ECONOMETRICA
JOURNAL OF THE ECONOMETRIC SOCIETY

Errors in Variables and Other Unobservables

Author(s): Zvi Griliches

Source: *Econometrica*, Nov., 1974, Vol. 42, No. 6 (Nov., 1974), pp. 971-998

Published by: The Econometric Society

Stable URL: https://www.jstor.org/stable/1914213

ECONOMETRICA

VOLUME 42 November, 1974 NUMBER 6

ERRORS IN VARIABLES AND OTHER UNOBSERVABLES[1]

BY ZVI GRILICHES

This lecture surveys the history and recent resurgence of interest in models with errors in variables and substantive unobservable variables. Among several examples of such models, special attention is paid to a schooling-occupation-income achievement model in which identification and estimation are based on the variance-components structure (across families) of the unobserved individual ability variable.

Then the officers of the children of Israel came and cried unto Pharaoh, saying, Wherefore dealest thou thus with thy servants?

There is no straw given unto they servants, and they say to us, Make brick: and behold thy servants are beaten; but the fault is in thine own people.

But he said, Ye are idle, ye are idle: Therefore ye say, Let us go and do sacrifice to the Lord.

Go therefore now, and work; for there shall no straw be given you, yet shall ye deliver the tale of bricks.

Exodus 5, 15–18

1. INTRODUCTION

IT IS CUSTOMARY in the introduction to these lectures to pay tribute to Henry Schultz or Irving Fisher and to make some connection between the topic at hand and their work. It is fair to say, however, that while both Fisher and Schultz were well aware of some of the problems to be discussed in this lecture, having looked the evil squarely in the eye, they moved on, and so did most of the profession with them. I wish to dedicate this lecture, instead, to the memory of Ragnar Frisch, who saw clearly most of the problems in this area and worked hard on their solution. He too, however, abandoned this area of work later in his life. Only recently, with the small revival of interest in models involving errors in variables and other unobservables has Frisch's work, and the work of other early pioneers such as Sewall Wright, begun to be rediscovered by a generation to whom all of this was nothing more than obscure footnote references in Cowles Commission monographs.[2]

I will try to summarize briefly what I think were the major econometric problems as Frisch saw them: (i) The observed econometric magnitudes are subject to *both* errors of measurement and to random disturbances. (ii) The economic relations of interest, which came to be known later on as the "structure," hold across the "systematic" parts of these variables and not across the observed magnitudes. (iii) In many cases there is more than one relation connecting subsets

[1] Henry Schultz memorial lecture presented at the Oslo meeting of the Econometric Society, August, 1973. I am indebted to Gary Chamberlain for many discussions and suggestions on the substance of this paper, to A. S. Goldberger and M. Nerlove for helpful comments on an earlier version, and to the National Science Foundation (grant no. G.S. 2762X) for financial support.

[2] See Arthur Goldberger's Schultz Lecture [**19**] given at the 1971 Barcelona meetings for a discussion of Wright's work. There is quite a bit of overlap between this lecture and Goldberger's, but a good story is worth telling twice.

of variables within a particular set, leading to the problem "confluence." The investigator, faced with a particular set of data, has to determine, somehow, how many independent relations obtain within this set of data and between which variables, and to estimate the coefficients of these relations as well as possible.

I confess that I found Frisch's major work on this subject [**10**] very hard to read and to understand, even with hindsight and knowledge of some of the solutions to the problems with which he was struggling. His most general model can be characterized schematically as follows (cf. [**10**], pp. 47–57):

$$y_k = x_k + e_k$$

where the x's are the relevant "true" economic magnitudes, the y's are the observables, and the e's are random errors of measurement. Each "true" economic variable x_k is in turn related to a number of "basic" variables $f_1, \ldots, f_n$, which are not observed directly. An unobservable variable which affects only one particular x, say x_k, is thought of as a disturbance, and we shall label it u_k. "Basic" variables which affect more than one x at a time are called "systematic." Thus,

$$x_k = Fa_k + u_k$$

where $F = (f_1, \ldots, f_n)$, and the whole system of observations can be written in matrix form as

$$Y = (y_1, \ldots, y_k) = FA + U + E.$$

I have deliberately altered Frisch's notation to stress the parallel with factor analysis. The u's are "specific factors", the f's are general factors, and the e's are measurement errors. Frisch's work can be read as an early contribution to the principal components and factor analysis literature though it seems to have had little effect on or recognition in psychometrics.[3]

The model as outlined is grossly underidentified. The road followed by psychologists towards identification was to assume that the f's, the u's and the e's were all mutually orthogonal and to let the interpretation of these f's emerge as the consequence of the statistical analysis. Both lines seem unpromising to an economist: our "basic" variables are unlikely to be orthogonal, and we believe that our theories impose more structure on the data and provide more definite ideas about what the f's should be and how they should enter into the model. Instead, Frisch concentrated on, first, ways for determining the approximate rank of A from the data, i.e., how many independent relations connect the systematic parts of the observables? Second, having decided on the number of relations and the variables belonging to them, how does one estimate the coefficients of interest or get *bounds* on them?

It is fair to say that, while his questions turned out to be very fruitful, his solutions were often rather unsatisfactory and inelegant, the latter being perhaps the ultimate sin in the eyes of the more rigorous generation that followed him. He tried to do too much; he tried to solve *simultaneously* the errors-in-variables problem, the *simultaneity* (confluence) problem, and the model *choice* problem.

[3] See Goldberger [**16**] for a detailed discussion of parallels between econometrics and psychometrics.

The last is still probably the most important methodological problem before us, only modest progress having been made on it in recent years. The progress that was to come in the 1940's and 1950's was based on handing the problem of model choice back to the theorist and limiting the role of the statistician to the development of estimation methods for a *given* model whose structure, including the maximal rank of A, was specified in *advance* of estimation. In their haste to solve the simultaneity (or "confluence") problem, the developers of estimation methods for simultaneous equations systems dropped overboard as unnecessary baggage the concept of "basic," "systematic," or underlying unobservable variables, and greatly downgraded the errors-in-variables problem.

The question of why the market for errors-in-variables models has been so poor is interesting, and I shall speculate a bit about it in the next section. It is surely not due to the absence of errors in our data. Unobservable variables are such a tempting and possibly useful concept that they have been slowly creeping back into our models and were probably never effectively banished from the scene. I shall discuss them in the third section of this paper. The fourth section will examine in some detail a particular income-generating model based on an unobservable ability variable with a variance-components structure. The fifth section presents two more examples of models containing unobserved human capital and other erroneously measured variables. The final section of the paper again touches on the earlier and current literature on this topic.

2. ERRORS IN ECONOMIC VARIABLES

That our data are subject to a variety of errors needs no reiteration or much documentation. Nevertheless, we seem to have a rather ambivalent attitude about it. We complain quite a bit but do very little about it. Much of the problem, I think, arises because of the separation in economics between data producers and data analyzers. By and large, we do not produce our own data and, hence, do not feel responsible for it. In most physical or biological sciences, the experimenter is also the analyzer, and since he generates the data that he analyzes, he is aware of the sources of error in his data and has a direct incentive to improve their quality. He would be embarrassed if, when the experiment were duplicated by somebody else, it turned out that his results were due to impurities in materials or other lapses in experimental technique. This is not just a matter of experimental science versus non-experimental science. In non-experimental sciences such as astronomy or archeology, observations are performed or collected directly by professionals themselves or under their immediate supervision, and analysis is closely tied to the generation of the observational material. Our data are largely "reported" by firms and persons who are not professional observers and who do not have any stake in the correctness and precision of the observations they report.[4] While economists have increased their use of surveys in recent years

[4] I am distinguishing here between "reporting" or generating the original observation and "collecting" it. Collection is often in the hands of professionals. The lack of professionalism in reporting is emphasized by S. Kuznets [**46**].

and even designed and commissioned a few special purpose ones of their own, in general, the data collection and thus the responsibility for the quality of the collected material is still largely delegated to census bureaus, survey research centers, and similar institutions, and is divorced from the direct supervision and responsibility of the analyzing team.

It is also fair to note that part of the problem arises from our own fuzziness about what it is we would like to observe and from the complexity of the phenomena which we are trying to measure. Ask yourself what the exact observations are that you would like to see collected for the problems in which you are interested, if cost is no object. Do not state just vague variables, but specify operationally the desired measurements. Thinking about such "alibi-removing" questions should make us a bit more humble and restrain our continuing attacks on the various official data producing agencies.

In any case, it is clear that the original data are subject to both random and systematic errors. The analysis of systematic errors is really a version of the more general problem of tests of alternative hypotheses. If we have a clear idea of the source of such errors, our models are susceptible to specification error tests.[5] I shall deal in the rest of this lecture with the random errors of measurement and observation that our textbooks assume and exposit.

The "random errors-in-variables" model has been analyzed extensively in the literature and estimation methods have been developed for it. Nevertheless, until recently, applied econometric work has paid little more than lip service to it. The question is, why? My answer is based primarily on my own experience with what was taught to me in the fifties and what I was teaching to my students in the sixties. Others were probably influenced by slightly different heritages and may have seen these same developments in a different light.

There were two major streams of applied econometrics work to be discerned when I began my studies: agricultural demand analysis and macro-economic models.[6] The agricultural economics tradition of demand and price analysis, including the Working brothers, Mordecai Ezekiel, Henry Schultz, and Fredrick Waugh, had reached its peak and probable end in the mid-fifties in the demand and price structure work at the United States Department of Agriculture under the direction of Fox and Foote. The macro-economic model-building tradition, begun by the work of Tinbergen and Klein and buttressed by the newly developed econometric methodology of the Cowles Commission, was just beginning to hit its stride. The two traditions had come together repeatedly, in the presence of Tintner at Iowa State College, in the Girshick-Haavelmo article [**15**], in the person of Lawrence Klein who had been a student of George Kuznets at Berkeley and who had worked on the demand for lemons before moving on to do graduate work at Massachusetts Institute of Technology, and on many other occasions.

[5] See, for example, the literature on quality change and the measurement of prices summarized in Griliches [**27**].

[6] Clearly, there were other important contemporaneous lines of work, such as the work of Paul Douglas and his students, the National Bureau tradition, the work of the DAE at Cambridge, England, and many others.

For both traditions, simultaneity and identification were the major intellectual problems, not errors in variables. Both traditions dealt largely with *aggregated data*, either at the industrial (crop) or national level, and did not feel that *random* errors of measurement were a major problem, at least relative to the other specification problems faced by them.

Another good reason for ignoring errors in variables was the absence of any good cure for this disease. The statistical–theoretical discussion of these problems was obscure, largely published abroad and inaccessible, both physically and intellectually. The work of Frisch and Koopmans was known only secondhand, and taking their suggestions seriously would have required more information than was usually available and computations which were very complicated and laborious by the standards of those times. The solution in terms of bounds derived from "alternative" regressions was part of the econometric folklore and discussed explicitly in Henry Schultz's work, but was associated and often confused with the identification problem. In 1940 Tintner published his book on the variate-difference method [**70**], based on the earlier work of the Bulgarian statistician Oscar Anderson. The computations suggested in that volume were based on a set of assumptions that were not that easy to accept ("smoothness" of the underlying "true" variables), were also laborious, and required statistical justification and associated apparatus that were far from elegant. Since the Cowles Commission breakthrough, in the seminal work of Trygve Haavelmo, came only a few years later, it both overshadowed it and made it largely obsolete. By the time Tintner wrote his *Econometrics* textbook [**71**], published in 1952, he himself gave the variate-difference method little space and only perfunctory attention.

The central idea that cracked the simultaneity problem was contained in the recognition that by specifying *explicitly* the *rest* of the model generating the observed variables, one may have enough variables and restrictions to resolve the confluence problem posed in the estimation of a single equation. That is, the Cowles Commission literature brought in more equations to solve the more equations problem. Now, the developers of these new methods, Haavelmo, Koopmans, and T. W. Anderson, among others, were well aware of the errors-in-variables literature. They knew that bringing in additional equations and constraints could also solve the errors-in-variables problem, not just the simultaneity problem, but they did not pursue this very far. The topic had some coverage in Monograph 10 [**31**], but had largely disappeared from sight by the time the subject was codified in Monograph 14 [**45**]. Actually, a major set of papers on the interrelation of errors-in-variables and errors-in-equations models was written in 1946 by Hurwicz and Anderson [**38**] but never published. The clientele for the new estimation methods was not too interested in errors in variables, and the computational problems associated with the new methods were bad enough without additional complications. At the time there was much worry about lack of identification in models of this type in general, and the introduction of errors in variables would have only aggravated this problem. Moreover, it was easy to show that errors in the dependent variables could be absorbed into the shock terms for estimation purposes (but not for prediction or simulation), and that admitting errors in exogenous variables was tantamount to making them endogenous and reducing the problem to what looked like the previous case.

Be that as it may, by the late 1950's the simultaneous equations literature was developing rapidly without paying any more attention to the errors-in-variables problem. The latter was still discussed in a single-equation context but with very little constructive content. The situation was surveyed in two important papers written by statisticians rather than practicing econometricians: Durbin in 1954 [**6**] and Madansky in 1959 [**50**]. It is probably a fair comment to note that textbooks written at the end of the 1960's summarize this literature quite well but, except for the use of the permanent versus transitory income literature as an illustration, they contain little that advances the subject beyond the Durbin survey.

In 1950 Oscar Morgenstern published a horrifying little book [**54**], which was branded alarmist, and largely ignored. It seemed to be indiscriminate in its criticisms, and did not distinguish adequately between errors in the levels of the variables and errors in their movements. Also, it was largely destructive. It was not until there was a serious return to the analysis of micro data that errors in variables began to be taken seriously again. The desire to analyze data based on observations on individuals stripped away the "errors-cancel-out-in-aggregates" justification and forced us to face the issue squarely. Also, the relative success of applying rather primitive errors-in-variables methodology to the permanent-transitory income problem heightened the consciousness of the distinction between observed and true magnitudes and pointed towards possible ways of solving them. Thus, by the early 1970's grouping methods spread from the permanent-transitory income literature to production function estimation (Griliches and Ringstad [**29**]); replicated observations and a priori information on error variances (Bowles [**2**]) and instrumental variables (Griliches and Mason [**28**]) were used to estimate an error-ridden recursive income-generating model, and at the same time Zellner and Goldberger were resurrecting the theoretical study of this subject. Since this story is inseparable from that of "unobservable" variables, I shall turn to a brief discussion of the latter, before returning to illustrate the new developments in greater detail.

3. UNOBSERVABLES

> The "true" variables are variables such that, if their behavior should contradict a theory, the theory would be rejected as false; while "observational" variables, when contradicting the theory, leave the possibility that we might be trying out the theory on facts for which the theory was not meant to hold, the confusion being caused by the use of the same names for quantities that are actually different [**30**, p. 7].

There are at least three types of unobservables in econometric models: (i) Fixed parameters such as an elasticity of demand to be estimated from the data. (ii) Variables which affect the observable variables but which are not themselves directly observable either because the observed magnitudes are subject to measurement error *or* because these variables do not correspond directly to anything that is likely to be measured. Examples are concepts such as "permanent income," "expected price," "human capital," or "ability." Their existence is postulated, and they are defined implicitly by the specification of the model and the methods used to estimate it. (iii) Disturbances, either as errors of measurement or as errors in equations. There is no clear dividing line between these three categories.[7]

Disturbances are unobservables which we assume to be random and often independent from a lot of things, but otherwise we make few a priori statements about them. They are not part of the content of our theoretical presuppositions, but rather auxiliary magnitudes introduced to justify our estimation procedures and reconcile the observed data with our theoretical constructs. Substantive unobservable variables, the kind on which we shall focus, may or may not be random variables, but are variables about which we are willing to make many more a priori assumptions. They are the carriers of some of the content of our theories, and we are willing to specify both which other variables affect them and which are affected by them in turn.

Such variables made their first appearance in economics as the "trend," "cycle," "seasonal," and "irregular" components of time series analysis. They underlay much of the National Bureau's business cycle research and were one of the major topics of Frisch's book on confluence analysis [**10**]. That literature, however, had an air of "fishing-in-the-data" and was subject to the criticism of "measurement without theory." In his comment in Monograph 10 on Frisch's work, Haavelmo [**31**] objected to the use of error-in-variables models and the notion of "systematic" unobservables on the grounds that the relationships specified in such models should contain *also* shocks, not just errors, in which case the same identification problems would arise. In his view "unobservables" should be connected to the other variables by relations which fit the "true" variables exactly, with no need for a disturbance; these relations should be autonomous (i.e., structural); and there should be some hope of explicitly observing such variables, so that the whole system may have some use for prediction purposes. Since he had little hope that these requirements would be satisfied, he saw little use in formulating econometric models in such terms.

In retrospect, it seems to me that his requirements were too stringent. Adding disturbances to equations connecting or specifying unobservable variables raises no special new problems. It does affect identification and esimation but does not preclude them. Also, potential observability seems to me to be an unnecessarily severe requirement. Hypothetical constructs such as "permanent income" or "human capital" are useful even if one cannot think of any procedure which would measure them directly. In many cases such variables can be thought of as simplifications, specifying which other variables should be in what clusters and how their influence is transmitted from one cluster to another.

The rise of "theory" and the development of simultaneous equations estimation methods drove such "unobservable" variables underground, but not for long. They quickly reappeared, providing themselves, however, with a "theory" to justify their own independent existence. The great revival came with Friedman's

[7] Holte [**37**] suggests that these distinctions are often arbitrary and misleading.

[**9**] permanent income and Cagan's [**3**] expected price formulations.[8] The latter was influential in starting the late 1950's boom in distributed lag models. Most of that literature, and in particular Nerlove's work [**58** and **59**], was in the tradition of "unobservables" but did not use a stochastic specification for the equation defining the unobservables nor make an explicit connection to the errors-in-variables literature.[9] It was not until later, with the reinterpretation of distributed lags as a signal-extraction problem (see Muth [**57**], Nerlove [**60**], and Grether and Nerlove [**24**]) that the errors-in-variables formulation crept back into this literature.

The permanent-transitory income model of consumption has served as a paradigm for this literature, and we might as well look at it again here. The simplest version of this model can be written as

$$c = \beta y^* + u$$

and

$$y = y^* + e,$$

where c and y are the observed magnitudes of consumption and income, y^* is permanent income, and u and e are "transitory" components of consumption and income respectively; all variables are measured around their respective population means. Assuming that u, e, and y^* are independent of each other leaves the problem underidentified, there being only three observables $(\sigma_y^2, \sigma_c^2, \sigma_{yc})$ to estimate four unknown parameters $(\beta, \sigma_{y^*}^2, \sigma_u^2$, and $\sigma_e^2)$. Friedman suggested two solutions for this problem: grouping for cross-sectional data and an expectation-formation equation for time-series data. The grouping suggestion, using the means of these variables across cities or occupations, is equivalent to assuming that the distribution of the unobservable y^* has a particular city or occupation component structure and using city or occupation dummies as instrumental variables. The time-series formulation postulated y^* as a geometrically declining weighted average of past observed values of y. Later on, Watts [**73**] and others suggested a cross-sectional specification of y^* in terms of observables such as education, age, race, and other socio-demographic variables. In either case, it consisted of adding an equation defining y^* as

$$y^* = X\gamma$$

where X was a set of observable variables with the γ's possibly constrained to decline geometrically in the time-series interpretation. Note the lack of a disturbance in the original formulation of this equation.

That is roughly where the topic stood in the mid-sixties when I was doing my survey of distributed lag models (Griliches [**25**]). Since then, there have been two major related developments: First, in an important paper circulated in 1967, Zellner [**75**] noticed that in the above model β can be estimated more efficiently

[8] These works circulated in preliminary form in the early and mid-fifties, several years before their official publication date.

[9] Except for Klein's 1958 paper on distributed lag estimation.

than had been the case to date. The usual approach was to use the X's as instrumental variables for y^* in the first equation as part of a two-stage least squares procedure. No advantage was taken of the fact that the X's appeared also in the *other* equation, the one connecting the observed y to the unobservable y^*, thus imposing additional constraints on the model. In essence, Zellner reformulated the problem as

$$c = X\gamma\beta + u$$

and

$$y = X\gamma + e$$

and suggested a full rather than a *limited* information solution to it. Subsequently, Goldberger [**18** and **19**] developed maximum likelihood estimation methods for this model, including the case where the equation defining y^* also contains a disturbance of its own. The relationship between this type of model and canonical correlation analysis was shown by Hauser and Goldberger [**34**], canonical correlation being used as the method for imposing proportionality restrictions on coefficients of the X's across the two reduced form equations for c and y.

Goldberger, who had been working on parallels between psychometrics, sociometrics, and econometrics, noticed that this kind of model becomes much more interesting if the unobservable y^* enters into more than one substantive equation. The model is then subject to many more constraints which may allow the identification of parameters of interest even in the absence of any observable X's. Since this is discussed at some length in his Barcelona lecture, I will not elaborate on it further here, except to note that if the original model is extended to two different components of consumption with *independent* transitory components, the model is identified and estimable without further ado and without the necessity for bringing in additional X's.

There are several useful ideas here, as I see it. The first is that specifying a model for an unobservable may impose additional constraints across equations which may be used to generate more efficient parameter estimators than those used heretofore. The second is that these same constraints provide, on the other hand, additional tests on the validity of the original model. And the third is that thinking about unobservables in a multi-equation context may allow one to formulate and estimate models in which the unobservables are even "more" unobservable than has usually been the case. That is, one may formulate models in which there are no explicit X's determining or identifying y^* and no one specific direct even if error ridden measure of it, identification being the consequence of the imposed cross-equational constraints.

The constraints-across-equations-to-improve-efficiency idea should prove very useful in distributed lag estimation. This was noticed originally by Nerlove [**59**] but the limited computational resources of the time did not encourage pursuing such ideas further. For example, a demand for money model which depends on the expected rate of inflation and the assumption that expectations are formed in some sense optimally on the basis of past information suggests that one should estimate *jointly* both the optimal predictor of inflation and the demand for money equations. Similarly, an analysis of the term structure of interest rates could be improved by connecting the estimated lag structures of past inflation rates across the separate equations for different interest rates.

That more equations allow one to ask more questions about the same phenomenon is not surprising. It was also noticed by Nerlove [**58**] when he suggested that the permanent income hypothesis implied that the demand for meat and the demand for other food items both depend on the *same* unobservable permanent income and that this is a test for a particular formulation of it. What is slightly more novel is the idea of an "unobservable" without any bona-fide exogenous instruments to identify it, but that too has some precursors. Fairly early in the Cowles Commission days, Hurwicz [**38**] noted that in models with lags, a model with errors in variables *and* errors in equations may be identified by using lagged values of the dependent variables as instruments. Also what is suggested here is old hat to psychometricians, whose factor analytic models do just that, estimating underlying unobservable factors from sets of observable dependent variables without including any observable exogenous variables in their systems at all.

It is time, however, to stop reminiscing and to turn instead to the consideration of some actual examples of such models.

4. AN APPLICATION[10]

> What happens (when one adds one variable too many, Z.G.) is very much the same as what happened to the man who climbed up a ladder, and insisted on taking still another step after he had reached the top of the ladder [**10**, p. 80].

In this section we invade the borderland between economics and sociology and consider a model in which observable measures of an individual's success such as his income and occupational standing depend on a general set of exogenous variables X, schooling S, and an alleged unobserved ability variable A. Observations are available for individuals (j) belonging to the same family (i).

That is,

$$y_{kij} = X_{ij}\alpha_k + \beta_k S_{ij} + \gamma_k A_{ij} + u_{kij},$$

where k identifies the particular dependent variable and equation, i is the family, j is the individual, and u_{kij} is a disturbance assumed to be distributed independently from X_{ij}, S_{ij}, A_{ij}, and u_{hij}, for $h \neq k$, and for all i and j. In addition, we assume that the unobservable A and the observable exogenous variable X are independent of each other. This assumption may require a reinterpretation of the X's occasionally. Also, the X's need not be the same across equations. The simultaneity problem and associated least squares bias problem arises because we cannot assume that schooling and ability are independent. If schooling does, in fact, depend on A_{ij}, and on a subset of the same X's, then

$$S_{ij} = X_{ij}\alpha_s + \gamma_s A_{ij} + w_{ij}$$

[10] This section is based on Chamberlain and Griliches [**5**] where further details and data source descriptions can be found.

where w_{ij} is a disturbance assumed to be distributed independently of the other disturbances in the system and of A_{ij} and the X_{ij}'s.

Note that there is no observable variable in this model that just equals the unobservable plus measurement error. It differs thus from Griliches and Mason [**28**] where a set of test scores (T) were available and the model could be expanded by adding the equation $T = A + t$ where t is a random measurement error. Nor is it likely that we can solve the identification problem by exclusion restrictions. Exogenous variables such as race, age, or region are likely to affect both the levels of schooling and the subsequent success indicators such as occupation and income.

The unobservable part of each equation consists of the contribution of a common left-out "disturbance" (or factor) A_{ij} and an independent equation-specific disturbance (factor) u or w. This is a straight one-common-factor model but, because S is endogenous, we cannot observe these disturbances directly and have to move to reduced form equations instead. Solving S out from the y equations, specializing to the $K = 2$ case, i.e., two success indicator variables (income and occupation), relabeling S as y_3, and suppressing the ij subscripts for typographical ease, we get

$$y_1 = X(\alpha + \beta_1\alpha_3) + [(\gamma_1 + \beta_1\gamma_3)A + u_1 + \beta_1 w],$$

$$y_2 = X(\alpha_2 + \beta_2\alpha_3) + [(\gamma_2 + \beta_2\gamma_3)A + u_2 + \beta_2 w],$$

and

$$y_3 = X\alpha_3 + [\gamma_3 A + w].$$

In this form the model is more complicated, with the reduced form residuals connected by more than just one common factor. Without specifying any constraints on the α's and in the presence of only two success measures, the model is not identified.

Identification is achieved by postulating a family variance-components structure for A. That is, we assume that

$$A_{ij} = F_i + G_{ij}$$

where F_i is the family component of ability (and background) common to brothers, while G_{ij} is the level of their individual ability differences, independent of F_i by construction, with the indexes i and j going from 1 to q and from 1 to p respectively.

To simplify notation we order all our observations by families. Within a family the p individual observations on the first equation are followed by the p observations on the second equation, and so on. The whole system is then written as one stacked multi-variate regression

$$y = X\delta + \varepsilon$$

where y runs over families, family members, and all the dependent variables, X includes all the x's in all the equations, and δ is the vector of the reduced form coefficients of the observable x's. A particular δ, say δ_1, equals $\alpha_1 + \beta_1\alpha_3$, while a particular reduced form disturbance ε, equals, for example,

$$\varepsilon_{2ij} = (\gamma_2 + \beta_2\gamma_3)(F_i + G_{ij}) + \beta_2 w_{ij} + u_{2ij}.$$

Letting

$$\mathscr{E}(\varepsilon\varepsilon') = I_q \otimes \Omega$$

be the variance-covariance matrix of these disturbances, where I_q is a $q \times q$ identity matrix, and Ω is a $3p \times 3p$ matrix, it is clear that the peculiar structure of the ε's together with the assumption of no correlation among the F's, G's, u's, and w, impose a number of constraints on it. Since it turns out that δ is unconstrained in our model, we shall use the constraints on Ω, which is estimable from the variance-covariance matrix of the reduced form residuals, to identify the interesting parameters (β's and γ's) of our model.

For our model, Ω can be written as

$$\Omega = dd' \otimes l_p l_p' + \Sigma \otimes I_p$$

where l_p is the unit vector with p terms, while I_p is a $p \times p$ identity matrix. The unknown components of Ω are given by $dd' + \Sigma$ and can be estimated using the 3×3 variance-covariance matrix of reduced-form residuals R fitted across all individuals and families since

$$\mathscr{E}(R) = dd' + \Sigma$$

with d and Σ related to the original parameters of the model in the following manner:

$$d = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = \sigma_f \begin{pmatrix} \gamma_1 + \beta_1\gamma_3 \\ \gamma_2 + \beta_2\gamma_3 \\ \gamma_3 \end{pmatrix},$$

$$\Sigma = \tau dd' + V,$$

where

$$\tau = \sigma_g^2/\sigma_f^2$$

and

$$V = \begin{pmatrix} \beta_1^2\sigma_w^2 + \sigma_{u_1}^2 & \beta_1\beta_2\sigma_w^2 & \beta_1\sigma_w^2 \\ & \beta_2^2\sigma_w^2 + \sigma_{u_2}^2 & \beta_2\sigma_w^2 \\ & & \sigma_w^2 \end{pmatrix}.$$

Thus, dd' represents the contribution of the family component of the unobservable to Ω, $\tau dd'$ the contribution of the individual (within-family) component of the unobservable, while V encompasses the rest of the within-family individual effects, the u's and w. It is clear that if we could estimate V directly, we could easily identify the β's since, for example, $\beta_1 = v_{13}/v_{33}$. We cannot do that, but we can estimate dd', and we do have enough restrictions to estimate V in turn and proceed from there.

I will show that the model is identified as far as the β's are concerned (the γ's are identified only up to a scale factor since their scale cannot be separated from the somewhat arbitrary scale of the unobservable itself) by outlining an inefficient but straightforward estimation procedure based on the sufficient statistics in our hands, R and $\bar{R}$, where $\bar{R}$ is the variance-covariance matrix of *averaged* reduced form residuals, averaged separately for each variable over each family, summarizing the between-families relationships of these variables.[11] Given our model,

$$\mathscr{E}(\bar{R}) = dd' + \frac{1}{p}\Sigma$$

where p is the number of individuals per family. That is, averaging within families does not eliminate the variance components associated with the family effects (F), but does reduce the role of the individual effects (G, u, and w) on the order of $1/p$. Let us call $dd' = \theta$, remembering that rank $\theta = 1$. Then $\mathscr{E}(R) = \theta + \Sigma$ and $\Sigma \doteq \tau\theta + V$. Now, an estimate of θ is given by

$$\hat{\theta} = \frac{p}{p-1}\left(\bar{R} - \frac{1}{p}R\right),$$

and an associated estimate of Σ is

$$\hat{\Sigma} = \frac{p}{p-1}(R - \bar{R}) = \frac{p}{p-1}W,$$

where $W = R - \bar{R}$ is the "within" families variance-covariance matrix of the sample residuals. If we knew τ, we could get $\hat{V}$ directly from $\hat{\theta}$ and $\hat{\Sigma}$. Let us write down the estimate of $\hat{V}$ conditional on τ:

$$\hat{V} = \hat{\Sigma} - \tau\hat{\theta},$$

and hence,

$$\hat{\beta}_1 = (\hat{\sigma}_{13} - \tau\hat{\theta}_{13})/(\hat{\sigma}_{33} - \tau\hat{\theta}_{33})$$

and

$$\hat{\beta}_2 = (\hat{\sigma}_{23} - \tau\hat{\theta}_{23})/(\hat{\sigma}_{33} - \tau\hat{\theta}_{33}),$$

but also

$$\widehat{\beta_1\beta_2} = (\hat{\sigma}_{12} - \tau\hat{\theta}_{12})/(\hat{\sigma}_{33} - \tau\hat{\theta}_{33}).$$

Multiplying the first two expressions together and setting the result equal to the third gives us a quadratic expression in observables to determine $\hat{\tau}$ and, therefore, $\hat{\beta}_1$, $\hat{\beta}_2$ and the $\hat{\gamma}$'s (up to a scale and sign normalization).

[11] Arranging the estimated reduced form residuals $\hat{\varepsilon}_{kij}$ in a matrix $E = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \hat{\varepsilon}_3)$ and their averages over families ($\bar{\hat{\varepsilon}} = (1/p)\Sigma\hat{\varepsilon}_{kij}$) in a matrix $\bar{E}$, $R = E'E/pq$ and $\bar{R} = \bar{E}'\bar{E}/q$.

I have described the derivation of these estimates in some detail not because this is the best way of generating estimators, but because it makes it more transparent where and how the information on the β's emerges from the sample. The actual procedure outlined is inefficient for two reasons. First, in defining $\hat{\theta}$ as $[p/(p-1)][\bar{R} - R/p]$ we did not impose the restriction that the rank of $\theta = dd'$ equal unity. Thus there are more restrictions still to be imposed on the data. Second, R and $\bar{R}$ were computed from least squares residuals which did not use the information that Ω is not proportional to an identity matrix. Given estimates of dd' and Σ, we have an estimate of Ω, and therefore Ω^{-1}, and can compute GLS estimates of the reduced form coefficients (δ), and hence an improved set of residuals and associated R and $\bar{R}$ matrices and in turn better estimates of the structural coefficients (β's and γ's).

The maximum likelihood estimation procedure for this model will be sketched only briefly here.[12] It differs somewhat depending on whether the unobservable is viewed as a random or fixed variable and whether or not the X's are the same in all the equations. However, both of these variants affect only the starting values of the reduced form residuals (the ε's) and the exact procedure for deriving GLS estimators of δ given $\hat{\Omega}$. They do not affect the *procedure* for estimating $\hat{\Omega}$ or for deriving estimates of β and γ from it, though they do change the numerical conclusions reached.

The essence of the ML procedure can be grasped by noting that the family component of the unobservable can be represented by a set of family dummy variables subject to the constraint that they be proportional to each other across the various equations. Thus, the reduced form disturbance ε_k can be rewritten as

$$\varepsilon_k = Dfc_k + v_k$$

where ε_k is a $1 \times pq$ vector of reduced form disturbances, D is a $pq \times q$ matrix of q family dummy variables, f is a $1 \times q$ vector of unknown family coefficients, c_k is the factor of proportionality across equations, proportional to our original d coefficients, and v_k contains all of the other individual components of ε. Arranging the observable estimators of the disturbance vectors, the reduced form residuals, in a matrix

$$E = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \hat{\varepsilon}_3),$$

the estimation problem is equivalent to finding a 1×3 vector g and a $1 \times q$ vector f such that the correlation between Eg and Df is maximized. This is a straightforward canonical correlation problem whose solution will yield both a weighting scheme g to optimally aggregate the reduced form residuals across equations, and a set of f coefficients, multiplying each of the family dummy variables, to be interpreted as estimates of the unknown family "ability" levels. Noting that $D'D = pI_q$ and $D'E = p\bar{E}$, where $\bar{E}$ is the matrix of residuals averaged over families,

[12] The ML procedures for this special model were developed largely by Gary Chamberlain and are described in Chamberlain and Griliches [5] in some detail.

and that $E'E = pqR$ and $\bar{E}'\bar{E} = q\bar{R}$, the canonical correlation is just

$$\frac{g'\bar{R}g}{g'Rg},$$

and is maximized by letting g be the eigenvector of $\bar{R}$ in the metric R corresponding to the largest eigenvalue μ of $\bar{R}g = \mu Rg$, with μ being the square of the maximal canonical correlation coefficient between E and D. The reduced form ability coefficients d are then (after normalization) obtained from $\hat{d} = \bar{R}g$.

This gives an estimate of θ that satisfies the requirement $\hat{\theta} = \hat{d}d'$. Given $\hat{d}$ and $\hat{\Sigma}$ we can derive the desired β and γ coefficients as before.[13]

Having an estimate of Ω, we use an explicit expression for Ω^{-1} in terms of Σ^{-1} and gg' and get the GLS estimators of δ, the reduced-form coefficients, as a matrix weighted average of the within and between-families estimates of δ:

$$\hat{\delta} = (H_w + H_B)^{-1}(H_w\hat{\delta}_w + H_B\hat{\delta}_B).$$

where H_w and H_B are the respective precision matrices of these estimators:

$$H_w^{-1} = \Sigma \otimes W_{xx}^{-1}$$

and

$$H_B^{-1} = p\left(dd' + \frac{1}{p}\Sigma\right) \otimes B_{xx}^{-1},$$

where W_{xx} and B_{xx} are within and between variance-covariance matrices of the x's.[14] The joint ML estimates of δ, Σ, and d, based on the assumption that A, F, u, and w are all normal i.i.d. variables, are obtained by iterating these equations.[15] Starting with a consistent estimator of δ, say δ_T (the unconstrained total sample least squares estimator of the reduced form slope coefficients), one forms the residual matrix E, obtains d and Σ from the canonical correlation analysis outlined above, forms the H_w and H_B and obtains a new estimate of δ, computes a new E, and performs the canonical correlation analysis for a second time to obtain a new d and Σ. At this point the estimates are asymptotically (as $q \to \infty$) efficient. To get ML estimators these iterations are continued until satisfactory convergence has occurred.

We use the 1927 income (y_1), 1928 occupational standing (y_2), and the highest grade of schooling attained (y_3) for 156 pairs of brothers from Indiana (U.S.A.)

[13] Except that now we have a unique solution for τ and do not have to choose between the two roots of a quadratic equation.

[14] This parallels estimation in single-equation variance-component models. See Maddala [**51**] and Mazodier [**53**] for similar formulae.

[15] Note that we are here in the *random* unobservable effects context. A fixed unobservable family effects version of our model would assume that there are unknown family (dummy) level parameters to be estimated and would not constrain Σ otherwise. Its estimation procedure is not equivalent to just using the "within"-families information because the necessity of imposing the constraint that the estimated levels of these dummies be proportional to each other across equations. It turns out, surprisingly, that for a given δ the estimation procedure in this case is the same as for the random-effects estimator outlined above. The difference occurs in the formula for δ, with the fixed-effects estimate of δ being closer to the within-families estimator (δ_w).

to illustrate this approach.[16] The data are from Gorseline [**23**], and are described more fully in Chamberlain and Griliches [**5**]. The X set contains a constant, age, and age squared, with age squared appearing only in the income equation.[17]

Table I lists different estimates of the structural equations coefficients. The first two columns are ordinary least squares estimates ignoring the unobservable ability variable, based on the total and within-families samples respectively. The

TABLE I

PARAMETER ESTIMATES: INCOME–OCCUPATION–SCHOOLING MODEL. 156 PAIRS OF BROTHERS, 1928, INDIANA, U.S.A.[a]

Coefficients of the structural equations	Method: Biased least squares, Total sample		Within families		Maximum likelihood systems estimate
Age in the					
income equation	.088	(4.2)	.061	(2.0)	.080
occupation equation	.005	(1.6)	.011	(1.2)	.006
schooling equation	−.066	(3.6)	.029	(0.6)	−.067
Age squared in the					
income equation	−.001	(2.5)	−.001	(1.5)	−.001
Schooling in the					
income equation β_1	.082	(8.2)	.080	(7.3)	.088
occupation equation β_2	.104	(10.1)	.135	(9.0)	.107
"*Ability*" in the					
income equation γ_1					−.416
occupation equation γ_2					−.214
schooling equation γ_3					.092

[a] Original data from Gorseline [**23**]. The γ coefficients are scaled by assuming that $\sigma_f^2 = 1$ and $\gamma_3 > 0$. The numbers in parentheses are the computed t-coefficients.

last column lists our full-information maximum likelihood estimates of this system. Table II reports some of the intermediate calculations and the estimates of other parameters of this model.[18]

The model is computable (the iterations converged rather rapidly) and some of the estimated parameters make sense. For example, the estimated family component variance ratio in the unobservable variable, $\sigma_f^2/\sigma_A^2 = .66$, is quite reasonable. A pure genetic heredity model would predict a ratio of .5 to .6 (see Jinks and Fulker [**39**]). Adding common financial wealth to the interpretation of the unobservable makes the estimated value quite plausible. The fit is also quite good, the largest

[16] Data were also available on a number of 3, 4, and 5 brother families. Since the computations for the "unbalanced" (unequal p's) case are somewhat more complicated, they are not described here. See Chamberlain [**4**] for a solution. The empirical results are similar.

[17] The results are quite similar when experience (age minus age stopped school) is substituted for age.

[18] The numerical results differ from those given in an earlier draft of this lecture because I am limiting myself now to the consideration of the "balanced" equal number of brothers per family case and because of the correction of a punching error in one of the observations.

TABLE II

GORSELINE [23] BROTHERS: INTERMEDIATE DATA AND CALCULATIONS

$$R = \begin{bmatrix} .437 & .231 & .928 \\ & .488 & 1.168 \\ & & 11.193 \end{bmatrix}, \quad \bar{R} = \begin{bmatrix} .313 & .157 & .551 \\ & .243 & .532 \\ & & 6.512 \end{bmatrix}.$$

Sample size $N = p \cdot q = 312$, $q = 156$, $p = 2$.

Canonical weights $g' = (1.76 \quad .30 \quad -.19)$.
Squared canonical correlation coefficients:
$\mu_1 = .75$, $\mu_2 = .63$, $\mu_3 = .45$.

$$\hat{d} = \begin{bmatrix} .408 \\ .204 \\ -.092 \end{bmatrix}, \quad \hat{\Sigma} = \begin{bmatrix} .270 & .148 & .966 \\ & .446 & 1.187 \\ & & 11.184 \end{bmatrix}.$$

$$\hat{\sigma}^2_{u_1} = .098, \qquad \sigma^2_{u_2} = .297, \qquad \sigma^2_w = 11.18.$$

$$\sigma^2_f/\sigma^2_a = .66.$$

$$\operatorname{plim} \bar{R} = dd' + \frac{1}{p}\Sigma = \begin{pmatrix} .302 & .158 & .445 \\ & .265 & .575 \\ & & 5.600 \end{pmatrix}.$$

canonical correlation accounting for about 50 per cent of the unexplained variance ((.75–.5)/(1–.5)). (In our model, in the absence of any common unobservable variables, the expected canonical correlation squared is .5.) However, as illustrations of the "superiority" of the proposed methodology, these results are quite unsatisfactory. The estimated schooling coefficients (β's) are almost unaffected by the choice of estimation method, and the estimated negative coefficients for the unobservable in the income and occupation equations make its interpretation as an "ability" variable untenable.[19]

The source of our problem can be seen already in the total and within-families least squares estimates of the schooling coefficients. They change very little when we eliminate family effects from the data. Since the maximum likelihood procedure is essentially equivalent to extrapolating the changes that occur when we look at within-families relationships as compared to the total data set to their hopefully correct population values, it is not surprising that not much is happening, since there is very little to extrapolate. Moreover, since at least for the occupation equation the schooling coefficient becomes higher when family effects are eliminated, it is easy to see that the unobservable may have to have opposite

[19] Since the scale of A is arbitrary, we normalize it so that its coefficient in the schooling equation is positive ($\gamma_3 > 0$) and set the variance of its family component (F) equal to unity.

signs in the schooling and later success (occupation and income) equations to rationalize these results.[20]

At this point we can move in two directions. We can complicate our model or simplify it. We can complicate it by relaxing some of the restrictive assumptions of the model. We have investigated two of these: (i) the assumption that u_1 and u_2, the disturbances in the income and occupation equations, are uncorrelated, and (ii) the assumption that there is only *one* common unobservable variable connecting all the structural equations. It turns out that letting u_1 and u_2 be correlated does not change our results significantly, and I shall not, therefore, discuss this generalization further here.[21] The more-factors case is more complicated and merits some further attention.

Under the null hypothesis that there is only one common factor, the square of the second canonical correlation coefficient should be close to its expected value of .5; $\hat{\mu}_2 = .63$ is not very close. The likelihood ratio test statistic for the hypothesis that $\mu_2 = .5$ is 11, a not very likely value for a $\chi^2(2)$ variate.

There are also several substantive reasons for suspecting that more than one factor is at work here. We have lumped the family's socio-economic status and the children's native intelligence into one measure A. These two different kinds of "inheritance" may not have the same coefficients in the different equations. Moreover, there may be more than one type of "intelligence," including a kind (such as "test-wiseness") which may lead to scholastic but not necessarily to material success (in addition to its effect via schooling).

We can expand our model to the two-factor case, but the parameters are not identifiable without further restrictions on the two unobservables. If we interpret the first factor as genetic ability and the second factor as family wealth and status, then one would expect

$$\sigma_{f_1}^2/\sigma_{a_1}^2 \simeq .6,$$

while

$$\sigma_{f_2}^2/\sigma_{a_2}^2 \simeq 1.0,$$

and a positive correlation between the family components of those factors on the order of .5 or less. Such restrictions are enough to yield sensible ranges for the

[20] It is worth noting that such results may not be all that foolish. (The human facility for rationalization is boundless.) It is conceivable that above average ability at learning or above average family wealth leads to an over-investment in schooling, resulting, in a sense, in less marketable human capital than is implied by observed formal schooling levels. There may be, in effect, "errors" in our schooling measure (as a measure of human capital) which are positively correlated with ability.

[21] An equivalent generalization of the model is achieved by introducing y_2 into the y_1 equation,

$$y_1 = X\alpha_1 + \beta_1 s + \eta y_2 + \gamma_1 A + u_1,$$

and keeping the $Eu_1u_2 = 0$ assumption. Surprisingly, this leads to exactly the same reduced form equations as before and the same ML estimation procedure, except that now, given d and Σ, β and γ are unidentified. We can, however, estimate β, γ, and η conditional on the σ_f^2/σ_a^2 variance ratio for an a priori reasonable range for this ratio. In the reasonable range of $.5 \leqslant \sigma_f^2/\sigma_a^2 \leqslant 1.0$, the semi-reduced form total effect of schooling on income ($\beta_1 + \eta\beta_2$) (including its effect via occupation) does not change by more than 4 per cent, from .090 to .086. Whatever it is we do not like about our results cannot be eased by allowing u_1 and u_2 to be correlated.

β's and γ's.[22] The resulting β's are not all that different, but the estimated factor structure is still not easy to interpret. For reasonable values of the other parameters it is possible to find rotations such that the first factor, "ability" has positive coefficients in all the equations while the second factor, "family wealth," has positive coefficients in the schooling equation and negative coefficients in the income and occupation equations. However, the "ability" factor accounts for very little of the schooling variance (less than 4 per cent). The same is also true for the earlier one factor results listed in Tables I and II. The contribution of the unobservable to the schooling variance is nil, $d_3^2 = \gamma_3^2$ being on the order of .01. It is, therefore, more appropriate to simplify the model rather than complicate it.

The simplification that suggests itself is $d_3 = 0$, implying that little of the variance of schooling among men in Indiana at the beginning of this century is related to "ability."[23] This would, of course, greatly simplify our model, eliminating the problem of simultaneity and leaving only the second-order problem of block-heteroscedasticity in the disturbances of the income and occupation equations. The likelihood ratio test for the hypothesis $d_3 = 0$ is essentially equivalent to asking whether the variance-covariance matrix of structural disturbances is block diagonal, and whether the canonical correlation squared, μ_1^2, is increased by adding y_3 (schooling) to the set of indicators. The computed test statistic is .27, which is not surprising for a $\chi^2(1)$ variate if $d_3 = 0$. A similar test can be also computed in the two-factor case (conditional on our restriction $\sigma_{f_2}^2/\sigma_{a_2}^2 = 1$) with similar results. There is some evidence for separate family effects in the success (income and occupation) equations and in the schooling equation, but these effects are not strongly related, resulting in little bias in the least squares estimates of the schooling coefficients which ignore such connections.

Our first attempt to apply such a model to real data has produced more questions than it has answered, but that is the fate of much of substantive research and is part of the learning process. Too much is being asked of a rather limited set of data. We hope to have access, in the near future, to a more recent data set on brothers which will provide direct measures of the family's economic position and contain also some direct, though not error-free, measures of an individual's ability. This will allow us to refine our model and explore some of the same issues again.

Before we leave this model, however, I would like to draw your attention to the fact that it is not limited to a families–individuals interpretation. It can also be applied to repeated observations on the same individuals or enterprises with i being the individual index and j being the "replicated" time index. For example, it could be used to solve the Marschak and Andrews [**52**] and Mundlak and Hoch [**56**] problem of estimating production functions in the absence of observed variation in factor prices, provided that repeated observations over time are available on the same cross-section of enterprises. Assuming that the production function disturbance has a variance-components structure, $u_{it} = \mu_i + e_{it}$, and

[22] See Chamberlain and Griliches [**5**] for details.

[23] This is not inconsistent with some scattered data on this topic collected by Taubman and Wales [**68**], indicating a much lower correlation between schooling and test scores at the turn of the century.

that only the permanent or management disturbance μ_i is transmitted to the factor demand equations leads to a model that is similar to the one outlined above and is equivalent to Mundlak's [**55**] specification. Our estimation procedure is more efficient, however, since it utilizes the additional information contained in the between-enterprise variances and covariances and is not based solely on the within-enterprises variation.

5. OTHER EXAMPLES

In this section I will discuss briefly two additional applications of such models. The first, taken from Rosen [**66**], refers to growth in earnings as the result of on-the-job training. In terms of true variables, Rosen's equation can be written as

$$x_{it} = a_i + \gamma x_{it-1}$$

where x_{it} is the permanent change in the earnings of individual i during the time period t, a_i is an unobserved initial human capital (or ability) level, and γ is related to the discount rate used in the optimization process $[\gamma = (1 + r)^2]$. We can observe only actual earnings, not permanent ones, and they are affected by random and time independent transitory components v_t. Then, in terms of the *change* in observable earnings,

$$y_{it} = a_i + \gamma x_{it-1} + v_t - v_{t-1},$$

where the v's are transitory errors in earnings *levels*, implying that the errors in the earnings change equation are negatively correlated. The estimation problem here is two-fold: First, the a_i's are unobservable and, second, the y's are subject to correlated measurement error. If we had a long enough time series on individuals this would present little problem. We could difference out the a_i's (i.e., use within-individuals estimates) and use y_{t-3} as an instrumental variable, since the $v_t - v_{t-1}$ transitory errors are moving averages which are uncorrelated three terms apart. The problem, however, is that one does not usually have a long enough consistent time series on the same individuals. Moreover, it is probably unwise to assume that such a simple second-order difference equation with fixed coefficients is likely to hold for more than a few years at a time. That is, one would not use this type of model on a really long series anyway. In any case, at the moment the tapes we are interested in have only four years of data on individuals and, hence, only three consecutive income changes. For that data set, the reduced form equations can be written as

$$y_2 = \gamma^2 x_0 + (1 + \gamma)a + v_2 - v_1,$$

$$y_1 = \gamma x_0 + a + v_1 - v_0,$$

and

$$y_0 = x_0 + v_0 - v_{-1}.$$

The variance-covariance matrix of the observables is then

$$Eyy' = \operatorname{var}(y) = \gamma_1\gamma_1'\sigma_{x_0}^2 + \gamma_2\gamma_2'\sigma_a^2 + \begin{bmatrix} 2 & -1 & 0 \\ & 2 & -1 \\ & & 2 \end{bmatrix}\sigma_v^2$$

where $\gamma_1' = (\gamma^2, \gamma, 1), \gamma_2' = (1 + \gamma, 1, 0)$, and we have assumed that a and x_0 are independently distributed. Relaxing the last assumption would add a $(\gamma_1\gamma_2' + \gamma_1'\gamma_2)\sigma_{x0a}$ term. There are only four $(\gamma, \sigma_{x_0}^2, \sigma_a^2, \sigma_v^2)$ or five (adding $\sigma_{x_{0a}}$) unknown parameters to be estimated on the basis of six observed moments. The model is highly non-linear but should be susceptible to estimation.[24] If more time-series observations were available, the model would be much more over-identified and some of the restrictive assumptions could be relaxed and tested.

A second application is related to the human capital accumulation model outlined by Griliches and Mason [**28**]. It is similar to the model sketched out in Section 4, except that it does not use familial information and assumes the availability of additional data on test scores and other background variables for individuals, which were not available in the Section 4 context.

It can be outlined in its most stripped down version as follows:

$$A_0 = B + h,$$

$$S = \alpha A_0 + \beta B + u,$$

$$A_1 = \gamma S + A_0,$$

$$T = A_1 + t,$$

and

$$y = rA_1 + e,$$

where A stands for the unobservable ability, achievement, or human capital concept. The model assumes that this unobservable individual "capacity" measure is determined by initial conditions based on observed social background variables (B) and on unobserved heredity *and* environment differences (h); that both one's initial capacity and one's background affect the level of schooling attained (S); that schooling expands the individual's capacity above and beyond its initial levels (i.e., $\gamma \neq 0$); that the test scores available to us (T) are unbiased but error-ridden measures of this capacity at a particular point of time, and that observed earnings are proportional to unobserved capacity (human capital level A_1) at that time. The subscripts 0 and 1 identify initial and current levels respectively. This is a very simple model which could be extended in several additional directions. In particular, the availability of parallel data on other family members (fathers and brothers) would allow a better distinction to be made between social

[24] It is somewhat similar to the Holbrook and Stafford [**36**] model. Cf. also Goldberger [**17**]. Note that all of the variables are measured around their means and that no information is extracted from these means because of the possibility of exogenous common time effects.

background and heredity effects on ability. Also, the availability of data on changes in schooling and additional training and other experience variables would introduce additional equations between the S, T, and y equations, yielding a better resolution of the various schooling effects. Moreover, if data were available for several years on the same individuals, this would allow the estimation of "aging" effects, and a better treatment of transitory sources of income variation. However, the abbreviated model outlined above will suffice for our illustrative purposes.

The usual single equation approach to the estimation of the effect of schooling on income ($r\gamma$) either ignores the unobserved ability variable or substitutes test scores for it. The first is unsatisfactory because the omitted unobserved component of ability (h) may be correlated with the included schooling variable. Substituting observed test scores directly for A_1 is also unsatisfactory, since they are subject to large random measurement error. Substituting a predicted test score ($\hat{T}$) from a two-stage procedure based only on the background variables B is also unsatisfactory, since it still leaves the h component unaccounted for. We must, therefore, turn to a simultaneous equations systems approach. The reduced form equations for the three observed endogenous variables (S, T, and y) as functions of the observed pre-determined background variables (B) and the unobserved random components of the various equations (h, u, t, and e) can be written as

$$S = (\alpha + \beta)B + [\alpha h + u],$$

$$T = [1 + \gamma(\alpha + \beta)]B + [(1 + \gamma\alpha)h + \gamma u + t],$$

and

$$y = r[1 + \gamma(\alpha + \beta)]B + [r(1 + \gamma\alpha)h + r\gamma u + e],$$

or more compactly as

$$x_k = \pi_k B + w_k.$$

The π's can be estimated consistently by a direct application of least squares to this system. First round estimates of the interesting parameters γ and r can be had from

$$\hat{\gamma} = (\pi_2 - 1)/\pi_1$$

and

$$\hat{r} = \pi_3/\pi_2.$$

The variance-covariance matrix of the reduced form residuals, W, imposes additional restrictions on the data which can be used to retrieve several of the remaining interesting parameters of the following model:

$$W = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ & w_{22} & w_{23} \\ & & w_{33} \end{bmatrix} = dd'\sigma_h^2 + cc'\sigma_u^2 + n_2 n_2'\sigma_t^2 + n_3 n_3'\sigma_e^2,$$

where

$$d' = [\alpha, 1 + \gamma\alpha, r(1 + \gamma\alpha)], \qquad c' = (1, \gamma, r\gamma), \qquad n_2' = (0, 1, 0),$$

and

$$n_3' = (0, 0, 1).$$

Using the previously estimated $\hat{\gamma}$, we have

$$w_{12} - \hat{\gamma} w_{11} = \alpha\sigma_h^2$$

and

$$w_{12}\left(\frac{w_{i3}}{w_{i3}} - \hat{\gamma}\right) = (1 + \gamma\alpha)\sigma_h^2.$$

These can be manipulated further to get an estimate of $\alpha/(1 + \gamma\alpha)$ and therefore, given an estimate of γ, also estimates of α and σ_h^2. One can similarly derive estimates of σ_t^2, σ_e^2, and another estimate of $r(w_{13}/w_{12})$.

These estimates are, of course, not fully efficient, since γ and r were estimated in the first round without taking the restrictions connecting the components of W into account. Moreover, B will usually contain several variables and their coefficients should be constrained to be proportional to each other across equations. The estimation problem raised by the necessity of imposing nonlinear restrictions across both π (the slope coefficients of the reduced form equations) and W (the variance-covariance matrix of the reduced form residuals) has not been solved yet, but its solution seems to be within reach.

6. DISCUSSION

We have been discussing identification and estimation in a peculiar type of simultaneous equation model, with the structural disturbances having both a factor-analytic and a variance-components structure.[25] As one gets interested in this type of model and starts reading the literature, both old and new, one discovers many pre-cursors and much parallel work. The following comments are not intended to constitute a complete literature survey nor even an adequate account of all the parallel work that is going on, but I do want to highlight some of the major converging streams of thought on this problem.

We noted before that imposing some underlying structure on the unobservable as an aid in its identification was an idea that has been around for a long time. It is really a generalization of the idea of estimability of the errors-in-variables model in the context of replicated observations. Besides the earlier mentioned variate-differences method and grouping as a solution to the permanent-transitory income problem, it is interesting to recall Koopmans' [**45**] discussion in Monograph 14 of Ezekiel's attempt to identify a macro-model by sectoring the investment

[25] In factor analysis one usually thinks of repeated observations on different variables for the same individual. In variance-components, one thinks of a grouping of individual observations into regions or families. Formally the two approaches are equivalent.

equation and adding a separate trend and cycle structure to it. Similarly, in later years, a variety of dummy variables have been used as instruments in various simultaneous equation models, but usually with little discussion about the resulting structure of the model.

There has been little explicit discussion of identification in such models between the unpublished Hurwicz and Anderson papers [**38**] and the recent work of Geraci and Goldberger [**14**]. Besides items already mentioned, one may also recall Liviatan's [**48** and **49**] suggestions to use other error-ridden variables in the rest of the system to help in the identification of a particular equation and the Marschak and Andrews [**52**] and Mundlak and Hoch [**56**] discussion of identification and estimation in the production function–factor demand context. Such models are often partially under-identified, resulting in answers in terms of likely ranges on the coefficients and requiring additional prior information in terms of plausible ranges for some parameters for the identification of other parameters of particular interest. Results in terms of "bounds" were one of the major contributions of Frisch's work. This type of approach has lain largely unexploited except for an unpublished paper by Harberger [**33**], in the simultaneous equation context. It has been superseded, I believe, by the Bayesian approach to these problems (see particularly Zellner [**76**, Ch. 5] and Florens, Mouchart and Richard [**8**]), but an operational application of informative-priors in a combined errors-in-variables and simultaneous-equations model does not appear to be around the corner just yet.

One of the major differences between this type of model and the earlier "straight" regression and simultaneous equations literature is the necessity of imposing a set of constraints on the variance-covariance matrix of the disturbances of the system (in the structural or reduced form). The estimation problem that arises was already discussed by Frisch [**10**] under the label of "compatibility smoothing." In his case, having convinced himself that a particular model was in fact applicable, that the estimated bunch-map was "tight-enough", he faced the problem of producing a unique regression estimate from the various not-too-far-apart slope estimates. He wanted, effectively, to fit a matrix, whose rank was equal to unity, to the observed correlation matrix of the selected subset of variables. We faced the same problem in Section 4, trying to find that d vector which would make dd' come close to θ in some sense. Frisch developed an ad hoc procedure for doing this but did not employ a formal optimality criterion. Such smoothing techniques are inherently arbitrary unless an appropriate metric is specified. A solution in terms of minimizing the Euclidean distance between two matrices is given, among other places, in Rao [**62**, p. 52]. For our model, however, we need a solution which is equivalent to GLS rather than OLS and that is provided by Anderson [**1**], leading us to smooth θ towards dd' in the metric of R.

As far as estimation is concerned, there is quite a bit of relationship to the single equation variance-components literature discussed by Nerlove [**61**], Wallace and Hussein [**72**], Maddala [**51**], and Mazodier [**53**], among others, except that we have to generalize it to a multi-equation context.

There is also much in common between this type of model and the factor-analytic literature and the newly resurgent path-analysis literature in sociology. A fixed-factors model was discussed by Whittle already in [**74**]. In a series of important papers Jöreskog [**40**, **41**, and **43**] developed a maximum-likelihood procedure for fitting a covariance matrix-subject-to-restrictions model of which much of factor analysis and econometrics can be viewed as special cases. In fact, the model outlined in Section 4 can be thought of as an application of his "factoring the multitest–multioccasion correlation matrix" [**40**] procedure. The difference between us is in the algorithm used. His is a general non-linear Fletcher-Powell type solution to the fitting problem, while we take advantage of the specific structure of the likelihood function for our problem to get a simpler and more direct estimation procedure.

Applications of Jöreskog's scheme and path analysis models are included in the recent volume edited by Goldberger and Duncan [**22**], and in the Hauser and Goldberger [**34**] paper, which is very close to the models discussed here. Since one of the characteristic features of such models is the presence of cross-equation proportionality constraints, the algorithms often incorporate canonical correlation computations. Sargan's [**67**] and Hannan's [**32**] papers provide a connection between the simultaneous equations and canonical correlations literatures.

As we solve the estimation and computability problems for such models, we will be faced again with the problem of model choice, a problem to which much of Frisch's earlier work was devoted, and which still remains largely open. Testing in and discriminating between non-nested, non-linear models is far from a settled statistical subject. Some progress has been made in recent years, largely along Bayesian lines (see Leamer [**47**] and Zellner and Geisel [**77**]), but much remains to be done. It would also be helpful to have more relevant and less erroneous data. In any case:

> (No) technique ... will, like a stone of the wise, solve all the problems of testing "significance" with which the economic statistician is confronted .... The ultimate test of significance must consist in a network of conclusions and cross-checks where theoretical economic considerations, intimate and realistic knowledge of the data, and a refined statistical technique concur [**10**, p. 192].

Harvard University

REFERENCES

[**1**] ANDERSON, T. W.: "Statistical Inference for Covariance Matrices With Linear Structure," in *Multivariate Analysis II*, ed. P. R. Krishnaiah. New York: Academic Press, 1969.

[**2**] BOWLES, S.: "Schooling and Inequality from Generation to Generation," *Journal of Political Economy*, 80 (1972), 219–251.

[**3**] CAGAN, P.: "The Monetary Dynamics of Hyper-Inflation" in *Studies in the Quantity Theory of Money*, ed. M. Friedman. Chicago: University of Chicago Press, 1956.

[**4**] CHAMBERLAIN, G.: "Unobservables in Econometric Models," unpublished Ph.D. dissertation, Harvard University, 1973.

[5] CHAMBERLAIN, G., AND Z. GRILICHES: "Returns to Schooling of Brothers and Ability As An Unobservable Variance Component," Harvard Institute of Economic Research, Discussion Paper No. 340, Cambridge, 1974.
[6] DURBIN, J.: "Errors in Variables," *Review of the International Statistics Institute*, 1 (1954), 23–32.
[7] FELDSTEIN, M.: "Errors in Variables: a Consistent Estimator with Smaller MSE in Finite Samples," Harvard Institute of Economic Research, Discussion Paper No. 275, Cambridge, 1973.
[8] FLORENS, J. P., M. MOUCHART, AND J. F. RICHARD: "Bayesian Inference in Error-in-Variables Models," CORE Discussion Paper No. 7221, Heverlee, 1972.
[9] FRIEDMAN, M.: *A Theory of the Consumption Function*. New York: NBER Gen. Ser. 63, 1957.
[10] FRISCH, R.: *Statistical Confluence Analysis by Means of Complete Regression Systems*. Oslo: University Economics Institute, Publication No. 5, 1934.
[11] FULLER, W. A.: "Some Properties of Estimators for the Errors-in-Variables Model," mimeograph, Iowa State University, 1972.
[12] GEARY, R. C.: "Determination of Linear Relations Between Systematic Parts of Variables with Errors of Observations the Variances of Which are Unknown," *Econometrica*, 17 (1949), 30–58.
[13] GERACI, V.: "Simultaneous Equation Models with Measurement Error," unpublished Ph.D. Thesis, University of Wisconsin, Madison, 1973.
[14] GERACI, V. J., AND A. S. GOLDBERGER: "Simultaneity and Measurement Error," SSRI Workshop Series 7125, University of Wisconsin, Madison, 1971.
[15] GIRSHICK, M. A., AND T. HAAVELMO: "Statistical Analysis of the Demand for Food: Examples of Simultaneous Estimation of Structural Equations," *Econometrica*, 15 (1947), 79–110.
[16] GOLDBERGER, A. S.: "Econometrics and Psychometrics: A Survey of Communalities," *Psychometrika*, 36 (1971), 83–107.
[17] ———: "The Generalized Permanent Income Hypothesis: A Factor-Analytic Interpretation," University of Wisconsin, Madison, unpublished, ditto, 1971.
[18] ———: "Maximum-Likelihood Estimation of Regression Models Containing Unobservable Variables," *International Economic Review*, 13 (1972), 1–15.
[19] ———: "Structural Equation Methods in the Social Sciences," *Econometrica*, 40 (1972), 979–1002.
[20] ———: "Structural Equation Models, An Overview" in *Structural Equation Models in the Social Sciences*, ed. A. S. Goldberger and O. D. Duncan. New York: Seminar Press, 1973.
[21] ———: "Unobservable Variables in Econometrics," in *Frontiers of Econometrics*, ed. P. Zarembka. New York: Academic Press, 1974, pp. 193–213.
[22] GOLDBERGER, A. S., AND O. D. DUNCAN, EDS.: *Structural Equation Models in the Social Sciences*, New York: Seminar Press, 1973
[23] GORSELINE, D. E.: *The Effect of Schooling Upon Income*. Bloomington: Indiana University Press, 1932.
[24] GRETHER, D. M., AND M. NERLOVE: "Some Properties of 'Optimal' Seasonal Adjustment," *Econometrica*, 38 (1970), 682–703.
[25] GRILICHES, Z.: "Distributed Lags: A Survey," *Econometrica*, 35 (1967), 16–49.
[26] ———: "Notes on the Role of Education in Production Functions and Growth Accounting," in *Education, Income and Human Capital*, ed. W. L. Hansen. NBER *Studies in Income and Wealth*, Vol. 35, 1970, pp. 71–127.
[27] ———, ED.: *Price Indexes and Quality Change*. Cambridge, Mass.: Harvard University Press, 1971.
[28] GRILICHES, Z., AND W. M. MASON: "Education, Income, and Ability," *Journal of Political Economy*, 80 (1972), 74–103.
[29] GRILICHES, Z., AND V. RINGSTAD: *Economies of Scale and the Form of the Production Function*. Amsterdam: North-Holland Publishing Co., 1971.
[30] HAAVELMO, T.: "The Probability Approach in Econometrics," *Econometrica*, 12 (1944), Supplement.
[31] ———: "Remarks on Frisch's Confluence Analysis and its Uses in Econometrics," in *Statistical Inference in Dynamic Economic Models*, ed. T. C. Koopmans. Cowles Commission Monograph No. 10. New York: Wiley, 1950.
[32] HANNAN, E. J.: "Canonical Correlation and Multiple Equation Systems in Economics," *Econometrica*, 35 (1967), 123–138.
[33] HARBERGER, A. C.: "On the Estimation of Economic Parameters," unpublished Cowles Commission Discussion Paper No. 2088, Chicago, 1953.
[34] HAUSER, R. M., AND A. S. GOLDBERGER: "The treatment of Unobservable Variables in Path Analysis" in *Sociological Methodology 1971*, ed. H. L. Costner. San Francisco: Jassey-Bass, 1971.
[35] HILDRETH, C., AND J. P. HOUCK: "Some Estimators for a Linear Model with Random Coefficients," *Journal of the American Statistical Association*, 63 (1968), 584–595.

[36] HOLBROOK, R., AND F. STAFFORD: "The Propensity to Consume Separate Types of Income: A Generalized Permanent Income Hypothesis," *Econometrica*, 39 (1971), 1–22.
[37] HOLTE, F. C.: *Economic Shock-Models*. Oslo: Norwegian Universities Press, 1962.
[38] HURWICZ, L., AND T. W. ANDERSON: "Statistical Models with Disturbances in Equations and/or Disturbances in Variables," unpublished Cowles Commission memoranda in four parts: I. Introduction by L. Hurwicz, II. Contemporaneous Systems by T. W. Anderson, III. Lagged Systems by L. Hurwicz, and IV. Some Notes on Tintner's Statistical Methods by T. W. Anderson, Chicago, 1946.
[39] JINKS, J. L., AND D. W. FULKER: "Comparison of the Biometrical Genetical, MAVA, and Classical Approaches to the Analysis of Human Behavior," *Psychological Bulletin*, 73 (1970), 311–349.
[40] JÖRESKOG, K.: "Factoring the Multitest-Multioccasion Correlation Matrix," in *Current Problems and Techniques in Multivariate Psychology*, ed. C. E. Lunneborg. Seattle: University of Washington, 1970.
[41] ———: "A General Method for Analysis of Covariance Structures," *Biometrika*, 57 (1970), 239–251.
[42] ———: "Statistical Analysis of Sets of Congeneric Tests," *Psychometrika*, 36 (1971), 109–133.
[43] ———: "A General Method for Estimating a Linear Structural Equation System," in *Structural Equation Models in the Social Sciences*, ed. A. S. Goldberger and O. D. Duncan. New York: Seminar Press, 1973.
[44] JÖRESKOG, K., AND A. S. GOLDBERGER: "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable," SSRI Workshop Paper 7328, University of Wisconsin, Madison, 1973.
[45] KOOPMANS, T. C.: "Identification Problems in Economic Model Construction" in *Studies in Econometric Method*, ed. W. C. Hood and T. C. Koopmans, Cowles Commission Monograph 14. New York: Wiley, 1953.
[46] KUZNETS, S.: "Data for Quantitative Economic Analysis: Problems of Supply and Demand," lecture delivered at the Federation of Swedish Industries. Stockholm: Kungl Boktryckeriet P.A. Norsted and Soner, 1971.
[47] LEAMER, E. E.: "Bayesian Model Selection with Applications," Harvard Institute of Economic Research Discussion Paper No. 151. Cambridge, Massachusetts, 1970.
[48] LIVIATAN, N.: "Errors in Variables and Engel Curve Analysis," *Econometrica*, 29 (1961), 336–362.
[49] ———: "Tests of the Permanent-Income Hypothesis based on a Reinterview Savings Survey," in *Measurement in Economics*, Studies in Memory of Yehuda Grunfeld, ed. Carl F. Christ et al. Palo Alto, California: Stanford University Press, 1963.
[50] MADANSKY, A.: "The Fitting of Straight Lines When Both Variables are Subject to Error," *Journal of American Statistical Association*, 54 (1959), 173–206.
[51] MADDALA, G. S.: "The Use of Variance Component Models in Pooling Cross-Section and Time Series Data," *Econometrica* 39 (1971), 341–358.
[52] MARSCHAK, J., AND W. H. ANDREWS: "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 12 (1944), 143–206.
[53] MAZODIER, P.: "L'Estimation des Modéles à Erreurs Composées," *Annales de l'I.N.S.E.E.*, 7 (1971), 43–72, and "The Econometrics of Error Components Models," unpublished Ph.D. dissertation, Harvard University, 1971.
[54] MORGENSTERN, O.: *On the Accuracy of Economic Observations*. Princeton: Princeton University Press, 1950.
[55] MUNDLAK, Y.: "Estimation of Production and Behavioral Functions from a Combination of Cross-Section and Time-Series Data" in *Measurement in Economics*, Studies in Memory of Yehuda Grunfeld, ed. Carl F. Christ et al. Palo Alto, California: Stanford University Press, 1963.
[56] MUNDLAK, Y., AND I. HOCH: "Consequences of Alternative Specifications in Estimation of Cobb-Douglas Production Functions," *Econometrica*, 33 (1965), 814–828.
[57] MUTH, J.: "Optimal Properties of Exponentially Weighted Forecasts with Permanent and Transitory Components," *Journal of the American Statistical Association*, 55 (1960), 299–306.
[58] NERLOVE, M.: "The Implications of Friedman's Permanent Income Hypothesis for Demand Analysis," *Agricultural Economics Research*, 10 (1958), 1–14.
[59] ———: *Distributed Lags and Demand Analysis*, USDA Agriculture Handbook No. 14. Washington: Government Printing Office, 1958.
[60] ———: "Distributed Lags and Unobserved Components in Economic Time Series," in *Ten Economic Studies in the Tradition of Irving Fisher*, ed. W. Fellner et al. New York: Wiley, 1967.

[61] ———: "A Note on Error Component Models," *Econometrica*, 39 (1971), 383–396, and "Errata." *Econometrica*, 40 (1972), 218.
[62] RAO, C. R.: *Linear Statistical Inference and its Applications*. New York: Wiley, 1965.
[63] REIERSOL, O.: "Confluence Analysis by Means of Lag Moments and Other Methods of Confluence Analysis," *Econometrica*, 9 (1941), 1–24.
[64] ———: "Identifiability of a Linear Relation Between Variables Which are Subject to Error," *Econometrica*, 18 (1950), 375–389.
[65] ———: "A Reconsideration of Confluence Analysis," unpublished Memorandum of the Universitetets Sociola Konomiske Institute, Oslo, 1955.
[66] ROSEN, S.: "Income Generating Functions and Capital Accumulation," Harvard Institute of Economic Research Discussion Paper No. 306, Cambridge, Massachusetts, 1973.
[67] SARGAN, J. D.: "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26 (1958), 393–415.
[68] TAUBMAN, P., AND T. WALES: "Mental Ability and Higher Educational Attainment in the 20th Century," NBER Occasional Paper 118, 1972.
[69] THEIL, H.: *Principles of Econometrics*. New York: Wiley, 1971.
[70] TINTNER, G.: *The Variate Difference Method*. Cowles Commission Monograph No. 5. Bloomington, Ind.: Principia Press, 1940.
[71] ———: *Econometrics*. New York: Wiley, 1952.
[72] WALLACE, T. D., AND A. HUSSAIN: "The Use of Error Components Models in Combining Cross-Section With Time Series Data," *Econometrica*, 37 (1969), 55–72.
[73] WATTS, H. W.: "An Objective Permanent Income Concept for the Household," unpublished Cowles Foundation Discussion Paper 99, Yale University, New Haven, 1960.
[74] WHITTLE, P.: "A Principal Components and Least Squares Method of Factor Analysis," *Skandinavisk Aktuarietidskrift*, 35 (1953), 223–239.
[75] ZELLNER, A.: "Estimation of Regression Relationships Containing Unobservable Independent Variables," *International Economic Review*, 11 (1970), 441–454.
[76] ———: *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley, 1971.
[77] ZELLNER, A., AND M. GEISEL: "Analysis of Distributed Lag Models with Application to Consumption Function Estimation," *Econometrica*, 38 (1970), 865–888.