K. Kazanoglu

NBER WORKING PAPER SERIES

THE COVARIANCE STRUCTURE OF EARNINGS AND THE ON THE JOB TRAINING HYPOTHESIS

John C. Hause

Working Paper No. 25

CENTER FOR ECONOMIC ANALYSIS OF HUMAN BEHAVIOR AND SOCIAL INSTITUTIONS
National Bureau of Economic Research, Inc.
261 Madison Avenue, New York, N.Y. 10016

December 1973

The Covariance Structure of Earnings and The On the Job Training Hypothesis

John C. Hause*
University of Minnesota
National Bureau of Economic Research

Note to the reader:

There are two modest modifications or additions that will (hopefully) be made to this study. The first is an attempt to understand the somewhat anomalous result mentioned on page 27, where the GLS estimates of $\hat{\sigma}_w$ exceed the corresponding OLS estimates. I hope to have a few additional calculations made in Uppsala to clarify this puzzle.

I also plan to include in an appendix a few more results from the computations, including the individual OLS and GLS estimates.

*This research was partially funded by a grant from The National Science Foundation to NBER for research on the determinants of the distribution of income and earnings, GS-31334. The opinions expressed herein are those of the author and do not necessarily reflect the views of The National Science Foundation. Additional acknowledgments are found with the footnotes.

1. INTRODUCTION

The determination and explanation of statistical properties of the earnings profile is becoming a focal point of research by economists interested in the distribution of earnings and in the life cycle of earnings. The first order statistics (profiles of mean earnings as a function of labor force experience or age) have been studied extensively, both because of their central role in human capital models and because cross sectional and mean time series have been relatively easy to obtain for carrying out such studies. Profiles of earnings for groups with different levels of schooling attainment, corrected for differences in personal characteristics, have been the primary empirical basis for estimating rates of return from schooling. Economists have generally interpreted the inverted U-shape of the profiles as reflecting the growth of initial productivity with experience (and investment) and the ultimate deceleration or decline in earnings with depreciation and obsolescence, and continue to study them in order to understand these processes more clearly. Work by Mincer [9] has given strong emphasis to the potential significance of post-school investments by persons to increase earnings capacity, with primary focus on the implications of on the job training.

Data on the mean earnings profile of a cohort, combined with knowledge about the variance (as a function of experience or age) are adequate for raising and resolving some significant questions. However, this information leaves unanswered how individual profiles are distributed about the group mean. This issue is important for several reasons. 1) Investments in human capital usually modify the entire subsequent earnings profile. The dispersion of earnings for a single year is due to a combination of systematic, transitory, and compensatory components that are difficult to disentangle and to relate to investment decisions by the individual. The risk due to different payoffs from such investments is much more adequately reflected by the dispersion of lifetime earnings (e.g., as measured by the variance of the logarithm of the discounted earnings stream) than by the dispersion for one year.[1] In principle, one can imagine partitioning the variance of lifetime earnings into systematic elements perceived by a person prior to an investment in human capital, systematic elements not anticipated, and "random" factors. These questions cannot be attacked directly without information on the autocovariance structure of earnings. An upper bound on the variability of lifetime earnings can be obtained consistent with knowledge of the standard deviation of earnings as a function of age, since the supremum of the standard deviation of (discounted) lifetime earnings is the discounted value of the standard deviation of earnings as a function of age. Unfortunately, the lower bound of discounted earnings, based on this information is zero, and these bounds are far too wide to be of much use.[2] Fragmentary data reveal that earnings over time have substantial positive correlation, but the correlation is considerably less than unity, so neither bound is a very satisfactory estimate of the standard deviation of (discounted) lifetime earnings.

2) Information on individual earnings profiles can also be useful for attempting to determine what specific human capital investments do that increase productivity and earnings. It is difficult to identify transitory fluctuations in earnings and employment from cross sectional data, yet the distribution of these elements may be significantly determined by specific investments, such as schooling.

3) Data on individual earnings profiles can help unravel the effect of human capital investments on the life cycle of earnings. At present, "optimal programs of human capital investment" are not well understood, partly because of very limited information of the substitutability and complementarity of different human capital investments. Studies of American data have established that the mean profile of earnings of those with more schooling rises more rapidly and for more years, than the profile of earnings for those with less schooling. The mean profiles do not reveal the extent to which schooling and on the job training may be partial substitutes. This information is necessary in order to determine the differences in human capital investments by individuals due to differences in the investment opportunities they face.

A direct empirical attack on these topics is possible with longitudinal data from which autocovariances can be computed. Although scattered pieces of evidence on autocorrelations of earnings and of income have been reported by others, e.g., Friedman and Kuznets [3], Hanna [4], Mendershausen [7], Thatcher [12].

the populations on which they are based are usually heterogeneous in age and education. Hence, these correlations are weighted averages of different segments of earnings (or income) profiles from cohorts with different schooling attainment and are difficult to interpret. Finally, there seems to be little literature that systematically exploits the covariance structure to estimate or explain lifetime earnings, aside from some work on the random walk hypothesis, see Fase [2].

This study examines the on the job training hypothesis and takes full advantage of the covariance structure. This hypothesis is developed in a human capital framework and asserts that systematic differences in on the job training lead to systematic differences in earnings profiles. The hypothesis has been discussed extensively by Mincer [9], whose statistical work has been based exclusively on first order statistics. The hypothesis is important in its own right, and is a good example of a topic in which the covariance structure of earnings can play a powerful role in developing and testing statistical hypotheses and for further theoretical analysis of human capital investments and their returns. The substantive conclusion from a study of one sample of Swedish income statistics yields upper bound estimates of OJT effects for low levels of schooling attainment that are empirically significant. One standard deviation variation of earnings profile slopes taken over a five year period yields a differential change in income which exceeds 20% of mean income at age 26.

The next section discusses the on the job training hypothesis and a statistical specification for studying it. The third section develops this model to provide possible tests for the existence of on the job training effects and estimates of earning profile differences that may be due to on the job training. The final section applies the parametric model to time series income data from a Swedish cohort of males with several levels of schooling attainment. Much future work remains to be carried out in this framework.

2. THE "ON THE JOB TRAINING HYPOTHESIS"

Mincer [9] has suggested that a significant source of differential human capital investment arises from differences in on the job training (denoted OJT in the following discussion), which in turn lead to differences in earnings profiles. Some jobs require relatively long times for new workers to acquire normal levels of job skill. Workers entering such jobs may initially receive low earnings, corresponding to their low net productivity. As they acquire more experience and skill, earnings rise. Since capital market exchange opportunities make current dollars more valuable than future dollars, a simple model suggests that earnings at later points in time must be high enough to offset the low initial earnings from these jobs (in the sense of equalizing present values). Apprenticeships for certain crafts or the establishment of a professional reputation in medicine or law are examples of jobs requiring significant post-school investment.

Consider a model in which there is a perfect capital market, perfect foresight, no nonpecuniary occupational preferences, and equal earnings potential for a cohort of entrants to the labor force, whose members have the same schooling attainment. Figure 1 illustrates the earnings profiles for three jobs, a, b, and c, requiring increasing levels of post-school OJT investments. Assume that there are no additional out-of-pocket investments required for any of these jobs. Under these assumptions, the earnings profiles will not be in equilibrium unless they have the same present value; i.e., unless $\int_{\tau_o}^{\tau_f} e^{-\rho\tau} y_i(\tau) d\tau$ are equal for i=a,b,c; where $y_i(\tau)$ is the earnings profile for job i, ρ is the market interest rate, and τ_o and τ_f are the date of entry to and exit from the labor force, respectively. There is little theory to guide us in selecting a particular function, $y_i(\tau)$, for the earnings profile, and there is no necessity for the family of profiles to be linear in the rising portion. Like cost functions, the earnings profiles presumably depend on technology and prices. Mincer's own work with simple, analytically convenient earnings functions often leads to a "crossover" point τ_c (or to a relatively narrow crossover interval) for given schooling attainment, and we retain this assumption in the following development. The linear profile assumption is a convenient fiction which is probably not unduly restrictive. Since we will be concerned primarily with the extent to which differential on the job training is important, for most of the statistical work, we can regard the linear profile segments as deviations from the mean earnings profile for the cohort (which has the usual skewed, inverted U-shape with the ascent in the early years considerably steeper than the descent in the later years). Furthermore, the time series data analyzed in this paper span much shorter time intervals than the entire earnings profile.

The empirical relevance of this model clearly depends on the ex-

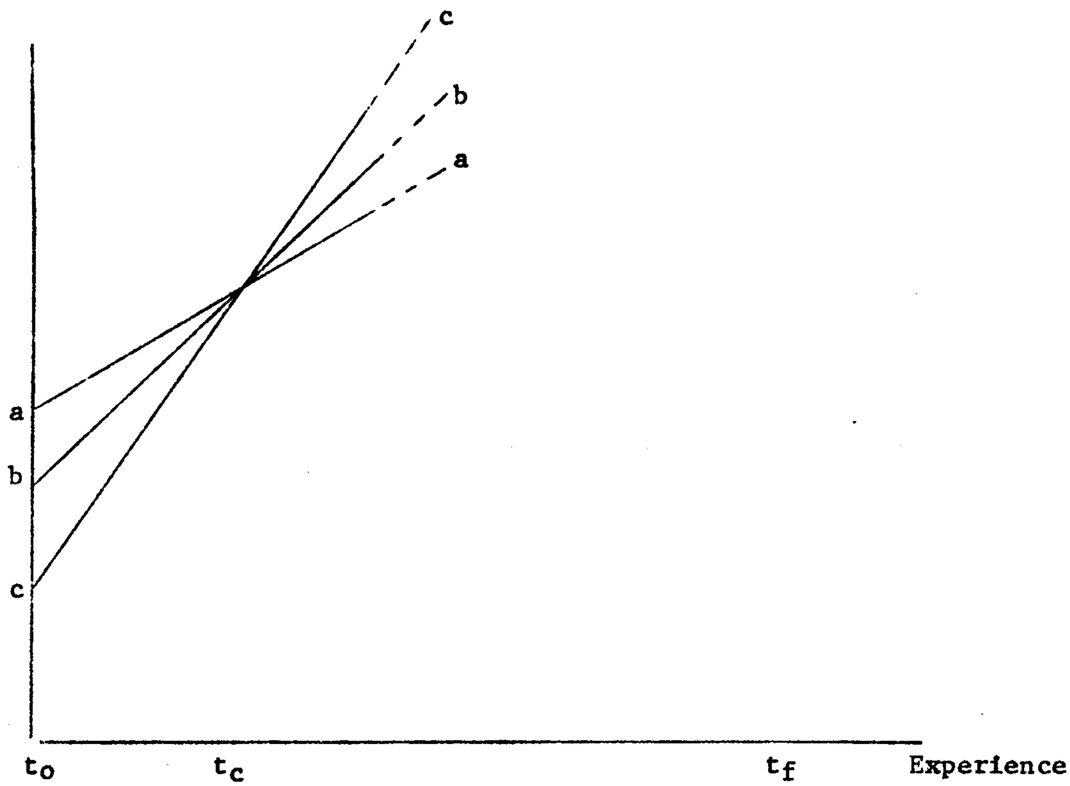

Fig. 1 - Linearized Rising Sector of Earnings Profile of Equal Ability Individuals

tent to which there are systematic differences in the earnings profiles due to economic choices made by workers. If the profiles merely reflect inevitable increases in productivity that accompany job experience and physical and psychological maturation and aging, there is little theoretical gain from describing the profile as if people are making investment decisions in on the job training once they enter the labor force. The scenario that accompanies Mincer's formal on the job training model [9] might be interpreted as if this form of human capital investment requires a worker to make optimal period by period investment decisions, and if the worker decides to hold his stock of human capital constant, observed earnings will equal potential earnings and the earnings profile will be horizontal. It is not possible to infer the range of on the job investment choices available to or actually chosen by workers from the mean earnings profile of a cohort with given schooling attainment. (However, if mean profiles of earnings for different occupations, given schooling, start out at significantly different levels, and cross each other after some years, with the initially lower profile remaining higher after the crossover, this fact would suggest that the choice of a job carries with it an implied choice of earnings profile, and competition for jobs should tend to equalize returns at the margin so that there is no net advantage of one profile over the other for workers with similar tastes and economic ability. Even this information would not indicate the extent to which workers have substantial latitude for differential on the job investment once a job is chosen.

Most of the remainder of this paper is devoted to developing and testing a model that can indicate the maximum economic significance that on the job training could have by determining the extent to which there are differences in the growth rate of earnings with experience, and to developing an indirect test of the possible net effect on on the job training based on partial correlations of earnings. There may, of course, be other reasons for differences in the slope of earnings profiles other than OJT. For example, people with higher ability may simply increase their productivity from job experience more rapidly than others as a joint product of the job experience. A more completely specified model that includes personal characteristics (that may themselves be determined in part by previous investments in human capital) in addition to the time series data on earnings or income could probably be used to disentangle other distinct factors from OJT effects related to differences in the slopes of earnings profiles. In the absence of this additional study, the dispersion in slopes we find in this paper provide an upper bound on OJT effects for the sample cohorts studied.

The simple model of earnings profiles illustrated in Figure 1 has several implications for correlations of earnings from different points along the profile. In the following discussion we assume that i and j both lie in the time interval where the earnings profiles are rising and that $i < j$. If i and j are both less than t_c, or if they are both greater than t_c, $r_{ij} > 0$. If either i or j is equal to t_c, $r_{ij} = 0$ (in a degenerate sense, since the variance of $y(t_c) = 0$ in this model). And if $i < t_c < j$, $r_{ij} < 0$.

The model in this form is clearly inadequate for empirical work, since there is no t_c such that var $y(t_c) = 0$ and steadily increases in going in either direction in time from t_c. Furthermore, fragmentary evidence suggests that cohorts of individuals who have the same schooling attainment and who entered the labor force simultaneously have positive earnings correlations for all relevant i and j. These objections can be evaded by modifying the model to allow for a distribution of earnings potential rather than assuming that everyone has the same initial economic ability. This feature can be incorporated into the model by assuming a distribution of earnings at the crossover time t_c, and that those with the highest earnings at that time have the highest (discounted) lifetime earnings. If this distribution of economic ability has a large enough variance, this factor could mask the simple patterns of earnings correlations implied by the initial OJT model. Finally, there is presumably substantial residual variability in earnings that must be taken into account in the statistical study of time series data on earnings.

These considerations lead us to consider the following statistical specification of the (linearized) earnings profiles:

(1) $$x_t = m + tw + u_t \ .$$

In this equation, x_t represents earnings in period t, and time has been translated so that the crossover period, $t_c = 0$. The random variables m, w, and u_t are assumed to have the following properties. m and w are the systematic characteristics that distinguish individuals in the model, while u_t, the earnings residual in period t, is assumed to be uncorrelated with m and w, for all t. The variable m is assumed to yield the distribution of earnings at the crossover period t_c (=0) in the complete absence of u_t, and represents the distribution of earnings differentials due to differences in economic ability (or earnings potential). The variable w determines the slope of the rising portion of the linearized earnings profile, and is the key element in this model linking it with the on the job training hypothesis. If there are large, systematic differences in the OJT obtained by different workers in a sample, these differences should be reflected by a correspondingly large dispersion of the distribution of w for the sample members. The absence of a time subscript, t, on m and w reflects the assumption that m and w are specific to the individual, and do not vary along the profile. The variable m is presumably a predetermined characteristic of each individual at the time he enters the labor force, and is supposed to capture genetic and environmental differences in background as well as earlier investments in human capital. The variable w is determined by the individual by the type of occupation he enters as well as by the extent to which he decides to acquire OJT (to the extent that the OJT is a choice variable, given the occupation he enters).

There is little theory to guide us in specifying the statistical properties of the earnings residual, u_t, and in the following, we will assume that $u_t = y_t + z_t$. y_t is a random walk component $y_t = \sum_{i=t_0}^{t} \varepsilon_i$, with ε_i the independent (nontransitory) random shock, with $\text{cov}(\varepsilon_i,\varepsilon_j) = 0$ if $i \neq j$; $= \sigma_{\varepsilon_i}^2$ for $j=i$. z_t represents a transitory earnings residual in period t, and we assume $\text{cov}(z_i,z_j) = 0$ if $i \neq j$; $= \sigma_{z_i}^2$ if $j=i$ and $\text{cov}(\varepsilon_i,z_j) = 0$; $\text{cov}(\varepsilon_j,z_i) = 0$ for all i and j. Some students of earnings profiles have assumed that changes in earnings along the profile can be represented by a random walk, which corresponds to random, but nontransitory factors that permanently change the expected level of earnings (e.g., see Fase [2] for extensive work with this model). The element y_t captures this aspect. The concept of "transitory" variations in earnings (and income) has long been used in the theoretical and econometric specification of earnings profiles, and z_t represents this component. It incorporates minor accidents, incidental unemployment, and the like, which are assumed to exert an effect on earnings for only a single period.

We have not yet made any assumptions about cov (m,w), the relation between the distribution of earnings at the crossover time, t_c (in the absence of residual earnings variations from u_t) and the distribution of slopes of the earnings profiles. It seems plausible that high ability people will have higher earnings at t_c and may also have a greater slope to their earnings profile, due to a capacity to acquire certain job-related skills more rapidly than those with low ability. Previous work by Hause [5] provides some empirical evidence that a direct measure of ability (from test scores) is positively correlated with the slope of the earnings profile. At the very least, it seems reasonable to assume $\text{cov}(m,w) \geq 0$. If we have precise knowledge of the crossover time, it would be possible to estimate this covariance directly from longitudinal data. But an identification problem arises in the specification (1), which makes it impossible to identify cov (m,w) and t_c simultaneously. The consequences of misspecifying t_c can be determined from the structure of the covariance matrix implied by (1). Suppose we have external information about the true t_c, and T is the column time vector with origin at $t_c = 0$. Then:

$$(2)\ [\sigma_{xx}] = [1]\sigma_{mm} + TT'\sigma_{ww} + [1T' + T1']\sigma_{mw} + [\sigma_{uu}]$$

In this matrix equation, $[1]$ is a matrix of 1's, and 1 is a column vector of 1's. If we now shift the time vector T by adding the sector τ to each component, we obtain:

$$(2')\ [\sigma_{xx}] = [1]\sigma^*_{mm} + (T + \tau 1)(T + \tau 1)'\sigma^*_{ww} + [1(T+\tau 1)' + (T+\tau 1)1']\sigma^*_{mw} + [\sigma^*_{uu}]$$

By equating the components of (2) and (2') we obtain:

$$(3a)\ [\sigma^*_{uu}] = [\sigma_{uu}]$$

$$(3b)\ \sigma^*_{mm} = \sigma_{mm} + \tau^2\sigma_{ww}$$

$$(3c)\ \sigma^*_{ww} = \sigma_{ww}$$

$$(3d)\ \sigma^*_{mw} = \sigma_{mw} - \tau\sigma_{ww}$$

If the crossover is misspecified by assuming that it takes place at too young of age, τ is positive. Equations 3 then imply that the covariance matrix of residuals is unchanged, as is the scalar σ^*_{ww}. However, the apparent crossover variance σ^*_{mm} is too large (from 3b) and the apparent covariance of m and w, σ^*_{mw} is too small (from 3d). Indeed, a sufficiently large underestimate of the crossover age could make the apparent σ^*_{mw} negative, even if the true $\sigma_{mw} \geq 0$, as we assumed earlier. On the contrary, if the crossover age is overestimated (τ negative), the apparent σ^*_{mw} is too large.

There is no direct way to observe t_c. However, Mincer has shown with one specification of the on the job investment function that $1/r$ is an upper bound for t_c with that specification. In this bound, r is the rate of return for investment in on the job training.

If one has a sample of individuals, with a direct measure of ability as well as time series data on earnings, correlations of estimated m's with measured ability should be interpreted with caution, since an error in specifying t_c will lead to an error in estimating the "true" crossover m's. However, an error in specifying t_c will have no effect on the estimated slope parameter of the individual earnings profiles, w. Therefore, correlations between directly measured ability and estimated earnings profile slopes w are not affected by this specification.

3. STATISTICAL TESTING OF THE ON THE JOB TRAINING HYPOTHESIS

This section considers two procedures for determining the possible existence and empirical significance of investment in on the job training. The first procedure is an extension of the test of looking at the pattern of simple correlations of earnings that would be appropriate in the absence of differential ability and residual variations in earnings. The second is an attempt to estimate the dispersion of the w's which themselves are estimated from individual time series of earnings.

A. Testing OJT Effects by Partial Correlations of Earnings

In the second section, we considered the distribution of m as reflecting differences in economic ability which would be observed at t_c in the absence of residual earnings variation, u_t. This feature of the model suggests looking at the partial correlation of earnings $r_{ik \cdot t_c}$, where $i < t_c < k$. If we think of observed earnings at t_c as measuring (with error) economic ability, then we would expect this partial correlation to be negative, in analogy with the simple correlation of earnings, $r_{ik} < 0$ for a cohort of equal ability people, if $i < t_c < k$. Since we have no direct information on the precise years of experience at which t_c occurs, this leads us to consider the sign of the partial correlation of earnings $r_{ik \cdot j}$, where $i < j < k$. This problem is equivalent to determining the sign of the regression coefficient of x_i in a regression of x_k on x_j and x_i, which in turn depends on the sign of the determinant:

$$D = \begin{vmatrix} cov(x_i, x_k) & cov(x_i, x_j) \\ cov(x_j, x_k) & var(x_j) \end{vmatrix} .$$

From equation (1), one calculates:

$$(4a)\ cov(x_q, x_r) = M + \tau_q \tau_r W + C(\tau_q + \tau_r) + U_{qr}$$

$$(4b)\ var(x_j) = M + \tau_j^2 W + C(2\tau_j) + U_{jj}$$

In these equations, M, W, and U_{jj} are the variances of the random variables m, w, and u_j, respectively; C and U_{qr} are the covariances cov(m,w) and $cov(u_q, u_r)$; and the subscripted τ's are the periods corresponding to the time subscript, where the time origin $\tau_c = 0$. We obtain:

$$(5)\quad D = \underset{(a)}{M(U_{jj}+U_{ik}-U_{ji}-U_{jk})} + \underset{(b)}{(MW-C^2)[\tau_j(\tau_j-\tau_i-\tau_k)+\tau_i\tau_k]}$$

$$+\underset{(c)}{W[\tau_i\tau_k U_{jj}+\tau_j(\tau_j U_{ik}-\tau_k U_{ij}-\tau_i U_{jk})]}$$

$$+\underset{(d)}{C[2\tau_j U_{ik}+(\tau_i+\tau_j)U_{jj}-(\tau_j+\tau_k)U_{ij}-(\tau_i+\tau_j)U_{jk}]}+\underset{(e)}{(U_{jj}U_{ik}-U_{ij}U_{jk})}.$$

In the absence of residual earnings variation (so that all the U's are 0), only term (b) in equation (5) is nonzero. Since $C^2/MW = r^2_{mw}$, the squared simple correlation coefficient of m and w, $(MW-C^2) > 0$ unless the correlation is perfect. If $\tau_j = \tau_c (\equiv 0)$, the coefficient of $(MW-C^2)$ is $(\tau_i\tau_k)$. Furthermore, since $\tau_i < \tau_j = 0 < \tau_k$; $(\tau_i\tau_k) < 0$, and so the partial correlation of earnings $r_{ik\cdot j} < 0$ under these assumptions. This conclusion verifies the intuitive argument for the negative sign of $r_{ik\cdot j}$, based on analogy with the condition $r_{ik} < 0$ if $\tau_i < \tau_c < \tau_k$ when residual earnings variance and the variance of m are both zero. A closer inspection of the coefficient of $(MW-C^2)$ reveals that it is always negative for $i < j < k$, which means that this OJT model always statistically implies the partial correlation of earnings, $r_{ik\cdot j} < 0$ in the absence of residual earnings variation from u_t [3]. This result is an interesting and important implication of this version of the OJT model. I am unaware of any other seriously proposed theory of the earnings profile that leads to this conclusion, including theories that attempt to incorporate imperfect foresight or risk attitudes.

Some restrictions are necessary on the covariance matrix of residual earnings in order to draw further conclusions about the sign of equation (5), and we return to the specification in section 2, for $u_t = y_t + z_t$, where $\text{cov}(y_q, z_r) = 0$ for all q and r, y_t is a random walk, and z_t is a random component with zero autocorrelation. In this instance, $U_{qr} = Y_{qq}$ $(=\text{var}(y_q))$ for q<r; and $U_{jj} = Y_{jj} + Z_{jj}$ (where $Z_{jj}=\text{var}(z_j)$). We note with a pure random walk $Y_{qq} < Y_{rr}$ for q<r. With this specification, we have:

$$(5') \quad D = \underset{(a)}{MZ_{jj}} + (MW-C^2)\underset{(b)}{[\tau_j(\tau_j - \tau_i - \tau_k) + \tau_i\tau_k]} + W\underset{(c)}{[\tau_i(\tau_k - \tau_j)Y_{jj} + \tau_j(\tau_j - \tau_k)Y_{ii} + \tau_i\tau_k Z_{jj}]} + C\underset{(d)}{[(\tau_j - \tau_k)Y_{ii} + (\tau_k - \tau_j)Y_{jj} + (\tau_i + \tau_k)Z_{jj}]} + \underset{(e)}{Y_{ii}Z_{jj}}$$

Suppose for the moment that the transitory component of earnings (Z) is nonzero, but the random walk component (Y) is zero. Then to term (b) in (5'), which we have already determined to be negative, we include in following terms: $MZ_{jj} + \tau_i\tau_k WZ_{jj} + (\tau_i+\tau_k)CZ_{jj}$. The first term MZ_{jj} is positive, which tends to mask the negative value of (b) in (5'). However, if $\tau_i < \tau_j = 0 < \tau_k$, the second term is negative, and takes on its largest value for a given $(\tau_k-\tau_i)$ when they are centered on τ_c. For the values of τ_i and τ_k, $(\tau_i+\tau_k)C = 0$. Thus if C is of modest size, a negative partial correlation of earnings $r_{ik \cdot j}$ is reinforced by choosing $\tau_j = 0$ and τ_i and τ_k equal, but of opposite sign. This procedure should maximize the opportunity for detecting the influence of OJT.

On the other hand, if transitory variability (Z) is zero, but the random walk component (Y) is not, we consider the terms from (5'):

$$(6) \quad W[\tau_i(\tau_k - \tau_j)Y_{jj} + \tau_j(\tau_j - \tau_k)Y_{ii}] + C[(\tau_j - \tau_k)Y_{ii} + (\tau_k - \tau_j)Y_{jj}].$$

The coefficient of C is positive, since for a random walk $Y_{jj} > Y_{ii}$ for j>i. Since we expect $C \geqq 0$, the second term will make a non-negative contribution to the partial correlation of earnings $r_{ik \cdot j}$. If the uncorrelated random shocks ε_t that lie behind the random walk have constant variance σ_ε^2 along the earnings profile, then the coefficient of W in (6) can be written $[\tau_i(\tau_k-\tau_j)(\tau_j-\tau_o) - \tau_j(\tau_k-\tau_j)(\tau_i-\tau_o)]\sigma_\varepsilon^2$ where τ_o is the period of entry to the labor force and σ_ε^2 is the (common) constant variance of the random shock generating the random walk. This coefficient is clearly negative for $\tau_o < \tau_i < \tau_j < \tau_k$. If the variance of the random shock decreases with labor market experience, the coefficient becomes even more negative. Finally, if the residual earnings variation is nonzero both from transitory earnings z_t and a random walk component, there is an additional positive contribution to the partial correlation $r_{ik \cdot j}$ from the term $Y_{ii}Z_{jj}$.

This short catalog of the signs of the terms in (6) indicates the conditions under which a negative partial correlation of earnings, $r_{ik \cdot j}$ is most likely to occur. In the absence of an OJT effect in this model, W and C equal zero, and no term in equation (5') is negative. The most serious obstacles to observing a negative partial correlation in this model appear to stem from the positive terms MZ_{jj}, $Y_{ii}Z_{jj}$, and $C((\tau_j-\tau_i)Y_{ii} + (\tau_k-\tau_j)Y_{jj})$ if $\tau_i < \tau_c < \tau_k$. And the last of these terms may not be very important if C is small.

B. Testing OJT Effects by Estimating Var(w)

If time series data on individual earnings are available, one may attempt to estimate the slope parameter of the earnings profile, w, (and the constant m) from the simple regression:

(7) $$x_\tau = m + w\tau + u_\tau \quad .$$

One may then compute the empirical standard deviation (or variance) of these estimated slope parameters, and assess how large of influence one or two standard deviations of this empirical distribution would have in creating earnings differentials in the earnings profiles. Of course, this procedure leads to an upper bound estimate of the true standard deviation of w, since the individual $\hat{w}$ estimates have some error in them, and $var(\hat{w}) = var(w) + var(e_w)$, where the second term is the variance of the errors which we assume uncorrelated with the true w. We note that unlimited sample sizes will not make $var(\hat{w})$ converge to the true var(w), since a larger sample does not reduce the sampling variability of the $\hat{w}$'s.

Since data limitations often force one to estimate regression (7) on the basis of relatively short time series, it is important to obtain reasonably efficient estimates of the w's, in order to reduce the significance of the sampling variance when computing the empirical standard deviation of the w's. In general, we expect to find autocorrelation of the residual u_τ's in regression (7), and we should attempt to use this information for obtaining GLS estimates of the w's.

Norlén (in the appendix) has proposed the following resolution to this problem, based on a theorem by Rao [11] . In the context of this study, we consider the regression:

(8) $$\underset{Tx1}{x} = \underset{Tx2}{Z}\,\underset{2x1}{\beta} + \underset{Tx1}{u} \quad ;$$

where x is the time series of a person's earnings, Z is the matrix

$Z' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ T_0 & T_1 & \cdots & T_{T-1} \end{bmatrix}$ and $\beta' = (m,w)$ <u>Theorem I</u> (Rao [11]; also see Norlén [10] and in the appendix). Consider the covariance matrix Ω^* with form $\Omega^* = \Omega + Z\Phi Z'$, with Φ an arbitrary symmetric positive definite matrix. Then the GLS estimators with Ω^* and Ω are the same. A short proof and other details are in the appendix to this paper by Norlén.

In the absence of the true covariance matrix of the earnings disturbances in (7), [i.e., of $Euu' = \Omega$] for GLS estimation, we use the empirical covariance matrix[4] $\hat{\Omega}^* = [1/(n-1)]\sum_{1}^{n}(x_i - \bar{x})(x_i - \bar{x})'$

(where the sum is over individuals), since $E\hat{\Omega}^* = Z[\frac{1}{n-1}\sum_{1}^{n}$ [illegible] $]Z' + \Omega$,

and this expected matrix satisfies the conditions of the Rao theorem.

We thus obtain the approximate GLS estimates

$$(9) \qquad \hat{\beta}_i = b_i = (Z'\hat{\Omega}^* Z)^{-1} Z'\hat{\Omega}^*(x_i - \bar{x}) \ .$$

This procedure presumably gives us good estimates of the income profile slope parameter $\hat{w}_i$, but in the absence of restrictions on Ω, we are unable to say how much of observed variance of $\hat{w}$ is due to the true variance of w and how much is due to sampling variation from Ω. Without some such restriction, we are limited to the assertion that the empirical variance of $\hat{w}$ provides a reasonable upper bound estimate of the true variance of w. The lower bound estimate of var(w) is zero, which would occur if all of the observed variation in $\hat{w}$ is due to the sampling variation.

An alternative route by which one might make more precise estimates of var(w) is to impose <u>a priori</u> restrictions on the structure of Ω. The discussion in part A of this section assumed that u_t can be decomposed into the sum of the uncorrelated terms $y_t + z_t$, where y_t is a random walk and z_t is a transitory disturbance. If we observe the earnings profile of a sample for T periods (T > 4), then in the most general case, we can rewrite the covariance matrix in at most 2T parameters (instead of the (T+1)T/2 parameters possible for a covariance matrix) in the following form:

$$(10) \quad \Omega = \begin{bmatrix} c_0^2 & & & 0 \\ & c_1^2 & & \\ & & \ddots & \\ 0 & & & c_{T-1}^2 \end{bmatrix} + \begin{bmatrix} d_0^2 & d_0^2 & \cdots & d_0^2 \\ d_0^2 & d_1^2 & & d_1^2 \\ \vdots & & & \vdots \\ d_0^2 & d_1^2 & \cdots & d_{T-1}^2 \end{bmatrix}$$

where the d_i^2 are a nondecreasing sequence of positive numbers. (More restrictive assumptions on Ω are almost always made in actual calculations.) If we then test and accept the hypothesis that the true $\mathrm{var}(x_i) = Z\Phi Z' + \Omega$ with the 3 elements of Φ and the 2T (or less!) parameters of Ω as unknown, $\hat{\Phi}$ would be a good estimate of $\mathrm{var}(\beta)$.[5] Note that this procedure does not depend on first-round estimates of the $\hat{\beta}_i$.

4. RESULTS FROM STATISTICAL CALCULATIONS FOR ON THE JOB TRAINING EFFECTS

The empirical results discussed here all based on cohorts of males. Within each cohort, there is little variance in age (and presumably, in years of post-school employment experience), and the members have the same level of schooling attainment. Most of the calculations reported here utilize subsamples of the Swedish Low Income Commission Study (Låginkomstutredningen), for individuals born between 1936 and 1941.

The schooling attainment classes include: graduates of the Swedish elementary school (folkskola) terminating their formal education at approximately 14 years; those who then had some additional vocational training; graduates of secondary school (realskola) and realskola graduates with additional vocational training. Samples were too small to study gymnasium graduates or those with academic or professional degrees. Taxable income appears to be an adequate surrogate for earnings for this segment of the earnings profile. The taxable incomes are divided by the Swedish consumer price index, and the time index on which the earnings regressions are calculated are based on age rather than calendar year in which income was received. This procedure neglects the possible effect of annual increases in labor productivity for the Swedish economy. It is assumed this factor can be neglected for the six-year cohort. It is also assumed that within the schooling levels, age is a good proxy for post-school employment experience. Taxable income data from 1951-1966 (except for 1959) were obtained from official Swedish records. The calculations take age 22 as the first year for computing the earnings regressions on time, because of enormous noise in the taxable income data due to compulsory military service for the most part between ages 19 and 20 for those with lower levels of schooling attainment.

We consider first a series of alternative estimates of var(w), the variance of the slope parameter of the earnings (taxable income) profiles for deviations from the empirical mean profile for the elementary school graduate cohort in the regression equation:

$$(1') \qquad x_t = \bar{x}_t + m + wt + u_t ,$$

where $\bar{x}_t$ is the empirical mean at time t.

Set A consists of a single OLS regression, in which for each member of the sample, a simple OLS regression was run, and the mean and standard deviations of w was calculated across individuals from the estimated regression parameters. Taxable income in excess of 5000 Skr was required for this calculation. The major defect with OLS is that it takes no account of the covariance structure of u_t, and because this defect leads to larger errors in estimating the parameters m and w in the individual regressions, the calculated empirical standard deviation of w tends to be exaggerated.

Set B consists of three approximate GLS regressions using Rao's theorem, using the empirical covariance of earnings

$\hat{\Omega}^* = \frac{1}{n-1} \sum_{i=1} (x_i - \bar{x})(x_i - \bar{x})'$, which was shown to satisfy Rao's theorem in section 3. It was argued in section 3 that these estimates provide the basis for an upper bound estimate for the standard deviation of w across individuals, since it is not possible to distinguish the sampling variability of the estimated w's from the variation of the true individual w's. However, these estimates are superior to the OLS estimates, since the sampling variance of the estimated w's is less with GLS. The three regressions in this set require individuals to have taxable incomes $\geq$ 5000, $\geq$ 3000, and > 0 Skr respectively, in order to see whether a truncation point seriously modifies the estimated standard deviation of w across individuals.

Set C consists of four regressions that attempt to restrict the covariance matrix $E_{uu'} \equiv \Omega$ to four alternative parametric forms.

Specifically, we take $[1/(n-1)]\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})' = Z\Phi Z' + \Omega$,

where $Z' = \begin{bmatrix} 11\ldots1 \\ 01\ldots6 \end{bmatrix}$, $\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{bmatrix}$, and the four alternative specifications of Ω are:

(11) case

C1 $\Omega = \lambda I$ (i.e., just a constant transitory disturbance);

C2 $$\Omega = \lambda I + \theta \begin{bmatrix} 0 & 1 & 0 & . & . & . & 0 \\ 1 & 0 & 1 & & & & 0 \\ 0 & 1 & 0 & & & & 0 \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ 0 & & . & & . & . & 0 \end{bmatrix}$$ (i.e., tridiagonal);

C3 $$\Omega = C2 + \Psi \begin{bmatrix} 0 & 0 & 1 & 0 & & & & 0 \\ 0 & 0 & 0 & 1 & . & . & . & 0 \\ 1 & 0 & 0 & 0 & & & & 0 \\ 0 & 1 & 0 & 0 & & & & 0 \\ \cdot & & & & & & & \cdot \\ \cdot & & & & & & & \cdot \\ \cdot & & & & & & & \cdot \\ 0 & 0 & 0 & 0 & & & & 0 \end{bmatrix}$$ (i.e., pentagonal);

C4[6] $$\Omega = \lambda I + \theta \begin{bmatrix} 1 & 1 & 1 & . & . & . & 1 \\ 1 & 2 & 2 & & & & 2 \\ 1 & 2 & 3 & . & . & . & 3 \\ \cdot & & \cdot & & & & \\ \cdot & & \cdot & & & & \\ \cdot & & \cdot & & & & \\ 1 & 2 & 3 & . & . & . & 6 \end{bmatrix}$$ (i.e., transitory disturbance & random walk) .

We note that:

(11')
$$Z\Phi Z' = \phi_{11}\,\mathbf{1}\mathbf{1}' + \phi_{12} \begin{bmatrix} 0 & 1 & 2 & . & . & . & 6 \\ 1 & 2 & 3 & & & & 7 \\ 2 & 3 & 4 & & & & 8 \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ 6 & 7 & 8 & & & & 12 \end{bmatrix} + \phi_{22} \begin{bmatrix} 0 & 0 & 0 & . & . & . & 0 \\ 0 & 1 & 2 & & & & 6 \\ 0 & 2 & 4 & & & & 12 \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ 0 & 6 & 12 & & & & 36 \end{bmatrix}$$

We estimate the parameters of (11) and (11') by weighted OLS, regressing the 28 elements of the triangular portion of the matrix $\hat{\Omega}^*$, excluding those above the principal diagonal, on the corresponding elements of $Z\Phi Z' + \Omega$. The weights were the number of observations available for estimating each covariance term of $\hat{\Omega}^*$.

This quick-and-dirty procedure for estimating the parameters in the Set C specification is crude, and at some point, it will probably be worthwhile to develop a maximum liklihood procedure for better estimates with more manageable statistical properties.

We then take $(\hat{\phi}_{22})^{\frac{1}{2}}$ as an estimate of the standard deviation of w across individuals, unless the parameterization of Ω leads to an excessive difference between $\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})'$ and the estimated decomposition $Z\hat{\Phi}Z' + \hat{\Omega}$. The purpose of the Set C estimates is to try to reduce the noise from sampling error in the estimates of the individual w's which is present in the GLS estimates of Set B.

Table I shows the results from these different estimating procedures.

We consider first the estimates of the parameter of greatest interest in this study, the standard deviation of w over individuals, which reflects the deviation of the slope parameter of the taxable income profiles. The difference in the OLS and the corresponding GLS estimate (the second entry in Table I) is small, and it is unlikely that it is statistically significant at conventional significance levels for this cohort. However, the GLS estimate is slightly smaller, which is the expected direction of difference, because of the greater efficiency of GLS. The three alternative GLS estimates indicate that there is only a small difference in the estimated standard deviation of w, depending on truncation of the taxable income series at >0 and ≥5000 Skr.

The parametric estimates of Set C for this parameter are smaller for the four alternative parametric specifications of Ω, compared with the first two estimates. C4, the random walk and transitory distur-

TABLE I

Alternative Estimates of the On-the-Job Parameter* W
Standard Deviation and Related Parameters Across
Individuals with Elementary School Attainment[a]

	# of Cases	Standard Deviation of w	m	m^+	r_{wm}	r^+_{wm}	λ	θ	Ψ
Set A: OLS									
(Income ≥5000 Skr)	111	.60	2.11	1.74	-.60	- .05			
Set B: GLS									
(Income ≥5000 Skr)	111	.57	2.11	1.77	-.58	- .04			
(Income ≥3000 Skr)	128	.59	2.34	2.02	-.53	- .03			
(Income >0 Skr)	135	.60	2.53	2.23	-.46	+.015			
Set C: Parameterized Estimates									
(Income ≥5000 Skr)									
C1 Pure transitory	111	.53	1.47	1.02	-.37	+ .50	1.45		
C2 Triagonal	111	.51	1.38	.92	-.33	+ .62	1.62	.325	
C3 Pentagonal	111	.50	1.31	.85	-.32	+ .71	1.74	.437	.191
C4 Random walk & transitory	111	.48	1.35	.94	-.56	+ .22	1.11	.433	

*We remind the reader that w is the slope parameter of earnings profile deviations from the mean profile of earnings.

[a]The sample is composed of males from the Swedish Låginkomstutredningen (LIU) who were born 1936-1941, who had terminated formal schooling at the elementary level (folkskola). The time series data on taxable income were obtained from official statistics by Göran Ahrne for the present study. The time vector t' = (0,1,2,...6) has 0 occuring at age 22, and 6 at age 28. The sample is further restricted to those having valid taxable income statistics for 4 or more years.

bance model has the lowest value, .48, of all the estimates. The Set C estimates are intended to reduce the effect of sampling variations in $\hat{w}$, in trying to obtain a superior estimate of the true variance of the w's. Thus the smaller values of the estimated standard deviation of w in these estimates are not unexpected, compared with OLS and the corresponding GLS estimates. The sample means of taxable income for $t' = (0, 1, \ldots, 6)$ are $x' = (7.42, 8.08, 8.65, 9.24, 10.14, 10.27, 10.81) \times 1000$ Skr (Swedish crowns). If we take one standard deviation of difference in the profile slope for five years for the GLS case, restricted to those with taxable income $\geq$5000 Skr, we obtain a change in relative incomes of 2750 Skr, an empirically quite substantial difference. However, this estimate is a reasonable <u>upper bound</u>, because it incorporates the sampling variation in the $\hat{w}$'s. The random walk estimate, C4, provides a corresponding five-year effect of 2400 Skr, which is still a quite large empirical difference.

Two sets of estimates of the standard deviation of the regression constant (m and m^{+}) are provided. Section 3 points out that the value of this parameter depends on the origin of the time vector associated with taxable income. The first (m) corresponds to assigning $t = t_c = 0$ at age 22, the first term in the income profile data used in these calculations. There is no basis for assuming that "crossover" t_c, occurs at this young of age, especially since at this age, the covariance of w and m, cov(w,m), is substantial and negative, contrary to other theoretical considerations and direct empirical evidence that the correlation of ability and the slope of the earnings profile is greater than or equal to zero, see, e.g., Hause [5].

The standard deviation of m^+ is obtained with the assumption that $t_c = 0$ at age 24, two years later than the m estimate. The reason for not assuming $t_c = 0$ at an older age is because equation (3d) implies that the variance of m is less than zero for many of the regressions, which is not very appealing. In all cases the assumption that t_c occurs at age 26 or greater leads to negative estimates for the variance of m. Even with the assumption that $t_c = 0$ at 24, the covariance term is negative in Sets A and B, although very small (and probably not significantly different from zero). This result may be inferred from the correlation coefficient of r^+_{wm} for Sets A and B in Table I. This evidence points to $24 < t_c < 25$ as a plausible interval for t_c, consistent with this specification.

Set C estimates of the standard deviation of m and m^+ are substantially different from the corresponding estimates of A and B. This difference may reflect identification problems in an adequate parametric specification of Ω. For example, if there is a constant matrix that should be added to the C4 (random walk) model, one cannot distinguish between the parameter of this constant matrix from the parameter that was previously identified as the variance of m. Thus, one may have substantial doubt about the relevance of the Set C estimates of the standard deviation of m. It is interesting to note that the Set C estimates of r_{wm} and r^+_{wm} change the correlation from negative to large and positive values, as one compares the assumption that t_c occurs at 22 and 24, respectively. One notes in passing that the estimated r_{wm} from C4 is +.72 if we assume the true $t_c = 25$.

In the C4 estimate of θ, the variance of a constant random walk increment is .433, which implies a standard deviation of $(\theta)^{\frac{1}{2}} = .66$. This value is large, and suggests that the random, nontransitory increment to income is large. The standard error of this estimate of θ in the 28 observation regression is .305, which hints at substantial imprecision. The standard error, as estimated by this crude procedure, cannot be taken seriously. The estimated constant transitory variance in Ω, λ, is also large, and has t-values ranging from 3 to 4 over the Set C regressions. This result may be partly due to the weighted regression based on the number of observations in the covariance matrix elements. The t values of the estimates of σ_{ww} are even larger, ranging from 4.5 to 7.

The Set C regression results could be used to provide approximate GLS estimates of the simple regression parameters of individuals, by using the estimated covariance matrix $Z\hat{\Phi}Z' + \hat{\Omega}$, but these calculations have not been carried out.

We turn next to a comparison of OLS and GLS estimates of the standard deviation of w[7] for other levels of schooling. Table II presents the estimates.

These results are painful to behold, since they conflict with the *a priori* expectation that the GLS estimates would tend to yield smaller estimates of the standard deviation of w than OLS, because of greater sampling variation. Perhaps they are an unlikely outcome from sampling error. I have not had time to explore this result, although it may be related to the incomplete observations used for computing the covariance matrix used as the GLS matrix. In the E_2 and E_3 cohorts, there are several estimated $\hat{w}$'s that are 3 or 4 standard deviations from the GLS

TABLE II

OLS and GLS Estimates of the OJT Parameter W Standard Deviation for Three Levels of Schooling Attainment[a]

Educational Attainment	E_1	E_2	E_3
Number of Observations	111	33	21
OLS Estimate for Income ≥5000 Skr	.60	.76	.55
GLS Estimate for Income ≥5000 Skr	.57	.85	.77

[a]Same LIU source as in Table I. E_1 = Folkskola graduate (about 8 years of schooling), E_2 = Folkskola graduate with some vocational schooling, E_3 = Realskola graduate.

mean, whereas no deviations of this size occurred in the corresponding OLS estimates.

For those interested the empirical effect of income differentials implied by these estimates of the standard deviation of w, relative to the income profiles, the E_2 profile $x' = (8.31, 9.01, 9.40, 10.17, 9.58, 10.87, 11.51) \times 1000$ Skr. For the E_3 cohort, the taxable income profile $x' = (7.97, 8.34, 8.68, 9.75, 9.95, 10.91, 10.81) \times 1000$ Skr.

Finally, we consider a sprinkling of estimates of partial correlation coefficients of income or earnings, $r_{ik \cdot j}$, where the time indices satisfy the condition $i < j < k$. The purpose of Section 3A, in the discussion of statistical testing of the OJT hypothesis is to develop a procedure that could be used with fragmentary earnings profile data (consisting of earnings for at least three different points along the profile). The test is to determine whether any systematic OJT effects that might be present in the data are strong enough to overcome the stochastic residual earnings variables, u_t, thereby yielding a significant negative sign for the partial correlation. The discussion concluded that under several specifications, the best opportunity for observing a net OJT effect (i.e., a negative $r_{ik \cdot j}$) occurs if i and k are on opposite sides of t_c and $j \doteq t_c$. We consider several sets of partial correlations computed from a few alternative samples in Table III.

The partial correlations for the first two samples are in most cases based on very small numbers of observations, and are erratic. Most of the coefficients are positive, and they do not, as a whole,

TABLE III

Partial Correlations of Earnings (or Income) at Different Ages and Levels of Schooling Attainment[a]

Sample	Educational Attainment					
1	Folkskola	$r_{17,28.22}$	$r_{18,28.22}$	$r_{21,28.24}$	$r_{22,28.24}$	$r_{24,28.26}$
		-.004	.031	-.036	.016	.059
		[22]	[23]	[47]	[22]	[56]
	Folkskola & Vocational School	$r_{21,28.26}$	$r_{24,28.26}$	$r_{24,28.27}$		
		-.004	.173	-.504		
		[11]	[14]	[14]		
	Realskola	$r_{21,28.24}$	$r_{24,28.26}$			
		.176	.555			
		[7]	[10]			
	Realskola & Vocational School	$r_{21,28.25}$	$r_{24,28.26}$			
		.007	.397			
		[13]	[15]			
2	Folkskola and Folkskola & Vocational School	$r_{27,31.29}$	$r_{27,34.30}$	$r_{26,39.31}$	$r_{30,39.35}$	
		.385	.312	-.108	.190	
		[156]	[114]	[92]	[90]	
	Realskola and Realskola & Vocational School	$r_{28,34.31}$	$r_{26,39.31}$	$r_{26,39.35}$	$r_{30,39.35}$	$r_{34,39.37}$
		.220	.118	-.599	.008	.752
		[28]	[16]	[16]	[15]	[16]
3[b]		$r_{65,55.60}$	$r_{60,50.55}$			
	Non-H.S. Grad.	-.082	+.266			
		[60]				
	H.S. Graduate	-.346	-.023			
		[117]				
	Some College	-.046	-.014			
		[51]				
	College Graduate	+.049	+.020			
		[70]				
	College Grad., 2 or more degrees	-.374	-.185			
		[47]				

Table III (con'd)

[a]Sample 1 is from the subset of the LIU data already used in Tables I and II. Sample 2 is from another subset of the LIU data of males born 1924-1928, with the income data collected by Ingalill Eriksson for her study <u>Ålder och Inkomst</u>. Numbers in brackets are number of observations.

[b]Sample 3 is from data collected by Daniel C. Rogers, mostly Connecticut 8th graders in 1935. The partial correlations are from recalled full-time earnings in 1966 for 1950, 1955, 1960, and 1965. The subscripts on the partial correlations are for calendar year, not age. Assuming 8th graders are about 14 years old would make the partial correlation subscripts for age roughly $r_{44,34.39}$ and $r_{39,29.34}$, respectively.

provide support that gross OJT effects are strong enough to outweigh the random residual variability in earnings. Sample 3 provides stronger support for OJT effects within schooling attainment classes. Recalled full time earnings data may have some error in measurement, but they may effectively substantially reduce the transitory earnings variations which tend to mask OJT effects. The parametric approach to estimating earnings profile standard deviations is obviously more appealing for determining the potential empirical significance of OJT, but further application of the partial correlation test for the potential existence of net OJT effects is worthwhile for fragmentary cohort data on earnings.

The negative partial correlations do not prove the existence of OJT effects (including the possible effect of systematic differences in occupational profiles, of course), but in the absence of a plausible alternative hypothesis, it is very tempting to conclude that some closely related effect is present in the Rogers data.

Appendix:

AN APPLICATION OF A THEOREM FOR BEST LINEAR UNBIASED ESTIMATION OF REGRESSION COEFFICIENTS IN THE CORRELATED CASE

By Urban Norlén

An application of a theorem by Rao (1967) for Generalized Least Squares (GLS) estimation is shown to facilitate the estimation procedure in a situation where a set of regression coefficients is to be estimated.

GLS estimation in its original form involves the disturbance covariance matrix, which is unknown here. With this theorem as a point of departure we select another covariance matrix and use this matrix in the place of the disturbance covariance matrix. This matrix is selected from a class of covariance matrices as given by the theorem and with the property that they all leave the GLS estimates unaffected.

1. THEORY

We consider the generalized linear regression model

(1) $$\underset{T\times 1}{y} = \underset{T\times p}{Z}\,\underset{p\times 1}{\beta} + \underset{T\times 1}{\varepsilon} ,$$

where y is the observation vector on the dependent variable, Z is a given observation matrix on the independent variables of rank $p(<T)$, β is the vector of parameters to be estimated and ε a disturbance vector. y has the mean vector and positive definite covariance matrix

(2) $$E(y) = Z\beta$$

(3) $$E(y - Z\beta)(y - Z\beta)' = \Omega$$

respectively. We shall be concerned with the Generalized Least Squares estimator

(4) $$b = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y$$

of β, which is the Best Linear Unbiased Estimator (BLUE) with covariance matrix

$$(5) \qquad E(b - \beta)(b - \beta)' = (Z'\Omega^{-1}Z)^{-1}.$$

One difficulty in employing this estimator is that the matrix Ω must be known, at least up to a factor of proportionality. This knowledge is often hard to get. In practice, therefore, this problem has often led to the use of the in general less efficient Least Squares (LS) estimator

$$(6) \qquad b^0 = (Z'Z)^{-1}Z'y$$

with covariance matrix

$$(7) \qquad E(b^0 - \beta)(b^0 - \beta)' = (Z'Z)^{-1}Z'\Omega Z(Z'Z)^{-1},$$

which in general is larger than (5) in the sense that the difference between (7) and (5) is a positive definite matrix. In some of these problem situations, however, the following theorem by Rao (1967) may be used in obtaining GLS estimates of β although the disturbance matrix still is unknown. When Z and Ω are of full rank the theorem may be stated as follows.

THEOREM (Rao 1967): Let X be a $T\times(T - p)$ matrix of rank $T - p$ such that $X'Z = 0$, and let Ω^* be a matrix of the form

$$(8) \qquad \Omega^* = \Omega + Z\Phi Z' + \Omega X\Psi X'\Omega,$$

where Φ and Ψ are arbitrary. Then the GLS estimator with Ω^* is the same as that for Ω.

If the matrix

$$(9) \qquad A = I + Z'\Omega^{-1}Z\Omega$$

is nonsingular, a simple proof of the sufficiency may be given. Consider the identity

$$(10) \qquad \Omega^{-1} - \Omega^{*-1} = (\Omega^{-1}\Omega^* - I)\Omega^{*-1} = (\Omega^{-1}Z\Phi Z' + X\Psi X'\Omega)\Omega^{*-1}.$$

Premultiplication by Z' gives after simplification

(11) $$Z'\Omega^{-1} = (I + Z'\Omega^{-1}Z\Phi)Z'\Omega^{*-1} = AZ'\Omega^{*-1}$$

(12) $$Z'\Omega^{*-1} = A^{-1}Z'\Omega^{-1}.$$

Postmultiplication by Z gives

(13) $$(Z'\Omega^{*-1}Z)^{-1} = (Z'\Omega^{-1}Z)^{-1}A,$$

which combined with (12) establishes the relation

(14) $$(Z'\Omega^{*-1}Z)^{-1}Z'\Omega^{*-1} = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1},$$

i.e. the GLS estimator with Ω^* and that for Ω are the same.

In passing we notice that an important class of covariance matrices is defined by (8) with $\Omega = \sigma^2 I$. For these covariance matrices the GLS and the LS estimators are the same.

2. APPLICATION

In the following application one use of Rao's theorem is demonstrated in a multirelation situation where a set of regression coefficients is to be estimated with the GLS method.

In his time series analysis of differences of Swedish income profiles for people with different education backgrounds, Hause (1973) specifies the following set of N regression relations

(15) $$\underset{Tx1}{y_i} = \underset{Tx1}{\mu} + \underset{Tx2}{Z}\underset{2x1}{\beta_i} + \underset{Tx1}{\varepsilon_i}, \qquad i = 1, i = 1, \ldots, N$$

where y_i is the observation vector on individual i's taxed income (in thousands of Swedish Crowns at the price level of 1949) at T consecutive ages, μ is the mean income for the N individuals, Z is the matrix

(16) $$Z = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T \end{bmatrix}$$

implying that the two parameters in the vector β_i are the intercept and slope respectively in the linear trend function for individual i measuring his deviances from the mean incomes μ. The individual intercepts and slopes are identified by the restriction

(17) $$\sum_{i=1}^{N} \beta_i = 0.$$

The following assumptions are made

(18) $$E(y_i) = \mu + Z\beta_i$$

(19) $$E(y_i - \mu - Z\beta_i)(y_i - \mu - Z\beta_i)' = \Omega,$$

i.e. the disturbance covariance matrix Ω is the same for all individuals.

Since we are dealing with time series with correlated observations it is desirable to employ the GLS method for the estimation of the regression coefficients. To achieve this end we proceed as follows.

ESTIMATION OF μ: The GLS method gives those values of the parameters that minimize the function

(20) $$f = \sum_{i=1}^{N} (y_i - \hat{\mu} - Zb_i)'\Omega^{-1}(y_i - \hat{\mu} - Z\beta_i).$$

Differentiation with respect to $\hat{\mu}$ gives

(21) $$\frac{1}{2}\frac{\partial f}{\partial \hat{\mu}} = -\sum_{i=1}^{N} \Omega^{-1}(y_i - \hat{\mu} - Zb_i) = -N\Omega^{-1}(\bar{y} - \hat{\mu}),$$

where we have used the restriction (19). The GLS estimator of μ is thus simply the sample mean income vector.

ESTIMATION OF THE β_i's: The GLS estimators of the β_i's are

(22) $$b_i = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}(y_i - \bar{y}) \qquad i = 1, \ldots, N.$$

The restriction (16) obviously holds true for these estimators. The disturbance covariance matrix Ω is, however, not known, which prohibits the straightforward use of this estimator. We may instead construct an approximate estimator by inserting

(23) $$S = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})(y_i - \bar{y})'$$

in the place of Ω in (22), i.e. by using the observed income covariance matrix. For the expectation of this matrix

(24) $$E(S) = Z\left(\frac{1}{N-1}\sum_{i=1}^{N}\beta_i\beta_i'\right)Z' + \Omega$$

belong to the equivalence class (8) that was shown to leave the GLS estimates unaffected. Approximate GLS estimates of the set of regression coefficients may thus be obtained from

(25) $$b_i = (Z'S^{-1}Z)^{-1}Z'S^{-1}(y_i - \bar{y}) \qquad i = 1, \ldots, N.$$

The expected difference in the variability of the OLS and approximate GLS parameter estimates can be shown analytically. One measure of the difference is

(26) $$D = \frac{1}{N-1}\sum_{i=1}^{N} b_i^0 b_i'^0 - \frac{1}{N-1}\sum_{i=1}^{N} b_i b_i' = (Z'Z)^{-1}Z'SZ(Z'Z)^{-1} - (Z'S^{-1}Z)^{-1},$$

where b_i^0 is the LS estimate (6) of the regression coefficients for individual i. This matrix is in general a positive definite matrix.

FOOTNOTES

*The initial phases of this research were supported by the National Bureau of Economic Research, and further work has been supported by a National Institute of Education (U.S. Office of Education) Grant No. OEG-0-72-1569, NIE Project No. 2-0680. Neither institution has reviewed or approved this work, or the conclusions. I am greatly indebted to Chris Sims for comments that have deeply influenced my thinking about the theoretical and statistical implications of the basic model, and to Urban Norlén for resolving much of the GLS estimation procedures and carrying out the GLS and parametric Ω calculations. It is a pleasure to acknowledge the cooperation of Sten Johansson and Ingalill Ericksson for making it possible for me to use a special data tape created by Ericksson for Låginkomstutredningen study in Sweden. Thanks are also due to Göran Ahrne who assembled the income data for those born 1936-1941. The study also has benefited from the valuable research assistance of Karlis Goppers, and from work by Eilon Amit, Lars Anlér, Lars Elve Larson, and Edward Fagerlund, who carried out computations. at different stages. The author is responsible for conclusions and errors, but the complexity of earnings profiles makes the data partly responsible for empirical ambiguity.

[1]This issue is discussed in greater detail in Hause [5].

[2]The lower bound estimate of zero is approached, for example, if working lifetime is divided into a large number of periods, in which there is no correlation of earnings between periods. The upper bound cited in the text is obtained in Hause [ ], based on a proof sketched by Sims. Let x_t be earnings as a function of age, t, μ_t the corresponding mean earnings as a function of age, and ρ the constant discount factor.

$$(\int e^{-\rho t}[E(x_t - \mu_t)^2]^{1/2}dt)^2$$

$$= \int\int (E[e^{-\rho t}(x_t - \mu_t)]^2)^{1/2}(E[e^{-\rho t'}(x_{t'} - \mu_{t'})]^2)^{1/2}dtdt'$$

$$\geq \int\int E[e^{-\rho t}(x_t - \mu_t)e^{-\rho t'}(x_{t'} - \mu_{t'})]dtdt'$$

(by Schwarz's inequality),

$$= E\int\int e^{-\rho t}(x_t - \mu_t)e^{-\rho t'}(x_{t'} - \mu_{t'})dtdt'$$

$$= E(\int e^{-\rho t}(x_t - \mu_t)dt)^2.$$

Taking square roots of the first and last steps yields the upper bound asserted in the text.

[3]Proof: The coefficient of $(MW - C^2)$ is negative, given $i < j < k$ if $j = 0$, regardless of the (permissible) values of the other two time indices, i and k. Now consider a translation in time of the three time indices by the same amount, τ, with τ_j initially 0. Then the coefficient of $(MW - C^2)$ is $\tau(-\tau - \tau_i - \tau_k) + (\tau_i + \tau)(\tau_k + \tau)$. The derivative of this translated coefficient with respect to τ is 0. Thus a constant translation of the time indices does not affect the value of the coefficient. Hence the coefficient is negative for all permissible values of the indices.

[4]This formula for $\hat{\Omega}^*$ assumes that all observation vectors are complete. In most available time series on earnings or income, the data base is modest, and one cannot afford the luxury of restricting the computations to complete observations. When there are missing observations each element of $\hat{\Omega}^*$ has the form

$$\frac{\sum_{k=1}^{n_{ij}} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{(n_{ij} - 1)}$$

where the sum is only over the n_{ij} observations for which x_{ik} and x_{jk} is present (k is the subscript for an individual), and the means $\bar{x}_i$ and $\bar{x}_j$ are also based on this same set of observations. We retain the slightly misleading notation of the text even when there are missing observation elements for simplicity, remembering, however, that covariance matrices constructed as described in this note are not necessarily positive definite.

[5]I am indebted to U. Norlén and C. Sims for discussion of this procedure.

[6]One complication arises in the case in which Ω is assumed to be the combination of a random walk and a transitory disturbance. Since the data do not pick up the taxable income profile at the beginning of labor force participation, it seems plausible that the random walk for the initial data point will have already progressed to a level where the variance from the first (and subsequent) years should augment Ω by $\mu 11'$. But μ cannot be identified, since $\mu 11'$ cannot be distinguished from $\phi_{11}11'$. Hence the estimated ϕ_{11} cannot be assumed a good estimate across individuals of the variance of m, the constant parameter in the regressions of (1'), in specification C4.

[7]The corresponding Set C estimates are not presented because of complications stemming from small sample sizes. The covariance matrix whose elements below (and including) the principle diagonal are used as dependent variables in a regression on the assumed structure of Ω is based on incomplete observations and need not be positive definite. A number of peculiarities arose in the estimated unconstrained parameters in preliminary calculations.

REFERENCES

[1] ERIKSSON, I.: *Ålder och Inkemst* (monograph for Låginkamstutredningen), Stockholm: Allmänna Förlaget, 1970.

[2] FASE, M. M. G.: *An Econometric Model of Age-Income Profiles*, Rotterdam: Rotterdam University Press, 1970.

[3] FRIEDMAN, M. AND S. KUZNETS: *Income from Independent Professional Practice*, New York: National Bureau of Economic Research, 1954.

[4] HANNA, F. "The Accounting Period and the Distribution of Income," Part 3 of F. Hanna, J. Peckman, and S. Lerner, *Analysis of Wisconsin Income*, Vol. 9, Conference on Research in National Income and Wealth, New York: NBER, 1948.

[5] HAUSE, J. C.: "Earnings Profile: Ability and Schooling," *Journal of Political Economy* (1972), Supplement, S108 - S138.

[6] HAUSE, J. C.: "The Risk Element in Occupational and Educational Choices: Comment," (to appear) *Journal of Political Economy* 82 (1974).

[7] MENDERSHAUSEN, H.: *Changes in Income Distribution During the Great Depression*, Vol. 7, Studies in Income and Wealth, New York: NBER, 1946.

[8] MINCER, J.: "The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach," *Journal of Economic Literature* 8 (1970), 1 - 26.

[9] MINCER, J.: *Schooling, Experience, and Earnings*. (Preliminary, but revised manuscript.) NBER, 1972 **[forthcoming, NBER, 1974]**.

[10] NORLÉN, U.: "The Structure of Covariance Matrices for which Least Squares is Best Linear Unbiased and its Use," (preliminary manuscript), Uppsala Institute of Statistics, 1973.

[11] RAO, C. R.: "Least Squares Theory Using an Estimated Dispersion Matrix and its Application to Measurement of Signals," in Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967.

[12] THATCHER, A. R., "Year-to-Year Variations in the Earnings of Individuals, *Journal of the Royal Statistical Association*, Series A, 134 (1971), 374 - 382.

NATIONAL BUREAU OF ECONOMIC RESEARCH, INC.

WORKING PAPER SERIES *

Number	Author	Title	Date
1	Finis Welch	Education, Information, and Efficiency	6/73
2	Barry R. Chiswick	Hospital Utilization: An Analysis of SMSA Differences in Hospital Admission Rates, Occupancy Rates and Bed Rates	6/73
3	Swarnjit S. Arora	Error Components Regression Models and their Applications	6/73
4	L. A. Lillard	Human Capital Life Cycle of Earnings Models: A Specific Solution and Estimation	7/73
5	James P. Smith	A Life Cycle Family Model	7/73
6	Victor Zarnowitz	A Review of Cyclical Indicators for the United States: Preliminary Results	7/73
7	Lewis C. Solmon	The Definition and Impact of College Quality	8/73
8	Robert E. Lipsey and Merle Yahr Weiss with the assistance of Barbara Rotenberg	Multinational Firms and the Factor Intensity of Trade	9/73
9	L. A. Lillard	From Age-Earnings Profiles to the Distribution of Earnings and Human Wealth	9/73
10	Paul W. Holland	Monte Carlo for Robust Regression: the Swindle Unmasked	9/73
11	Paul W. Holland	Weighted Ridge Regression: Combining Ridge and Robust Regression Methods	9/73
12	Melvin Reder	Citizen Rights and the Cost of Law Enforcement	10/73
13	Reuben Gronau	Wage Comparisons - A Selectivity Bias	10/73
14	Lewis C. Solmon and Paul Wachtel	The Effects on Income of Type of College Attended	10/73
15	Ray C. Fair	A Comparison of FIML and Robust Estimates of a Nonlinear Macroeconomic Model	10/73

* Selected titles. Copies obtainable from authors.

Number	Author	Title	Date
16	David C. Hoaglin	Monte Carlo Techniques in Studying Robust Estimators	11/73
17	Paul Taubman	Schooling, Ability, Non Pecuniary Rewards, Socioeconomic Background and the Lifetime Distribution of Earnings	11/73
19	Edward F.X. Hughes, Eugene M. Lewit, Richard N. Watkins, Richard Handschin	Utilization of Surgical Manpower in a Prepaid Group Practice	12/73
20	Victor R. Fuchs	Short-Run and Long-Run Prospects for Female Earnings	12/73
21	Robert T. Michael and Robert J. Willis	Contraception and Fertility: Household Production under Uncertainty	12/73
22	Michael Grossman	The Correlation Between Health and Schooling	12/73
23	Warren Sanderson	What Happened During the Baby Boom? New Estimates of Age- and Parity-Specific Birth Probabilities for American Women	12/73

NATIONAL BUREAU OF ECONOMIC RESEARCH, INC.

WORKING PAPER SERIES *

Number	Author	Title	Date
1	Finis Welch	Education, Information, and Efficiency	6/73
2	Barry R. Chiswick	Hospital Utilization: An Analysis of SMSA Differences in Hospital Admission Rates, Occupancy Rates and Bed Rates	6/73
3	Swarnjit S. Arora	Error Components Regression Models and their Applications	6/73
4	L. A. Lillard	Human Capital Life Cycle of Earnings Models: A Specific Solution and Estimation	7/73
5	James P. Smith	A Life Cycle Family Model	7/73
6	Victor Zarnowitz	A Review of Cyclical Indicators for the United States: Preliminary Results	7/73
7	Lewis C. Solmon	The Definition and Impact of College Quality	8/73
8	Robert E. Lipsey and Merle Yahr Weiss with the assistance of Barbara Rotenberg	Multinational Firms and the Factor Intensity of Trade	9/73
9	L. A. Lillard	From Age-Earnings Profiles to the Distribution of Earnings and Human Wealth	9/73
10	Paul W. Holland	Monte Carlo for Robust Regression: the Swindle Unmasked	9/73
11	Paul W. Holland	Weighted Ridge Regression: Combining Ridge and Robust Regression Methods	9/73
12	Melvin Reder	Citizen Rights and the Cost of Law Enforcement	10/73
13	Reuben Gronau	Wage Comparisons - A Selectivity Bias	10/73
14	Lewis C. Solmon and Paul Wachtel	The Effects on Income of Type of College Attended	10/73
15	Ray C. Fair	A Comparison of FIML and Robust Estimates of a Nonlinear Macroeconomic Model	10/73

* Selected titles. Copies obtainable from authors.

Number	Author	Title	Date
16	David C. Hoaglin	Monte Carlo Techniques in Studying Robust Estimators	11/73
17	Paul Taubman	Schooling, Ability, Non Pecuniary Rewards, Socioeconomic Background and the Lifetime Distribution of Earnings	11/73
19	Edward F.X. Hughes, Eugene M. Lewit, Richard N. Watkins, Richard Handschin	Utilization of Surgical Manpower in a Prepaid Group Practice	12/73
20	Victor R. Fuchs	Short-Run and Long-Run Prospects for Female Earnings	12/73
21	Robert T. Michael and Robert J. Willis	Contraception and Fertility: Household Production under Uncertainty	12/73
22	Michael Grossman	The Correlation Between Health and Schooling	12/73
23	Warren Sanderson	What Happened During the Baby Boom? New Estimates of Age- and Parity-Specific Birth Probabilities for American Women	12/73