistical Models, I, New York: McGraw-Hill, 1961.

27. Grether, D. and G. Maddala, "On the Asymptotic Properties of Certain Two-Step Procedures Commonly Used in the Estimation of Distributed Lag Models,"Cowles Foundation Discussion Paper No. 301, Yale University, 1970 (mimeographed).

28. Griliches, Z., "Distributed Lags: A Survey," Econometrica, XXXV (January 1967), 16-49.

29. Haitovsky, Y., "Multicollinearity in Regression Analysis: Comment," The Review of Economics and Statistics, LI (November 1969), 486-9.

30. Harman, H., Modern Factor Analysis, second edition, Chicago: University of Chicago Press, 1967.

31. Hildreth, C., "Simultaneous Equations: Any Verdict Yet?," Econometrica, XXVIII (October 1960), 846-54.

32. Hotelling, H., "The Relation of the Newer Multivariate Statistical Methods to Factor Analysis," British Journal of Statistical Psychology, X (1957), 69-79.

33. James, R., University Mathematics, Belmont, California: Wadsworth, 1963.

34. Johnston, J., Econometric Methods, New York: McGraw-Hill, 1963.

35. Jöreskog, K.G., Statistical Estimation in Factor Analysis, Stockholm: Almquist & Wiksell, 1963.

36. Kadane, J., "Testing Overidentifying Restrictions When the Disturbances are Small," Journal of the American Statistical Association, LXV (March 1970), 182-5.

37. Kendall, M., A Course in Multivariate Statistical Analysis, third edition, New York: Hafner, 1965.

38. Kendall, M., and A. Stuart, The Advanced Theory of Statistics, III, New York: Hafner, 1966.

39. King, B., "Comment on 'Factor Analysis and Regression'," Econometrica, XXXVII (July 1969), 538-40.

40. Klein, L., "The Efficiency of Estimation in Econometric Models," in _Essays in Economics and Econometrics_, Chapel Hill: University of North Carolina Press, 1960.

41. _______, "Estimation of Interdependent Systems in Macro-economics," _Econometrica_, XXXVII (April 1969), 171-92.

42. Klein, L., and M. Nakamura, "Singularity in the Equation Systems of Econometrics," _International Economic Review_, III (September 1962), 274-99.

43. Kloeck, T., and L. Mennes, "Simultaneous Equations Estimation Based on Principal Components of Predetermined Variables," _Econometrica_ XXVIII (January 1960), 45-61.

44. Koerts, J., and A. Abrahamse, "On the Power of the BLUS Procedure," _Journal of the American Statistical Association_, LXIII (December 1968), 1227-36.

45. _______, "A Comparison Between the Power of the Durbin-Watson Test and the Power of the BLUS Test," _Journal of the American Statistical Association_, LXIV (September 1969), 938-48.

46. Koopmans, T., and W. Hood, _Studies in Econometric Method_, New York: John Wiley & Sons, 1953.

47. Lawley, D., and A. Maxwell, _Factor Analysis as a Statistical Method_, London: Butterworths, 1963.

48. Liu, T., "Underidentification, Structural Estimation, and Forecasting," _Econometrica_, XXVIII (October 1960), 855-65.

49. Liviatan, N., "Consistent Estimation of Distributed Lags," _International Economic Review_, IV (January 1963), 44-52.

50. Madansky, A., "Instrumental Variables in Factor Analysis," _Psychometrika_, XXIX (June 1964), 105-13.

51. Maddala, G., and A. Rao, "On Durbin's Test for Serial Correlation in Distributed Lag Models," New Haven: Cowles Foundation Discussion Paper No. 302, 1970.

52. Malinvaud, E., "Estimation and prévision dans les modèles économiques autoregressifs," _Review of the International Statistical Institute_, XXIX (1961), 1-32.

53. Marschak, J., "Economic Measurements for Policy and Prediction," in W. Hood and T. Koopmans, eds., _Studies in Econometric_

Method, New York: John Wiley & Sons, 1953, 1-26.

54. Massy, W., "Principal Components in Exploratory Statistical Research," Journal of the American Statistical Association, LX (March 1965), 234-56.

55. Nerlove, M., and K. Wallis, "Use of the Durbin-Watson Statistic in Inappropriate Situations," Econometrica, XXXIV (January 1966), 235-8.

56. Parks, R., "Efficient Estimation of a System of Regression Equations When Disturbances are Both Serially and Contemporaneously Correlated," Journal of the American Statistical Association, LXII (June 1967), 500-9.

57. Powell, A., "Aitken Estimators as a Tool in Allocating Predetermined Aggregates," Journal of the American Statistical Association, LXIV (September 1969), 913-22.

58. Quandt, R., "On Certain Small Sample Properties of K-Class Estimates," International Economic Review, VI (January 1965), 92-104.

59. Rao, C., Linear Statistical Inference and its Applications, New York: John Wiley & Sons, 1965.

60. Sargan, J., "The Maximum Likelihood Estimation of Economic Relationships With Autoregressive Residuals," Econometrica, XXIX (July 1961), 414-26.

61. ———, "A Comparison of the Asymptotic Efficiencies of Estimates of Stochastic Models With Autoregressive Errors," mimeographed, 1968.

62. Scott, J., "Factor Analysis and Regression," Econometrica, XXXIV (July 1966), 552-62.

63. ———, "Factor Analysis Regression Revisited," Econometrica, XXXVII (October 1969), 719.

64. Stone, R., "On the Interdependence of Blocks of Transactions," Journal of the Royal Statistical Society, Supplement, IX (1947), 1-32.

65. Taylor, L., and T. Wilson, "Three-Pass Least Squares: A Method for Estimating Models with a Lagged Dependent Variable," The Review of Economics and Statistics, XLVI (November 1964), 329-46.

66. Theil, H., Economic Forecasts and Policy, Amsterdam: North-

Holland, 1961.

67. Tintner, G., Econometrics, New York: John Wiley & Sons, 1952.

68. Wallis, K., "Lagged Dependent Variables and Serially Correlated Errors: A Reappraisal of Three-Pass Least Squares," The Review of Economics and Statistics, XLIX (November 1967), 555-67.

69. Wickens, M., "The Consistency and Efficiency of Generalized Least Squares in Simultaneous Equation Systems with Autocorrelated Errors," Econometrica, XXXVII (October 1969), 651-9.

70. Wonnacott, R., and T. Wonnacott, Econometrics, New York: John Wiley & Sons, 1970.

71. Zellner, A., "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association, LVII (June 1962), 348-68.

72. Zellner, A., and D. Huang, "Further Properties of Efficient Estimators for Seemingly Unrelated Regression Equations," International Economic Review, III (September 1962), 300-13.

73. Zellner, A., and H. Thornber, "Computational Accuracy and Estimation of Simultaneous Equation Economic Models," Econometrica, XXXIV (July 1966), 727-9.

IV. ESTIMATION OF THE MODEL

1. The Data Used

1.1. Assets

Seasonally unadjusted, current dollar quarterly stocks of the financial assets were obtained from the Federal Reserve Board's flow of funds accounts.[1] I found it convenient and desirable to assemble the asset data by market, in order to be completely certain about how the clearing of each market was accomplished. In most cases, the public was given the role of absorbing any residual holdings, because: (a) it is the most conglomerate sector, having been defined to include those residual sectors which were not specifically included in my other four sectors; (b) its holdings are generally too large to be significantly distorted by the addition of residual claims; and (c) in the compilation of the Reserve Board's data, "every transaction of the households is a residual, since every one...is derived from the books of other sectors." [6, p. I.44]

Currency plus demand deposits

The compilation here was straightforward, with the following series readily obtained: C2+DD2, C4+DD4,

---

[1]A full, consistent set of this data has recently been published in annual form [6], and can be obtained from the Board in quarterly form.

DD5, DD3, NFR, and C (currency outside banks). The public's holdings were residually calculated

$$C1 + DD1 = C + DD3 - DD5 - (C2 + DD2) - (C4 + DD4)$$

Time deposits and savings accounts

I included all corporate time deposits in CD2 ; all other time deposits and all savings accounts were credited to the public.

Short-term bonds

Included in this series are short-term marketable U.S. government securities and open market paper. The composition is displayed below:

	1952-I		1961-I		1969-IV	
	U.S. Govt	OMP	U.S. Govt	OMP	U.S. Govt	OMP
Public	12.0	-1.4	28.1	-4.4	53.4	-21.7
Corporations	9.5	.3	15.4	1.4	8.4	15.2
Banks	16.2	1.1	22.3	3.0	24.0	6.4
Government	-37.7	.0	-65.8	.0	-85.8	.0
Total	.0	.0	.0	.0	.0	.0

Savings Institutions' holdings of short-term government securities were included with their long-term bonds. Their short-term holdings amounted to .7, 1.8, and 2.9 billions in the three years displayed above. The public absorbed this as well as other discrepancies, since its holdings of both short- and long-term bonds were residually calculated so as to clear the markets. The above table makes clear the compositional shift that occurred with the growth of the OMP market.

Long-term bonds

In this series I included state and local government securities,[2] corporate and foreign bonds, and U.S. government securities (other than short-term marketable ones). The composition is displayed below:

	1952-I			1961-I			1969-IV		
	U.S. Govt	State & Local	Corp. & Frgn.	U.S. Govt	State & Local	Corp. & Frgn.	U.S. Govt	State & Local	Corp. & Frgn.
Public	88.2	-10.5	35.4	89.2	-21.7	70.9	108.8	-66.7	141.5
Corp.	10.3	.8	-40.0	4.1	2.5	-75.7	4.6	6.3	-147.6
Banks	45.3	9.5	2.3	39.8	18.5	1.0	40.6	60.2	-.8
Savings Inst.	11.5	.2	2.3	12.6	.7	3.8	16.1	.2	6.9
Govt.	-155.3	.0	.0	-145.7	.0	.0	-170.0	.0	.0

Federal long-term bonds have grown very slowly, while there has been a rapid expansion in public holdings of corporate bonds and bank holdings of state and local bonds.

[2]Between five and eight percent of these obligations are in fact short term, but sectoral holdings could not be obtained.

Public loans

Here I included consumer and security credit from banks and all noncorporate bank loans "not elsewhere classified." This composition is as follows:

	1952-I	1961-I	1969-IV
Bank loans N.E.C.	11.2	22.5	55.8
Consumer credit	7.6	21.1	48.2
Security credit	2.1	4.1	10.6

Consumer credit grew somewhat more rapidly than the other items in the first half of the period and slightly less rapidly in the second half.

Corporate loans

This series involved no special problems.

Mortgages

This was also straightforward, with the public's holdings again residually calculated.

Equity

This was calculated net of the value of open-end investment company shares.

Other

After compiling the above market data, I turned to the sectoral data to see what items had been implicitly left for the "other" category. The public's holdings of all items were once again calculated residually, so that they absorbed all final discrepancies.

To put the "other" items in perspective, the size in 1969-IV of the various items in the portfolios of the four private sectors is displayed in the balance sheets appearing on the next page.

The composition of the corporate sector's other holdings is presented in the following table:

OTH2

	1952-I	1961-I	1969-IV
Consumer credit	4.5	7.8	19.2
Trade credit	49.3	100.7	198.2
Misc. assets	15.5	39.7	78.5
-Finance Co. loans	-1.3	-2.2	-9.5
-U.S. govt. loans	-.6	-.8	-1.8
-Profit taxes payable	-20.2	-11.9	-21.2
-Trade debt	-36.0	-66.3	-117.5
-Misc. liabilities	-21.6	-47.5	-80.8
OTH2	-10.4	19.5	65.1

One obviously large item is the trade credit and debt data, which was not considered to be part of the portfolio decisions made by any sector. Even if it had been thought more reasonable to do so, it would have been difficult to use the data because of a large discrepancy

PORTFOLIO HOLDINGS IN 1969-IV
(Billions of dollars)

PUBLIC

Assets		Liabilities	
C1+DD1	183.4	Oth1	-32.2
TD1	175.8	W1-Oth1	1360.7
SA1	202.6		
Short1	31.7		
Long1	183.6		
-PL1	-114.6		
-Mort1	-217.6		
E1	883.6		
W1	1328.5		1328.5

CORPORATIONS

Assets		Liabilities	
C2+DD2	28.6	Oth2	65.1
CD2	17.0	W2-Oth2	-287.6
Short2	23.6		
-Long2	-136.7		
-CL2	-93.3		
-Mort2	-71.7		
W2	-232.5		-232.5

BANKS

Assets		Liabilities	
NFR	-.261	DD3	172.5
Short3	30.1	TD3	175.8
Long3	100.0	CD3	17.0
PL3	114.6	-RR	-39.4
CL3	93.3	-Oth3	72.1
Mort3	70.3		
W3	408.0		408.0

SAVINGS INSTITUTIONS

Assets		Liabilities	
C4+DD4	2.2	SA4	202.6
Long4	23.2	-Oth4	16.5
Mort4	193.7		
W4	219.1		219.1

created by timing factors.[3] If this discrepancy is allocated according to the reported liability proportions, then the net intersectoral claims are small though not insignificant:

Trade Data for 1969-IV

	Reported			Allocated Discrepancy	New Net
	Trade Credit	Trade Debt	Net		
Public	7.5	-24.3	-16.8	-11.0	-27.8
Corp.'s	198.2	-117.5	80.7	-53.2	27.5
Gov't.	7.3	-4.8	2.5	-2.2	.3
Total	213.0	-146.6	66.4	-66.4	.0

Consumer credit and finance company loans to corporations are somewhat more troublesome. There is a fair case for taking them out of the "other" category; but my model has no wholly satisfactory "asset" categories in which to put them, and creating new asset classifications did not seem warranted by these relatively small and to some extent offsetting items.

The composition of the banking sector's other

[3]Specifically, sellers write down receivables before buyers record payables, and buyers cross off payables before sellers cross off receivables.

holdings is displayed below:

OTH3

	1952-I	1961-I	1969-IV
Hypothecated depreciation	.5	.8	.0
Misc. assets	1.9	3.9	11.6
-Federal Reserve float	-.7	-.9	-3.4
-Loans from affiliates	-.0	-.0	-.6
-Taxes payable	-.7	-1.4	-.8
-Misc. liabilities	-3.8	-8.3	-49.6
-Net worth	-11.1	-18.2	-29.3
	-13.9	-24.1	-72.1

The only notable item is the extraordinary growth in miscellaneous liabilities. A substantial though not dominant part of this growth consisted of liabilities to foreign branches and floats in the banking system:

	1960-IV	1965-IV	1966-IV	1967-IV	1968-IV	1969-IV
Liabilities to frgn. branches	.9	1.4	4.0	4.2	6.0	13.0
Floats in comm. banking system	1.2	3.1	4.1	4.7	6.3	12.4
Unallocated	7.5	13.5	14.8	16.9	20.1	25.0
Misc. liabilities	9.7	17.9	22.9	25.8	36.0	49.6

The liabilities to foreign branches might have been made a separate item; but I decided that they were too small, too stable in the sample period for the associated coefficients to be reliably estimated, and not well enough understood for those coefficients to be subjectively specified.

The other items for savings institutions are listed below:

OTH4

	1952-I	1961-I	1969-IV
Consumer credit	.3	1.2	2.5
Time deposits	.2	.1	.1
Corporate stock	.3	.8	2.3
Other loans	.1	.2	.9
Misc. assets	1.3	4.1	9.2
-Borrowing from FHLB	-.6	-1.5	-9.3
-Bank loans N.E.C.	-.1	-.2	-.6
-Taxes payable	-.0	-.0	-.1
-Misc. liabilities	-.6	-2.5	-4.7
-Net worth	-3.9	-8.6	-16.8
	-3.1	-6.2	-16.5

The most important entries are net worth and borrowing from the Federal Home Loan Board. The latter item has been handled as if it were an exogenous instrument administered by the government, though it might have been better to treat it as an endogenous asset whose rate alone is set by the government.

The government's other items are displayed in the following table:

OTH5

	1952-I	1961-I	1969-IV
Time deposits	.3	.3	.3
Loans to coops (BC)	.3	.7	1.7
Loans to farmers (FICB)	.9	1.6	4.3
Loans to S & L (FHLB)	.6	1.5	9.3
Other loans to R. of W.	10.4	13.6	26.6
Other loans to others	3.8	7.1	18.0
Taxes receivable	20.6	13.3	20.7
Trade credit	1.6	1.8	7.3
F.R. float	.6	.9	3.4
Misc. assets	1.1	4.1	4.6
-Trade debt	-2.5	-3.2	-4.8
-Life insurance reserves	-6.2	-6.4	-7.3
-Retirement fund reserves	-7.6	-14.1	-25.2
-Misc. liabilities	-4.2	-2.6	-3.2
	20.6	18.6	55.7

The various loans are of substantial size and have grown significantly. However, it is probably most correct to think of them as exogenous policy items whose size and rates are set by the government--which is equivalent to how they have been treated.

The residually-calculated other holdings of the public are summarized below:

OTH1

	1952-I	1961-I	1969-IV
Life insurance reserves	6.2	6.4	7.3
Retirement fund reserves	7.6	14.1	25.2
Net taxes receivable	.3	.0	1.4
Net worth of banks and savings institutions	15.0	26.8	46.1
Misc. assets	30.2	60.9	138.3
-Savings inst. and gov't. time deposits attributed to public	-.5	-.4	-.4
-Savings inst. corp. stock attributed to public	-.3	-.8	-2.3
-Consumer credit	-4.8	-9.0	-21.7
-Other loans	-14.2	-21.7	-39.0
-Net trade debt	-12.4	-33.0	-83.2
-Misc. liabilities	-20.3	-52.6	-103.9
	6.8	-9.3	-32.2

1.2. Rates

All of the interest rates (other than r_{PL}) were taken from the FRB-MIT data bank. The treasury bill rate was used for r_{short} , and the corporate bond rate was used for r_{long} . The effective rate on savings and loan shares was used for r_{SA} .

Since I was unable to locate a reputable series of interest rates on personal loans, I constructed one from the table of "bank rates on short-term business loans," which is printed regularly in the Federal Reserve Bulletin. By judiciously merging the data for $1,000-10,000 and $10,000-100,000 loans, I hoped to approximate the behavior of the rates applicable to households, unincorporated businesses, and the other groups which make up the public sector.

In the 35 centers surveyed since 1967,[4] the dollar volume of the $1,000-10,000 loans was about 1/10 the volume of the $10,000-100,000 loans, but I used a 3/1 weighting in constructing my series, because most though not all of the loans that we are interested in would seem to fit in the smaller loan category. The rate series on the smaller loans has in fact exhibited more of the stability that is associated with personal loan rates. In particular, during the secular rise in rates between 1952 and 1970, the rate on the smaller loans began 60 or 70 basis points above the rate on larger loans, and ended less than 10 basis points above that rate.

---

[4] In 1967 the survey was broadened and its timing revised, but the net effects were apparently minor [7, pp. 721-7].

1.3. Miscellaneous

Following the Brookings model [4, p. 507], I let

$$\lambda_1 = .84$$

$$\lambda_2 TD3\ k_{TD} + \lambda_3 CD3\ k_{CD} = .82\ k_T(TD3 + CD3)$$

where .84 and .82 are the average[5] ratios of member bank to all commercial bank demand and time deposits.

I constructed an expected rate of inflation series

$$\begin{aligned}\dot{P}^E = {} & .14\dot{P} + .13\dot{P}_{-1} + .12\dot{P}_{-2} + .11\dot{P}_{-3} \\ & + .09\dot{P}_{-4} + .08\dot{P}_{-5} + .07\dot{P}_{-6} + .06\dot{P}_{-7} \\ & + .05\dot{P}_{-8} + .05\dot{P}_{-9} + .05\dot{P}_{-10} + .05\dot{P}_{-11}\end{aligned}$$

where

$$\dot{P} = (PXOBE/PXOBE_{-1})^4 - 1$$

is the actual annual rate of inflation in the implicit price deflator of GNP (OBE definition). PXOBE was taken from the FRB-MIT data bank.

r_E was generated from the model discussed in Section 2 of Chapter II.

[5]The FRB-MIT model uses two "blow-up" factors. The demand deposit factor has steadily grown from .84 to .89, and is endogenously explained; the exogenous time deposit factor was generally between .81 and .82 until 1969, after which it fell to .77.

2. The Estimation Methods Employed

The shortest data source was the Reserve Board's flow of funds data, which covered the period 1952-I through 1969-IV. One quarter (1952-I) was used to give initial values to all lagged variables; the remainder of the period was divided into 55 in-sample quarters (1952-II through 1965-IV) and 16 out-of-sample quarters (1966-I through 1969-IV).

I used five basic single equation estimation techniques, each of which was employed both with and without the use of a priori restrictions. The first technique was ordinary least squares. The second method was a "strict" application of instrumental variables, which will be discussed below. In the third method, I assumed a first order autoregressive structure and searched (to two places, between -1.00 and +1.00) for the autoregressive parameters which minimized the sum of the squared residuals. In the fourth method, I applied strict instruments and then searched for the minimizing autoregressive parameters. The fifth technique was a "liberal" instrumental variables approach that is discussed below.

In the strict instrumental method, I replaced only those variables which I considered to be directly dependent on the LHS variables in each sector's demand equations. The market clearing rates-- r_{short}, r_{long}, and r_{mort} --were replaced in all equations. r_E and El were replaced in the public's equations; and r_{pl} and r_{cl}

were replaced in the banking equations. This implicitly treats the remaining variables in each equation as if they were predetermined; these remaining variables, their lagged values, the lagged LHS variable and the lagged values of the replaced variables form the minimal set of instruments needed for the consistent estimation of the parameters in a first order autoregressive equation.[6] Since this group of variables was always larger than the number of instruments suggested by Dhrymes and by Kloeck and Mennes,[7] I confined myself to this set of instruments.

Unfortunately, the set of instruments was often very large. In the banking sector, for example, there were a total of 40 suggested instruments. In the spirit of Fair's arguments,[8] I eliminated those variables that were very collinear with the variables that were left in the set of instruments. This reduced the set to the following 16 instruments: Lagged $r_{long}h_1$, $r_{pl}h_1$, $r_{cl}h_1$, and $r_{mort}h_1$; and current and lagged $r_d h_1$, $\dot{P}^E h_1$, $long3_{-1}/W3$, $mort3_{-1}/W3$, $PL3_{-1}/W3$, and $CL3_{-1}/W3$. Sixteen was the minimum number of instruments that could have been used, since there are ten variables to be replaced and six of the instruments are themselves explana-

[6] See 5.2 in Chapter III and the references cited there.

[7] See 2.4.6 in Chapter III.

[8] See 5.2 above.

tory variables.[9] Similar condensations were made for the public and corporate sectors.

In the liberal instrumental method, I replaced all of the variables which would be endogenous in a complete system, and all lagged endogenous variables since the disturbance terms may be serially correlated. This left the discount rate and the four seasonal dummies as the only explanatory variables which were not replaced in this method. All 25 explanatory variables in the banking sector's demand equations are either endogenous or are the product of two variables, at least one of which is endogenous. Because I had dim hopes of finding 25 reasonably satisfactory instruments, I replaced the individual endogenous variables rather than the products which appear as explanatory variables. Thus, $r_{short}h_1$ was replaced with $\hat{r}_{short}\hat{h}_1$ rather than with $(\widehat{r_{short}h_1})$. This reduced the minimal number of instruments needed to 18, a requirement that I was able (with some effort) to meet with the following variables: four seasonal dummies, time, noninstitutional population over the age of sixteen, current and lagged military prime contracts for defense goods, manhours in excess of ten million idle due to major strikes, employment of federal government, urban population, armed forces, lagged government demand deposits,

[9] We could not have formed 16 linearly independent combinations of less than 16 variables.

lagged government mortgages, the lagged federal reserve discount rate, the lagged reserve requirements against demand and time deposits, and the lagged ceiling rate on passbook savings deposits.

3. Reduced Sample Results

3.1. The Application of a priori Information

In each sector, I first examined the amount of variance captured by the principal components of the explanatory variables, in order to get a rough idea of the effective dimensionality of the data set. Then I examined the multiple correlation coefficients between each variable and the remaining explanatory variables and the partial correlation coefficients between all pairs of explanatory variables, in order to discern which were the most troublesome collinearity patterns. I've only presented the F- and τ-statistics associated with these correlation coefficients, since these are monotonic transformations of the correlation coefficients and can be easily guaged against critical points. The critical point for τ at the .01 level was almost invariably 2.7, and values which exceeded this have been circled. The critical points for F appear where appropriate.

3.1.1. Public

An immediate problem is that the estimated coefficient of $\Delta E1/W1$ in the demand equation for $\Delta E1/W1$ will be one and all other estimated coefficients in that equation will be zero, which are not likely to be useful estimates. If we assume the coefficient of $\Delta E1/W1$ to be one, then we can subtract $\Delta E1/W1$ from both sides of the equation and move another variable over to the LHS for regression purposes. This will generally yield non-zero estimates of the remaining parameters, but the estimates will depend on which variable is put on the LHS and will not in general enforce the balance sheet constraints. Similarly, if we constrain the coefficient of $\Delta E1/W1$ in this equation to be other than 1.0, then single equation estimates will violate the balance sheet constraints, since the estimated coefficients of $\Delta E1/W1$ in the remaining equations will still sum to -1.0. I consequently decided to constrain all of the coefficients of $\Delta E1/W1$:

$$F_1 = \begin{bmatrix} .0 \\ -.15 \\ -.25 \\ -.10 \\ -.20 \\ .0 \\ .0 \\ .70 \end{bmatrix}$$

The following table displays the variance captured by the components of the remaining explanatory variables:

% of Variance	Cumulative %
89.3311653	89.3311653
5.8474630	95.1786280
3.0681210	98.2467489
.6504684	98.8972168
.4614633	99.3586798
.2868083	99.6454878
.1703550	99.8158426
.0822704	99.8981123
.0455272	99.9436388
.0237135	99.9673519
.0133022	99.9806538
.0090235	99.9896765
.0044607	99.9941368
.0024825	99.9966192
.0017466	99.9983654
.0008353	99.9991999
.0007277	99.9999275

The table on the following page displays along a diagonal the F-statistics and off-diagonal the τ-statistics associated with the correlations among the explanatory variables.

The coefficients of $CIDD1_{-1}/W1$, $SA1_{-1}/W1$, $Longl_{-1}/W1$, and $Mortl_{-1}/W1$ are the most obvious candidates for a priori specification, though there are clearly many additional places in which only approximately correct a priori information can be beneficial. I decided to completely specify the adjustment matrix E_1 because the other lagged assets are also substantially collinear and because knowledge of E_1 makes it much

	$\dot{p}^E$	r_{TD}	r_{SA}	r_{short}	r_{long}	r_{PL}	r_{mort}	r_E	1.0	$\frac{Y_d}{WI}$
$\dot{p}^E$	5.1	1.1	-3.5	-2.3	3.2	-.2	-1.9	-.2	2.4	.6
r_{TD}		262.0	3.4	2.1	-.4	-.5	-.8	-.6	-.4	.9
r_{SA}			467.2	-2.8	.5	2.9	-.6	2.3	2.2	-.7
r_{Short}				30.0	4.9	.2	-3.2	-1.5	2.0	1.3
r_{Long}					347.3	3.7	5.1	3.4	-4.3	-.5
r_{PL}						264.9	.2	-1.8	2.4	-.4
r_{Mort}							176.4	-.9	5.1	1.5
r_E								735.6	-1.3	1.7
1.0										.3
$Y_d/W1$										51.4
$C1DD1_{-1}/W1$										
$TD1_{-1}/W1$										
$SA1_{-1}/W1$										
$Short1_{-1}/W1$										
$Long1_{-1}/W1$										
$-PL_{-1}/W1$										
$-Mort1_{-1}/W1$										
$E1_{-1}/W1$										

	$\frac{C1DD1_{-1}}{W1}$	$\frac{TD1_{-1}}{W1}$	$\frac{SA1_{-1}}{W1}$	$\frac{Short1_{-1}}{W1}$	$\frac{Long1_{-1}}{W1}$	$\frac{-PL1_{-1}}{W1}$	$\frac{-Mort1_{-1}}{W1}$	$\frac{E1_{-1}}{W1}$
$\dot{p}^E$	1.9	-2.1	-.9	1.0	-.9	-.9	-2.3	-1.7
r_{TD}	-.2	-.2	2.2	-.4	-1.5	.8	1.1	.5
r_{SA}	-.4	-2.7	.1	-.0	-2.6	-.9	-1.0	-1.2
r_{Short}	1.3	.3	-2.5	2.8	.0	-2.7	-1.6	.9
r_{Long}	-1.1	-.8	2.0	-5.3	-2.2	2.4	1.4	.6
r_{PL}	-.8	1.5	-1.0	2.5	2.5	-1.0	-.1	.8
r_{Mort}	.6	-.6	-2.3	5.4	-.1	-.2	-2.9	-.2
r_E	2.7	4.1	.2	2.4	5.9	-1.9	2.0	-.4
1.0	2.0	1.5	2.0	-5.7	1.0	1.0	2.5	.4
$Y_d/W1$	-2.2	.6	-.4	-.9	1.6	-1.3	.0	-1.6
$C1DD1_{-1}/W1$	482.5	1.5	-.5	1.2	1.0	-2.2	.9	1.3
$TD1_{-1}/W1$		114.2	2.2	-2.1	-5.8	1.7	-1.1	-2.0
$SA1_{-1}/W1$			1513.6	2.2	1.4	.6	-8.4	1.1
$Short1_{-1}/W1$				37.5	-1.5	2.7	.4	-1.4
$Long1_{-1}/$					667.2	2.1	-1.0	.4
$-PL_{-1}/W1$						88.2	3.3	.2
$-Mort1_{-1}/W1$							2049.5	-.3
$E1_{-1}/W1$								24.9

$P[F_{16,38} > 2.5] = .01$

easier to get at the troublesome interest rates.[10]

$$E_1 = \begin{bmatrix} 1.0 & .0 & .0 & .01 & .02 & .06 & .05 & .05 \\ .0 & 1.0 & .0 & .04 & .05 & .12 & .15 & .10 \\ .0 & .0 & 1.0 & .05 & .15 & .12 & .20 & .20 \\ .0 & .0 & .0 & .90 & .08 & .0 & .05 & .05 \\ .0 & .0 & .0 & .0 & .70 & .0 & .05 & .10 \\ .0 & .0 & .0 & .0 & .0 & .70 & .05 & .0 \\ .0 & .0 & .0 & .0 & .0 & .0 & .40 & .0 \\ .0 & .0 & .0 & .0 & .05 & .0 & .05 & .05 \end{bmatrix}$$

The τ- and F-statistics for the remaining explanatory variables are

	$\dot{p}^E$	r_{TD}	r_{SA}	r_{short}	r_{long}	r_{PL}	r_{mort}	r_E	1.0	$Y_d/W1$
$\dot{p}^E$	4.3	2.3	-3.0	-.3	1.1	-.1	2.0	-.8	.7	.6
r_{TD}		336.2	16.0	2.9	.3	-1.5	-2.6	-.3	-1.1	.8
r_{SA}			702.0	-3.6	.1	3.3	.9	-.2	.6	-.5
r_{short}				22.9	2.8	1.9	-1.6	-2.2	.5	1.8
r_{long}					189.9	3.6	.6	.6	-4.9	-.6
r_{PL}						463.0	5.0	.3	4.8	-.4
r_{mort}							118.9	-.8	-1.1	1.2
r_E								95.8	.1	14.7
1.0										1.6
$Y_d/W1$										80.0

[10]The two-step procedure suggested in Chapter II was not employed here because of the time involved in applying it to ten estimation methods.

$$P[F_{8,46} > 2.9] = .01$$

I subsequently constrained the following coefficients:

	r_{SA}	r_{long}	r_{PL}	$Y_d/W1$
C1DD1*/W1	-2.0	-.5	-1.0	.20
TD1*/W1	-8.0	-1.0	.0	-.04
SA1*/W1	15.0	-3.0	-1.0	-.04
Short1*/W1	.0	-.5	.0	-.02
Long1*/W1	-2.0	10.0	.0	-.08
-PL1*/W1	.0	.0	4.0	-.06
-Mort1*/W1	.0	.0	.0	.0
E1*/W1	-3.0	-5.0	-2.0	.04

The τ- and F-statistics are now

	$\dot{p}^E$	r_{TD}	r_{short}	r_{mort}	r_E	1.0
$\dot{p}^E$	4.3	-3.5	2.2	2.0	.2	-.8
r_{TD}		39.4	3.1	2.9	-3.0	-.0
r_{short}			18.5	1.0	-1.2	.0
r_{mort}				21.1	-2.0	11.0
r_E					23.1	7.4
1.0						

$$P[F_{4,50} > 3.7] = .01$$

At this point, I decided that further a priori specification was not warranted by my faith in such specification when compared to the magnitude of the remaining collinearity. Reinforcing this, the eigenvalues no longer indicate a critical collinearity problem:

% of Variance	Cumulative %
79.2259884	79.2259884
10.2210859	89.4470739
7.1663198	96.6133938
2.5998898	99.2132835
.7866960	99.9999790

The final equations with a priori specification are:

$$\frac{\Delta a_1}{W1} - \overline{F}_1 \frac{\Delta E1}{W1} + E_1 \frac{a_1(-1)}{W1} - \overline{E}_1 \overline{A}_1^{(1)} \begin{bmatrix} r_{SA} \\ r_{long} \\ r_{PL} \end{bmatrix} - \overline{E}_1 \overline{B}_1^{(1)} \frac{Y_d}{W1}$$

$$= E_1 A_1^{(2)} \begin{bmatrix} \dot{p}^E \\ r_{TD} \\ r_{short} \\ r_{mort} \\ r_E \end{bmatrix} + E_1 B_1^{(2)} [1.0]$$

3.1.2. Corporations

The variance captured by each component is listed below:

% of Variance	Cumulative %
99.6063089	99.6063089
.1896666	99.7959747
.0971080	99.8930826
.0398826	99.9329643
.0329315	99.9658957
.0202618	99.9861565
.0081689	99.9943247
.0030962	99.9974203
.0011813	99.9986010
.0006434	99.9992437
.0005656	99.9998093
.0000848	99.9998932
.0000375	99.9999302
.0000204	99.9999504
.0000162	99.9999657

The patterns underlying this severe collinearity are displayed in the following table of τ- and F-statistics.

The lagged asset holdings are each highly correlated with the remaining explanatory variables, are highly intercorrelated with each other, and are intercorrelated with variables (such as $\dot{P}^E$, $\Delta ATL/W2$, and the rates) whose coefficients I would rather not specify. I therefore completely specified the adjustment matrix

$$E_2 = \begin{bmatrix} 1.0 & .0 & .01 & .02 & .02 & .03 \\ .0 & 1.0 & .14 & .14 & .13 & .15 \\ .0 & .0 & .85 & .14 & .13 & .15 \\ .0 & .0 & .0 & .60 & .12 & .09 \\ .0 & .0 & .0 & .10 & .60 & .08 \\ .0 & .0 & .0 & .0 & .0 & .50 \end{bmatrix}$$

	$\dot{p}^E$	r_{CD}	r_{short}	r_{long}	r_{CL}	r_{mort}	$\frac{Y-G}{W2}$	q	$\frac{C2DD2_{-1}}{W2}$
$\dot{p}^E$	7.7	-.6	1.9	1.2	-1.3	-.5	1.5	-1.1	-1.2
r_{CD}		57.0	.2	.3	.4	-1.4	1.5	-2.7	-1.0
r_{short}			26.7	1.2	4.2	-.8	-2.4	1.7	-1.2
r_{long}				105.5	2.3	1.8	-.3	1.1	.9
r_{CL}					203.2	2.8	1.8	-.4	1.0
r_{mort}						74.5	1.6	-1.9	.5
$\frac{Y-G}{W2}$							374.1	1.0	1.9
q								20.7	- .2
$\frac{C2DD2_{-1}}{W2}$									729.1
$\frac{CD2_{-1}}{W2}$									
$\frac{short2_{-1}}{W2}$									
$\frac{-long2_{-1}}{W2}$									
$\frac{-CL2_{-1}}{W2}$									
$\frac{-mort2_{-1}}{W2}$									
$\frac{\Delta ATL}{W2}$									

	$\frac{CD2_{-1}}{W2}$	$\frac{short2_{-1}}{W2}$	$\frac{long2_{-1}}{W2}$	$\frac{-CL2_{-1}}{W2}$	$\frac{-Mort2_{-1}}{W2}$	$\frac{\Delta ATL}{W2}$	1.0
$\dot{p}^E$	-1.2	-1.2	.4	1.0	-2.9	-1.0	2.3
r_{CD}	-4.1	-1.5	-1.3	-1.9	.8	-1.2	2.2
r_{short}	-2.6	-1.4	-.6	-.2	-2.8	-.7	-2.1
r_{long}	.8	2.8	.4	.8	1.0	.3	-1.1
r_{CL}	2.4	-1.3	-1.5	.2	2.3	-1.1	1.7
r_{mort}	.4	.1	1.1	.5	.6	.6	3.5
$\frac{Y-G}{W2}$	-.1	1.7	.3	-1.1	-2.0	1.1	-3.3
q	-1.2	-1.5	-.9	-2.9	1.2	-.9	3.9
$\frac{C2DD2_{-1}}{W2}$	-5.8	-4.3	-3.9	-4.2	-4.2	-3.9	1.5
$\frac{CD2_{-1}}{W2}$	135.2	-3.1	-1.2	-2.8	-5.4	-1.9	-.2
$\frac{short2_{-1}}{W2}$		69.6	-5.7	-3.4	-1.7	-5.4	3.1
$\frac{-long2_{-1}}{W2}$			179.9	-4.8	-.1	-8.1	3.9
$\frac{-CL2_{-1}}{W2}$				134.0	-.7	-2.7	2.1
$\frac{-mort2_{-1}}{W2}$					143.6	-1.4	-.9
$\frac{\Delta ATL}{W2}$						8.3	4.1

$$P[F_{14,40} > 2.6] = .01$$

This also allowed another troublesome variable, (Y-G)/W2 , to be eliminated from the RHS by fixing its coefficients in the "desired" equations

$$B_{23}\,\frac{Y-G}{W2} = \begin{bmatrix} .05 \\ .0 \\ .0 \\ -.02 \\ -.01 \\ -.02 \end{bmatrix} \frac{Y-G}{W2}$$

with these restrictions, the τ- and F-statistics were

	$\dot{p}^E$	r_{CD}	r_{short}	r_{long}	r_{CL}	r_{mort}	q	$\frac{\Delta ATL}{W2}$	1.0
$\dot{p}^E$	3.7	-2.4	2.7	1.0	-1.6	.9	-1.1	.4	-.1
r_{CD}		16.1	.1	2.6	.7	-3.1	1.8	.6	2.4
r_{short}			26.2	.8	4.3	-4.2	-.1	-.7	2.7
r_{long}				173.8	3.2	2.2	1.3	-.8	-2.7
r_{CL}					294.2	6.6	1.1	.6	-4.4
r_{mort}						146.9	-1.4	.0	19.4
q							17.5	-.4	2.4
$\frac{\Delta ATL}{W2}$								.4	0
1.0									

$$P[F_{7,47} > 3.0] = .01$$

r_{long}, r_{CL}, and r_{mort} remain so extremely collinear (with each other and with the remaining explanatory variables) that there still should be gains to be made with approximately correct a priori information. Accordingly, I constrained the coefficients of r_{long} and r_{CL} in the "desired" relations to be[11]

$$A_{24} r_{long} + A_{25} r_{CL} = \begin{bmatrix} .0 \\ 1.5 \\ 1.5 \\ -8.0 \\ 3.0 \\ 2.0 \end{bmatrix} r_{long} + \begin{bmatrix} .0 \\ .5 \\ .5 \\ 1.3 \\ -3.0 \\ .7 \end{bmatrix} r_{CL}$$

The eigenvalues are now such that

% of Variance	Cumulative %
77.8853273	77.8853273
16.2606568	94.1459837
3.8564144	98.0023975
1.5157173	99.5181141
.4017315	99.9198456
.0801390	99.9999838

and the τ- and F-statistics are:

[11] The signs are the opposite of their usual values because W2 is negative.

	$\dot{p}^E$	r_{CD}	r_{short}	r_{mort}	q	$\frac{\Delta ATL}{W2}$	1.0
$\dot{p}^E$	4.5	-2.8	2.7	.4	-1.4	.5	.9
r_{CD}		16.8	2.6	-1.3	3.8	.5	-.2
r_{short}			17.8	3.2	1.9	.9	-3.9
r_{mort}				9.9	2.6	.3	12.9
q					21.8	-.5	.7
$\frac{ATL}{W2}$						.5	.2
1.0							

$$P[F_{5,49} > 3.4] = .01$$

There obviously remains some significant collinearity, but none of it is extraordinary and any subjective constraints would have to be fairly accurate to exploit it.

The final equations are therefore

$$\frac{\Delta a_2}{W2} + E_2 \frac{a_2(-1)}{W2} - \overline{E}_2\overline{A}_2^{(1)} \begin{bmatrix} r_{long} \\ r_{CL} \end{bmatrix} - \overline{E}_2\overline{B}_2^{(1)} \frac{Y-G}{W2}$$

$$= E_2A_2^{(2)} \begin{bmatrix} \dot{p}^E \\ r_{CD} \\ r_{short} \\ r_{mort} \end{bmatrix} + E_2B_2^{(2)} \begin{bmatrix} q \\ 1.0 \end{bmatrix} + F_2 \frac{\Delta ATL}{W2}$$

3.1.3. Banks

The variance captured by each of the components is given below:

% of Variance	Cumulative %
80.9496002	80.9496002
12.2754458	93.2250452
3.0120911	96.2371359
2.7991198	99.0362549
.3448348	99.3810892
.2217488	99.6028376
.1471581	99.7499952
.0793817	99.8293762
.0620062	99.8913822
.0475697	99.9389515
.0352851	99.9742365
.0085499	99.9827862
.0078848	99.9906702
.0041892	99.9948587
.0020286	99.9968872
.0013853	99.9982719
.0005523	99.9988241
.0005272	99.9993505
.0003747	99.9997244
.0001152	99.9998388
.0000378	99.9998760
.0000176	99.9998932
.0000094	99.9999018
.0000035	99.9999046
.0000018	99.9999056

Quite obviously, many of the 25 parameters in each equation are not susceptible to accurate estimation. The τ- and F-statistics for the full set of explanatory variables revealed that for each of the variables

$$x = r_d,\ r_{short},\ r_{long},\ r_{PL},\ r_{CL},\ r_{mort},\ \dot{p}^E$$

the partial correlation between xh1 and xh2 is highly significant. In response to this, I constrained the associated coefficients in the "desired" equations to in each case be equal--i.e., $A_{31} = A_{32}$. This reduced the demand equations to

$$\Delta \begin{bmatrix} NFR \\ short3 \\ long3 \\ mort3 \end{bmatrix} \frac{1}{W3} = E_3 A_3 \begin{bmatrix} \dot{p}^E \\ r_d \\ r_{short} \\ r_{long} \\ r_{PL} \\ r_{CL} \\ r_{mort} \end{bmatrix} + \sum_{i=1}^{2} E_3 B_{3i} h_i - E_3 \begin{bmatrix} NFR \\ short3 \\ long3 \\ mort3 \\ PL3 \\ CL3 \end{bmatrix} \frac{1}{W3}$$

$$+ F_3 \frac{\Delta \ell_1}{W3} + \theta_{PL} \frac{\Delta PL3}{W3} + \theta_{CL} \frac{\Delta CL3}{W3} + U_3$$

The τ- and F-statistics for these explanatory variables are presented on the following page. Because of the high correlation between r_d and r_{short} , I now added the constraint that their coefficients in the desired relations be equal, with opposite signs. That is, the relevant part of each equation is changed from $\psi_{i2} r_d + \psi_{i3} r_s$ to $\psi_{i3}(r_s - r_d)$ by the restriction $\psi_{i2} = -\psi_{i3}$.

To combat the high collinearity between $\Delta \ell_3 / W3$ and several other variables I constrained its coefficients to be zero:

	$\dot{p}^E$	r_d	r_s	r_{long}	r_{PL}	r_{CL}	r_{mort}	h1	h2
$\dot{p}^E$	3.4	-1.6	.8	1.1	-1.8	1.3	.5	1.1	1.0
r_d		141.1	7.7	-3.3	1.4	.1	-1.4	-.2	-1.4
r_s			74.4	5.1	-4.1	2.7	.5	.2	1.6
r_{long}				156.8	4.4	-2.0	.1	1.9	-2.5
r_{PL}					918.0	10.2	2.0	2.1	4.0
r_{CL}						586.0	.0	-2.2	-3.7
r_{mort}							80.1	.7	-.7
h1								1315.7	-14.2
h2									1356.3
$\frac{NFR_{-1}}{W3}$									
$\frac{short3_{-1}}{W3}$									
$\frac{long3_{-1}}{W3}$									
$\frac{mort3_{-1}}{W3}$									
$\frac{PL3_{-1}}{W3}$									
$\frac{CL3_{-1}}{W3}$									
$\frac{\Delta \ell_1}{W3}$									
$\frac{\Delta PL1}{W3}$									
$\frac{\Delta CL1}{W3}$									

	$\frac{NFR_{-1}}{W3}$	$\frac{short3_{-1}}{W3}$	$\frac{long3_{-1}}{W3}$	$\frac{mort3_{-1}}{W3}$	$\frac{P13_{-1}}{W3}$	$\frac{CL3_{-1}}{W3}$	$\frac{\Delta\ell_1}{W3}$	$\frac{\Delta PL1}{W3}$	$\frac{\Delta CL2}{W3}$
$\dot{p}^E$	.9	-.9	-1.0	-.3	-.2	-.8	-.8	-.1	-.9
r_d	.7	.2	-1.0	3.7	-.7	.3	-.2	-1.1	-.8
r_s	.1	.3	1.4	-3.2	1.3	1.0	.6	.8	-.4
r_{long}	-.4	1.3	.4	3.2	-.8	.5	.8	-.5	-.7
r_{PL}	.4	-1.2	-.3	-2.9	.2	.4	-.7	.8	-.1
r_{CL}	-1.2	1.4	.8	-2.3	.0	-.0	1.1	-.4	.3
r_{mort}	.9	-.2	-1.1	2.2	-.6	-.8	-.4	-.3	.6
h1	2.6	26.2	18.2	1.6	5.7	9.9	16.1	-1.4	-.9
h2	2.0	12.7	7.9	3.3	4.6	5.8	9.3	-1.9	-.3
$\frac{NFR_{-1}}{W3}$	47.3	-2.6	-2.5	1.6	-5.4	-2.8	-2.1	3.7	.8
$\frac{short3_{-1}}{W3}$		346.9	-20.7	-1.8	-6.1	-10.8	-20.6	1.4	.7
$\frac{long3_{-1}}{W3}$			743.5	-.3	7.2	-11.6	-18.6	1.1	.6
$\frac{mort3_{-1}}{W3}$				571.4	2.2	-.6	-1.6	-1.2	.1
$\frac{PL3_{-1}}{W3}$					684.2	-5.4	-5.5	-1.2	-.1
$\frac{CL3_{-1}}{W3}$						103.4	-10.4	1.1	1.8
$\frac{\Delta\ell_1}{W3}$							103.9	.3	.2
$\frac{\Delta PL1}{W3}$								11.7	.5
$\frac{\Delta CL2}{W3}$									4.8

$$P[F_{16,38} > 2.5] = .01$$

$$F_3 = \begin{pmatrix} .0 \\ .0 \\ .0 \\ .0 \end{pmatrix}$$

This states that there are no short term adjustment effects of an increase in that part of demand deposits which are not covered by required reserves, offset exactly by a fall in -Oth3 and time deposits which are not covered by required reserves.

It is simplest to consider the typical quarter in which reserve requirements are constant:

$$0 = \Delta[(1 - \lambda_1 k_{DD})DD3] + \Delta[(1 - \lambda_2 k_{TD})TD3 - Oth3]$$

$$= (1 - \lambda_1 k_{DD})\Delta DD3 + (1 - \lambda_2 k_{TD})\Delta TD3 - \Delta Oth3$$

The change in required reserves is therefore

$$DRR = \lambda_1 k_{DD}\Delta DD3 + \lambda_2 k_{TD}\Delta TD3$$

$$= \lambda_1 k_{DD}\Delta DD3 + \lambda_2 k_{TD} \frac{\Delta Oth3 - (1 - \lambda_1 k_{DD})\Delta DD3}{1 - \lambda_2 k_{TD}}$$

$$= \frac{\lambda_1 k_{DD} - \lambda_2 k_{TD}}{1 - \lambda_2 k_{TD}} \Delta DD3 + \frac{\lambda_2 k_{TD}}{1 - \lambda_2 k_{TD}} \Delta Oth3$$

Even in this special case, we cannot say with certainty whether required reserves are rising or falling. An increase in demand deposits not covered by required reserves will generally involve an increase in required reserves; the offsetting fall in TD3 - Oth3 may raise or may lower required reserves.

The partial correlation between $Short3_{-1}/W3$ and $Long3_{-1}/W3$ is very significant and each of these variables is highly correlated with the remaining explanatory variables as a whole. If the relevant parts of the equations are

$$-\begin{pmatrix}\varepsilon_{12}\\ \varepsilon_{22}\\ \varepsilon_{32}\\ \varepsilon_{42}\end{pmatrix}\frac{Short3_{-1}}{W3} - \begin{pmatrix}\varepsilon_{13}\\ \varepsilon_{23}\\ \varepsilon_{33}\\ \varepsilon_{43}\end{pmatrix}\frac{Long3_{-1}}{W3}.$$

from

$$\begin{pmatrix}\varepsilon_{12}\\ \varepsilon_{22}\\ \varepsilon_{32}\\ \varepsilon_{42}\end{pmatrix}(Short3^* - Short3_{-1}) + \begin{pmatrix}\varepsilon_{13}\\ \varepsilon_{23}\\ \varepsilon_{33}\\ \varepsilon_{43}\end{pmatrix}(Long3^* - Long3_{-1})$$

in the adjustment equations, then I constrained

$$\begin{pmatrix}\varepsilon_{12}\\ \varepsilon_{22}\\ \varepsilon_{32}\\ \varepsilon_{42}\end{pmatrix} - \begin{pmatrix}\varepsilon_{13}\\ \varepsilon_{23}\\ \varepsilon_{33}\\ \varepsilon_{43}\end{pmatrix} = \begin{pmatrix}-.05\\ .65\\ -.55\\ -.05\end{pmatrix}$$

which is the net effect on NFR, short3, long3, and mort3 of desiring to hold \$1 more of short3 and \$1 less of long3 .

The net effect of a rise in corporate loans exactly offset by a fall in public loans is given by

$$\frac{\Delta PL3}{W3} = \frac{-\Delta CL3}{W3} \Longrightarrow \theta_{PL}\frac{\Delta PL3}{W3} + \theta_{CL}\frac{\Delta CL3}{W3} = (\theta_{CL} - \theta_{PL})\frac{\Delta CL3}{W3}$$

I constrained

$$\theta_{CL} - \theta_{PL} = \begin{pmatrix} .01 \\ .10 \\ -.07 \\ -.04 \end{pmatrix}$$

The demand equations are now

$$\Delta\begin{bmatrix} \text{NFR} \\ \text{Short3} \\ \text{Long3} \\ \text{Mort3} \end{bmatrix}\frac{1}{W3} - \begin{bmatrix} .01 \\ .10 \\ -.07 \\ -.04 \end{bmatrix}\frac{\Delta CL3}{W3} - \begin{bmatrix} -.05 \\ .65 \\ -.55 \\ -.05 \end{bmatrix}\frac{\text{Long3}_{-1}}{W3} = \sum_{i=1}^{2} E_3 B_{3i} h_i$$

$$+ E_3 A_3 \begin{bmatrix} \dot{p}^E \\ r_s - r_d \\ r_{long} \\ r_{PL} \\ r_{CL} \\ r_{mort} \end{bmatrix} - E_3^{(1)} \begin{bmatrix} \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{32} \\ \varepsilon_{42} \end{bmatrix} \left(\frac{\text{Short3}_{-1}}{W3} + \frac{\text{Long3}_{-1}}{W3} \right)$$

$$+ \theta_{PL}\left(\frac{\Delta PL3}{W3} + \frac{\Delta CL3}{W3} \right)$$

The τ- and F-statistics for these explanatory variables are given on the next page. The collinearity still seems substantial, and a potential source of gains through subjective constraints. Since h1 and h2 are so highly collinear with each other and with the remaining explanatory variables, I imposed the constraint

$$B_{31} - B_{32} = \begin{pmatrix} .005 \\ .060 \\ -.015 \\ -.025 \\ -.015 \\ -.010 \end{pmatrix}$$

As with the public and corporate sectors, I now decided to specify E_3 completely, since the lagged assets were troublesome and this would also allow me to get at the rates more easily:

$$E_3 = \begin{bmatrix} 1.0 & .0 & .05 & .05 & .05 & .05 \\ .0 & .85 & .20 & .10 & .50 & .65 \\ .0 & .15 & .70 & .20 & .30 & .20 \\ .0 & .0 & .05 & .65 & .15 & .10 \end{bmatrix}$$

Each of the last two columns displays the effect of a rise in wealth accompanied by an equal rise in desired loan holdings which cannot be immediately accommodated. Finally, I constrained the coefficients of r_{long} and r_{PL} to be

	$\dot{p}^E$	$r_s - r_d$	r_{long}	r_{PL}	r_{CL}	r_{Mort}	h1	h2
$\dot{p}^E$	4.1	.9	1.2	-1.5	.3	1.0	.9	.7
$r_s - r_d$		5.8	4.2	-3.4	2.9	2.2	2.1	2.0
r_{long}			173.7	3.4	-.9	1.9	-3.7	-2.2
r_{PL}				782.0	8.5	1.4	4.7	8.6
r_{CL}					466.4	.8	-3.6	-6.9
r_{mort}						89.1	1.7	.1
h1							232.4	-9.0
h2								300.0
$\frac{NFR_{-1}}{W3}$								
$\frac{short3_{-1}+long3_{-1}}{W3}$								
$\frac{mort3_{-1}}{W3}$								
$\frac{PL3_{-1}}{W3}$								
$\frac{CL3_{-1}}{W3}$								
$\frac{\Delta PL3+\Delta CL3}{W3}$								

	$\frac{NFR_{-1}}{W3}$	$\frac{S_{-1}+L_{-1}}{W3}$	$Mort3_{-1}$	$PL3_{-1}$	$CL3_{-1}$	$\frac{\Delta PL+\Delta CL}{W3}$
$\dot{p}^E$	1.1	-.9	-.7	.5	2.6	-.2
r_s-r_d	.1	.8	-.4	.6	.3	-.4
r_{long}	-.6	.9	.7	-.9	.4	-.6
r_{PL}	.2	-.8	-.5	-.6	.6	1.2
r_{CL}	-1.3	.3	1.2	.3	-.0	-1.1
r_{mort}	1.1	-.8	-.1	.3	-.9	1.0
h1	1.9	8.3	1.1	2.5	3.0	-6.0
h2	.4	4.3	-.3	2.8	.9	-4.0
$\frac{NFR_{-1}}{W3}$	55.5	-1.9	5.2	-5.3	-1.9	3.8
$\frac{short3_{-1}+long3_{-1}}{W3}$		87.0	2.1	-3.7	-3.7	6.8
$\frac{mort3_{-1}}{W3}$			243.5	10.9	2.1	-4.7
$\frac{PL3_{-1}}{W3}$				363.5	-1.8	6.8
$\frac{CL3_{-1}}{W3}$					26.2	4.6
$\frac{PL3+CL3}{W3}$						16.1

$P[F_{12,42} > 2.6] = .01$

	r_{Long}	r_{PL}
NFR*/W3	.02	.01
Short3*/W3	-2.02	-3.0
Long3*/W3	20.00	-3.0
PL3*/W3	-3.00	+10.00
CL3*/W3	-5.00	-3.0
Mort3*/W3	-10.00	-1.01

The demand equations are now

$$\Delta\begin{bmatrix} NFR \\ Short3 \\ Long3 \\ Mort3 \end{bmatrix}\frac{1}{W3} + \bar{E}_3\begin{bmatrix} NFR \\ Short3 \\ Long3 \\ Mort3 \\ PL3 \\ CL3 \end{bmatrix}\frac{1}{W3} - \begin{bmatrix} .01 \\ .10 \\ -.07 \\ -.04 \end{bmatrix}\frac{\Delta CL3}{W3} - \begin{bmatrix} .00175 \\ .03150 \\ -.01300 \\ -.02025 \end{bmatrix}h_1 - \begin{bmatrix} .1595 \\ -4.3065 \\ -1.6420 \\ 5.7990 \end{bmatrix}r_{PL}$$

$$- \begin{bmatrix} .1200 \\ -7.017 \\ 9.597 \\ -2.700 \end{bmatrix}r_{Long} = E_3A_3^{(2)}\begin{bmatrix} \dot{p}^E \\ r_s - r_d \\ r_{CL} \\ r_{Mort} \end{bmatrix} + E_3B_{32}1.0$$

$$+ \theta_{PL}\left[\frac{\Delta PL3}{W3} + \frac{\Delta CL3}{W3}\right].$$

The τ- and F-statistics for these RHS variables are:

	$\dot{p}^E$	$r_s - r_d$	r_{CL}	r_{Mort}	$\frac{\Delta PL3+\Delta CL3}{W3}$	1.0
$\dot{p}^E$	3.5	3.0	-2.9	2.9	.7	-2.0
$r_s - r_d$		6.6	3.9	-4.1	-.9	3.5
r_{CL}			186.2	27.0	-2.1	-12.4
r_{Mort}				187.3	1.9	25.3
$\frac{\Delta PL3+\Delta CL3}{W3}$					2.9	-2.0
1.0						

$$P[F_{4,50} > 3.7] = .01$$

The collinearity is still fairly serious--especially for r_{CL} and r_{Mort} --but I was not confident enough to further specify coefficients. The eigenvalues for this final set of RHS variables are such that:

% of Variance	Cumulative %
54.3297767	54.3297367
22.9947313	77.3244581
18.9876554	96.3121128
3.4395156	99.7516279
.2483545	99.9999819

(The equations for r_{PL} and r_{CL} were also subjected to a component analysis and estimated with and without constraints. However, since I ended up not using these equations in my reduced sample forecasting tests, I have omitted discussion of their estimation.)

3.1.4. Savings Institutions

The following table displays the variance captured by each of the 11 components for the savings institutions sector.

% of Variance	Cumulative %
33.5442781	33.5442781
33.5429463	67.0872240
31.7118297	98.7990532
1.1881871	99.9872398
.0086828	99.9959221
.0024013	99.9983225
.0008278	99.9991493
.0005407	99.9996891
.0001556	99.9998446
.0001274	99.9999710
.0000004	99.9999714

The F- and τ- statistics are displayed in the table on the next page.

The most serious problem is clearly the intercorrelation among $Long4_{-1}/W3$, $Mort4_{-1}/W3$ (to a lesser extent $C4DD4_{-1}/W3$), and the four seasonal dummies. The statistics here are somewhat misleading in that they appear to indicate that the seasonal dummies are very highly correlated with one another, when the problem is actually that $Long4_{-1}/W4$ and $Mort4_{-1}/W4$ are nearly

	r_{Long}	r_{Mort}	$\dot{p}^E$	S1	S2	S3	S4	$\frac{C4DD4_{-1}}{W4}$	$\frac{Long4_{-1}}{W4}$	$\frac{Mort4_{-1}}{W4}$	$\frac{\Delta SA4}{W4}$
r_{Long}	64.7	5.66	-.1	1.7	1.7	1.7	1.7	-1.5	-2.0	-1.6	-2.5
r_{Mort}		39.8	1.6	-1.0	-1.0	-1.0	-1.0	-1.1	1.7	1.6	1.5
$\dot{p}^E$			2.5	-1.5	-1.5	-1.6	-1.5	2.8	1.4	1.3	-1.3
S1				131814.	-664.6	-612.8	-480.5	6.1	41.0	54.1	4.9
S2					138552.	-484.	-746.7	6.1	42.3	55.5	5.0
S3						136131.	-379.1	6.1	41.9	55.4	4.8
S4							138513.	6.0	43.0	56.3	5.0
$\frac{C4DD4_{-1}}{W4}$								80.6	-4.6	-5.2	-1.5
$\frac{Long4_{-1}}{W4}$									2966.5	-62.1	-4.8
$\frac{Mort4_{-1}}{W4}$										3566.6	-4.7
$\frac{\Delta SA4}{W4}$											40.3

$P[F_{9,45} > 2.8] = .01$

constant and the manifestation of this correlation depends heavily on all four seasonal dummies being present. That this is indeed the case is obvious from the τ- and F-statistics when the four seasonal dummies are replaced by a constant and three seasonal dummies (see table next page).

Considering $\mathrm{Long4}_{-1}/\mathrm{W4}$, $\mathrm{Mort4}_{-1}/\mathrm{W4}$, and the constant term, we can only hope to obtain a good estimate of one of the three associated parameters, and to accomplish this we will need two a priori restrictions. $\mathrm{C4DD4}_{-1}/\mathrm{W4}$ is also a serious problem, and is pairwise correlated with many variables; it's coefficients are therefore good candidates for a subjective constraint.

The next most urgent problem seems to be the correlation between r_{CB} and r_{Mort}; here we can get one good estimate, but will need one restriction. Finally, $\Delta\mathrm{SA4/W4}$ seems to be a serious problem and is correlated with many variables.

The actual constraints that I imposed are as follows: The adjustment matrix was assumed to be

$$E_4 = \begin{bmatrix} 1.0 & .1 & -.1 \\ .0 & .8 & .2 \\ .0 & .1 & .7 \end{bmatrix}$$

The coefficients of $\Delta\mathrm{SA4/W4}$ were set at

	r_{Long}	r_{Mort}	$\dot{p}^E$	Const	S_2	S_3	S_4	$\frac{C4DD4_{-1}}{W4}$	$\frac{Long4_{-1}}{W4}$	$\frac{Mort4_{-1}}{W4}$	$\frac{\Delta SA4}{W4}$
r_{Long}	64.7	5.7	-.1	1.7	1.0	-1.5	.8	-1.5	-2.0	-1.6	-2.5
r_{Mort}		39.8	1.6	-1.0	-2.2	.3	-2.3	-1.1	1.7	1.6	1.5
$\dot{p}^E$			2.5	-1.5	1.5	-.4	2.5	2.8	1.4	1.3	-1.3
Const				.	.1	5.5	.2	6.1	41.0	54.1	4.9
S2					9.1	-.8	-8.0	-3.3	.4	.2	2.5
S3						22.3	-.1	-3.6	-4.8	-5.1	-8.3
S4							22.0	-4.7	.5	.2	3.4
$\frac{C4DD4_{-1}}{W4}$								80.6	-4.6	-5.2	-1.5
$\frac{Long4_{-1}}{W4}$									2966.5	-62.1	-4.8
$\frac{Mort4_{-1}}{W4}$										3566.6	-4.7
$\frac{\Delta SA4}{W4}$											40.2

$$\bar{F}_4 = \begin{bmatrix} .08 \\ -.02 \\ -.04 \end{bmatrix}$$

and it was assumed that

$$\bar{A}_4 = \begin{bmatrix} \alpha_{11} & -.2\gamma_1 & -.2\gamma_2 \\ \alpha_{21} & 3.0\gamma_1 & -2.8\gamma_2 \\ \alpha_{31} & -2.8\gamma_1 & 3.0\gamma_2 \end{bmatrix}$$

With these constraints imposed, the demand equations are

$$\frac{\Delta a_4}{W4} + E_4 \frac{a_4(-1)}{W4} - F_4 \frac{\Delta SA4}{W4} = E_4 \begin{bmatrix} 11 \\ 21 \\ 31 \end{bmatrix} \dot{P}^E + \gamma_1 \begin{bmatrix} -.18r_{Long} \\ 1.84r_{Long} \\ -1.66r_{Long} \end{bmatrix}$$

$$+ \gamma_2 \begin{bmatrix} -.18r_{Mort} \\ -1.64r_{Mort} \\ 1.84r_{Mort} \end{bmatrix} + E_4 B_4 \begin{bmatrix} S1 \\ S2 \\ S3 \\ S4 \end{bmatrix} + U_4$$

3.2. Forecasting Tests

In accordance with my arguments in Chapter I, I calculated mechanical ex post forecasts for the 16 quarters 1966-I through 1969-IV. In practice, we would not normally make quarterly forecasts so far from the data used to estimate the model; but this seems to generally be the best way to arrive at a test which is severe enough to discriminate between models and to weed out those models that rely on the limited dimensionality of the sample data. Forecasting one quarter extrapolations of accidential correlations is generally trivial and uninteresting. What we all seek is a model that will be able to forecast major swings in the variables and the effects of substantive policy changes. If we had a test forecast period which was full of fortuitous shocks, then letting repeatedly reestimated models forecast one quarter away from the in-sample period could provide a satisfactory test. In the absence of cooperative data, forecasting several quarters away from the sample data appears to be a good way to thwart temporary correlations.

One caveat to the results reported here is Christ's argument [1, pp. 546-9] that ex post predictive tests are unfair since the model may have been chosen with the data in mind. He advocates fitting one's model to all available data and then waiting for future data to test its forecasting ability. This is of course not very practical advice. It also seems unduly strict in this case since the model and the estimation methods seem to be too firmly rooted in economic and econometric theory to have been invented for the 1966-1969 experience. The a priori constraints on the other hand do not predate the out-of-sample period and may have been tainted by that experience. How serious this problem may be is unknowable; the only assurance I can provide is that the data was not consciously mined in any way.

A quadratic loss function implies [12, Ch. II] that we should look at squared errors. This harsher penalizing of larger errors is very attractive, and is consistent with the minimization of squared residuals in all of the regression techniques used here.[12] Whereas in parameter estimation there is substantive debate over which residuals should be minimized, the most interesting measure of forecasting accuracy is clearly the errors in the reduced form equations--i.e., the errors resulting when all currently endogenous variables are simultaneously forecast.

On the other hand, most models are incomplete, and even with complete models it is often desirable to concentrate on certain variables by assuming that the remaining variables are forecast perfectly. This should help to pinpoint trouble spots but may give a misleading portrait of a model's forecasting powers. In comparisons between estimators, this technique should be biased against estimators which treat as endogenous variables that are

[12]On the other hand, mean absolute or mean absolute percent errors seem to be much more easily put into perspective. Steckler [10, 11] has twice found that squared and absolute errors ranked forecasts in approximately the same order.

assumed exogenous in the forecasting tests.

In the simulations made with this model, the non-financial variables were taken as given and I had hoped to be able to solve the system for all of the endogenous financial variables. The solution[13] of this moderately large nonlinear system turned out however to be quite difficult. One reason for this is that most of the unconstrained adjustment matrices are unstable and instability is generally devastating to iterative solution techniques. The other major problem was the intense simultaneity created by the four market clearing rates and the two loan rates which "cleared" the banks' demand equations.

After a sufficient amount of frustration, I decided to run the forecasting tests with all of the interest rates exogenous. An obvious alternative would have been to begin again with all of the rates determined by simple rate adjustment mechanisms. Besides requiring a great deal of additional time and effort, this would have left me wide open to Christ's criticism of models which are altered to make their results more attractive.

On the other hand, in order to make true ex ante forecasts with a final model, I will undoubtedly have to make some concessions. Thus, in the single model that

[13]The program that I used was an adaptation by Thad Mirer of a simulation package for the FRB-MIT model.

I estimate in Section 4 below, I have purposely made assumptions which this experience had indicated would be necessary.

In evaluating a model's forecasting ability, it is usually helpful to consider an alternative "naive" model. One problem with such a comparison is that there are various levels of naiveté, which produce benchmarks of uneven difficulty. For example, the "no change" naive model is a tougher standard for relatively random variables than for trend-dominated variables. Similarly, it has been argued [2, 3] that the "no change" and "same change" naive models are not stringent enough, and that one should instead use the more general autoregressive scheme

$$\hat{Y} = \sum_{i=1}^{p} \alpha_i Y_{-i}$$

where the weights are those which result from an ordinary least squares regression of Y on its past values. Others have argued [8] that this is too tough a criteria, since Y's lagged values are likely to be highly correlated with the true explanatory variables and will therefore act as proxies for those variables. This is not a very convincing argument however, since there is a real question of whether theoretical models outperform "frankly statistical" models.[14]

---

[14] See Kendall, et.al. [5] and the ensuing discussion, including the remark by Durbin that, "As far as short

The most naive forecaster doesn't need to use any regression programs, though he might glance at the data. A more industrious straw man would take a careful statistical look at each variable's own history. The most ambitious skeptic of economic theory would mine all of the available data, by searching all possible sets of explanatory variables to find the best fit. Not having the time or temperment for exhaustive naiveté, I limited myself to an eight quarter autoregressive benchmark, weakened to the extent necessary to insure consistent forecasts.[15]

There are 35 endogenous variables (not counting interest rates); but since there are only 14 stochastic equations, there are only 14 linearly independent[16] forecast errors. With 14 independent pieces of information

term economic forecasting is concerned, my feeling is that it is not clear at present whether one does better to fit economic models based on postulated relationships between the variables, or to use statistical forecasting of a frankly ad hoc character."

[15] In deference to the market clearing identities, I constrained the estimated distributed lag parameters to be equal for all assets in the same market.

[16] Here only, I mean linearly independent in the sense of not being related by linear relations which contain no estimated parameters. This distinction is made because if there were fewer predetermined variables than stochastic relations, we would still presumably be interested in forecasts of as many variables as there are stochastic relations. No distinction is necessary if the forecasts are made with nonzero disturbances.

about forecasting accuracy, we should concentrate on exactly 14 linearly independent endogenous variables.

The tables on the following page display the Root Mean Square Errors (RMSE), for 14 such variables for each of the various estimation techniques. The last row of these tables displays the average RMSE's, a summary measure which weights larger variables more heavily than smaller variables. This can be seen by rewriting the variables as

$$\overline{RMSE} = \frac{1}{14} \sum_i RMSE_i = \sum_i \frac{\overline{Y}_i}{14} \frac{\sqrt{\sum_t (Y_i - \hat{Y}_i)^2}}{\overline{Y}_i} = \sum_i \frac{\overline{Y}_i}{14} \phi_i$$

ϕ_i is a dimensionless measure of forecasting accuracy; the $\overline{Y}_i/14$ are the weights used to construct the aggregate measure of forecasting accuracy. This is a reasonable technique for items within the same balance sheet[17] or in the same market, but would be a dubious technique for combining forecasts of rates with forecasts of assets.

Instead, we could apply equal weights to the ϕ_i constructed above. Another alternative is to rank the 11 forecasts of each variable and then compute the average

[17]For items in the same balance sheet, this measure is analogous to Theil's elegant measure of the "information inaccuracy" of the $\hat{Y}_i$:

$$\Sigma\, Y_i \, \mathrm{Log}(Y_i/\hat{Y}_i)$$

where $\Sigma \hat{Y}_i = \Sigma Y_i = 1$. Unfortunately, this measure is not defined for negative Y_i and therefore cannot be employed here.

In-Sample RMSE's

	I	II	III	IV	V	VI	VII	VIII	IX	X	Naive
TD1	.528	.620	.528	.602	1.141	1.358	1.351	1.162	1.173	1.402	.669
SA1	.497	.618	.497	.626	5.437	2.199	2.183	1.803	1.849	2.157	.347
Short1	.695	.970	.695	.877	2.046	1.614	1.620	1.724	1.735	1.623	1.142
Long1	.798	.852	.798	.889	1.312	3.756	3.737	3.314	3.349	3.796	1.021
-PL1	.758	.870	.758	.841	5.052	1.982	1.996	2.252	2.231	1.982	.977
-Mort1	.233	.247	.233	.250	.600	1.054	1.060	.936	.964	1.064	.341
CD2	.304	.304	.319	.305	.411	1.083	1.077	.547	.531	1.069	.409
Short2	.762	.819	.838	.825	1.184	1.326	1.296	1.087	1.041	1.321	1.167
-Long2	.728	.735	.758	.815	1.166	1.329	1.315	1.091	1.084	2.991	1.232
-CL2	.801	.803	.701	.861	.863	1.513	1.545	1.135	1.173	1.959	.732
-Mort2	.245	.252	.275	.370	.390	.428	.424	.441	.436	2.986	.264
DD3	1.868	2.098	2.354	2.873	4.441	3.397	3.306	7.436	7.410	2.757	1.481
NFR	.407	.783	.426	.860	3.057	.886	.912	.740	.875	.875	.644
Mort4	.295	.499	.382	.512	4.692	1.977	1.832	1.657	1.684	2.718	.283
AVE.	.637	.748	.683	.822	2.271	1.707	1.690	1.809	1.824	2.050	.765

Out-of-Sample RMSE's

	I	II	III	IV	V	VI	VII	VIII	IX	X	Naive
TD1	5.485	8.256	5.761	7.397	13.972	11.461	11.480	8.012	8.832	13.904	2.734
SA1	9.222	3.990	8.905	3.980	54.871	12.469	12.924	8.610	9.155	12.172	1.350
Short1	10.114	4.116	10.273	26.849	43.672	12.056	11.569	8.148	8.026	10.225	2.836
Long1	12.461	11.776	12.315	22.384	6.989	12.173	11.977	10.485	10.416	14.318	2.559
-PL1	11.647	14.529	11.630	10.206	48.339	17.247	16.628	12.679	13.526	18.613	2.821
-Mort1	4.960	4.164	4.993	.711	6.388	7.768	6.910	4.774	4.383	7.825	.918
CD2	1.636	1.563	2.893	2.219	6.209	18.265	18.573	2.032	2.180	14.917	1.689
Short2	1.382	2.584	1.992	5.696	5.297	21.875	24.741	1.989	1.291	19.433	1.471
-Long2	17.927	15.243	17.962	5.494	16.422	38.190	38.963	2.240	1.949	14.931	2.709
-CL2	7.085	6.466	8.258	6.132	7.568	30.843	31.387	4.032	4.899	15.132	1.236
-Mort2	3.086	3.726	3.288	1.953	5.230	3.646	3.908	1.129	.954	30.063	.421
DD3	24.870	31.493	7.683	7.126	36.148	64.324	63.156	14.221	14.305	55.008	3.474
NFR	13.710	11.700	11.775	36.271	46.230	2.055	2.516	1.091	2.530	4.041	.546
Mort4	4.294	2.524	6.276	1.903	46.245	15.666	13.360	8.035	8.579	25.132	.939
AVE.	9.147	8.724	8.143	9.880	24.842	18.262	19.156	6.248	6.502	18.265	1.836

Notes to Tables

I = OLS

II = strict instruments

III = first order autoregressive

IV = II plus III

V = liberal instruments

VI = I with a priori constraints

VII = II with a priori constraints

VIII = III with a priori constraints

IX = IV with a priori constraints

X = V with a priori constraints

rank assigned to each forecasting technique. The in-sample and out-of-sample ranks are displayed on the following pages. It can be seen that this changed the order of some of the methods which fell into the middle of the aggregate ranking.

A detailed discussion of the individual forecasting results could take a great deal of time and space. I've instead decided to state without proof some summary conclusions:

(i) The results overwhelmingly illustrate the fragility of estimates based on very collinear data and the inadequacy of in-sample fit as a guide to forecasting prowess.

(ii) The unconstrained estimates generally overstated the interest sensitivity of C1DD1, SA1, PL1,Short1 vs Long1, C2DD2, Long2, CL2, Mort2, NFR, Short3 vs Long3, C4DD4. This was apparently due to the presence of omitted pressures and enticements.

(iii) Putting in maxima and minima constraints could have been rewarding. In method I, for example, forecast C2DD2 and NFR fell to -$20 billion and -$29 billion, respectively. Constraints would however have necessitated some mechanism for distributing unallowed holdings to keep them from all being thrown into residual assets.

(iv) The liberal instruments methods (V and X) were decidedly unsuccessful. This was apparently because

In-Sample Ranks of RMSE's

	I	II	III	IV	V	VI	VII	VIII	IX	X	Naive
TD1	1.5	4	1.5	3	6	10	9	7	8	11	5
SA1	2.5	4	2.5	5	11	10	9	6	7	8	1
Short1	1.5	4	1.5	3	11	6	7	9	10	8	5
Long1	1.5	3	1.5	4	6	10	9	7	8	11	5
-PL1	1.5	4	1.5	3	11	6.5	8	10	9	6.5	5
-Mort1	1.5	3	1.5	4	6	9	10	7	8	11	5
CD2	1.5	1.5	4	3	6	11	10	8	7	9	5
Short2	1	2	4	3	8	11	9	6	5	10	7
-Long2	1	2	3	4	7	10	9	6	5	11	8
-CL2	3	4	1	5	6	9	10	7	8	11	2
-Mort2	1	2	4	5	6	8	7	10	9	11	3
DD3	2	3	4	6	9	8	7	11	10	5	1
NFR	1	5	2	6	11	9	10	4	7.5	7.5	3
-Mort4	2	4	3	5	11	9	8	6	7	10	1
AVE.	1.607	3.250	2.500	4.214	8.214	9.036	8.714	7.429	7.750	9.286	4.000

	I	II	III	IV	V	VI	VII	VIII	IX	X	Naive
TD1	2	6	3	4	11	8	9	5	7	10	1
SA1	7	3	5	2	11	9	10	4	6	8	1
Short1	5	2	7	10	11	9	8	4	3	6	1
Long1	9	5	8	11	2	7	6	4	3	10	1
-PL1	4	7	3	2	11	9	8	5	6	10	1
-Mort1	6	3	7	1	8	10	9	5	4	11	2
CD2	2	1	7	6	8	10	11	4	5	9	3
Short2	2	6	5	8	7	10	11	4	1	9	3
-Long2	8	7	9	4	6	10	11	2	1	5	3
-CL2	6	5	8	4	7	10	11	2	3	9	1
-Mort2	5	8	6	4	10	7	9	3	2	11	1
DD3	6	7	3	2	8	11	10	4	5	9	1
NFR	9	7	8	10	11	3	4	2	5	6	1
-Mort4	4	3	5	2	11	9	8	6	7	10	1
AVE.	5.357	5.0	6.0	5.0	8.714	8.714	8.929	3.859	4.143	8.5	1.5

the twenty instruments that I used contained a great deal less explanatory power than did the seventeen original explanatory variables in the public sector, and this loss of power was transmitted to other sectors.

(v) The a priori constraints that I imposed generally overstated the interest responsiveness of asset holdings. It is unclear whether this was a misspecification of the desired holdings or the adjustment equations.

(vi) For some selected variables such as NFR, the constraint methods (VI through X) did very well. Methods VI and VII were not however generally impressive, and the success of methods VIII and IX was largely dependent on an unanticipated virtue of applying a priori information. In the unconstrained estimation, the autoregressive parameters, $\hat{\rho}_i$, were all close to zero--apparently because any incorrectly excluded explanatory variables in the sample period were small relative to the random noise in the equations. With subjective constraints, however, a linear combination of explanatory variables of the form $\Sigma(\beta_i - \bar{\beta}_i)X_i$ is put into the error term, resulting in significant positive $\hat{\rho}_i$. Large $\hat{\rho}_i$ can in turn be very useful in compensating for repeated under- or over-predictions. Since this out-of-sample period can be characterized as a steady extrapolation into unprecedented interest rate levels, many of the errors were repeated quarter after quarter.

4. Full Sample Results

I estimated the final model, using all of the available data, by method VIII--single equation OLS with a priori constraints and residual-minimizing autoregressive parameters. This method was modified somewhat in that I was this time more modest in my application of subjective information.

4.1. The Application of a priori Information

The correlation patterns among the full set of explanatory variables for each sector are very similar to those for the reduced sample period, and I have consequently omitted the rather large initial tables of associated τ- and F-statistics. One consistent difference from the earlier statistics is that the extraordinarily high values of most variables in the previously out-of-sample years boosted the calculated dispersion of these variables about their means and consequently raised most of the correlation coefficients.

In all sectors the lagged asset holdings seemed to be especially good candidates for a priori specification. They were generally among the most troublesomely collinear of the variables; I am relatively confident about what their coefficients should be; and specification of these coefficients makes it easy to constrain the interest rate coefficients.

4.1.1. Public

After constraining the adjustment matrix, E_1, to have the same values I assigned it in the reduced period, the τ- and F-statistics were

	$\dot{p}^E$	r_{TD}	r_{SA}	r_{Short}	r_{Long}	r_{PL}	r_{Mort}	r_E	1.0	$\frac{Y_d}{W1}$
$\dot{p}^E$	12.1	.8	-2.2	.1	2.6	-1.4	1.9	-.6	.1	.8
r_{TD}		522.2	19.6	4.0	-2.6	1.5	-3.9	-.5	-3.3	.7
r_{SA}			1056.5	-4.4	4.8	-1.0	2.0	-.3	5.3	-.6
r_{Short}				98.3	5.0	1.5	-.6	-2.7	1.9	2.3
r_{Long}					512.0	1.3	1.6	.8	-5.5	-.6
r_{PL}						373.9	8.3	.6	-1.3	.0
r_{Mort}							437.0	-.7	2.8	.5
r_E								134.1	.6	16.1
1.0										1.7
$Y_d/W1$										86.4

$$P[F_{8,62} > 2.8] = .01$$

I now specified the coefficients of r_{SA} and r_{Long} to be 1/2 the values I assumed in the reduced sample period. The τ- and F-statistics became:

	$\dot{p}^E$	r_{TD}	r_{Short}	r_{PL}	r_{Mort}	r_E	1.0	$\frac{Y_d}{W1}$
$\dot{p}^E$	13.6	-3.1	2.3	-1.0	2.4	-.4	-2.5	.8
r_{TD}		62.8	2.1	2.2	-1.0	-2.0	2.3	.0
r_{Short}			84.1	3.1	-1.2	-3.0	-1.9	2.9
r_{PL}				498.3	17.6	.8	-3.1	-.0
r_{Mort}					401.2	-.6	6.1	-.2
r_E						182.5	.3	16.8
1.0								2.4
$Y_d/W1$								115.9

$$P[F_{6,64} > 3.1] = .01$$

The eigenvalues were such that

% of Variance	Cumulative %
86.7471304	86.7471304
9.8683101	96.6154404
1.8263723	98.4418125
.6885396	99.1303520
.4767431	99.6070948
.3673132	99.9744072
.0255627	99.9999685

I decided to stop at this point, whereas in the reduced sample period I went on to specify the coefficients of r_{PL} and $Y_d/W1$.

4.1.2. Corporations

I first specified the adjustment matrix and the coefficients of $(Y-G)/W2$ identically to the constraints I imposed for the reduced sample regressions. The τ- and F-statistics were than

	$\dot{p}^E$	r_{CD}	r_{Short}	r_{Long}	r_{CL}	r_{Mort}	q	$\frac{\Delta ATL}{W2}$	1.0
$\dot{p}^E$	12.1	-1.3	3.2	1.1	-2.3	2.1	-3.1	.7	-1.3
r_{CD}		49.5	.8	2.7	2.5	-3.2	.4	.8	1.3
r_{Short}			109.5	1.8	5.1	-3.9	.6	-1.1	2.2
r_{Long}				397.9	.4	4.6	4.0	-.5	-5.0
r_{CL}					692.4	10.0	-.5	.5	-6.0
r_{Mort}						568.9	-1.3	.0	19.2
q							23.2	-.3	3.5
$\frac{\Delta ATL}{W2}$								.5	.1
1.0									

$$P[F_{7,63} > 2.9] = .01$$

I now constrained

$$A_{25} r_{CL} + A_{26} r_{Mort} = \begin{bmatrix} .0 \\ .5 \\ .5 \\ 1.3 \\ -3.0 \\ .7 \end{bmatrix} r_{CL} + \begin{bmatrix} .0 \\ .1 \\ .1 \\ .5 \\ .3 \\ -1.0 \end{bmatrix} r_{Mort}$$

The τ- and F-statistics became

	$\dot{p}^E$	r_{CD}	r_{Short}	r_{Long}	q	$\frac{\Delta ATL}{W2}$	1.0
$\dot{p}^E$	15.0	-2.5	2.3	3.2	-3.5	.7	1.1
r_{CD}		59.6	3.0	2.8	1.6	.9	-5.6
r_{Short}			107.4	5.3	.6	-1.0	-2.4
r_{Long}				126.0	2.8	.1	4.1
q					27.9	-.6	6.6
$\frac{\Delta ATL}{W2}$						.6	.8
1.0							

$$P[F_{5,65} > 3.3] = .01$$

The eigenvalues were such that

% of Variance	Cumulative %
54.3425846	54.3425846
39.7112055	94.0537901
4.2782654	98.3320551
1.1605833	99.4926376
.4228832	99.9155207
.0844593	99.9999800

4.1.3. Banks

As in the reduced sample, variables of the form Xh1 and Xh2 were highly collinear with one another. In the asset demand and rate adjustment equations, I again responded by constraining the associated coefficients to be equal. This resulted in the replacement $\beta_1 Xh1 + \beta_2 Xh2 = \beta_1 X(h_1 + h_2) = \beta_1 X$. I decided to specify F_1 , E_3 , and the coefficients of r_{PL} identically with the reduced sample period; I also set the coefficients of r_{Long} at half their assigned values for the smaller sample. The τ- and F-statistics were then

	$\dot{P}^E$	r_d	r_s	r_{CL}	r_M	1.0	$\frac{\Delta PL3}{W3}$	$\frac{\Delta CL3}{W3}$
$\dot{P}^E$	12.0	-2.6	3.7	-3.0	4.5	-4.0	.9	-.2
r_d		266.7	7.7	.4	1.7	-1.9	.3	-.9
r_s			258.9	4.4	-4.1	2.6	.9	-1.7
r_{CL}				726.0	16.4	-9.2	-.6	1.0
r_M					488.7	26.0	.1	.1
1.0							-.0	-.7
$\frac{\Delta PL3}{W3}$							5.9	5.7
$\frac{\Delta CL3}{W3}$								10.4

$$P[F_{6,64} > 3.1] = .01$$

The eigenvalues were such that

% of Variance	Cumulative %
73.9197731	73.9197731
13.8587244	87.7784967
7.5408993	95.3193960
2.8072481	98.1266441
1.2576137	99.3842573
.516282	99.9005384
.0994243	99.9999619

No further constraints were imposed.

The explanatory variables in the rate adjustment equations for r_{PL} and r_{CL} are very similar to the preceding variables and the intercorrelations are consequently not displayed here. The coefficients of r_{Long} in $PL3^*/W3$ and $CL3^*/W3$ were constrained to have the same values assumed in the asset demand equations

4.1.4. Savings Institutions

I assumed here that the mortgage rate was set by this sector according to a rate adjustment mechanism and that they absorbed whatever market surplus arose. The demand equations are therefore

$$\Delta \begin{bmatrix} C4DD4 \\ Long4 \end{bmatrix} \frac{1}{W4} = E_4 A_4 \begin{bmatrix} \dot{p}^E \\ r_{Long} \\ r_{Mort} \end{bmatrix} + E_4 B_4 \begin{bmatrix} S1 \\ S2 \\ S3 \\ S4 \end{bmatrix} - E_4 \begin{bmatrix} C4DD4 \\ Long4 \\ Mort4 \end{bmatrix}_{-1} \frac{1}{W4}$$

$$+ F_4 \frac{\Delta SA4}{W4} + \theta_{Mort} \frac{\Delta Mort4}{W4}$$

The $\mathfrak{r}$- and F-statistics for these variables are

	$\dot{p}^E$	r_{Long}	r_{Mort}	S1	S2	S3	S4	$\frac{C4DD4_{-1}}{W4}$	$\frac{Long4_{-1}}{W4}$	$\frac{Mort4_{-1}}{W4}$	$\frac{\Delta SA4}{W4}$	$\Delta Mort4$
$\dot{p}^E$	8.7	.1	1.9	.7	.7	.7	.7	.8	-.7	-.8	-.4	-2.4
r_{Long}		213.1	7.0	3.2	3.2	3.2	3.2	-5.0	-3.2	-3.1	-3.1	-1.8
r_{Mort}			133.3	-1.7	-1.7	-1.7	-1.7	-.1	2.2	2.0	.6	2.0
S1				$3x10^5$	-1128.	-890.	-976.	7.8	73.0	103.2	6.5	10.0
S2					$3x10^5$	-1059.	-1352.	7.8	75.2	107.0	6.5	10.1
S3						$3x10^5$	-871.	7.8	75.4	107.9	6.4	10.1
S4							$3x10^5$	7.7	75.9	108.2	6.5	10.1
$C4DD4_{-1}/W4$								228.6	-6.6	-6.9	-4.6	-3.2
$Long4_{-1}/W4$									7989.	-131.2	-6.1	-10.4
$Mort4_{-1}/W4$										$1x10^4$	-6.1	-10.4
$\Delta SA4/W4$											46.6	-2.5
$\Delta Mort4/W4$												55.0

$P[F_{10,60} > 2.6] = .01$

The high partial correlations among the Si are again due to all four seasonals being necessary to capture the constancy of other variables, most notably the lagged asset holdings. The adjustment matrix was consequently constrained to be

$$E_4 = \begin{bmatrix} 1.0 & .38 & .66 \\ .0 & .62 & .34 \\ .0 & .0 & .0 \end{bmatrix}$$

These are again the partial effects of disequilibria, given that mortgage holdings cannot be changed. Equivalently, any such short run change in mortgage holdings will create an offsetting market surplus that must be absorbed.

The τ- and F-statistics are:

	$\dot{P}^E$	r_{Long}	r_{Mort}	S1	S2	S3	S4	$\frac{\Delta SA4}{W4}$	$\frac{\Delta Mort4}{W4}$
$\dot{P}^E$	7.1	-1.6	2.4	-.8	-.7	-.7	-.7	-.2	-2.2
r_{Long}		166.4	17.3	-5.1	-4.8	-5.3	-4.6	-1.5	-1.0
r_{Mort}			167.1	11.4	10.6	11.6	10.4	-.5	.2
S1				889.0	-60.7	-42.6	-62.0	3.7	1.7
S2					961.9	-46.9	-63.8	3.6	2.1
S3						899.8	-39.9	2.5	2.6
S4							991.0	4.0	1.8
$\frac{\Delta SA4}{W4}$								38.1	3.2
$\frac{\Delta Mort4}{W4}$									19.4

$$P[F_{7,63} > 2.9]$$

No further constraints were applied.

The adjustment equation for r_{Mort} is

$$\Delta r_{Mort} = \phi_{Mort} \frac{Mort4^* - Mort4_{-1}}{W4}$$

which can be expanded to

$$r_{Mort} = \frac{\phi_{Mort} A_4^{(1)}}{\lambda} \begin{bmatrix} \dot{p}^E \\ r_{Long} \end{bmatrix} + \frac{1}{\lambda} r_{Mort}(-1) + \frac{\phi_{Mort} B_4^{Mort}}{\lambda} \begin{bmatrix} S1 \\ S2 \\ S3 \\ S4 \end{bmatrix}$$

$$- \frac{\phi_{Mort}}{} \frac{Mort4_{-1}}{W4}$$

where $\lambda = 1 - \phi_{Mort} A_4^{(2)}$; $A_4^{(2)}$ is the coefficient of r_{Mort} and $A_4^{(1)}$ consists of the coefficients of $\dot{p}^E$ and r_{Long} in the $Mort4^*$ equation.

The τ- and F-statistics are

	$\dot{p}^E$	r_{Long}	$r_{Mort}(-1)$	S1	S2	S3	S4	$\frac{Mort4_{-1}}{W4}$
$\dot{p}^E$	9.8	2.6	.8	1.9	1.8	1.8	1.9	-4.2
r_{Long}		157.5	10.7	-14.7	-14.6	-14.4	-14.5	4.6
$r_{Mort}(-1)$			115.1	8.7	8.8	8.7	8.7	.1
S1				936.6	-61.0	-60.1	-61.1	6.6
S2					943.9	-61.0	-61.5	6.5
S3						933.8	-61.3	6.5
S4							951.1	6.5
$\frac{Mort4_{-1}}{W4}$								33.4

$$P[F_{6,64} > 3.1] = .01$$

The eigenvalues were such that

% of Variance	Cumulative %
33.5377545	33.5377545
33.5357485	67.0735025
32.1192245	99.1927271
.7887775	99.9815044
.0152554	99.9967594
.0028355	99.9995947
.0003836	99.9999781

As with the asset demand equation, I did not specify any of the coefficients in the desired relations.

4.1.5. Government

In my attempts in Section 3 to let all rates be endogenous, I found that (due to the thinness of the market) the short term rate was extremely unstable. I consequently assumed here that this rate is an exogenous policy instrument which the government sets with whatever open market transactions are necessary.

4.2. Forecasting Tests

With the interest rates (other than r_{short}) permitted to be forecast imperfectly, there are 20 stochastic equations and hence 20 linearly independent forecast errors. The full sample RMSE's for 20 linearly independent variables are presented below.

	Rates Endogenous	Rates Exogenous	Autoregressive Naive
TD1	2.097	1.700	1.309
SA1	2.847	2.902	.546
Short1	2.030	2.091	1.487
Long1	4.251	4.098	1.474
-PL1	3.104	3.003	1.488
-Mort1	1.417	1.445	.472
E1	48.257		34.164
CD2	.906	.931	.782
Short2	1.373	1.215	1.261
-Long2	1.196	1.167	1.425
-CL2	2.188	1.921	.831
-Mort2	.495	.490	.324
DD3	6.249	8.053	2.025
NFR	.573	.575	.647
Mort3	3.555	3.412	.270
Long4	2.088	3.720	.511
r_{long}	.271		.133
r_{mort}	.100		.086
r_{PL}	.188		.129
r_{CL}	.260		.208

The second column of this table contains the RMSE's when all interest rates are assumed (as in Section 3) to be forecast perfectly. This assumption here does not have a major impact on the RMSE's, and in fact raises the average for the 15 comparable variables from 2.291 to 2.448.

REFERENCES

1. Christ, C., Econometric Models and Methods, New York: John Wiley and Sons, 1966.

2. Coen, P., E. Gomme, and M. Kendall, "Lagged Relationships in Economic Forecasting," (with discussion), Journal of the Royal Statistical Society, Series A, CXXXII (1969).

3. Cooper, R., "The Predictive Performance of Quarterly Econometric Models of the United States," Working Paper No. 113, University of California, Berkeley, 1969.

4. Cunningham, J., in The Tasks of Economics, NBER 45th Annual Report, June 1967, 60-64.

5. de Leeuw, F., "A Model of Financial Behavior," in The Brookings Quarterly Econometric Model of the United States, J. Duesenberry, G. Fromm, L. Klein, and E. Kuh, eds., Chicago: Rand McNally, 1965, 465-532.

6. Federal Reserve Board, Flow of Funds Accounts 1945-1968, Washington: Board of Governors of the Federal Reserve System, 1970.

7. __________, Federal Reserve Bulletin, LIII (May 1967)

8. Howrey, E., L. Klein, and M. McCarthy, "Notes on Testing the Predictive Performance of Econometric Models," University of Pennsylvania Discussion Paper No. 173.

9. Steckler, H., "Evaluation of Econometric Inventory Forecasts," Review of Economics and Statistics, LI (February 1969), 77-83.

10. __________, and S. Burch, "Selected Economic Data-Accuracy vs. Reporting Speed," Journal of the American Statistical Association, LXIII (June 1968), 436-44.

11. Theil, H., Applied Economic Forecasting, Amsterdam: North Holland, 1966.