1972-06

A study of Navy first-term reenlistments

Altergott, Dennis Albert

Monterey, California. Naval Postgraduate School

A STUDY OF NAVY FIRST-TERM REENLISTMENTS

Dennis Albert Altergott

NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

A Study of Navy First-Term Reenlistments

by

Dennis Albert Altergott

Thesis Advisor: R. S. Uhler

June 1972

Approved for public release; distribution unlimited.

A Study of Navy First-Term Reenlistments

by

Dennis Albert Altergott
Lieutenant, United States Navy
B.A., University of Southern California, 1965

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 1972

[illegible]
[illegible]
[illegible]

ABSTRACT

It occasionally happens in economic analyses that the correctly specified model contains variables for which no observed data has been collected. When the data in a linear regression model are cross-sectional it is possible, under certain conditions on the nature of the variables, to estimate the independent effects of a specific set of explanatory variables on the dependent variable. A procedure for doing this is presented.

A commonly used model of reenlistment behavior, for which the data base is cross-sectional, satisfies the requisite conditions. This permits the estimation of the independent effect of the military wage on reenlistment rate, as an illustration of the proposed procedure.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGMENT

The author wishes to express his gratitude to the personnel of the Center for Naval Analyses for furnishing the source of data used in this study, as well as for providing valuable guidance and assistance.

TABLE OF SYMBOLS AND ABBREVIATIONS

A'	the transpose of A
$r(A)$	the rank of A
$tr(A)$	the trace of A
$\overline{0}$	the vector of zero elements
$X \sim F$	the random variable X has distribution F
$X_{\cdot jt}$	the mean of X_{ijt} taken over subscript i
$[Z_{ij}]$	the vector with elements Z_{ij}
$A \to B$	$\lim A = B$

I. INTRODUCTION

A. PRELIMINARY

There is currently some concern about the enlistment and retention of men to serve in the armed forces in a draft-free environment. In defining the problem to be resolved, a number of studies (notably [1]) have attempted to describe the factors which affect enlistment and reenlistment behavior. A large part of this interest is directed toward the determination of a military wage structure which will ensure that civilians will enlist, and that servicemen will reenlist, in sufficient numbers to meet service manpower requirements. This paper will concentrate on a part of this latter problem. Specifically, the purpose here is to estimate the elasticity of reenlistment rate with respect to military wage for first-term reenlistees in the Navy. Though studies of this kind have already been conducted, there are a number of reasons for additional study. Among them is that a new source of data (previously unused data in the form of BuPers Report ED198A for fiscal years 1964 through 1970) is used here, which is more complete than that used in prior studies. As a consequence of the availability of the new data, some omissions of previous studies may be corrected. But, most importantly, a somewhat novel procedure is used to estimate the parameter of interest in what will later be introduced as the reenlistment model.

B. BACKGROUND; DESCRIPTION OF THE DATA

In the past, extensive reliance has been placed in the technique of gathering information about reenlistment behavior by the use of surveys over potential reenlistees. This technique depends on before-the-fact information, which is in the form of the stated intentions of men facing

the decision to reenlist. Typically these surveys seek to determine, by means of a question and response approach to the subjects, the factors which affect the reenlistment decision, and thus have value in indicating the lines along which quantitative research should be performed. That is, they serve primarily to identify those factors which should enter into an analytic model of reenlistment behavior. But once such a model is constructed, reliable quantitative results can only be obtained by investigating the observed behavior of potential reenlistees. This after-the-fact information, the revealed reenlistment behavior, is provided by the newly available data used in this paper.

Data extracted from BuPers Report ED198A for use here have the form of pooled time series and cross-sectional information. In particular, the numbers of men eligible to reenlist and the numbers of these that do in fact reenlist are provided for each combination of

(1) Pay grade: E-1 through E-9

(2) Rate (a Navy skill or job specialty classification): BM, QM, ST, TM, FT, MT, ET, DS, AT, AX, AQ, TD, SM, RD, RM, CT, AC, PT, HM, DT, DM, MU, EA, AG, PH, YN, PN, DP, SK, DK, JO, PC, AK, AZ, GM, MN, IM, OM, EN, BT, EM, IC, CM, AD, AO, AB, AE, AM, PR, LI, MR, SF, DC, PM, ML, CE, EO, BU, SW, MT, CS, SH, SD, MM, AV, SP, BR, EQ, CU, SO, AW, AS.

(3) Mental Group: I, II, upper III, lower III, IV.

(4) Fiscal year of reenlistment: 1964 through 1970. First-term reenlistments only are considered. (First-term reenlistments are those of servicemen completing their initial term of active obligated service.) Reenlistments beyond the first term are considerably less interesting, since these advanced-term reenlistments typically involve personnel already committed (psychologically) to a Navy career.

"Mental Group," a designation akin to IQ that is applied to enlisted personnel, is determined by testing as is intelligence quotient. As such it is not likely to be highly reliable. Aside from the facility with which personnel in the higher mental groups may enter certain more technical Rates, and the fact that it may be significant for an enlisted man who wishes to become an officer candidate, there is no special advantage or disadvantage accrued by designation as a member of any particular mental group. On the contrary, there is possibly even a tendency on the part of a certain group of men to score poorly, purposely, in the testing. This group would consist of some of the personnel of better than average education who have enlisted in the Navy, during the past few years of a high level of military activity in Vietnam, to fulfill military service obligation and to avoid more hazardous duties. It is likely that some part of this group, in merely wishing to serve their required time in the armed forces, would seek to escape prominence in their enlisted service. There is, as a consequence, seemingly little general incentive to score well in Mental Group testing. In addition, testing for Mental Group classification is subject to the same criticisms that have recently been directed at classical IQ testing: some minority groups may be put at a disadvantage by the biased (toward comprehensibility by white mid-Americans) nature of the test. In any case, classification by Mental Group is certainly less reliable than cross-sectional classification by pay grade or Rate, or time series classification by fiscal year of reenlistment. As a consequence, the Mental Group classification will not be of primary interest here.

Certain of the Rates included in the above report are unsuitable for inclusion in the analysis. Those Rates that are discarded from the data base are AV, SP, BR, EQ, CU, SO, AW, AS, MT, DS and SD. Any Rate not

included in the study was disallowed for one of the following reasons:

1. The Rate consisted of pay grades E-7 through E-9 only;
2. The Rate's membership consisted in large part of foreign nationals who could be expected to reenlist with high probability;
3. Data for the Rate were not available for each of the fiscal years 1964 through 1970.

The fact that the data consists of a time series of cross-sections of revealed reenlistment behavior allows the correction of an omission of previous research. To date little effort has been made to establish a relationship between the variation over time of reenlistment behavior and the variation over time of pecuniary considerations facing the potential reenlistee. The time series of cross-sectional data provides a basis on which such a relationship can be constructed. The term "constructed" is used advisedly, since the pecuniary factors considered here are those imbedded in a particular model of reenlistment behavior.

Another disadvantage of previous research has been that pecuniary factors for potential reenlistees have only been considered in coarse detail. The minuteness of the new cross-sectional data, on the other hand, permits a more precise formulation of the economic factors that face the individual potential reenlistee. These factors vary from man to man; they are dependent on the individual's level of proficiency (pay grade), job specialty (Rate), and fiscal year in which the reenlistment decision is made.

II. THEORY UNDERLYING THE REENLISTMENT MODEL

A. FOUNDATION

The aim in this paper is to determine the rate of change of first-term Navy reenlistments with respect to the rate of change in military compensation. Toward this end a model is presented to describe reenlistment behavior, quantitatively represented by reenlistment rate, in terms of those variables which affect the reenlistment decision. Then, using the model as a basis the pure effect of the military wage on reenlistment rate is determined. Necessarily, the influence of all other variables must be removed in order to estimate the independent effect of the military wage.

B. TASTE AND OPPORTUNITY FACTORS.

Consider an individual who is eligible to reenlist. The variables which affect his decision may be aggregated into three broad categories: pecuniary, personal non-pecuniary and general non-pecuniary. The first two of these categories are of interest in this section (the final category is discussed later). Within the first category are all factors which reflect opportunity (monetary) considerations. It includes such variables as expected basic military wage, benefits to servicemen which may be expressed equivalently in monetary terms, and the alternative civilian wage. Elements in the personal non-pecuniary class include such factors as military job satisfaction, agreeability with the quality of home life offered by Navy service, adaptability to the military hierarchy, and attitude towards sea or shipboard duty. Variables which are described as non-pecuniary are difficult to

quantify. However, by employing the concept of reservation wage (for a more complete discussion, see, for example, Gray [2]), the effect of these purely individual non-pecuniary factors on the reenlistment decision can be incorporated in a variable with analytic expression. The qualifying phrase "purely individual" is to be stressed. Just as factors which affect the reenlistment decision and which are unique to each individual can be identified, so can be recognized non-pecuniary factors affecting the reenlistment decision which are unique to each Rate, or to each pay grade, or to each year. Variables of this sort are the general non-pecuniary factors and will be introduced and treated later. This is accomplished by considering the pecuniary compensation that will just induce an individual to reenlist. The variables in the class of personal non-pecuniary factors can be viewed as elements which contribute to the determination of the value of compensation required to induce reenlistment. Knowledge of this level of compensation for an individual makes knowledge of the personal non-pecuniary factors affecting his reenlistment behavior redundant (at least in a study where interest centers on macroscopic reenlistment behavior). As a consequence, the personel non-pecuniary variables need not be explicitly considered[1] since they are imbedded into the individual's reservation wage, which will now be defined. Suppose that an individual deliberating reenlistment is capable of estimating the expected present value of his alternative courses of action: to

---

[1]This is an advantage of the use of data describing revealed reenlistment behavior: and individual's personal non-pecuniary attitudes are inconsequential; the fact of his reenlistment displays that any personal dislikes of the service were overcome by sufficient compensation.

reenlist or not to reenlist. Let WM represent the present value of all pecuniary returns if his choice is to reenlist, and let WC represent the present value of all pecuniary returns if he chooses not to reenlist. WM consists of two types of pecuniary returns. Most obviously there are those whose dollar value is fixed and is not subject to individual interpretation: basic pay, variable reenlistment bonus, basic allowance for subsistence, clothing allowance. There are also pecuniary returns whose dollar value is in large part subjectively determined by the individual: free medical services for the serviceman and his dependents, Navy exchange and commissary privileges and others. This distinction is not negligible, and will be treated explicitly later. For a serviceman on active duty, the determination of WC is not as straightforward as that of WM. Typically the serviceman may have little more than a rough estimate, in the year in which the reenlistment decision is made, of the mean wage received by civilians working in a job category similar to that of the serviceman and located in the geographical area of interest to him. Now define $\frac{WM}{WC}$ as the relative wage. Then the reservation relative wage is defined as the value of the above ratio which will just induce the serviceman to reenlist. The individual will reenlist if his actual relative wage is greater than or equal to his reservation relative wage. Similarly, among the entire cohort of eligible reenlistees, those that reenlist will be those whose actual relative wage is greater than or equal to their reservation relative wage. Now consider the domain of possible values of reservation relative wage. For each number in this domain, some portion of the eligible population will reenlist. As a consequence, the reenlistment rate (over the eligible population) has some functional expression over the

domain of reservation relative wage. This introduces a variable of fundamental importance in constructing an analytic expression for reenlistment rate.

The form of the functional dependence will be discussed later. It is worth noting here than an individual's reservation relative wage is some fixed value of the ratio $\frac{WM}{WC}$. Presumably, an individual considering reenlistment is able to estimate the expected present value of pecuniary returns for not reenlisting, so his reservation relative wage can be equivalently expressed as the ratio of a sufficiently large value of expected present value of returns for reenlisting to his estimate of returns for not reenlisting. This says of course that for each individual the reservation wage uniquely determines a value of WM sufficiently large to induce reenlistment. As a consequence reenlistment rate, for fixed WC, has a functional representation over the domain of WM: for each value of WM a certain fraction of the eligible population with given WC will reenlist. The implications of these obvious comments are meant as a preliminary to later work. In order to assure proper statistical control of the variables in the model, it is necessary to be able to match observations of reenlistment rate with corresponding relative wage. That is, a particular set of men eligible to reenlist faces a given relative wage (the members of this set who reenlist in the face of this relative wage are those for whom this relative wage is the reservation relative wage). This set of men eligible to reenlist must be identifiable, for each observed relative wage, in order to be able to perform significant statistical analysis. By the preceeding remarks, an equivalent necessary condition for proper statistical control is that for any fixed value of WC it is possible to identify the set of men eligible to reenlist which corresponds to any value

of WM. Or, for any value of WC and any value of WM, it is necessary to be able to identify the appropriate corresponding eligible population. Now just as the purpose of this section was to eliminate the necessity of identifying, and including in the model, variables which are in the class of personal non-pecuniary factors, a purpose of later section will be to remove the requirement that the value of WC for a potential reenlistee be known. What will in effect be accomplished is that the variable WC will be removed from the model, so that a correspondence between reenlistment rate and WM only need be made in order to satisfy the functional requirement that reenlistment rate depends on relative wage and the statistical requirement that the appropriate eligible population be identifiable for given WM and WC.

C. THE REENLISTMENT MODEL IN CROSS-SECTION AND TIME SERIES; OTHER FACTORS AFFECTING REENLISTMENT RATE

In the preceeding section, a model of the form R = f(WM/WC) was postulated, where WM and WC are as previously defined and R represents reenlistment rate. Fisher [3] and [4] first concluded that a model of the form R = f(ln (WM/WC)) was indicated. Specifically, Fisher concluded that the appropriate model was expressed by:

$$R = \alpha + \beta \ln (WM/WC) + \varepsilon,$$

a linear expression for R in ln(WM/WC), with disturbance term ε. Later work, for example Nelson [5], employed a relation of the form:

(a) $$\ln R = \alpha + \beta \ln(WM/WC) + Z + \varepsilon,$$

where the term Z represents an additional set of variables which are included in the model. The variables in Z depend, of course, on the

author of the study employing the model. A similar model in Logit[2] form,

(b) $$\ln\left(\frac{R}{1-R}\right) = \alpha + \beta \ln(WM/WC) + Z + \varepsilon,$$

has also been considered by, for example, Gray [2] and Wilburn [6].

In this paper models of both forms (a) and (b) will be considered for comparative purposes. Note that equations (a) and (b) may be rewritten as:

(a') $$\ln R = \alpha + \beta \ln WM - \beta \ln WC + Z + \varepsilon,$$

(b') $$\ln\left(\frac{R}{1-R}\right) = \alpha + \beta \ln WM - \beta \ln WC + Z + \varepsilon.$$

Or:

(a") $$R = \alpha' \left(\frac{WM}{WC}\right)^{\beta} Z' \varepsilon' ,$$

(b") $$\frac{R}{1-R} = \alpha' \left(\frac{WM}{WC}\right)^{\beta} Z' \varepsilon',$$

where:

$$\alpha' = \exp(\alpha), \quad Z' = \exp(Z), \text{ and } \varepsilon' = \exp(\varepsilon).$$

These equations imply that, depending on which of the models (a) or (b) is used, either ln R or $\ln(\frac{R}{1-R})$ is linear in the natural log of the ration WM/WC (neglecting for the moment the effect of the variables in Z). The implicit assumption is made, then, that the potential reenlistee values the dollars in WM and in WC in constant ratio. That is, the potential reenlistee is indifferent to an equal percentage change in WM and in WC: his reenlistment decision remains the same whether the relative wage offered him is the ratio WM_1/WC_1, or the ration $(1 + a)WM_1/(1 + a)WC_1$, for any a (a may be positive, negative or zero, repreenting an increase, decrease or lack of change

---

[2] Note that just as reenlistment rate R can be considered to be the sample estimate of the probability of reenlisting, the ratio R/(1-R) may be interpreted as the sample estimate of the odds of reenlisting.

respectively in each of WM_1 and WC_1). This may not actually reflect the candidate reenlistee's utility of dollars in WM and WC. The man may in fact value a percentage increase in his civilian alternative wage WC more highly (or even less than) the same percentage increase in WM.

To relieve this possibly erroneous assumption, the following revisions to models (a) and (b) will be used:

(c) $$R = \alpha' \left(\frac{WM}{WC^{\delta}}\right)^{\beta} Z' \, \varepsilon' \quad ,$$

(d) $$\frac{R}{1-R} = \alpha' \left(\frac{WM}{WC^{\delta}}\right)^{\beta} Z' \, \varepsilon' \ .$$

The parameter δ reflects the possibility that a potential reenlistee values a percentage change in WM and the same percentage change in WC differently. Presumably, the value of β is positive. If this is the case, then: if $\delta > 1$ a percentage change in WC is valued more highly than the same percentage change in WM; if $\delta = 1$ equations (c) and (d) become (a) and (b); if $0 < \delta < 1$ a percentage change in WM is valued more highly than the same percentage change in WC; if $\delta = 0$ the decision to reenlist is independent of the candidate reenlistee's civilian alternative wage; a value of $\delta < 0$ indicates an aversion to civilian dollars. These equations may be rewritten as:

(c') $$\ln R = \alpha + \beta \ln WM + \gamma \ln WC + Z + \varepsilon,$$

(d') $$\ln \left(\frac{R}{1-R}\right) = \alpha + \beta \ln WM + \gamma \ln WC + Z + \varepsilon,$$

where: $$\gamma = -\beta\delta.$$

If $\gamma = -\beta$, then the equations (c') and (d') become (a') and (b').

The coefficient β in the equations (c') and (d') is the parameter of interest. In equation (c'), β is the military wage elasticity of reenlistment rate since application of the partial differential operator

∂ to (c'), while neglecting the disturbance term ε, yields:

$$\partial(\ln R) = \beta\ \partial(\ln WM) + \gamma\ \partial(\ln WC) + \partial Z\ ;$$

or

$$\partial R/R = \beta(\partial WM/WM) + \gamma\ \partial(\ln WC) + \partial Z.$$

Similarly, in equation (d') β represents the elasticity of the odds of reenlistment with respect to military wage.

It is now appropriate to consider some assumptions about the nature of the cross-section and time series data. First, consider reenlistment behavior of cohorts of eligible reenlistees over time. It seems reasonable to assume that an individual deliberating reenlistment is unaffected by the past reenlistment behavior of others, and that his decision is also unaffected by past values of relative wage. Stated equivalently, this assumption is that the model contains no lagged values of reenlistment rate or relative wage. This is a simplified assumption; it is of course also possible to postulate and use a model which contains lagged values of relative wage. Now consider the effect of the war in Vietnam on initial enlistments or of general civilian unemployment on reenlistments in the Navy. These are examples of temporal factors that can be expected to have a significant effect on initial enlistments (in the first case) or reenlistments (in the second case) in the Navy. It seems reasonable, then, that a variable reflecting such temporal factors should be included in the model. Similarly, a potential reenlistee who is a member of a certain Rate and is in a certain pay grade may be affected by factors peculiar to his Rate and pay grade, as well as to factors unique to the year in which the reenlistment decision is made. In particular, since enlisted men in higher pay grades typically enjoy greater prestige and increased personal liberty than men in the lower pay grades, it may be hypothesized

that pay grade affects reenlistment rate in ways not expressible in terms of pecuniary compensation, as well as in its contribution to WM. It cannot, then, be fairly assumed that factors which depend on Rate, pay grade or year of eligibility to reenlist do not separately influence the reenlistment decision. As a consequence, variables representing the influence of such factors will be included in the model. [Such variables are, in general, unobservable or not quantifiable. Their inclusion in the model is a formalism for the sake of completeness.] These factors are the general non-pecuniary factors whose existence was previously hypothesized.

Note that nothing has yet been said about the influence of Mental Group on the reenlistment decision. It seems likely that personnel in different Mental Groups will reenlist at different rates. But designation of an individual as a member of a particular Mental Group is somewhat less accurate, hence less meaningful for statistical purposes, distinction than classification of personnel by Rate, pay grade or year of reenlistment. Additionally WM for a candidate reenlistee does not depend on his Mental Group. [An individual's expected WC may, however, depend on his Mental Group. If this is the case, it should emerge in comparison of results for separate Mental Groups.] Hence, Mental Group classification will not be used to define any of the variables of the model. Instead, the model to be constructed will be applied to all personnel in each of the Mental Groups separately. The results for the Mental Groups will then be statistically compared.

Now consider a potential reenlistee viewing his military and civilian pecuniary alternatives. WM depends (in a manner to be made explicit later) on his Rate and pay grade and on the year in which his current enlistment expires. But typically the potential reenlistee's view of

his civilian alternatives is limited; he has been efficiently isolated from the civilian world and civilian labor market by the requirements of his military service. And, typically, it is likely that he has been unable to go job-seeking in the geographical area of interest to him for civilian life. So it may be realistic to suppose that the alternative civilian wage perceived by the potential reenlistee can be considered to be the median wage (or average wage) of the civilian population working in his skill category (craftsman, mechanical, electrical, clerical and so on) in the year in which he is eligible to reenlist. This will be taken as a formal assumption: the civilian alternative wage perceived by an individual in a given Mental Group depends only upon his Rate and the year in which the reenlistment decision is made. [This assumption may be faulty in that the alternative civilian wage may also depend on the potential reenlistee's military pay grade. That is, an advanced rank status in the military may promise higher pay in the civilian economy, since it may be interpreted as being equivalent to advanced expertise.]

Since the assumption has been made that variables representing R, WM and WC are not lagged in the model, the time series data in R, WM and WC may be considered as another cross-section. Make, for the moment[3], the stronger assumption that the model contains no lagged variables at all. Then the time series, represented by year in which observations are made, may be considered as another cross-section. Let the

[3]This assumption is made for the sake of simplicity of representation. Later it will be seen that the assumption is not necessary; equivalent results are obtained if it is not made. At the same time it will be seen that the analagous assumption for the variables R, WM and WC may be weakened somewhat: identical results will be achieved even if the model contains lagged values of the variable WC.

subscripts i, j and t represent Rate, pay grade and year of reenlistment eligibility. Then the equations (c') and (d') can be represented in cross-section data as

(e) $\ln R_{ijt} = \alpha + \beta \ln WM_{ijt} + \gamma \ln WC_{it} + A_i + B_j + C_t + \varepsilon_{ijt}$,

(f) $\ln \left(\frac{R_{ijt}}{1-R_{ijt}}\right) = \alpha + \beta \ln WM_{ijt} + \gamma \ln WC_{it} + A_i + B_j + C_t + \varepsilon_{ijt}$,

where:

R_{ijt} is observed reenlistment rate for Rate i, pay grade j, year t;

WM_{ijt} is military wage for Rate i, page grade j, year t;

WC_{it} is alternative civilian wage for Rate i in year t;

The variables A_i, B_j, and C_t represent all factors which influence reenlistment in, respectively, Rate i, pay grade j, or year t uniquely; ε_{ijt} is the disturbance term for the observation of R_{ijt}. A_i, B_j, and C_t are the variables whose introduction into the model was promised earlier. Note that these variables are invariant over subscripts not included in their notational expression. For example, the factors represented by C_t depend only on the year of reenlistment, and are invariant over Rate and pay grade.

Note that a crucial assumption implicit in equations (e) and (f) is that the variables R_{ijt} and WM_{ijt} are the only variables in the model which are not invariant over at least one cross-sectional dimension (for convenience, the set of all Rates considered in the analysis will be referred to as a cross-sectional "dimension"; similarly for the set of all years and the set of all pay grades considered). Later work relies heavily on this assumption.

The models represented by equations (e) and (f) seem reasonably complete with the introduction of the variables A_i, B_j and C_t as "catch-all" categories to reflect all factors which influence reenlistment

depending on Rate, pay grade and year separately. But it is clear that the inclusion of these variables creates a problem: quantification of A_i, B_j and C_t is difficult if not impossible. Note that this problem is indissoluble. The influence of such variables as C_t and WC_{it} on the decision of a potential reenlistee is almost certainly non-trivial. Their effects cannot reasonably be ignored in any rational model of first-term reenlistment behavior. One possible approach to resolving this problem is to construct a model using dummy variables to represent Rate, pay grade and year. But in the face of 61 rates, nine pay grades and seven years this may yield results too minutely specialized to be interesting unless a certain amount of arbitrary aggregation (over Rates, pay grades and years) is done. In any case, an alternative procedure for ridding the models (e) and (f) of the effects of the variables A_i, B_j and C_t will be used here. Use of this procedure is also motivated by a desire to rid the model of the variable WC_{it}, the civilian alternative wage, the method of measurement of which may be subject to dispute.

To specify the procedure, consider:

(e) $$\ln R_{ijt} = \alpha + \beta \ln WM_{ijt} + \gamma \ln WC_{it} + A_i + B_j + C_t + \varepsilon_{ijt} ,$$

in "observed" data.

Taking the mean, for Rate i and pay grade j , over all years:

(e1) $$\ln R_{ij.} = \alpha + \beta \ln WM_{ij.} + \gamma \ln WC_{i.} + A_i + B_j + C_{.} + \varepsilon_{ij.}$$

Where, for example,

$$R_{ij.} = \frac{1}{T} \sum_{t=1}^{T} R_{ijt}$$

and

$$WC_{i.} = \frac{1}{T} \sum_{t=1}^{T} WC_{it} ,$$

for T = number of years considered in the data.

Taking the mean, for Rate i in year t, over all pay grades:

$$\text{(e2)} \quad \ln R_{i.t} = \alpha + \beta \ln WM_{i.t} + \gamma \ln WC_{it} + A_i + B_. + C_t + \varepsilon_{i.t} \;.$$

Taking the mean, for pay grade j in year t, over all Rates:

$$\text{(e3)} \quad \ln R_{.jt} = \alpha + \beta \ln WM_{.jt} + \gamma \ln WC_{.t} + A_. + B_j + C_t + \varepsilon_{.jt}$$

Taking the mean, for year t, over all Rates and pay grades:

$$\text{(e4)} \quad \ln R_{..t} = \alpha + \beta \ln WM_{..t} + \gamma \ln WC_{.t} + A_. + B_. + C_t + \varepsilon_{..t}$$

Taking the mean, for pay grade j, over all Rates and years:

$$\text{(e5)} \quad \ln R_{.j.} = \alpha + \beta \ln WM_{.j.} + \gamma \ln WC_{..} + A_. + B_j + C_. + \varepsilon_{.j.}$$

Taking the mean, for Rate i, over all pay grades and years:

$$\text{(e6)} \quad \ln R_{i..} = \alpha + \beta \ln WM_{i..} + \gamma \ln WC_{i.} + A_i + B_. + C_. + \varepsilon_{i..}$$

Taking the grand mean:

$$\text{(e7)} \quad \ln R_{...} = \alpha + \beta \ln WM_{...} + \gamma \ln WC_{..} + A_. + B_. + C_. + \varepsilon_{...} \;.$$

Adding and subtracting,

(e) - (e1) - (e2) - (e3) + (e4) + (e5) + (e6) - (e7)

yields the equation:

$$\ln R_{ijt} - \ln R_{ij.} - \ln R_{i.t} - \ln R_{.jt} + \ln R_{i..} + \ln R_{.j.} + \ln R_{..t} - \ln R_{...} =$$
$$\beta(\ln WM_{ijt} - \ln WM_{ij.} - \ln WM_{i.t} - \ln WM_{.jt} + \ln WM_{i..} + \ln WM_{.j.} + \ln WM_{..t} - \ln WM_{...}) +$$
$$\varepsilon_{ijt} - \varepsilon_{ij.} - \varepsilon_{i.t} - \varepsilon_{.jt} + \varepsilon_{i..} + \varepsilon_{.j.} + \varepsilon_{..t} - \varepsilon_{...} \;.$$

A similar result holds for the model represented by equation (f).

This is the form of the data that will be used in a linear regression to estimate the coefficient β. For want of more convenient terminology, data in the form above will often be referred to as "normalized

data", while the initial values of each ln R_{ijt} and ln WM_{ijt} will be called the "original data." In addition, the procedure of obtaining normalized data from the original data will sometimes be called "the model" when no ambiguity is possible. Some features of "the model" in this sense are investigated in Section IV.

Now note that any variable which has fewer than three subscripts in its notational expression disappears from the normalized form of the data. A little reflection shows that lagged values of any such variable are also purged in the normalized data. In particular this holds for the variable WC_{it}. As a consequence, it is only necessary, in order to obtain the identical equation in normalized data, to assure that the model contains no lagged values of R_{ijt} and WM_{ijt}.

The question of the nature of the normalized disturbance term:

$$\varepsilon_{ijt} - \varepsilon_{ij.} - \varepsilon_{i.t} - \varepsilon_{.jt} + \varepsilon_{i..} + \varepsilon_{.j.} + \varepsilon_{..t} - \varepsilon_{...}$$

will be taken up later.

D. THE CONSTRUCTION OF WM

The measurement of WM used here is that proposed by Burton C. Gray in [13].

As mentioned previously, pecuniary compensation for reenlisting can be viewed as consisting of two types of remuneration: the actual wage received by the reenlistee and the value placed by the reenlistee on the peripheral benefits of military service. A component of the actual wage received by a reenlistee that is unique to first-term reenlistments is the Variable Reenlistment Bonus (VRB). This bonus is a multiple of the reenlistee's annual base pay (which in turn depends upon pay grade) and varies from year to year and from Rate to Rate (depending on the valuation placed on reenlistments in a given Rate in a given year).

VRB has since fiscal year 1965 been the primary tool used to selectively (by Rate) influence reenlistments. Prior to FY 1965 all reenlistees received a reenlistment bonus that was a fixed multiple of annual base pay. Ideally, one should wish to evaluate the effect of VRB on first-term reenlistment behavior. But since the determination of a single parameter of interest is intended simply as being illustrative of the fundamental goal of this paper, an investigation of the consequences of using normalized data, this is not done. VRB enters the construction of WM as merely another component.

Now consider the future of a reenlistee. He can reasonably expect promotion to a higher pay grade within his next term of enlistment, with a concurrent increase in pay. This expectation obviously influences the reenlistment decision (for it can be supposed that fewer men would reenlist without the promise of probable advancement in rank), but in a way difficult to specify. The simplifying assumption is made that this promise of increased future pay offsets the lesser valuation of future dollars. That is, in considering the present value of WM, the potential reenlistee employs a discount rate of zero.

A final assumption, due to the nature of the available data base, is made. For want of other information, it is assumed that all reenlistments are made for an obligation of four years.

With the preceeding paragraphs in mind, it is possible to postulate the following construction:

$$WM = 4C + P\left[\frac{1 + VRB}{3} + 4(1 + K)\right] ,$$

where: for a potential reenlistee WM is the present value of military wage for a four-year reenlistment (at a zero discount rate), P is the reenlistee's annual base pay, VRB is the appropriate Variable Reenlistment Bonus multiple,

C is a constant representing the monetary valuation of the peripheral benefits of military service for a four-year reenlistment,

K is a dimensionless multiplicative constant representing the the valuation of those benefits associated with military service that can be expected to increase with annual base pay. K is intended to reflect such elements as tax advantages, allowances and commissary and exchange benefits, whose value increases as base pay increases.

This may be rewritten, for Rate i, pay grade j and year t, as:

$$WM_{ijt} = 4C + P_{ijt} \left[\frac{1 + VRB_{ijt}}{3} + 4\ (1 + K) \right] .$$

The construction of WM allows freedom for parameterization of the constants C and K. In order to get an idea of the sensitivity of the coefficient β to changes in assumed C and K, regression analyses are performed for various presumably reasonable values of these constants.

III. APPLICATION

A. PRELIMINARY

Consider the consequences of applying the natural logarithm transformation to the variables R_{ijt} and $R_{ijt}/(1-R_{ijt})$. These variables have respective ranges of values of $[0,1]$ and $[0,\infty)$, which under the natural logarithm transformation become $(-\infty,0]$ and $(-\infty,\infty)$. Thus this transformation avoids the awkward situation of having a finite range of values on the dependent variable (in the case of R_{ijt}) in a linear regression analysis. But there is a limitation associated with the use of the logarithmic transformation: under this transformation a reenlistment rate of zero is undefined. Hence in the model represented by equation (e) of the preceeding section, no observations of zero reenlistment rate can be allowed. Additionally, in the model represented by equation (f), a reenlistment rate equal to one must be disallowed, since this corresponds to an infinitely large value of the odds of reenlistment. Accordingly, since it is desirable to use the same data base for each of the models (e) and (f), any observations of reenlistment rate equal to zero or one will be discarded. This is not felt to restrict the analysis too severely since reenlistment rates of zero or one, the extreme values of the data, typically correspond to extraordinary classes of reenlistees. In particular, reenlistment rates of zero are most common in very low pay grades and reenlistment rates of one are usually observed in the highest pay grades. This suggests that a zero reenlistment rate can usually be associated with a class of men who show an unsuitability for military service, while a reenlistment rate equal to one can usually be associated with the class of men who

thrive in the military. Neither of these classes is particularly interesting for a study of general reenlistment behavior.

Now suppose that in models (e) and (f) the error terms ε_{ijt} are independent, identically distributed Normal random variables, each with mean zero and variance σ^2. Then the application of ordinary least squares procedures to estimate the coefficient β in the normalized form of model (e),

$$\ln R_{ijt} - \ln R_{ij.} - \ln R_{i.t} - \ln R_{.jt} + \ln R_{i..} + \ln R_{.j.} +$$

$$\ln R_{..t} - \ln R_{...} =$$

$$\beta(\ln WM_{ijt} - \ln WM_{ij.} - \ln WM_{i.t} - \ln WM_{.jt} + \ln WM_{i..} +$$

$$\ln WM_{.j.} + \ln WM_{..t} - \ln WM_{...}) +$$

$$\varepsilon_{ijt} - \varepsilon_{ij.} - \varepsilon_{i.t} - \varepsilon_{.jt} + \varepsilon_{i..} + \varepsilon_{.j.} + \varepsilon_{..t} - \varepsilon_{...}$$

yields an unbiased estimator for this coefficient. The same is true for ordinary least squares estimation of β in the normalized form of model (f). These assertions will be proved in Section IV, where it will also be shown that the above assumption about the distribution of the disturbance terms ε_{ijt} may be relaxed somewhat.

B. VALUES FOR PARAMETERIZED C AND K

Regression analyses were performed for each combination of the following selected values of the constants C and K:

C	K
500	
1000	0.10
1500	0.15
2000	0.20

It is felt that these selected values represent a range broad enough to include realistic possible values of the constants.

C. THE REGRESSION ANALYSES

In addition to estimating the coefficient β in the normalized forms of the models (e) and (f), it may be interesting (for comparative purposes) to estimate β in the equations:

$$\text{(g)} \qquad \ln R_{ijt} = \alpha + \beta \ln WM_{ijt} + \varepsilon_{ijt} ,$$

$$\text{(h)} \qquad \ln\left(\frac{R_{ijt}}{1 - R_{ijt}}\right) = \alpha + \beta \ln WM_{ijt} + \varepsilon_{ijt} ,$$

where it is assumed that the ε_{ijt}'s are independent, identically distributed Normal random variables with mean zero and variance σ^2. Note that these latter equations are truncated forms of the models (e) and (f): the variables WC_{it}, A_i, B_j, C_t are neglected.

Four selections for the value of the constant C and three choices for the constant K yield 12 different constructions of WM. Regression analyses are conducted for each of these constructions of WM, using models (e) (normalized), (f) (normalized), (g) and (h) for each of five Mental Groups. This produces 240 least squares estimations to be considered. Results for one construction of WM for models (e) (normalized), (f) (normalized), (g) and (h) and each of the five Mental Group classifications are looked at in detail in this section. Less detailed regression analysis results for the remaining 11 constructions of WM are given in Appendix A in tabular form.

Now consider Table I, which gives summary results for the construction of WM using C = 500 and K = 0.10. Denote Mental Groups I, II, upper III, lower III and IV as Mental Groups 1, 2, 3, 4 and 5 respectively.

Table I

Normalized Model (e)	B	SE	t	$\hat{\sigma}^2$	R	N
MG 1	1.17260	0.26011	4.49983	0.19601	0.1904	720
MG 2	1.76626	0.17863	9.90073	0.15014	0.3070	1259
MG 3	1.84425	0.21828	8.44902	0.17024	0.2956	996
MG 4	1.34492	0.20119	6.68474	0.15299	0.2629	805
MG 5	1.50907	0.28158	5.35927	0.13337	0.2601	530
Normalized Model (f)						
MG 1	1.87660	0.36445	5.14912	0.38339	0.2167	720
MG 2	2.72210	0.24978	10.89793	0.29433	0.3346	1259
MG 3	2.61042	0.30134	8.66258	0.32445	0.3025	996
MG 4	2.00364	0.28072	7.13740	0.29784	0.2793	805
MG 5	2.16256	0.39745	5.44106	0.26571	0.2638	530
Model (g)						
MG 1	1.36861	0.12644	10.82445	0.59642	0.3746	720
MG 2	1.91656	0.09547	20.07401	0.65793	0.4927	1259
MG 3	1.58111	0.11230	14.07977	0.62178	0.4078	996
MG 4	1.44961	0.12798	11.32667	0.62849	0.3712	805
MG 5	1.54090	0.14984	10.28386	0.46204	0.4085	530
Model (h)						
MG 1	1.85354	0.17451	10.62108	1.13624	0.3685	720
MG 2	2.70828	0.13598	19.91696	1.33460	0.4898	1259
MG 3	2.05608	0.15295	13.44309	1.15332	0.3922	996
MG 4	1.93526	0.17588	11.00301	1.18701	0.3620	805
MG 5	2.11862	0.21826	9.70676	0.98037	0.3891	530

Let B denote the estimate for β, SE represent the standard error of the estimate of β, t represent the computed t-statistic, $\hat{\sigma}^2$ be the estimate of the variance σ^2, R be the multiple correlation coefficient and N represent the number of observations of R_{ijt}. [It will be shown in Section IV that $\hat{\sigma}^2$ is an unbiased estimator for σ^2.] Note that the computed values of the t-statistic indicate that in each of the twenty least squares estimations of β represented in Table I the estimated coefficient is significantly different from zero. But also note that in comparing results for the normalized models (e) and (f) and the corresponding truncated non-normalized models (g) and (H), the following differences are consistently true for each Mental Group:

1. The values of computed t-statistic for models (g) and (h) are greater than the values for models (e) and (f).
2. The standard error of the esimate is less for models (g) and (h) than for models (e) and (f)
3. The multiple correlation coefficient R is greater for models (g) and (h) than for models (e) and (f).

These considerations might seem to indicate that models (g) and (h) fit the data better than the corresponding normalized forms of models (e) and (f). But in reality the results 1., 2., and 3. are not particularly surprising, since the computed value of t is directly proportional to, and the computed value of SE inversely proportional to, the square root of the sum of squared deviations from the mean of the explanatory variable, while $1-R^2$ is inversely proportional to the sum of squared deviations from the mean of the dependent variable. That is, for a single explanatory variable with observed values x_i, i = 1, ...n, and a dependent variable with observed values y_i, i = 1, ...n,

$$SE = \left[\frac{\sigma^2}{\sum_{1}^{n} (X_i - \overline{X})^2} \right]^{\frac{1}{2}}$$

$$t = \frac{B}{SE},$$

and:

$$R^2 = 1 - \frac{\sum_1^n (y_i - BX_i)^2}{\sum_1^n (y_i - \bar{y})^2},$$

where:

$$\bar{y} = \frac{1}{n} \sum_1^n y_i, \quad \bar{x} = \frac{1}{n} \sum_1^n x_i,$$

B is the estimated regression coefficient, and $\hat{\sigma}^2$ is the estimate of σ^2. Hence as the sum of squared deviations from the mean of both the explanatory variable and the independent variable decrease, it is to be anticipated that SE and R^2 will increase and the computed t-statistic will decrease. To see how this fact yields the results in comparisons 1., 2., and 3. above, consider the explanatory and dependent variables of the models (e) (normalized) and (g). Dropping for a moment the logarithm symbol, model (e) (normalized) has dependent variable;

$$R_{ijt} - R_{ij.} - R_{i.t} - R_{.jt} + R_{i..} + R_{.j.} + R_{..t} - R_{...}$$

and explanatory variable;

$$WM_{ijt} - WM_{ij.} - WM_{i.t} - WM_{.jt} + WM_{i..} + WM_{.j.} + WM_{..t} - WM_{...},$$

both of which have mean zero, while model (g) has dependent variable R_{ijt} and explanatory variable WM_{ijt}. Taking squared deviations from the mean for the variable R_{ijt}:

$$\sum_i \sum_j \sum_t (R_{ijt} - R_{...})^2 =$$

$$\sum_i \sum_j \sum_t (R_{ijt} - R_{ij.} - R_{i.t} - R_{.jt} + R_{i..} + R_{.j.} + R_{..t} - R_{...})^2 +$$

$$T \sum_i \sum_j (R_{ij.} - R_{...})^2 + J \sum_i \sum_t (R_{i.t} - R_{...})^2 +$$

$$I \sum_j \sum_t (R_{.jt} - R_{...})^2 + I\ J \sum_t (R_{...} - R_{..t})^2 +$$

$$I\ T \sum_j (R_{...} - R_{.j.})^2 + J\ T \sum_i (R_{...} - R_{i..})^2 \geq$$

$$\sum_i \sum_j \sum_t (R_{ijt} - R_{ij.} - R_{i.t} - R_{.jt} + R_{i..} + R_{.j.} + R_{..t} - R_{...})^2 \ ,$$

since all terms in the above equation are non-negative. But the term on the right hand side of this inequality is the sum of squared deviations from the mean of the dependent variable in the normalized form of model (e). A similar result holds in the comparison of the sum of squared deviations from the mean of the explanatory variables in models (e) (normalized) and (g). And a similar result holds in the comparison of the models (f) (normalized) and (h) as well. As a consequence, the results of comparisons 1., 2., and 3. are not unexpected.

Now consider the estimates of β presented in Table I. All estimates of the military wage elasticity of the odds of reenlistment and the probability of reenlistment exceed one. In fact, the estimates of the elasticity of R with respect to WM cluster loosly about a value of 1.5, while the estimates of the elasticity of $\frac{R}{1-R}$ with respect to WM have a median value of approximately 2. Since these estimates are based on a single choice for the construction of WM no great import will be assigned to them, except to note that they are not appreciably different from

estimates of these quantities obtained in other studies. For example, estimates of the WM elasticity of R in previous studies are generally confined to the range 0.8 to 3, with the bulk of the estimates lying in a range of values between 1 and 2. Note that in the normalized forms of models (e) and (f) the estimates of β for Mental Groups II and upper III seem to be appreciably higher than estimates of this coefficient for Mental Groups, I, lower III and IV (this apparent difference is not so marked for models (g) and (h); in any case models (g) and (h) are of interest here only for a comparison of results with the corresponding normalized forms of models (e) and (f), so that the former models will not be treated further). This result agrees very well with prior expectations: it indicates that personnel in the highest and lowest Mental Groups are less inclined toward reenlistment than men in the median Mental Groups. It can be argued that this result is reasonable since men in Mental Group I, who presumably possess greater intellectual ability, may find greater rewards and challenges in civilian life than in enlisted military service, while men in Mental Groups lower III and IV may often find themselves unable to compete for advancement successfully with men in higher Mental Groups, and may sometimes be unable to meet demands of competence placed on them by military service. For both the highest and lowest Mental Groups, then, enlisted military service may be viewed as limited in opportunity. To establish the validity of these initial observations it is desirable to determine if the estimates B contained in Table I do in fact estimate different coefficients β for different Mental Groups (that is, whether the same coefficient β applies for all Mental Groups or whether different coefficients β_i apply for different Mental Groups).

Toward this end a statistical test, in which the estimates B may be compared for each pair of Mental Groups in each of the models (e) (normalized) and (f) (normalized), is in order. Concentrate now on the normalized form of model (e). For the regression analysis of Mental Group i, i = 1, ...5, let $\hat{\sigma}_i^2$ be the estimate of σ^2, B_i be the estimate of β_i, and n_i be the number of observations. Since the estimated intercept for each least squares estimation using the normalized form of model (e) is zero, testing for the equality of the coefficients β_i is equivalent to testing for the equality of the appropriate regression lines. Now if Mental Groups i and j yield the same regression line in the normalized form of model (e), then $\hat{\sigma}_i^2$ and $\hat{\sigma}_j^2$ both estimate the same variance σ^2. And in this case,

$$\left[\frac{(I-1)(J-1)(T-1)\ n_i}{IJT} - 1\right] \frac{\hat{\sigma}_i^2}{\sigma^2} \sim \chi^2 \text{ with } \frac{(I-1)(J-1)(T-1)\ n_i}{IJT}$$

- 1 degrees of freedom,

and

$$\left[\frac{(I-1)(J-1)(T-1)\ n_j}{IJT} - 1\right] \frac{\hat{\sigma}_j^2}{\sigma^2} \sim \chi^2 \text{ with } \frac{(I-1)(J-1)(T-1)\ n_j}{IJT}$$

-1 degrees of freedom,

where these two Chi-squared random variables are independent since they are derived from two different (and assumed independent) populations of random variables. [See Section IV for the development of this assertion. Here I = 61 is the number of Rates, J = 9 is the number of pay grades and T = 7 is the number of years considered.)

Hence as the sum of two independent χ^2 random variables, the quantity:

$$\frac{1}{\sigma^2}\left\{\left[\frac{(I-1)(J-1)(T-1)}{IJT}\right]\left[n_i\,\hat{\sigma}_i^2 + n_j\,\hat{\sigma}_j^2\right] - (\hat{\sigma}_i^2 + \hat{\sigma}_j^2)\right\}$$

has χ^2 distribution with:

$$\frac{(I-1)(J-1)(T-1)}{IJT}\;(n_i + n_j) - 2$$

degrees of freedom. Now if Mental Groups i and j yield the same regression line then $\beta_i - \beta_j = 0$, in which case $B_i - B_j$ is Normally distributed with mean zero (since B_i and B_j are unbiased estimators of $\beta_i = \beta_j$) and variance:

$$\text{Var}\,(B_i - B_j) = \text{Var}\,(B_i) + \text{Var}\,(B_j) = \frac{\sigma^2}{\sum_{k=1}^{n_i}(X_k^i - \bar{X}^i)^2} + \frac{\sigma^2}{\sum_{k=1}^{n_j}(X_k^j - \bar{X}^j)^2}\;,$$

where for convenience X_k^m represents the k^{th} observation on the explanatory variable for the normalized form of model (e), applied to Mental Group m = i,j. Hence:

$$\frac{B_i - B_j}{\sigma\left[\dfrac{1}{\sum_{k=1}^{n_i}(X_k^i - \bar{X}^i)^2} + \dfrac{1}{\sum_{k=1}^{n_j}(X_k^j - \bar{X}^j)^2}\right]^{\frac{1}{2}}} \sim N(0,1)\quad .$$

As a consequence, under the composite hypothesis that $\hat{\sigma}_i^2$ and $\hat{\sigma}_j^2$ estimate the same parameter σ^2 and that $\beta_i = \beta_j$, the quantity:

$$\frac{\dfrac{(B_i - B_j)\left[\dfrac{(I-1)(J-1(T-1)}{IJT}(n_i + n_j) - 2\right]^{\frac{1}{2}}}{\sigma\left[\dfrac{1}{\sum\limits_{k=1}^{n_i}(X_k^i - \overline{X}^i)^2} + \dfrac{1}{\sum\limits_{k=1}^{n_j}(X_k^j - \overline{X}^j)^2}\right]^{\frac{1}{2}}}}{\dfrac{1}{\sigma}\left\{\left[\dfrac{(I-1)(J-1)(T-1)}{IJT}\right]\left[n_i\,\hat{\sigma}_i^2 + n_j\,\hat{\sigma}_j^2\right] - (\hat{\sigma}_i^2 + \hat{\sigma}_j^2)\right\}^{\frac{1}{2}}}$$

$$= \frac{(B_i - B_j)\left[\dfrac{(I-1)(J-1)(T-1)}{IJT}(n_i + n_j) - 2\right]^{\frac{1}{2}}}{\left[\dfrac{1}{\sum\limits_{k=1}^{n_i}(X_k^i - \overline{X}^i)^2} + \dfrac{1}{\sum\limits_{k=1}^{n_j}(X_k^j - \overline{X}^j)^2}\right]^{\frac{1}{2}}\left\{\left[\dfrac{(I-1)(J-1)(T-1)}{IJT}\right]\left[n_i\hat{\sigma}_i^2 + n_j\hat{\sigma}_j^2\right] - (\hat{\sigma}_i^2 + \hat{\sigma}_j^2)\right\}^{\frac{1}{2}}}$$

has t-distribution with:

$$\frac{(I-1)(J-1)(T-1)}{IJT}(n_i + n_j) - 2$$

degrees of freedom. Computing this statistic, for the normalized forms of models (e) and (f) separately, for each pair of Mental Groups, I, II, upper III, lower III and IV yields the results given in Table II.

Note that for very high level of significance, none of the coefficients B_i, B_j (for either model (e) or (f)) test significantly different from each other, so that for high chosen level of significance the composite null hypothesis that $\hat{\sigma}_i^2$ and $\hat{\sigma}_j^2$ both estimate common σ^2 and that $\beta_i = \beta_j$ cannot be rejected. But note that the magnitudes of the computed t-statistics for the most part give credence (especially in the normalized form of model (f)) to the observations that prompted this test: the sets

TABLE II

i,j	t(R)	$t\left(\frac{R}{1-R}\right)$	df
1,2	1.95	1.98	1481
1,3	2.00	1.57	1284
1,4	0.53	0.28	1141
1,5	0.84	0.51	935
2,3	0.28	0.28	1688
2,4	1.57	1.91	1545
2,5	0.75	1.17	1339
3,4	1.68	1.47	1348
3,5	0.91	0.87	1142
4,5	0.47	0.32	999

(i,j) refers to the comparison of coefficients for Mental Groups i and j.

t(R) is the computed t-statistic for the normalized form of model (e).

$t\left(\frac{R}{1-R}\right)$ is the computed t-statistic for the normalized form of model (f)

df is the appropriate degrees of freedom,

$$\frac{(I-1)(J-1)(T-1)}{IJT}(n_i + n_j) - 2$$

of the t-distribution to the nearest integer.

$\{\beta 2, \beta 3\}$ and $\{\beta 1, \beta 4, \beta 5\}$ of coefficients may be accepted as being different from each other, and the coefficients within each of these sets may be accepted as being the same, at an appreciably higher level of significance than any other partition of the set $\{\beta 1, \beta 2, \beta 3, \beta 4, \beta 5\}$.

IV. FEATURES OF THE MODEL

A. A MORE GENERAL CROSS-SECTIONAL MODEL

Consider a slightly more general form of the reenlistment model. For simplicity in the derivation of results, suppose that three cross-sectional dimensions are involved. Let $Y = X\beta + Z\Omega + \varepsilon$, where Y is an n-vector of observations on the dependent variable, X is an n x k matrix of observations on k explanatory variables, each of which varies over all cross-sectional dimensions (as did WM_{ijt} in the reenlistment model), β is a k-vector of coefficients corresponding to the variables X, Z is an n x m matrix of observations on m explanatory variables, each of which varies over at most two cross-sectional dimensions (as did WC_{it} and C_t for example, in the reenlistment model), Ω is an m-vector of coefficients corresponding to the variables in Z. Then it is evident that, if the observations are "normalized" as in the reenlistment model, the variables Z will disappear from the normalized data. So the model in normalized form becomes $Y_n = X_n \beta + \varepsilon_n$, where, for example, the typical element of ε_n is:

$$\varepsilon_{ijt} - \varepsilon_{ij.} - \varepsilon_{i.t} - \varepsilon_{.jt} + \varepsilon_{i..} + \varepsilon_{.j.} + \varepsilon_{..t} - \varepsilon_{...} \quad .$$

The procedure of normalizing data in this manner, then, is advantageous when it is desirable to rid the model of one or more of the variables in Z. For example, theoretical or practical considerations may dictate that a variable in Z be included in the model, but this variable may in practice turn out to be unobserved (as was WC_{it} in the reenlistment model) or even unobservable (as was C_t in the reenlistment model). An obvious disadvantage is that all the variables Z disappear in the

normalized data, so that none of the coefficients in Ω can be estimated using normalized observations. The normalization procedure can also be used to advantage to rid the model of disturbance terms of a certain form. This is the subject of a later part of this section.

B. A NECESSARY IDEMPOTENT MATRIX

Consider the set of all ordered triples of three indices, i, j, t:

$$\{(i,j,t): i = 1, \ldots I,\ j = 1, \ldots J,\ t = 1, \ldots T\}$$

There are IJT unique such ordered triples. Construct an IJT x IJT matrix, the rows and columns of which are each indexed with one of the ordered triples (i, j, t), as follows: If the k^{th} row of this matrix, call it V, is indexed with (i_1, j_1, t_1), then the k^{th} column of V is also indexed with (i_1, j_1, t_1). For the row of V indexed with (i_1, j_1, t_1) and the column of V indexed with (i_2, j_2, t_2), let the corresponding element of V be equal to

$-(J-1)(T-1)/IJT$	if $i_1 \neq i_2,\ j_1 = j_2,\ t_1 = t_2$
$-(I-1)(T-1)/IJT$	if $i_1 = i_2,\ j_1 \neq j_2,\ t_1 = t_2$
$-(I-1)(J-1)/IJT$	if $i_1 = i_2,\ j_1 = j_2,\ t_1 \neq t_2$
$(T-1)/IJT$	if $i_1 \neq i_2,\ j_1 \neq j_2,\ t_1 = t_2$
$(J-1)/IJT$	if $i_1 \neq i_2,\ j_1 = j_2,\ t_1 \neq t_2$
$(I-1)/IJT$	if $i_1 = i_2,\ j_1 \neq j_2,\ t_1 \neq t_2$
$-I/IJT$	if $i_1 \neq i_2,\ j_1 \neq j_2,\ t_1 \neq t_2$
$(I-1)(J-1)(T-1)/IJT$	if $i_1 = i_2,\ j_1 = j_2,\ t_1 = t_2$.

Within each row and each column of V, then, there are (I-1) elements of the first type, (J-1) elements of the second type, (T-1) elements of the third type, (I-1)(J-1) elements of the fourth type, (I-1)(T-1) elements

of the fifth type, (J-1)(T-1) elements of the sixth type, (I-1)(J-1)(T-1) elements of the seventh type, and one element of the eighth type.

From the symmetrical construction of V, it is apparent that V is symmetric. That V is singular is also apparent, since VN = 0, where N is the n-vector with unit elements (that is, the sum of the elements in each row and each column of V is equal to zero) and n = IJT.

And it can be shown that V is indempotent as well: Let X be an arbitrary n x r matrix. For convenience of representation, let the m^{th} row of X be indexed with the same ordered triple (i, j, t) as the m^{th} row of V. Consider the k^{th} column of VX. If X^k is the k^{th} column of X, then VX^k is the k^{th} column of VX, so that without loss of generality it is necessary only to consider the case r = 1 in order to establish the form of VX. Let X_{i_1}, j_1 t_1 be a typical element of the n x 1 matrix X. The the $(i_1, j_1, t_1)^{st}$ element of VX is of the form:

$$\frac{1}{IJT}\Bigg[(I-1)(J-1)(T-1)\, X_{i_1j_1t_1} - (J-1)(T-1) \sum_{\substack{i=1\\ i\neq i_1}}^{I} X_{ij_1t_1} -$$

$$(I-1)(T-1) \sum_{\substack{j=1\\ j\neq j_1}}^{J} X_{i_1jt_1} - (I-1)(J-1) \sum_{\substack{t=1\\ t\neq t_1}}^{T} X_{i_1j_1t} +$$

$$(T-1) \sum_{\substack{i=1\\ i\neq i_1}}^{I} \sum_{\substack{j=1\\ j\neq j_1}}^{J} X_{ijt_1} + (J-1) \sum_{\substack{i=1\\ i\neq i_1}}^{I} \sum_{\substack{t=1\\ t\neq t_1}}^{T} X_{ij_1t} + (I-1) \sum_{\substack{j=1\\ j\neq j_1}}^{J} \sum_{\substack{t=1\\ t\neq t_1}}^{T} X_{i_1jt}$$

$$- \sum_{\substack{i=1\\ i\neq i_1}}^{I} \sum_{\substack{j=1\\ j\neq j_1}}^{J} \sum_{\substack{t=1\\ t\neq t_1}}^{T} X_{ijt} \Bigg] =$$

$$\frac{1}{IJT}\Bigg[IJT\, X_{i_1 j_1 t_1} - JT \sum_{i=1}^{I} X_{i j_1 t_1} - IT \sum_{j=1}^{J} X_{i_1 j t_1} -$$

$$IJ \sum_{t=1}^{T} X_{i_1 j_1 t} + I \sum_{j=1}^{J} \sum_{t=1}^{T} X_{i_1 j t} + J \sum_{i=1}^{I} \sum_{t=1}^{T} X_{i j_1 t} +$$

$$T \sum_{i=1}^{I} \sum_{j=1}^{J} X_{i j t_1} - \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} X_{ijt} \Bigg] =$$

$$X_{i_1 j_1 t_1} - \frac{1}{I} \sum_{i} X_{i j_1 t_1} - \frac{1}{J} \sum_{j} X_{i_1 j t_1} - \frac{1}{T} \sum_{t} X_{i_1 j_1 t} +$$

$$\frac{1}{JT} \sum_{j} \sum_{t} X_{i_1 j t} + \frac{1}{IT} \sum_{i} \sum_{t} X_{i j_1 t} + \frac{1}{IJ} \sum_{i} \sum_{j} X_{i j t_1} -$$

$$\frac{1}{IJT} \sum_{i} \sum_{j} \sum_{t} X_{ijt} =$$

$$X_{i_1 j_1 t_1} - X_{.j_1 t_1} - X_{i_1 . t_1} - X_{i_1 j_1 .} + X_{i_1 ..} + X_{.j_1 .} + X_{..t_1} - X_{...} \quad .$$

That is, the matrix V is the linear transformation which reduces the original data X to data in the normalized form.

Now consider the matrix product VVX. Let $X'_{i_1 j_1 t_1}$ be the typical element of VVX, and let $X^o_{i_1 j_1 t_1}$ represent the typical element of VX:

$$X^o_{i_1 j_1 t_1} = X_{i_1 j_1 t_1} - X_{.j_1 t_1} - X_{i_1 . t_1} - X_{i_1 j_1 .} + X_{i_1 ..} +$$

$$X_{.j_1 .} + X_{..t_1} - X_{...} \quad .$$

Analagous to the above derivation,

$$X'_{i_1j_1t_1} = X^o_{i_1j_1\cdot} - X^o_{\cdot j_1t_1} - X^o_{i_1\cdot t_1} - X^o_{i_1j_1t} + X^o_{i_1\cdot\cdot} +$$

$$X^o_{\cdot j_1\cdot} + X^o_{\cdot\cdot t_1} - X^o_{\cdot\cdot\cdot} \quad .$$

But:

$$X^o_{\cdot j_1t_1} = X_{\cdot j_1t_1} - X_{\cdot j_1t_1} - X_{\cdot\cdot t_1} - X_{\cdot j_1\cdot} + X_{\cdot\cdot\cdot} +$$

$$X_{\cdot j_1\cdot} + X_{\cdot\cdot t_1} - X_{\cdot\cdot\cdot} = 0$$

$$X^o_{i_1\cdot t_1} = X_{i_1\cdot t_1} - X_{\cdot\cdot t_1} - X_{i_1\cdot t_1} - X_{i_1\cdot\cdot} + X_{i_1\cdot\cdot} +$$

$$X_{\cdot\cdot\cdot} + X_{\cdot\cdot t_1} - X_{\cdot\cdot\cdot} = 0$$

$$X^o_{i_1j_1\cdot} = X_{i_1j_1\cdot} - X_{\cdot j_1\cdot} - X_{i_1\cdot\cdot} - X_{i_1j_1\cdot} + X_{i_1\cdot\cdot} +$$

$$X_{\cdot j_1\cdot} + X_{\cdot\cdot\cdot} - X_{\cdot\cdot\cdot} = 0$$

$$X^o_{i_1\cdot\cdot} = X_{i_1\cdot\cdot} - X_{\cdot\cdot\cdot} - X_{i_1\cdot\cdot} - X_{i_1\cdot\cdot} + X_{i_1\cdot\cdot} +$$

$$X_{\cdot\cdot\cdot} + X_{\cdot\cdot\cdot} - X_{\cdot\cdot\cdot} = 0$$

$$X^o_{\cdot j_1\cdot} = X_{\cdot j_1\cdot} - X_{\cdot j_1\cdot} - X_{\cdot\cdot\cdot} - X_{\cdot j_1\cdot} + X_{\cdot\cdot\cdot} +$$

$$X_{\cdot j_1\cdot} + X_{\cdot\cdot\cdot} - X_{\cdot\cdot\cdot} = 0$$

$$X^o_{\cdot\cdot t_1} = X_{\cdot\cdot t_1} - X_{\cdot\cdot t_1} - X_{\cdot\cdot t_1} - X_{\cdot\cdot\cdot} + X_{\cdot\cdot\cdot} + X_{\cdot\cdot\cdot} + X_{\cdot\cdot t_1} - X_{\cdot\cdot\cdot} = 0$$

$$X^o_{\cdot\cdot\cdot} = X_{\cdot\cdot\cdot} - X_{\cdot\cdot\cdot} - X_{\cdot\cdot\cdot} - X_{\cdot\cdot\cdot} + X_{\cdot\cdot\cdot} + X_{\cdot\cdot\cdot} + X_{\cdot\cdot\cdot} - X_{\cdot\cdot\cdot} = 0.$$

So that: $$X'_{i_1 j_1 t_1} = X^o_{i_1 j_1 t_1} .$$

In particular this holds for the vector X^o_k which has zeros in each element except the k^{th}, which is equal to one. That is $VVX^o_k = VX^o_k$. But VVX^o_k is the k^{th} column of VV, and VX^o_k is the k^{th} column of V. This holds for each k = 1 ... IJT, so that each column of VV is equal to the corresponding column of V. Hence VV = V, so that V is, by definition, idempotent.

The idempotency of V can be seen equivalently as follows. Consider the equation $VX = \lambda X$, where λ is any eigenvalue of V, and X is a corresponding eigenvector $(X \neq \bar{0})$ by assumption). Pre-multiplying both sides of this equation by V yields:

$$VVX = V\lambda X = \lambda VX = \lambda^2 X.$$

But $VVX = VX = \lambda X$, so that $\lambda X = \lambda^2 X$. So either $\lambda = 0$ or it is possible to divide by λ to get $X = \lambda X$. Or $X'X = X'\lambda X = \lambda X'X$, where $X'X$ is a strictly positive scalar. Hence if $\lambda \neq 0$, then $\lambda = X'X/X'X=1$. That is, for the matrix V, all eigenvalues are equal to 1 or to 0. Now the claim that V is indempotent can be made, since a sufficient condition for a symmetric matrix to be indempotent is that each of its non-zero eigenvalues be equal to unity.

Now since V is indempotent, its rank is equal to its trace. And the trace of V is equal to the sum of its diagonal elements. That is, tr(V) $= IJT\,[(I-1)(J-1)(T-1)/IJT] = (I-1)(J-1)(T-1)$. Hence the rank of V is $(I-1)(J-1)(T-1)$.

C. ORDINARY LEAST SQUARES ESTIMATION UNDER THE TRANSFORMATION V

Consider once again the model described in Section A, $Y = X\beta + Z\Omega + \varepsilon$ where Y, X, β, Z, Ω and ε are as defined there. Recall that the number of

cross-sectional dimensions involved was assumed, for purely illustrative purposes, to be three. Suppose that one cross-sectional dimension is resolved into I categories, the second dimension into J categories, and the third dimension into T categories. Then there are $n = IJT$ observations in Y, and to each observation in Y there can be assigned a unique ordered triple (i,j,t) which represents the appropriate category of each of the cross-sectional dimensions for that observation in Y. Obviously this same ordered triple is assigned to the corresponding observations of the variables in X and in Z, as well as to the corresponding element of ε. Now suppose that the matrix V has been constructed so that the index of the p^{th} row of V is equal to the index of the p^{th} observation in Y. Then pre-multiplying the above equation by V yields $VY = V X \beta + VZ\Omega + V\varepsilon$, where $VZ = \overline{0}$ and $VY \neq \overline{0} \neq VX$ since by assumption the dependent variable whose observations are represented by Y and the k explanatory variables whose observations are represented by X vary over all cross-sectional dimensions, while the variables whose observations are represented by Z vary over at most two cross-sectional dimensions. So the equation becomes $VY = V X \beta + V\varepsilon$.

Note that the above property provides a concise operational definition of the phrase "varies over all cross-sectional dimensions." A non-stochastic variable whose vector of observations, over all possible categories of the cross-sectional dimensions, is given by W may be said to vary over all cross-sectional dimensions if $VW \neq \overline{0}$. It will be shown in a later section that the element of VW which is indexed by (i,j,t) may be interpreted as the three-way interaction of the i^{th} category of one cross-sectional dimension, the j^{th} category of the second dimension, and the t^{th} category of the third dimension. Similarly, for a stochastic

variable whose vector of observations is given by W, the element of VW indexed by (i,j,t) may be interpreted as the sample estimate of this three-way interaction term.

Now in order to discuss the ordinary least squares estimator of β in the equation $VY = V X \beta + V\varepsilon$ it is necessary to consider the rank of VX. Suppose that $r(X) = k$ ($k < n$), so that $(X' X)^{-1}$ exists. If it were the case that $r(X) < k$, then the coefficient vector β in the equation $Y = X\beta + Z\Omega + \varepsilon$ would be inestimable in the original data, since a necessary condition for the ordinary least squares estimators, in the original data, of β and Ω to exist is that both $X' X$ and $Z' Z$ are nonsingular. That is, these estimators in the original data, in partitioned matrix form,

$$\begin{bmatrix} \hat{\beta} \\ \hat{\Omega} \end{bmatrix} = \begin{bmatrix} X' X & X'Z \\ Z' X & Z'Z \end{bmatrix}^{-1} \begin{bmatrix} X' Y \\ Z' Y \end{bmatrix} ,$$

exist only if $(X' X)^{-1}$ and $(Z' Z)^{-1}$ exist. So the assumption that $r(X) = k$ is no more restrictive in the ordinary least squares estimation of β using data in the form VY, VX than it was in the ordinary least squares estimation of β using the original data Y, X. [Note that this discussion applies only to estimation of the <u>originally specified</u> k-vector β of coefficients. It may of course be possible, even if $r(X) < k$, to estimate a linear combination of some of the coefficients in β. But this is not the goal here.] Now since $r(V) = (I-1)(J-1)(T-1)$, a necessary condition for $(VX)' (VX) = X' VX$ to be nonsingular is that $r(VX) = K$. So a necessary condition is that $K \leq (I-1)(J-1)(T-1)$. That is, that the matrix X represents observations on at most (I-1)(J-1)(T-1) explanatory variables. Consequently, in all discussion hereafter, the requirement that $K \leq (I-1)(J-1)(T-1) < IJT = n$ will be made.

Additionally, the requirement that $r(VX) = k$ means that the columns of VX must be linearly independent. But these are simply the vectors which represent the three-way interaction terms for each variable in X. This is a new restriction, not encountered when basing estimators upon the original observations. It may turn out, in some cases, to prohibit application of V in the model. It is certainly not prohibitive when X represents observations on only one explanatory variable (as was the case for WM_{ijt} in the reenlistment model). It may be worth noting that the circumstances in which $r(VX) < k$ can be stated more succinctly: $r(VX) < k$ if and only if some linear combination of the vectors in X is in the null space of the transformation V.

If $r(VX) = k$, then $X'VX$ is nonsingular, and the ordinary least squares estimator, under the transformation V, for β in $Y = X\beta + Z\Omega + \epsilon$ is $B = ((VX)'(VX))^{-1} (VX)'(VY) = (X'VX)^{-1}X'VY$.

A definition of terms should now be made. B, in the equation above, has been called an estimator for β under the transformation V. But it is clear that if B is linear in VY, then it is also linear in Y. That is, for any linear transformation A, $A(VY) = CY$ for some linear transformation C. The reason for this apparently unnecessary terminology is that this estimator B is the best linear unbiased estimator for β (it will be shown later) among all those unbiased estimators for β that are linear in VY. [The definition of "Best" used throughout this paper is that employed in the Gauss-Markov theorem. An estimator $\hat{\beta}$ for β in the equation $Y = X\beta + Z\Omega + \varepsilon$ is best linear unbiased if it is linear in Y, if it is unbiased and if any other estimator of β which is also linear in Y and unbiased has a covariance matrix which exceeds that of $\hat{\beta}$ by a positive semidefinite matrix.] That B can be the best unbiased estimator linear in VY and

yet not be the best unbiased estimator linear in Y is clear, since the transformation V is not invertible. That is, no linear transformation on VW can reproduce W. If this were possible, then there would exist some matrix A such that AVW = W for all W. But since V is singular, there must exist a vector W_1 (not identically zero) such that $VW_1 = 0$. Specifically, $W_1 = N$ can be the n-vector with unit elements. So $AVW_1 = A\,\overline{0} = \overline{0} \neq W_1$. [Equivalently, V is not isomorphic. It has null space $S = \{W:VW = 0\}$. Consequently, V maps all vectors of the form Z + cN, where c is a scalar and N the n-vector of unit elements, into the vector VZ.] In addition to being the best linear unbiased estimator for β under the transformation V, B is in many cases the best linear unbiased estimator for β as well. This is the subject of the next part of this section.

D. POOLED TIME SERIES AND CROSS-SECTION DATA: EFFECT OF THE COMPOSITION OF THE DISTURBANCE TERM ON THE MODEL

The ordinary least squares estimator for β, under V, shows a degree of insensitivity in its quality of "best linear unbiasedness under V" to the composition of the disturbance term of the model. The type of composition of the disturbance term for which the property of best linear unbiasedness, under V, of B is invariant is considered here.

It may happen that in a regression model involving time series and cross-section data the disturbance term for an observation is composed of effects due to the cross-section, an effect due to the time series, and a series of remainder terms (that is, components of the disturbance term which are due to the joint effects of cross-section and time series).[4] For example, the disturbance term ε_{ijt} for economic entity i,

---

[4] As postulated by, for example, Kuh [11] and Chetty [12].

subject to factor j at time t may be given by:

1. $\varepsilon_{ijt} = \eta_{ijt} + \alpha_i + \gamma_j + \delta_t + \lambda_{ij} + \omega_{it} + \pi_{jt}$, where

2. $E(\eta_{ijt}) = 0$, $i = 1, \ldots I$, $j = 1, \ldots J$, $t = 1, \ldots T$

3. Var $(\eta_{ijt}) = \sigma^2$ for all i, j, t

4. η_{ijt}'s are independent, Normally distributed random variables

5. No statements can be made concerning the distributions of the random variables α_i, γ_j, δ_t, λ_{ij}, ω_{it}, π_{jt}.

6. No statements can be made concerning the independence, or correlations, of the random variables η_{ijt}, α_i, γ_j, δ_t, λ_{ij}, ω_{it}, π_{jt} (other than as in 4. above)

7. Each random variable is invariant over any dimension not included as a subscript in its notational expression.

The disturbance structure hypothesized here is central to later work. For ease of reference, call the error structure formally assumed by statements 1. through 7. above "disturbance structure (A)."

Under the specifications of disturbance structure (A), no conclusion can be made about the form of E (ε) or Var (ε). Consequently no claims can be made regarding the unbiasedness of the ordinary least squares estimator for β in the original data. And the generalized least square estimator is unknown, since Var (ε) is unknown. But for $\varepsilon = [\varepsilon_{ijt}]$ and $\eta = [\eta_{ijt}]$ as specified above, $V\varepsilon = V\eta$, since $V\alpha = V\gamma = V\delta = V\lambda = V\omega = V\pi = \overline{0}$. Hence under disturbance structure (A) the ordinary least squares estimator, under V, for β is unbiased:

$$B = (X'VX)^{-1} X'VY$$

$$E(B) = E[(X'VX)^{-1}X'VY] = E[(X'VX)^{-1} X'(V X \beta + V\varepsilon)] =$$

$$\beta + (X'VX)^{-1} X'V\, E(\eta) = \beta + \overline{0} = \beta .$$

And the variance of B is given by:

$$\begin{aligned} Var(B) = \; & E[(B-\beta)(B-\beta)'] = \\ & E[(X'VX)^{-1} X'V\varepsilon\varepsilon' VX(X'VX)^{-1}] = \\ & E[(X'VX)^{-1} X'V\eta\eta' VX(X'VX)^{-1}] = \\ & (X'VX)^{-1} X'VE(\eta\eta') VX(X'VX)^{-1} = \\ & \sigma^2 (X'VX)^{-1} X'VIVX(X'VX)^{-1} = \\ & \sigma^2 (X'VX)^{-1} X'VX(X'VX)^{-1} = \sigma^2 (X'VX)^{-1}, \end{aligned}$$

since $E(\eta\eta') = \sigma^2 I$, and since V is idempotent.

It is now possible to show that, under disturbance structure (A), B is the best linear unbiased estimator, under V, for β. But it is first worthwhile to show that any linear transformation which has null space identical to that of V (that is, any linear transformation which maps precisely the same vectors onto the null vector) is itself a linear transformation, under a nonsingular matrix, of V. That is, that the matrix V which removes the stochastic variables α_i, γ_j, δ_t, λ_{ij}, ω_{it} and π_{jt} from the disturbance term, and under which the image of a vector $[\eta_{ijt}]$ which varies over all dimensions is non-null, is unique up to a nonsingular linear transformation C. Suppose there exists another linear transformation, say A, such that $A\varepsilon = A\eta$ ($A\alpha = A\gamma = A\delta = A\lambda = A\omega = A\pi = \overline{0}$), for all n-vectors ε. Then since A and V are to have the same null space, $AX = \overline{0}$ if and only if $VX = \overline{0}$. In particular, this must hold for the vector VX: $AVX = \overline{0}$, if and only if $VVX = VX = \overline{0}$. An equivalent statement is that the system $A(VX) = 0$ has only the trivial solution $VX = \overline{0}$. Hence either A is nonsingular or $A = CV$ for nonsingular C (in the latter case

AVX = CVVX = CVX and AX = CVX). But if A is nonsingular, then AX = $\bar{0}$ implies that X = $\bar{0}$. So, for nonsingular A, A and V could not have the same null space. Hence A = CV, for nonsingular C.

Now since CV, for nonsingular C, is the only linear transformation which removes stochastic variables α_i, β_j, γ_t, λ_{ij}, ω_{it}, π_{jt} from the model, any other unbiased estimator of β must be linear in CVY, hence in VY. Consider any other such estimator, say AVY, where A is a k x n matrix independent of Y.

Let $\quad D = A - (X'VX)^{-1}X'V.$

Then $\quad AVY = [D + (X'VX)^{-1}X'] VY =$

$$[D + (X'VX)^{-1}X'] [V X \beta + V\varepsilon] =$$

$$[DVX + I] \beta + [D + (X'VX)^{-1}X'] V\varepsilon.$$

But $\quad E(AVY) = (DVX + I) \beta + [D + (X'VX)^{-1}X'] E(V\varepsilon) =$

$$(DVX + I) \beta + [D + (X'VX)^{-1}X'V] E(\eta) =$$

$$(DVX + I) \beta.$$

So in order for AVY to be unbiased, it is necessary that DVX = $\bar{0}$. So the estimator becomes $\beta + [D + (X'VX)^{-1}X'] V\varepsilon$. The corresponding sampling error is $[D + (X'VX)^{-1}X'] V \varepsilon$, and the covariance matrix is:

$$E[\{DV + (X'VX)^{-1}X'V \}V\varepsilon\varepsilon'V\{VD' + VX(X'VX)^{-1} \}] =$$

$$[DV + (X'VX)^{-1}X'V] E(\eta\eta') [VD' + VX(X'VX)^{-1}] =$$

$$\sigma^2 [DV + (X'VX)^{-1}X'V] [VD' + VX(X'VX)^{-1}] =$$

$$\sigma^2 [DVD' + DVX(X'VX)^{-1} + (X'VX)^{-1}X'VD' + (X'VX)^{-1}X'VX(X'VX)^{-1}] =$$

$$\sigma^2 [DVD' + (X'VX)^{-1}].$$

So the covariance matrix of the estimator AVY exceeds the covariance matrix of $B = (X'VX)^{-1}X'VY$ by DVD', a positive semidefinite matrix. Hence B is the best linear unbiased estimator under V in the sense that its

covariance matrix is exceeded, by a positive semidefinite matrix, by the covariance matrix of any other linear unbiased estimator of β under V.

And, since B is the best linear unbiased estimator for β under V, and since only those estimators linear in VY can claim to be unbiased, the estimator B is the best linear unbiased estimator for β under disturbance structure (A).

The discussion of the hypothesized error structure has been couched in terms of pooled cross-section and time series data. But in any regression model involving cross-sectional data (no matter what the nature of the cross-sectional dimensions) it is clear that, if no more specific statement about the error structure can be made than that disturbance structure (A) applies, then $B = (X'VX)^{-1}X'VY$ is the best linear unbiased estimator for β.

E. AN UNBIASED ESTIMATOR FOR σ^2

Assume disturbance structure (A) from the preceeding section applies. The purpose of this section is to show that:

$$S^2 = e'e/[(I-1)(J-1)(T-1)-k]$$

is an unbiased estimator for σ^2 in

$$\text{Var}\ (B) = (X'VX)^{-1}\ \sigma^2\ .$$

Consider the estimator $B = (X'VX)^{-1}X'VY$ of β in the model:

$$Y = X\beta + Z\Omega + \varepsilon,\ VY = V\ X\ \beta\ + V\varepsilon.$$

The residual vector is $e = VY - VXB = VY - VX(X'VX)^{-1}X'VY = [V-VX(X'VX)^{-1}X'V]\ Y$. Let $M = V-VX(X'VX)^{-1}X'V$. Then $e = MY$ and M is an idempotent matrix with trace $(I-1)(J-1)(T-1)-k$. To see the idempotency of M:

$$MM = [V-VX(X'VX)^{-1}X'V]\,[V-VX(X'VX)^{-1}X'V] =$$

$$V-VX(X'VX)^{-1}X'V - VX(X'VX)^{-1}X'V + VX(X'VX)^{-1}X'VX(X'VX)^{-1}X'V =$$

$$V-VX(X'VX)^{-1}X'V - VX(X'VX)^{-1}X'V + VX(X'VX)^{-1}X'V =$$

$$V-VX(X'VX)^{-1}X'V = M.$$

To see $tr\ (M) = (I-1)(J-1)(T-1) - k$:

Since the trace of the difference of two matrices is equal to the difference of the traces,

$$tr(M) = tr(V) - tr(VX(X'VX)^{-1}X'V) =$$

$$(I-1)(J-1)(T-1) - tr(VX(X'VX)^{-1}X'V) .$$

And since for two matrices A, B, of compatible order, $tr\ (AB) = tr(BA)$,

$$tr(M) = (I=1)(J-1)(T-1) - tr((X'VX)^{-1}X'VX) =$$

$$(I-1)(J-1)(T-1) - tr(I_k) = (I-1)(J-1)(T-1) - k,$$

where I_k is the identity matrix of order k.

The residual vector may also be written, $e = MY = MVY = MV\ (X\beta + \varepsilon)$ $= MV\varepsilon$, since $MVX = VX - VX(X'VX)^{-1}X'VX = VX - VX = \bar{0}$.

So the error sum of squares is $e'e = \varepsilon'VM'MV\varepsilon = \varepsilon'VMV\varepsilon = \eta'VMV\eta =$ $\eta'M\ \eta$, since $V\varepsilon = V\eta$. And, since $\eta'M\eta$ is scalar, it is equal to its own trace: $e'e = tr(\eta'M\eta)$. And since $tr(AB) = tr(BA)$, $e'e = tr(\eta'M\eta) =$ $tr(M\eta\eta')$. And since the trace of a square matrix is a linear operation on the matrix, the expected value of the trace is equal to the trace of the expected value:

$$E(e'e) = E[tr(M\eta\eta')] = tr[E(M\eta\eta')] = tr[ME(\eta\eta')] = tr[\sigma^2 MI] =$$

$$tr[\sigma^2 M] = \sigma^2\ tr(M),$$

since for a scalar k and matrix A, $tr(kA) = k\ tr(A)$.

So $E(e'e) = \sigma^2 [(I-1)(J-1)(T-1) - k]$.

So, for $S^2 = e'e/[(I-1)(J-1)(T-1) - k]$, $E(S^2) = \sigma^2$.

F. THE JOINT DISTRIBUTION OF B AND S^2

A theorem with application in statistical analysis may be expressed as follows: If A is an idempotent matrix and μ is an n-variate Nromal random variable from a $N(0,\sigma^2)$ distribution, then the quadratic form $\frac{1}{\sigma^2}\mu'A\mu$ is distributed χ^2 with q degrees of freedom , where q = tr(A) = rank of A.[5] This theorem can be applied to the results of the preceeding section which showed that $e'e = \eta'M\eta$, where M is idempotent and the elements of η are independent identically distributed Normal random variables, each with mean zero and variance σ^2. By the theorem, $e'e/\sigma^2$ is distributed χ^2 with $(I-1)(J-1)(T-1) - k$ degrees of freedom.

Now consider the estimator B for β. It has already been shown that

$E(B) = \beta$ and

$Var(B) = \sigma^2(X'VX)^{-1}$.

$$\begin{aligned} \text{And } B &= (X'VX)^{-1}X'VY = (X'VX)^{-1}X'V(X\beta + \varepsilon) = \\ &(X'VX)^{-1}X'VX\beta + (X'VX)^{-1}X'V\varepsilon = \\ &\beta + (X'VX)^{-1}X'V\eta. \end{aligned}$$

So, since B is linear in the components of η, B has a multivariate normal distribution also

$$B \sim N(\beta, \sigma^2(X'VX)^{-1}).$$

It can now be shown that the Chi-square and Normal distributions described above are independent. Note that $e'e/\sigma^2 = \eta'M\eta/\sigma^2$ is an idempotent

---

[5]For a proff of this theorem, as well as of the converse implication, see Hogg, R., and Craig, A., Introduction to Mathematical Statistics, pp. 348-351, MacMillan, 1965.

quadratic form in η, and that $B = \beta + (X'VX)^{-1}X'V\eta$ is a vector whose elements are linear in η, where the components of η are independent identically distributed random variables. A sufficient condition for $e'e/\sigma^2$ and B to be statistically independent is that the product of $(X'VX)^{-1}X'V$ and M be equal to the null vector.[6] That this is so is easily verified:

$$[(X'VX)^{-1}X'V]\, M =$$

$$[(X'VX)^{-1}X'V]\,[V-VX(X'VX)^{-1}X'V] =$$

$$(X'VX)^{-1}X'V - (X'VX)^{-1}X'VX(X'VX)^{-1}X'V =$$

$$(X'VX)^{-1}X'V - (X'VX)^{-1}X'V = \bar{0} .$$

Hence $e'e/\sigma^2$ and B are independent.

Now since:

$$S^2 = \frac{e'e}{[(I-1)(J-1)(T-1) - k]} = \frac{\sigma^2}{[(I-1)(J-1)(T-1) - k]} \quad \frac{e'e}{\sigma^2}$$

is linear in $e'e/\sigma^2$, S^2 and B are independent as well.

As a consequence, it is now possible to get a joint distribution of S^2 and a linear combination of the components of B. Now $B - \beta \sim N(0, \sigma^2(X'VX)^{-1})$. Let W be a k-vector of constants.

Then $W'(B-\beta) \sim N(0, W'(X'VX)^{-1}W\sigma^2)$.

And $$\frac{W'(B-\beta)}{[\sigma^2\, W'(X'VX)^{-1}W]^{1/2}} \sim N(0,1).$$

---

[6]For a proof of this assertion, see Theil, H., Principles of Econometrics, pp. 83-84, Wiley, 1971.

So that, since B and S^2 are independent,

$$\frac{\dfrac{W'(B-\beta)}{[\sigma^2 W'(X'VX)^{-1}W]^{1/2}}[(I-1)(J-1)(T-1) - k]^{1/2}}{\{[(I-1)(J-1)(T-1) - k] S^2/\sigma^2\}^{1/2}} = \frac{W'(B-\beta)}{S[W'(X'VX)^{-1}W]^{1/2}} \quad \text{has}$$

t-distribution with (I-1)(J-1)(T-1)-k degrees of freedom.

So a confidence interval for $W'\beta$, is a linear combination of the elements of β, is given by

$$W'B \pm t_{1-\frac{\alpha}{2}} S \{W'(X'VX)^{-1}W\}^{\frac{1}{2}} \quad ,$$

where $t_{1-\frac{\alpha}{2}}$ is the 100 $(1-\alpha)^{th}$ percentile of a t-distribution with (I-1)(J-1)(T-1) - k degrees of freedom.

In particular this holds for a vector W_p which has zeros in each component, except for the p^{th} element which is equal to one. Application of this vector W_p will give a confidence interval for the p^{th} component of β, p = 1, ...k.

G. AN ALTERNATE DERIVATION OF V

The calculations which yield the elements of the matrix V, introduced in Section B , may not be apparent. The purpose of the present section is to delinate the sequence of steps that lead to the elements of V.

As a vehicle, consider a disturbance term of the form, once again,

(1) $\varepsilon_{ijt} = \eta_{ijt} + \alpha_i + \gamma_j + \delta_t + \lambda_{ij} + \omega_{it} + \pi_{jt}$, where nothing is known or can be reasonably assumed about the components of the ε_{ijt}'s except that the η_{ijt}'s are independent Normal random variables, each with mean zero and variance σ^2.

Now:

$$(2) \qquad \varepsilon_{ij.} = \frac{1}{T}\sum_{t} \varepsilon_{ijt} = \frac{1}{T}\sum_{t} \eta_{ijt} + \alpha_i + \gamma_j + \frac{1}{T}\sum_{t} \delta_t + \lambda_{ij} + \frac{1}{T}\sum_{t} \omega_{it} + \frac{1}{T}\sum_{t} \pi_{jt}$$

$$(3) \qquad \varepsilon_{i.t} = \frac{1}{J}\sum_{j} \varepsilon_{ijt} = \frac{1}{J}\sum_{j} \eta_{ijt} + \alpha_i + \frac{1}{J}\sum_{j} \gamma_j + \delta_t + \frac{1}{J}\sum_{j} \lambda_{ij} + \omega_{it} + \frac{1}{J}\sum_{j} \pi_{jt}$$

$$(4) \qquad \varepsilon_{.jt} = \frac{1}{I}\sum_{i} \varepsilon_{ijt} = \frac{1}{I}\sum_{i} \eta_{ijt} + \frac{1}{I}\sum_{i} \alpha_i + \gamma_j + \delta_t + \frac{1}{I}\sum_{i} \lambda_{ij} + \frac{1}{I}\sum_{i} \omega_{it} + \pi_{jt}$$

$$(5) \qquad \varepsilon_{i..} = \frac{1}{JT}\sum_{j}\sum_{t} \varepsilon_{ijt} = \frac{1}{JT}\sum_{j}\sum_{t} \eta_{ijt} + \alpha_i + \frac{1}{J}\sum_{j} \gamma_j + \frac{1}{T}\sum_{t} \delta_t + \frac{1}{J}\sum_{j} \lambda_{ij} + \frac{1}{T}\sum_{t} \omega_{it} + \frac{1}{JT}\sum_{j}\sum_{t} \pi_{jt}$$

$$(6) \qquad \varepsilon_{.j.} = \frac{1}{IT}\sum_{i}\sum_{t} \varepsilon_{ijt} = \frac{1}{IT}\sum_{i}\sum_{t} \eta_{ijt} + \frac{1}{I}\sum_{i} \alpha_i + \gamma_j + \frac{1}{T}\sum_{t} \delta_t + \frac{i}{I}\sum_{i} \lambda_{ij} + \frac{1}{IT}\sum_{i}\sum_{t} \omega_{it} + \frac{1}{T}\sum_{t} \pi_{jt}$$

$$(7) \qquad \varepsilon_{..t} = \frac{1}{IJ}\sum_{i}\sum_{j} \varepsilon_{ijt} = \frac{1}{IJ}\sum_{i}\sum_{j} \eta_{ijt} + \frac{1}{I}\sum_{i} \alpha_i + \frac{1}{J}\sum_{j} \gamma_j + \delta_t + \frac{1}{IJ}\sum_{i}\sum_{j} \lambda_{ij} + \frac{1}{I}\sum_{i} \omega_{it} + \frac{1}{J}\sum_{j} \pi_{jt}$$

$$(8) \qquad \varepsilon_{...} = \frac{1}{IJT}\sum_i\sum_j\sum_t \varepsilon_{ijt} = \frac{1}{IJT}\sum_i\sum_j\sum_t \eta_{ijt} + \frac{1}{I}\sum_i \alpha_i + \frac{1}{J}\sum_j \gamma_j +$$

$$\frac{1}{T}\sum_t \delta_t + \frac{1}{IJ}\sum_i\sum_j \lambda_{ij} + \frac{1}{IT}\sum_i\sum_t \omega_{it} + \frac{1}{JT}\sum_j\sum_t \pi_{jt} \quad .$$

Adding and subtracting (1)-(2)-(3)-(4)+(5)+(6)+(7)-(8), the disturbance term for the ijt^{th} observation in normalized data becomes:

$$\mu_{ijt} = \varepsilon_{ijt} - \varepsilon_{ij.} - \varepsilon_{i.t} - \varepsilon_{.jt} + \varepsilon_{i..} + \varepsilon_{.j.} + \varepsilon_{..t} - \varepsilon_{...} =$$

$$\eta_{ijt} - \eta_{ij.} - \eta_{i.t} - \eta_{.jt} + \eta_{i..} + \eta_{.j.} + \eta_{..t} - \eta_{...} =$$

$$\frac{1}{IJT}\Bigg\{ IJT\,\eta_{ijt} - JT\sum_i \eta_{ijt} - IT\sum_j \eta_{ijt} - IJ\sum_t \eta_{ijt} +$$

$$I\sum_j\sum_t \eta_{ijt} + J\sum_i\sum_t \eta_{ijt} + T\sum_i\sum_j \eta_{ijt} - \sum_i\sum_j\sum_t \eta_{ijt}\Bigg\} \quad .$$

The equations (2) through (8) above were written out in the inconvenient summative form to make obvious the fact that the variables α_i, γ_j, δ_t, λ_{ij}, ω_{it} and π_{jt} disappear completely from the disturbance term of the normalized model. This is so since the equations (1) through (8) are written in terms of the random variables themselves, not in terms of realizations of these random variables. These random variables also disappear, of course, in the event that one or more of them is degenerate, as might happen if an unobservable explanatory variable were implicitly included in the disturbance term ε_{ijt}.

The expression for μ_{ijt} consists of adding and subtracting various multiples of given random variables. But in this expression any random variable $\eta_{i_0 j_0 t_0}$ may be included under more than one summation sign. Concentrate on one normalized disturbance term, say $\mu_{i_1 j_1 t_1}$, and rearrange terms in the series of summations so that each random variable η_{ijt}

appears once and only once in the expression for $\mu_{i_1 j_1 t_1}$:

$$\mu_{i_1 j_1 t_1} = \frac{1}{IJT}\Bigg\{IJT\ \eta_{i_1 j_1 t_1} - JT\sum_i \eta_{i j_1 t_1} - IT\sum_j \eta_{i_1 j t_1} -$$

$$IJ\sum_t \eta_{i_1 j_1 t} + I\sum_j\sum_t \eta_{i_1 j t} + J\sum_i\sum_t \eta_{i j_1 t} + T\sum_i\sum_j \eta_{i j t_1}$$

$$-\sum_i\sum_j\sum_t \eta_{ijt}\Bigg\} =$$

$$\frac{1}{IJT}\Bigg\{(I-1)(J-1)(T-1)\ \eta_{i_1 j_1 t_1} - (J-1)(T-1)\sum_{\substack{i\\ i\neq i_1}} \eta_{i j_1 t_1} -$$

$$(I-1)(T-1)\sum_{\substack{j\\ j\neq j_1}} \eta_{i_1 j t_1} - (I-1)(J-1)\sum_{\substack{t\\ t\neq t_1}} \eta_{i_1 j_1 t} +$$

$$(T-1)\sum_{\substack{i\\ i\neq i_1}}\sum_{\substack{j\\ j\neq j_1}} \eta_{i j t_1} + (J-1)\sum_{\substack{i\\ i\neq i_1}}\sum_{\substack{t\\ t\neq t_1}} \eta_{i j_1 t} + (I-1)\sum_{\substack{j\\ j\neq j_1}}\sum_{\substack{t\\ t\neq t_1}} \eta_{i_1 j t}$$

$$-\sum_{\substack{i\\ i\neq i_1}}\sum_{\substack{j\\ j\neq j_1}}\sum_{\substack{t\\ t\neq t_1}} \eta_{ijt}\Bigg\}.$$

So that $\mu_{i_1 j_1 t_1}$ is a series of summations of independent, identically distributed Normal random variables.

Since each of these random variables η_{ijt} has mean zero and variance σ^2, it is clear that:

$$E(\mu_{i_1 j_1 t_1}) = 0$$

and

$$\mathrm{Var}\,(\mu_{i_1 j_1 t_1}) = \left(\frac{1}{IJT}\right)^2 \mathrm{Var}\Bigg\{(I-1)(J-1)(T-1)\ \eta_{i_1 j_1 t_1} -$$

$$(J-1)(T-1\sum_{\substack{i\\ i\neq i_1}} \eta_{i j_1 t_1} - (I-1)(T-1)\sum_{\substack{j\\ j\neq j_1}} \eta_{i_1 j t_1} - (I-1)(J-1)\sum_{\substack{t\\ t\neq t_1}} \eta_{i_1 j_1 t}$$

$$+ (T-1) \sum_{\substack{i \\ i \neq i_1}} \sum_{\substack{j \\ j \neq j_1}} \eta_{ijt_1} + (J-1) \sum_{\substack{i \\ i \neq i_1}} \sum_{\substack{t \\ t \neq t_1}} \eta_{ij_1t} + (I-1) \sum_{\substack{j \\ j \neq j_1}} \sum_{\substack{t \\ t \neq t_1}} \eta_{i_1jt}$$

$$\left. - \sum_{\substack{i \\ i \neq i_1}} \sum_{\substack{j \\ j \neq j_1}} \sum_{\substack{t \\ t \neq t_1}} \eta_{ijt} \right\} =$$

$$\left(\frac{1}{IJT}\right)^2 \Bigg\{ [(I-1)(J-1)(T-1]^2 \text{ Var } (\eta_{i_1j_1t_1}) + [(J-1)(T-1)]^2 \text{ Var}\Bigg(\sum_{\substack{i \\ i \neq i_1}} \eta_{ij_1t_1}\Bigg)$$

$$+ [(I-1)(T-1)]^2 \text{ Var}\Bigg(\sum_{\substack{j \\ j \neq j_1}} \eta_{i_1jt_1}\Bigg) + [(I-1)(J-1)]^2 \sum_{\substack{t \\ t \neq t_1}} \eta_{i_1j_1t} +$$

$$(T-1)^2 \text{ Var}\Bigg(\sum_{\substack{i \\ i \neq i_1}} \sum_{\substack{j \\ j \neq j_1}} \eta_{ijt_1}\Bigg) + (J-1)^2 \text{ Var}\Bigg(\sum_{\substack{i \\ i \neq i_1}} \sum_{\substack{t \\ t \neq t_1}} \eta_{ij_1t}\Bigg) +$$

$$(I-1)^2 \text{ Var}\Bigg(\sum_{\substack{j \\ j \neq j_1}} \sum_{\substack{t \\ t \neq t_1}} \eta_{i_1jt}\Bigg) + (1)^2 \text{ Var}\Bigg(\sum_{\substack{i \\ i \neq i_1}} \sum_{\substack{j \\ j \neq j_1}} \sum_{\substack{t \\ t \neq t_1}} \eta_{ijt}\Bigg) \Bigg\} =$$

$$\left(\frac{\sigma}{IJT}\right)^2 \Big\{ [(I-1)J-1)(T-1)]^2 + [(J-1)(T-1)]^2 (I-1) +$$

$$[(I-1)(T-1)]^2 (J-1) + [(I-1)(J-1)]^2 (T-1) + (T-1)^2(I-1)(J-1) +$$

$$(J-1)^2(I-1)(T-1) + (I-1)^2(J-1)(T-1) + (I-1)(J-1)(T-1) \Big\} =$$

$$\frac{\sigma^2 (I-1)(J-1)(T-1)}{IJT} .$$

Note that this applies for all μ_{ijt}. And since μ_{ijt} is a linear combination of independent, identically distributed Normal random variables, μ_{ijt} is also Normally distributed.

Note that the diagonal elements of the covariance matrix $E(\mu\mu')$ are each $\sigma^2(I-1)(J-1)(T-1)/IJT$. But also note that, since each of the μ_{ijt}'s is a linear combination of the same IJT random variables η_{ijt}, $i=1,\ldots I$, $j=1,\ldots J$, $t=1,\ldots T$, the μ_{ijt}'s are not independent.

The remainder of the covariance matrix may be found by straightforward but tedious calculations. Since $E(\mu_{ijt}) = 0$, these calculations (using the summative expression in the η_{ijt}'s for each $\mu_{i_1j_1t_1}$) yield

$$COV(\mu_{i_1j_1t_1}, \mu_{i_2j_2t_2}) = E(\mu_{i_1j_1t_1}\, \mu_{i_2j_2t_2}) =$$

$$\frac{-(J-1)(T-1)\sigma^2}{IJT} \quad \text{if } i_1 \neq i_2,\ j_1 = j_2,\ t_1 = t_2$$

$$\frac{-(I-1)(T-1)\sigma^2}{IJT} \quad \text{if } i_1 = i_2,\ j_1 \neq j_2,\ t_1 = t_2$$

$$\frac{-(I-1)(J-1)\sigma^2}{IJT} \quad \text{if } i_1 = i_2,\ j_1 = j_2,\ t_1 \neq t_2$$

$$\frac{(T-1)\sigma^2}{IJT} \quad \text{if } i_1 \neq i_2,\ j_1 \neq j_2,\ t_1 = t_2$$

$$\frac{(J-1)\sigma^2}{IJT} \quad \text{if } i_1 \neq i_2,\ j_1 = j_2,\ t_1 \neq t_2$$

$$\frac{(I-1)\sigma^2}{IJT} \quad \text{if } i_1 = i_2,\ j_1 \neq j_2,\ t_1 \neq t_2$$

$$\frac{-\sigma^2}{IJT} \quad \text{if } i_1 \neq i_2,\ j_1 \neq j_2,\ t_1 \neq t_2 \ .$$

So that, for the matrix previously defined, $\sigma^2 V = E(\mu\mu')$.

H. THE CASE WHEN FEWER THAN IJT OBSERVATIONS ARE USED

Suppose the components of the disturbance term are independent identically distributed Normal random variables with mean zero. Then

for ordinary least squares estimation in the original data the quantity $(IJT - k)\ S_0^2 / \sigma^2$ has χ^2 distribution with IJT - k degrees of freedom, where $S_0^2 = e'e/(IJT - k)$ is the estimator in the original data of σ^2.

When normalized data are used the quantity:

$$[(I-1)(J-1)(T-1) - k]\ \frac{S^2}{\sigma^2} = \left[\left\{\frac{(I-1)(J-1)(T-1)}{IJT}\ IJT\right\} - k\right] \frac{S^2}{\sigma^2}$$

has χ^2 distribution with $(I-1)(J-1)(T-1) - k = \frac{(I-1)(J-1)(T-1)}{IJT}\ IJT - k$ degrees of freedom, for S^2 the estimator of σ^2 previously derived. In addition, the latter distribution still applies when disturbance structure (A) is assumed. An analagous relationship holds when $n < IJT$ observations are used in the least squares estimation (such a case might arise when some observations must be discarded for one reason or another). In this case, for ordinary least squares estimation in the original data the quantity $(n - k)\ S_0^2 / \sigma^2$ has χ^2 distribution with n - k degrees of freedom. It is desired to show the analagous distribution (in S^2) when normalized data are used. But when not all observations are allowed, the method of "normalizing" the remaining observations is not obvious. The most straightforward approach is to take the appropriate means, in the normalization process, over those observations that are available. Then, for example, the normalization of the $(i,j,t)^{th}$ observation on the dependent variable (which is assumed to be used) still has the form:

$$y_{ijt} - y_{ij.} - y_{i.t} - y_{.jt} + y_{i..} + y_{.j.} + y_{..t} - y_{...}\ ,$$

where now

$$(*) \qquad y_{ij.} = \frac{1}{||T(i,j)||} \sum_{t \varepsilon T(i,j)} y_{ijt}$$

$$y_{i.t} = \frac{1}{||J(i,t)||} \sum_{j \varepsilon J(i,t)} y_{ijt}$$

$$Y_{.jt} = \frac{1}{||I(j,t)||} \sum_{i \varepsilon I(j,t)} y_{ijt}$$

$$y_{i..} = \frac{1}{||J(i,t)|| \cdot ||T(i,j)||} \sum_{j \varepsilon J(i,t)} \sum_{t \varepsilon T(i,j)} y_{ijt}$$

$$y_{.j.} = \frac{1}{||I(j,t)|| \cdot ||T(i,j)||} \sum_{i \varepsilon I(j,t)} \sum_{t \varepsilon T(i,j)} y_{ijt}$$

$$y_{..t} = \frac{1}{||I(j,t)|| \cdot ||J(i,t)||} \sum_{i \varepsilon I(j,t)} \sum_{j \varepsilon J(i,t)} y_{ijt}$$

$$y_{...} = \frac{1}{||I(j,t)|| \cdot ||J(i,t)|| \cdot ||T(i,j)||} \sum_{i \varepsilon I(j,t)} \sum_{j \varepsilon J(i,t)} \sum_{t \varepsilon T(i,j)} y_{ijt} ,$$

where, for example, T(i,j) is the set of all years in which the observations of y_{ijt}, for Rate i and pay grade j, are used and ||T(i,j)|| is the number of elements in T(i,j). The normalized value of any observation which is not used in the least squares estimation is taken to be zero. The same form applies for normalization of the explanatory variables in X. With a little reflection it is seen that, in effect, this normalization process implicitly takes the value of an unused observation of any variable to be the sum of the appropriate means over observations which are in fact used. That is, an unused observation y_{ijt} is taken to be equal to:

$$y_{ijt} = y_{ij.} + y_{i.t} + y_{.jt} - y_{i..} - y_{.j.} - y_{..t} + y_{...} ,$$

where the terms on the right hand side of this equation are as given in (*) above. In particular, this modified normalization process is applied to the disturbance terms ε_{ijt} as well. Let μ represent the n-vector ($n < IJT$ is the number of observations used) of disturbance terms under

the modified normalization. Define, as in the preceeding section,

$$V_0 = \frac{1}{\sigma^2} E(\mu\mu') \quad ,$$

where the matrix V_0 has order $n < IJT$. Note that the diagonal element of V_0 which corresponds to observation (i,j,t) is equal to:

$$\frac{(||I(j,t)||-1)(||J(i,t)||-1)(||T(i,j)||-1)}{||I(j,t)||\cdot||J(i,t)||\cdot||T(i,j)||}$$

since it represents the variance of a component of μ derived through the modified normalization specified in (*) above. Thus, the trace of V_0 is equal to:

$$\sum_{i\varepsilon UI(j,t)} \sum_{j\varepsilon UJ(i,t)} \sum_{t\varepsilon UT(i,j)} \frac{(||I(j,t)||-1)(||J(i,t)||-1)(||T(i,j)||-1)}{||I(j,t)||\cdot||J(i,t)||\cdot||T(i,j)||} .$$

Note also that V_0 is symmetric and that for an arbitrary n-component disturbance vector ε, $V_0V_0\varepsilon = V_0\varepsilon$, so that V_0 is idempotent. That this is so is clear since for $\varepsilon_{ij.}$, $\varepsilon_{i.t}$, $\varepsilon_{.jt}$; $\varepsilon_{i..}$, $\varepsilon_{.j.}$, $\varepsilon_{..t}$ and $\varepsilon_{...}$ as specified in the equations (*), $V_0\, \varepsilon_{ij.} = V_0\, \varepsilon_{i.t} = V_0\, \varepsilon_{.jt} = V_0\, \varepsilon_{i..} = V_o\, \varepsilon_{.j.} = V_o\, \varepsilon_{..t} = V_o\, \varepsilon_{...} = \bar{0}$. The matrix V_o has properties analagous to the matrix V considered previously, and represents the linear transformation which projects an n-vector of observations into the modified normalization of that vector.

Now let $N(n) = tr(V_o) =$

$$\sum_{i\varepsilon UI(j,t)} \sum_{j\varepsilon UJ(i,t)} \sum_{t\varepsilon UT(i,j)} \frac{(||I(j,t)||-1)(||J(i,t)||-1)(||T(i,j)||-1)}{||I(j,t)||\cdot||J(i,t)||\cdot||T(i,j)||}$$

and let $M_o = V_o - V_o\, X(X'V_0X)^{-1}X'V_0$, where X is now the n x k matrix of observations which results from removing the IJT-n unused observations from the original IJT x k matrix of observations X. Then the error sum

of squares for the least squares estimation in modified normalized form of the data (with unused observations removed) is $e'e = \varepsilon' M_0 \varepsilon$ where M_0 is an idempotent matrix of rank $N(n) - k$. That M_0 is idempotent is clear since $M_0 M_0 = [V_0 - V_0 X(X'V_0 X)^{-1} X'V_0][V_0 - V_0 X(X'V_0 X)^{-1} X'V_0] =$ $V_0 - V_0 X(X'V_0 X)^{-1} X'V_0 - V_0 X(X'V_0 X)^{-1} X'V_0 + V_0 X(X'V_0 X)^{-1} X'V_0 X(X'V_0 X)^{-1} X'V_0 =$ $V_0 - V_0 X(X'V_0 X)^{-1} X'V_0 = M_0$. And M_0 has trace (hence rank) $N(n) - k$ since:

$$
\begin{aligned}
tr(M) &= tr[V_0 - V_0 X(X'V_0 X)^{-1} X'V_0] = \\
tr(V_0) &- tr[V_0 X(X'V_0 X)^{-1} X'V_0] = \\
tr(V_0) &= tr[X'V_0 X(X'V_0 X)^{-1}] =
\end{aligned}
$$

$N(n) - k$. Hence for disturbance term ε specified by:

$$\varepsilon_{ijt} = \eta_{ijt} + \alpha_i + \gamma_j + \delta_t + \lambda_{ij} + \omega_{it} + \pi_{jt} \quad ,$$

where η_{ijt}'s are independent identically distributed Normal random variables with mean zero and variance σ^2, $\frac{1}{\sigma^2} \varepsilon' M_0 \varepsilon$ has χ^2 distribution with $N(n) - k$ degrees of freedom. Thus, for the estimator:

$$S^2 = \frac{e'e}{N(n) - k} = \frac{\varepsilon' M_0 \varepsilon}{N(n) - k} \quad \text{of} \quad \sigma^2 \quad ,$$

$[N(n) - k]\, S^2/\sigma^2$ has χ^2 distribution with $N(n) - k$ degrees of freedom.

For those cases in which the removal of observations is not systematic (that is, when observations are discarded in no regular pattern), computation of $N(n)$ may involve many computations and may require that one keep track of a large number of values of $||I(j,t)||$, $||J(i,t)||$ and $||T(i,j)||$. It may therefore, be beneficial to derive the distribution of an alternative random variable linear in S^2. The quantity:

$$\frac{\left[\frac{(I-1)(J-1)(T-1)\, n}{IJT} - k\right]}{N(n) - k} \; \frac{\left[N(n) - k\right] S^2}{\sigma^2} = \left[\frac{(I-1)(J-1)(T-1)n}{IJT} - k\right] \frac{S^2}{\sigma^2}$$

is linear in $\dfrac{[N(n) - k]\, S^2}{\sigma^2}$

hence has χ^2 distribution, with degrees of freedom given by:

$$E\left\{\left[\frac{(I-1)(J-1)(T-1)n}{IJT} - k\right]\frac{S^2}{\sigma^2}\right\} =$$

$$\left[\frac{(I-1)(J-1)(T-1)n}{IJT} - k\right] E\left(\frac{S^2}{\sigma^2}\right) =$$

$$\frac{(I-1)(J-1)(T-1)n}{IJT} - k \quad .$$

Thus the analogy is completed.

I. GENERALIZATION TO q CROSS-SECTIONS

There is a natural generalization of all of the preceeding sections to the case in which q cross-sectional dimensions are involved. Previously, recall, all was described in terms of three cross-sectional dimensions.

Suppose q cross-sectional dimensions are being considered in the model $Y = X\beta + Z\Omega + \varepsilon$. Analagously to the case for $q = 3$, let the variables whose observations are represented by X and Y vary over all q dimensions, and let each variable in Z vary over at most $q - 1$ dimensions. Also let the disturbance term ε be constructed analagously to the previously considered case, $q = 3$. That is, for q cross-sectional dimensions, with respective numbers of categories $I_1, \ldots I_q$, ε is a linear combination of:

$$\prod_{k=1}^{q} I_k$$

random vectors, one of which varies over q cross-sectional dimensions

(let this single random vector be denoted as η, as before, where the elements of η are written with q subscripts) and the remaining

$$\prod_{k=1}^{q} I_k - 1$$

of which vary over at most q - 1 dimensions (that is, the elements of each of these remaining random vectors are written with fewer than q subscripts). Also, the elements of η are independent, identically distributed Normal random variables, each with mean zero and variance σ^2, and the remaining

$$\prod_{k=1}^{q} I_k - 1$$

random variables are subject to any unknown distributions, and to any unknown conditions of stochastic non-independence.

All the properties that have been derived in preceeding sections flowed naturally from a knowledge of the idempotent matrix V. Thus, in order to characterize the general case for q cross-sectional dimensions, it is only necessary to find the appropriate matrix V_q whose properties are analagous to those of the previously defined V. To this end, let C_{ij} be the subscript (in the notational expression for the elements of η; there are q such subscripts in the notational expression for each element of η) representing the i^{th} category of the j^{th} cross-sectional dimension, $j = 1 \ldots q$, $i = 1, \ldots I_j$.

Then the elements of $V_q = \frac{1}{\sigma^2} E(\eta\eta')$ are given, for $i_p = 1, \ldots I_p$, $j_t = 1, \ldots I_t$, by $E({}^{\eta}C_{i_1 1}, \ldots C_{i_q q} \; {}^{\eta}C_{j_1 1}, \ldots C_{j_q q}) = (-1)^{p+q} \prod_{m \varepsilon S} (I_m - 1)$, where:

$$S = \left\{ m : C_{i_m m} = C_{j_m m},\ m = 1, \ldots q \right\},$$

and p is the number of elements in S. When S is empty, define $\prod_{m \varepsilon S} (I_m - 1) = 1$.

That is: S is the set of all cross-sectional dimensions for which the subscripts $C_{i_m m}$ and $C_{i_m m}$ are equal in the variables $^{\eta}C_{i_1 1}, \ldots C_{i_q q}$ and $^{\eta}C_{j_1 1}, \ldots C_{j_q q}$, whose covariance is an element of V_q. Or: S is the set of all cross-sectional dimensions for which the above two random variables correspond to the same category. Note that the set S depends on the two elements of η whose covariance is being considered.

To complete the analogy to the case q = 3, V_q is an idempotent matrix of order

$$\prod_{k=1}^{q} I_k$$

and trace (=rank)

$$\prod_{k=1}^{q} (I_k - 1) \quad .$$

J. THE INAPPROPRIATELY APPLIED MODEL: A CASE IN WHICH DISTURBANCE STRUCTURE (A) DOES NOT APPLY

Before proceeding with this section, it may be instructive to amplify on the derivation of the transformation V. Note that the originally stated purpose of the transformation V was to rid the model $Y = X\beta + Z\Omega + \varepsilon$ of the effects of certain unobserved or unobservable explanatory variables. The disturbance structure (A) hypothesized in Part D was constructed, more or less artifically, to take advantage of the properties of V. Disturbance structure (A) is simply the most general case of the original problem: it contains all possible sources of error which the transformation V is able to remove. Consider a model of the form $Y = X\beta + Z\Omega + \varepsilon$ as previously introduced. Then the following statements are equivalent:

a. ε obeys disturbance structure (A):

b. The elements of ε are independent, identically distributed Normal random variables, each with mean zero and variance σ^2, and included in the specification of the model (specifically, in Z)

is any variable (observed or not) which may be written as varying over fewer than q cross-sectional dimensions (q is the total number of dimensions involved in the data).

c. No knowledge or information about the disturbance term ε may reasonably be assumed except that at least one component of each ε_{ijt} is a sample from a Normal population with mean zero and variance σ^2.

This situation suggests two useful observations. The first concerns the unobserved or unobservable explanatory variables which, by the dictates of theory (that is, theory relating to the subject being modeled) or other considerations, are necessarily included in some model of the form considered here. Note that, since the transformation V rids the model of these variables (as long as each of these variables varies over fewer than q cross-sectional dimensions, where q is the total number of dimensions involved) in any case, it is conceptually and practically equivalent whether these variables are explicitly included in the formal form of the model, or whether they are implicitly "thrown into" the disturbance term. This is a trite observation, but it is well worth noting for the following reason: some studies and analyses (see, for example, Nerlove [8]), when implicitly including an unobserved or unobservable explanatory variable as a component of the disturbance term, make a strong and possibly erroneous[7] assumption in order to complete the regression analysis (that is, in order to be able to claim an unbiased

---

[7] The term "erroneous" should be seen in context. The case of interest here is that in which there exists some unobserved explanatory variable which is expected to have a significant effect on the dependent variable. In addition, it is supposed that the analyst has no (or does not care to get any) information about the values of this variable. Such a variable may indeed not even by quantifiable.

estimator of the regression coefficients) without using some transformation such as V to purge the model of the offending variable. Specifically, the required[8] assumption is that the disturbance term (which now implicity includes unobserved and unobservable explanatory variables) has known mean, usually zero. [It is further typically assumed that the disturbance term is Normally distributed, although this assumption is not necessary if all one wishes to do is ensure that the estimator is unbiased.] That this assumption may be erroneous can be seen in two approaches to the assumption. One may simply make this assumption with no justification. But since theory, or other consideration, has dictated that the unobserved explanatory variables does have an effect on the dependent variable, the original problem still remains. And the resolution to that problem is still to remove the offending explanatory variable (whether explicitly included in the model or implicitly included as a component of the disturbance term) by some transformation such as V.[9] Alternatively, one may attempt to justify the assumption by means of some device such as the Central Limit Theorem, in this case making the additional assumption that the components of the disturbance term, which now includes the unobserved explanatory variables, are independent. Ignoring for the moment

---

[8] This assumption is characterized as "required" since unless it is made, some unobserved explanatory variable is, in effect, still being considered an explicit term in the model.

[9] Note that V may not be unique in this respect. For example, in the model

$$y_{it} = \alpha + \beta X_{it} + \gamma Z_i + \varepsilon_{it} \quad ,$$

where one wishes to purge Z_i, the transformation W may be used, where:

$$W[y_{it}] = [y_{it} - y_{i.}],\ W[X_{it}] = [X_{it} - X_{i.}],\ W[Z_i] =$$

$$[Z_i - Z_i] = \bar{0},\ W[\varepsilon_{it}] = [\varepsilon_{it} - \varepsilon_{i.}],\ W[\alpha] = [\alpha - \alpha] = \bar{0} \ .$$

Here [Pit] is an n-vector whose elements are Pit.

the fact that this latter assumption is contrary to the assumptions of disturbance structure (A), this sort of argument may be reasonable in some cases. But in justifying the application of the Central Limit Theorem, in order to approximate a Normal random variable of known mean by a sum of random variables, one typically assumes that the disturbance term represents the net effect of numerous individually unimportant but collectively significant variables. But this is clearly not the case (at least this latest assumption cannot reasonably be made) when disturbance structure (A) pertains. And, more generally, it can be said that there are certainly studies of interest where this is not the case: the unobserved explanatory variable whose inclusion in the model was a necessity cannot in general be assumed not to dominate the disturbance term in which it is incorporated. In summary, there exist studies for which the use of a transformation such as V, to rid the model of undesired variables, is unavoidable if an unbiased estimator of the regression coefficients is to be obtained. Simply discarding an undesired variable as a component of a disturbance term with known mean should be viewed cautiously. As an example, in the reenlistment model, the inclusion of the terms WC_{it} and C_t in the disturbance term can be expected to have a large effect on the disturbance term.

The second observation concerns the best linear unbiasedness of the estimator $B = (X'VX)^{-1}X'VY$ for β in $Y = X\beta + Z\Omega + \varepsilon$. Recall that when disturbance structure (A) is assumed, B is the best linear unbiased estimator for β. Note that since, in disturbance structure (A), the random variables α, γ, δ, λ, ω and π may assume any (unknown) distribution, and since any error terms in the model (except the η_{ijt}'s) may be interdependent, disturbance structure (A) is more general than that typically assumed (specifically, that error structure in which the

elements of the disturbance term ε are independent, identically distributed Normal random variables, each with mean zero and variance σ^2). But it is not a generalization of this latter error structure: the latter is not a special case of disturbance structure (A). This is so since disturbance structure (A) is based on a certain lack of specific information or knowledge about the characteristics of the components of the disturbance term. As a consequence, if the error structure which one wishes to assume is not that specified by disturbance structure (A), then $B = (X'VX)^{-1}X'VY$ is not necessarily the best linear unbiased estimator for β in $Y = X\beta + Z\Omega + \varepsilon$.

This latest observation leads into the proper subject of this section: a consideration of a common case in which B is not the best linear unbiased estimator for β. For consistency of approach, suppose that the model is written in the form $Y = X\beta + \varepsilon$, where any unobserved or unobservable explanatory variables (if any), which were previously included in Z, are now included in the disturbance term ε. As has been seen, $B = (X'VX)^{-1}X'VY$ is the best linear unbiased estimator for β when ε obeys disturbance structure (A). Consider the asymptotic properties of the matrix V in three cross-sectional dimensions. As the number of categories, I, J, and T, in each cross-sectional dimension goes to infinity, the elements of V behave as follows:

$$\frac{(I-1)(J-1)(T-1)}{IJT} = \frac{1-\frac{1}{I}}{1}\;\frac{1-\frac{1}{J}}{1}\;\frac{1-\frac{1}{T}}{1} \rightarrow 1 ,$$

$$\frac{-(I-1)(J-1)}{IJT} = -\frac{1-\frac{1}{I}}{1}\;\frac{1-\frac{1}{J}}{1}\;\frac{1}{T} \rightarrow 0 ,$$

$$\frac{-(I-1)(T-1)}{IJT} = -\frac{1-\frac{1}{I}}{1}\;\frac{1-\frac{1}{T}}{1}\;\frac{1}{J} \rightarrow 0 ,$$

$$\frac{-(J-1)(T-1)}{IJT} = -\frac{1-\frac{1}{J}}{1}\frac{1-\frac{1}{T}}{1}\frac{1}{I} \rightarrow 0 ,$$

$$\frac{I-1}{IJT} = \frac{1-\frac{1}{I}}{1}\frac{1}{J}\frac{1}{T} \rightarrow 0 ,$$

$$\frac{J-1}{IJT} = \frac{1-\frac{1}{J}}{1}\frac{1}{I}\frac{1}{T} \rightarrow 0 ,$$

$$\frac{T-1}{IJT} = \frac{1-\frac{1}{T}}{1}\frac{1}{I}\frac{1}{J} \rightarrow 0 ,$$

$$\frac{-1}{IJT} \rightarrow 0 .$$

[Note that when q cross-sectional dimensions are considered, the number of unique elements in V is 2^q, since each element of V depends on the comparison of the subscripts of two random variables, each of which has q subscripts. These two random variables may either agree or disagree in each subscript. For q = 3, then, V has $2^3 = 8$ unique elements.] That is, the diagonal elements of V approach unity and all other elements of V approach zero. Or, as I,J, and T increase without bound, V tends to the identity matrix. As a consequence, $(X'VX)^{-1}X'VY$ approaches $(X'X)^{-1}X'Y$ as I, J and T become infinitely large. Hence, in the case that ε obeys disturbance structure (A), the ordinary least squares estimator $\hat{\beta} = (X'X)^{-1}X'Y$ is in the limit (in I, J and T) an unbiased

estimator for β, since it is the limit of a sequence of unbiased estimators.[10] This suggests that, for sufficiently large I, J and T, the ordinary least squares estimator for β, $\hat{\beta} = (X'X)^{-1}X'Y$ could serve to approximate the best linear unbiased estimator B when disturbance structure (A) holds. This line of thought will not be pursued: it is the converse suggestion, that B can serve to approximate $\hat{\beta}$ for sufficiently large I, J and T, that is more interesting here. Suppose that the transformation V was inappropriately applied to the model $Y = X\beta + \varepsilon$. Specifically, suppose that the components of ε are independent, identically distributed Normal random variables with mean zero and variance σ^2. Call this disturbance structure (B). Then the ordinary least squares estimator $\hat{\beta} = (X'X)^{-1}X'Y$ is the best linear unbiased estimator for β. Note that $B = (X'VX)^{-1}X'VY$ is still an unbiased estimator for β, but it is no longer best. But since V approaches the identity matrix as I, J and T increase, the less efficient estimator B approaches $(X'X)^{-1}X'Y$ as well. This suggests a pragmatic comparative scheme for the two estimators B and $\hat{\beta}$:

---

[10] In treating a subject related to that considered here, Wallace and Hussain [9] have shown the asymptotic equivalence of the Aitken estimator and an estimator derived under a linear transformation (much as B was derived from the linear transformation V) for a particular error structure. In the disturbance structure considered in their paper, the disturbance term was assumed to be a sum of independent random variables (in a combined time series and cross-section analysis),

$$\varepsilon_{it} = \alpha_i + \gamma_t + \eta_{it}, \text{ for which } E(\alpha_i) = E(\gamma_t) = E(\eta_{it}) = 0$$

and $Var(\alpha_i) = \sigma_1^2$, $Var(\gamma_t) = \sigma_2^2$, $Var(\eta_{it}) = \sigma_3^2$ for all i, t,

where σ_1^2, σ_2^2, and σ_3^2 were known.

The paper also showed the equivalence of the iterative Aitken estimator and the estimator derived under a linear transformation for the disturbance structure as above with

σ_1^2, σ_2^2, and σ_3^2 unknown.

1. Suppose disturbance structure (A) applies. Then $\hat{\beta}$ is biased, and B is the best linear unbiased estimator and should reasonably be used.

2. Suppose on the other hand that disturbance structure (B) is assumed to hold. Then $\hat{\beta}$ and B are both unbiased estimators, although B is less efficient than $\hat{\beta}$. But note that B has an advantage which may offset (on a case-by-case basis) its lesser efficiency: it guarantees to purge all random variables which are invariant over at least one cross-sectional dimension. That is, if one is unsure of the validity of the assumption that disturbance structure (B) holds, then one may see some value in applying the transformation V in order to rid the model of all such possible sources of error.

Two concluding observations should now be made. First, it is clear that application of the transformation V is equally inappropriate in all other cases where disturbance structure (A) does not hold in the model $Y = X\beta + \varepsilon$. An important special case is that in which the generalized least squares estimator for β is appropriate. Just as the ordinary least squares estimator $\hat{\beta} = (X'X)^{-1}X'Y$ is the best linear estimator for β when $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2 I$, the Aitken estimator $\tilde{\beta} = (X' \Omega^{-1}X)^{-1} X' \Omega^{-1}Y$ is the best linear unbiased estimator for β for the case in which $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2\Omega$.

Finally, it is worth repeating the crucial condition which underlies the specification of the case in which the transformation V is effective. In the model $Y = X\beta + Z\Omega + \varepsilon$ (or in the equivalent, under the transformation V, model $Y = X\beta + \varepsilon$, where the variables in Z are thrown into the disturbance term ε) V is effective in removing unobserved or unobservable variables (stochastic or deterministic) only if these variables

are invariant over at least one cross-sectional dimension. Accordingly, all work in this paper is performed under the assumption that each variable in X (those variables which vary over all cross-sectional dimensions) has been observed.

K. INTERPRETATION OF TERMS UNDER THE TRANSFORMATION V

Consider the model in the form $Y = X\beta + \varepsilon$, in three cross-sectional dimensions. The equation representing the data in the i^{th} category of the first cross-sectional dimension, the j^{th} category of the second dimension and the t^{th} category of the third dimension is $y_{ijt} = x_{ijt}\ \beta + \varepsilon_{ijt}$, where x_{ijt} is a k-vector of observations on the k explanatory variables in X. The categories of the cross-sectional dimensions corresponding to the observations y_{ijt} and x_{ijt} may be considered to be "treatments" which affect the values of the observations of y_{ijt} and x_{ijt} in the $(i, j, t)^{th}$ "cell". With this in mind, assume that each y_{ijt} and x_{ijt} can be represented as a sum of common mean, effects due to single treatments (here i, j, t represent the "treatments"), two-way interaction effects of pairs of treatments, and a three way interaction effect of the three treatments. [Note that since there is only one observation (on each of y_{ijt} and x_{ijt}) per "cell", it is generally not possible to discern between the effect of the three-way interaction term and the error term ε_{ijt}. In this case, however, it is known that a three-way interaction term does in fact exist. That this is so can be seen as follows: since x_{ijt} is deterministic, one can calculate the exact three-way interaction effect for cell (i, j, t) as $x_{ijt} - x_{.jt} - x_{i.t} - x_{ij.} + X_{i..} + X_{.j.} + X_{..t} - X_{...}$, subject only to roundoff error (this expression is the same as that of a sample estimate of the three-way interaction effect for the case of stochastic x_{ijt}). This is not identically zero

(for all cells), by previous hypothesis about the variables in X, so a three-way interaction effect is present. And since y_{ijt} is a linear function of x_{ijt}, y_{ijt} also includes a three-way interaction effect.]
That is, that:

$$y_{ijt} = \mu + \Theta_{ijt} + A_i + B_j + C_t + D_{ij} + E_{it} + F_{jt} + \varepsilon_{ijt}$$

$$x_{ijt} = \mu^0 + \Phi_{ijt} + A_i^0 + B_j^0, + C_t^0 + D_{ij}^0 + E_{it}^0 + F_{jt}^0,$$

where Θ_{ijt} and Φ_{ijt} are the three-way interaction terms mentioned above.

Substituting these into the model:

$$y_{ijt} = \mu + \Theta_{ijt} + A_i + B_j + C_t + D_{ij} + E_{it} + F_{jt} + \varepsilon_{ijt} =$$

$$(\mu^0 + \Phi_{ijt} + A_i^0 + B_j^0 + C_t^0 + D_{ij}^0 + E_{it}^0 + F_{jt}^0)\ \beta\ +$$

$$\varepsilon_{ijt} = x_{ijt}\ \beta + \varepsilon_{ijt}.$$

These effects can be equated term by term to give:

$$\mu = \mu^0\beta$$

$$A_i = A_i^0\beta$$

$$B_j = B_j^0\beta$$

$$C_t = C_t^0\beta$$

$$D_{ij} = D_{ij}^0\beta$$

$$E_{it} = E_{it}^0\beta$$

$$F_{jt} = F_{jt}^0\beta$$

and: $$\Theta_{ijt} = \Phi_{ijt}\ \beta \quad (*)$$

Now consider the data under the transformation V: $VY = V X \beta + V\varepsilon$. In the $(i, j, t)^{th}$ cell this gives:

$$y_{ijt} - y_{ij.} - y_{i.t} - y_{.jt} + y_{i..} + y_{.j.} + y_{..t} - y_{...} =$$

$$(x_{ijt} - x_{ij.} - x_{i.t} - x_{.jt} + x_{i..} + x_{.j.} + x_{..t} - x_{...}) \beta$$

$$+ (V\varepsilon)_{ijt} , \text{ where } (V\varepsilon)_{ijt} \text{ is the } (i, j, t)^{th} \text{ element of } V\varepsilon.$$

Note that the left hand side of this equation is the sample estimate of the three-way interaction term Θ_{ijt}. And the term in parentheses on the right hand side is the three-way interaction term Φ_{ijt}. This is the relationship specified in (*) above, with a sample estimate for Φ_{ijt} replacing Φ_{ijt} and with a disturbance term $(V\varepsilon)_{ijt}$ included. That is, under the assumption that y_{ijt} and x_{ijt} can each be represented as a sun of common mean, effects due to single treatments, two-way interaction effects of pairs of treatments, and a three-way interaction effect, it is true that $\Theta_{ijt} = \Phi_{ijt} \beta$. Hence β can be estimated by regressing the sample estimate of the three-way interaction term Θ_{ijt} on the three-way interaction term Φ_{ijt}. This is precisely what the estimator $B = (X'VX)^{-1} X'VY$ accomplishes.

APPENDIX A

Summary Data For Alternative WM

	β	SE	t	$\hat{\sigma}^2$	R	N
Normalized Model (e)						
MG 1	1.1925	.2263	5.2694	.1467	.1930	720
MG 2	1.[illegible]	.1549	11.5222	.3351	.3091	1259
MG 3	1.8595	.1892	9.8305	.3569	.2977	996
MG 4	1.3576	.1746	7.7756	.1145	.2646	805
MG 5	1.5192	.2441	6.2194	.0998	.2613	530
Normalized Model (f)						
MG 1	1.9060	.3164	6.0225	.2869	.2193	720
MG 2	2.7494	.2168	12.6820	.4692	.3368	1259
MG 3	2.6348	.2611	10.0914	.4927	.3048	996
MG 4	2.0213	.2436	8.2973	.2229	.2810	805
MG 5	2.1748	.3446	6.3119	.1990	.2649	530
Model (g)						
MG 1	1.3514	.1263	10.6972	.7736	.3709	720
MG 2	2.1173	.0984	21.5111	.6352	.5187	1259
MG 3	1.7451	.1163	15.0006	.6082	.4296	996
MG 4	1.5527	.1313	11.8232	.6209	.3851	805
MG 5	1.5034	.1485	10.1566	.4639	.4043	530
Model (h)						
MG 1	1.8317	.1743	10.5061	.1397	.3650	720
MG 2	2.9873	.1403	21.2218	1.2905	.5148	1259
MG 3	2.2643	.1587	14.2699	1.1315	.4123	996
MG 4	2.0695	.1805	11.4596	1.1741	.3749	805
MG 5	2.0733	.2163	9.5852	.9942	.3850	530

C = 500
K = 0.15

APPENDIX A

Normalized Model (e)	B	SE	t	$\hat{\sigma}^2$	R	N
MG 1	1.2115	.2270	5.3375	.1465	.1954	720
MG 2	1.8015	.1552	11.6053	.1121	.3111	1259
MG 3	1.8736	.18[illegible]	9.89[illegible]	.1272	.2995	996
MG 4	1.3690	.1750	7.8238	.1144	.2661	805
MG 5	1.5260	.2444	6.2440	.0998	.2622	530
Normalized Model (f)						
MG 1	1.9338	.3174	6.0926	.2866	.2217	720
MG 2	2.7753	.2173	12.7719	.2198	.3389	1259
MG 3	2.6576	.2612	10.1728	.2423	.3071	996
MG 4	2.0374	.2441	8.3453	.2227	.2825	805
MG 5	2.1853	.3449	6.3344	.1958	.2658	530
Model (g)						
MG 1	1.4973	.1321	11.3273	.5896	.3894	720
MG 2	2.0903	.0981	21.3038	.6394	.5151	1259
MG 3	1.7038	.1154	14.7649	.6117	.4241	996
MG 4	1.6037	.1330	12.0543	.6172	.3914	805
MG 5	1.6154	.1524	10.5952	.6763	.4187	530
Model (h)						
MG 1	2.0308	.1824	11.1309	1.1214	.3836	720
MG 2	2.9514	.1398	21.1080	1.2964	.5116	1259
MG 3	2.2160	.1572	14.0901	1.1363	.4080	996
MG 4	2.1367	.1829	11.6767	1.1678	.3810	805
MG 5	2.2209	.2222	9.9943	.9716	.3989	530

C = 500
K = 0.20

APPENDIX A

Normalized Model (e)	B	SE	t	$\hat{\sigma}^2$	R	N
MG 1	1.2707	.2505	5.0725	.1471	.1860	720
MG 2	1.9647	.1727	11.3710	.1126	.3054	1259
MG 3	2.0504	.2107	9.7303	.1276	.2949	996
MG 4	1.4965	.1943	7.6989	.1146	.2622	805
MG 5	1.6066	.2593	6.1952	.0999	.2603	530
Normalized Model (f)						
MG 1	2.0417	.3503	5.8270	.2878	.2125	720
MG 2	3.0393	.2418	12.5681	.2206	.3341	1259
MG 3	2.0047	.2009	9.9855	.2432	.3019	996
MG 4	2.3501	.2711	8.2419	.2231	.2793	805
MG 5	2.4372	.3864	6.3059	.1990	.2646	530
Model (g)						
MG 1	1.3958	.1359	10.2704	.6049	.3579	720
MG 2	1.9450	.1032	18.8491	.6774	.4694	1259
MG 3	1.5704	.1203	13.1206	.6357	.3842	996
MG 4	1.4816	.1382	10.7192	.6377	.3538	805
MG 5	1.6241	.1742	9.2988	.4679	.3953	530
Model (h)						
MG 1	1.8968	.1874	10.1189	1.1508	.3533	720
MG 2	2.7579	.1468	18.7861	1.3713	.2190	1259
MG 3	2.0587	.1637	12.5740	1.1762	.3704	996
MG 4	1.9806	.1898	10.4323	1.2030	.3455	805
MG 5	2.2335	.2390	9.3421	.9915	.3766	530

C = 1000
K = 0.10

APPENDIX A

Normalized Model (e)	B	SE	t	$\hat{\sigma}^2$	R	N
MG 1	1.2900	.2505	5.149[illegible]	.1469	.1887	720
MG 2	1.9793	.1726	11.4625	.1124	.3076	1259
MG 3	2.0616	.2102	9.8056	.1274	.2970	996
MG 4	1.5061	.1942	7.7545	.1145	.2639	805
MG 5	1.7014	.2732	6.2264	.0998	.2615	530
Normalized Model (f)						
MG 1	1.0696	.3504	5.9062	.2874	.2153	720
MG 2	3.0609	.2416	12.6663	.2202	.3364	1259
MG 3	2.9239	.2902	10.0756	.2427	.3044	996
MG 4	2.2479	.2709	8.2966	.2229	.2810	805
MG 5	2.4428	.3856	6.3343	.1988	.2658	530
Model (g)						
MG 1	1.4027	.1308	10.6753	.5987	.3701	720
MG 2	2.1729	.1068	20.3461	.6536	.4977	1259
MG 3	1.7946	.1260	14.2394	.6195	.4604	996
MG 4	1.7127	.1460	11.7296	.6223	.3825	805
MG 5	1.7180	.1672	10.2729	.4622	.4081	530
Model (h)						
MG 1	2.024	.1929	10.4926	1.400	.3646	720
MG 2	3.0742	.1520	20.2180	1.3250	.4954	1259
MG 3	2.3360	.1716	13.6097	1.1491	.3963	996
MG 4	2.2857	.2007	11.3862	1.1762	.3728	805
MG 5	2.3576	.2436	9.6745	.9814	.3880	530

C = 1000
K = 0.15

	B	SE	t	$\hat{\sigma}^2$	R	N
Normalized Model (e)						
MG 1	[illegible]	.2505	5.2208	.1468	.1912	720
MG 2	[illegible]	.1725	11.5500	.1122	.3098	1259
MG 3	[illegible]	.2007	9.8784	.1272	.2990	996
MG 4	1.5150	.1940	7.8081	.1144	.2656	805
MG 5	[illegible]	[illegible]	6.2542	.0997	.2626	530
Normalized Model (f)						
MG 1	2.0951	.3503	5.9806	.2871	.2178	720
MG 2	[illegible]	[illegible]	12.7605	.2108	.3386	[illegible]
MG 3	[illegible]	[illegible]	[illegible]	[illegible]	.3068	[illegible]
MG 4	[illegible]	[illegible]	[illegible]	.2227	.2806	[illegible]
MG 5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Model [illegible]						
MG 1	1.5415	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MG 2	[illegible]	.1093	21.0550	.6423	.5106	[illegible]
MG 3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MG 4	1.6880	.1449	11.6465	.6236	.3801	[illegible]
MG 5	1.6534	.1645	10.0470	.4656	.4006	530
Model (h)						
MG 1	2.1240	.1971	10.7744	1.1310	.3731	720
MG 2	3.2308	.1542	20.9402	1.3018	.5086	1259
MG 3	2.4771	.1753	14.1299	1.3526	.4080	996
MG 4	2.2555	.1992	11.3228	1.1780	.3710	805
MG 5	2.2735	.2396	9.4879	.9871	.3817	530

C = 1000
K = 0.20

APPENDIX A

	B	[illegible]	[illegible]	$\hat{\sigma}^2$	[illegible]	[illegible]
Normalized Model (a)						
MG 1	1.3667	.2752	4.9663	.1473	[illegible]	720
MG 2	2.1564	.1998	11.3005	.1127	.2032	1250
MG 3	2.2499	.2322	9.6802	.1222	.2232	996
MG 4	1.6454	.2143	7.6762	.1117	.2615	805
MG 5	1.8821	.3037	6.1969	.0999	.2604	530
Normalized Model (b)						
MG 1	2.2035	.3849	5.7241	.2502	.2029	720
MG 2	3.3465	.2670	12.5228	.2207	.3333	1250
MG 3	3.1806	.3205	9.9509	.2433	.3010	996
MG 4	2.4623	.2990	8.2352	.2222	.2791	805
MG 4	2.7083	.4235	6.3106	.1989	.2652	530
Model (a)						
MG 1	1.4147	.1445	9.2898	.6121	.3432	720
MG 2	1.0654	.1103	17.9150	.6037	.4490	1250
MG 3	1.5544	.1268	12.2506	.6430	.3622	996
MG 4	1.4505	.1452	9.9894	.6494	.3325	805
MG 5	1.5357	.1710	8.9769	.4812	.2639	530
Model (b)						
MG 1	1.0260	.1092	9.6675	1.1634	.3394	720
MG 2	2.7013	.1568	17.2028	1.4026	.4496	1250
MG 3	2.0286	.1724	11.7653	1.1966	.3496	996
MG 4	1.9408	.1993	9.7361	1.2218	.3249	805
MG 5	2.1138	.2436	8.5019	1.0163	.3470	530

C = 1500
K = 0.10

APPENDIX A

Normalized Model (e)	B	SE	t	$\hat{\sigma}^2$	R	N
MG 1	1.3857	.2745	5.0492	.1471	.1851	720
MG 2	2.1685	.1901	11.4022	.1125	.3062	1259
MG 3	2.2584	.2310	9.7726	.1275	.2961	996
MG 4	1.6521	.2136	7.7344	.1146	.2633	805
MG 5	1.8817	.3022	6.2261	.0998	.2615	530
Normalized Model (f)						
MG 1	2.2302	.3832	5.8088	.2978	.2119	720
MG 2	3.3636	.2660	12.6418	.2203	.3359	1259
MG 3	3.2051	.3189	10.0500	.2439	.3037	996
MG 4	2.4207	.2979	8.2925	.2229	.2809	805
MG 5	2.7064	.4264	6.3459	.1988	.2662	530
Model (g)						
MG 1	1.4284	.1452	9.8372	.6114	.3447	720
MG 2	2.0226	.1112	18.2414	.6870	.4574	1259
MG 3	1.6149	.1284	12.5721	.6435	.3704	996
MG 4	1.4889	.1464	10.1604	.6458	.3378	805
MG 5	1.6359	.1745	9.3745	.4755	.3777	530
Model (h)						
MG 1	1.9434	.2002	9.7072	.1623	.3406	720
MG 2	2.8814	.1582	18.2116	1.3893	.4569	1259
MG 3	2.1099	.1745	12.0862	1.1986	.3580	996
MG 4	1.9946	.2009	9.9246	1.2168	.3305	805
MG 4	2.2484	.2538	8.8579	1.0059	.3597	530

C = 1500
K = 0.15

APPENDIX A

	B	SE	t	$\hat{\sigma}^2$	R	N
Normalized Model (e)						
MG 1	1.4[illegible]	.2738	5.1229	.1470	.1878	720
MG 2	2.1[illegible]	.1895	11.4910	.1123	.3083	1259
MG 3	2.2[illegible]	.2299	9.8473	.1273	.2981	996
MG 4	1.6577	.2129	7.7864	.1144	.2650	805
MG 5	1.[illegible]12	.3[illegible]	6.2555	.0997	.2627	530
Normalized Model (f)						
MG 1	2.2[illegible]	.3[illegible]	5.8863	.2875	.2146	720
MG 2	3.3774	.2651	12.7361	.2199	.3381	1259
MG 3	3.2176	.3173	10.1384	.2425	.3061	996
MG 4	2.4775	.2969	8.3430	.2227	.2824	805
MG 5	2.7044	.4243	6.3728	.1987	.2673	530
Model (g)						
MG 1	1.5812	.1514	10.4379	.6024	.3620	720
MG 2	2.1937	.1139	19.2522	.6710	.4772	1259
MG 3	1.8396	.1349	13.6425	.6292	.3271	996
MG 4	1.6761	.1531	10.9479	.6343	.3604	805
MG 5	1.6917	.1763	9.5910	.4723	.3852	530
Model (h)						
MG 1	2.1510	.2089	10.2960	1.1457	.3587	720
MG 2	3.1147	.1620	19.2226	1.3570	.4766	1259
MG 3	2.3982	.1832	13.0960	1.1629	.3934	996
MG 4	2.2422	.2102	10.6627	1.1966	.3522	805
MG 5	2.3316	.2565	9.0893	.9991	.3678	530

C = 1500
K = 0.20

APPENDIX A

Normalized Model (e)	B	SE	t	$\hat{\sigma}^2$	R	N
MG 1	1.4623	.2996	4.8798	.1475	.1793	720
MG 2	2.3441	.2085	11.2291	.1129	.2922	1259
MG 3	2.4467	.2534	9.6531	.1278	.2028	996
MG 4	1.7923	.2342	7.6521	.1147	.2607	805
MG 5	2.0657	.3334	6.1952	.0999	.2603	530
Normalized Model (f)						
MG 1	2.3644	.4191	5.6405	.2996	.2060	720
MG 2	3.6475	.2918	12.4999	.2208	.3325	1259
MG 3	3.4706	.3498	9.9200	.2434	.3001	996
MG 4	2.6865	.3266	8.2246	.2232	.2787	805
MG 5	2.9769	.4704	6.3280	.1980	.2655	530
Model (g)						
MG 1	1.5736	.1588	9.9058	.6104	.1202	720
MG 2	2.2997	.1236	18.5965	.6814	.4645	1259
MG 3	1.8342	.1434	12.7859	.6405	.3758	996
MG 4	1.7325	.1651	10.4892	.6411	.3471	805
MG 5	1.7316	.1998	9.1191	.4791	.3689	530
Model (h)						
MG 1	2.1388	.2190	9.7635	1.1607	.3424	720
MG 2	3.2712	.1757	18.6116	1.3766	.4648	1259
MG 3	2.3940	.1950	12.2758	1.1838	.3628	996
MG 4	2.3173	.2268	10.2167	1.2029	.3392	805
MG 5	2.3818	.2760	8.6282	1.0126	.3515	530

C = 2000
K = 0.10

APPENDIX A

Normalized Model (e)	B	SE	t	$\hat{\sigma}^2$	R	N
MG 1	1.4800	.2[illegible]2	4.9[illegible]	.1473	.1821	720
MG 2	2.3526	.2[illegible]4	11.3[illegible]	.1126	.3046	1259
MG 3	2.4510	.2517	9.73[illegible]	.1276	.0251	996
MG 4	1.7960	.2328	7.71[illegible]	.1146	.2626	805
MG 4	2.0612	.3310	6.2267	.0998	.2616	530
Normalized Model (f)						
MG 1	2.3886	.4172	5.7250	.2882	.2089	720
MG 2	3.6586	.2902	12.6068	.2204	.3350	1259
MG 3	3.4804	.3474	10.0177	.2430	.3028	996
MG 4	2.6903	.3247	8.2833	.2230	.2806	805
MG 5	2.9688	.4670	6.3561	.1987	.2666	530
Model (g)						
MG 1	1.558[illegible]	.1583	9.8440	.6113	.3448	720
MG 2	2.1934	.1210	18.1156	.6890	.4550	1259
MG 3	1.8422	.1430	12.[illegible]12	.6391	.3782	996
MG 4	1.6630	.1617	10.2825	.6441	.3411	805
MG 5	1.6823	.1866	9.0130	.4806	.3652	530
Model ([illegible])						
MG 1	2.1230	.2183	9.7242	1.1618	.3411	720
MG 2	3.1213	.1720	18.1392	1.3916	.4555	1259
MG 3	2.4097	.1943	12.3971	1.1807	.3659	996
MG 4	2.2292	.2220	10.0410	1.2137	.3340	805
MG 5	2.3190	.2712	8.5504	1.0149	.3487	530

C = 2000
K = 0.15

Normalized Model (e)	B	SE	t	$\hat{\sigma}^2$	R	N
MG 1	1.[illegible]	.2069	5.0421	.1471	.1849	720
MG 2	2.3603	.2063	11.4371	.1124	.3070	1259
MG 3	2/4539	.2500	9.8156	.1274	.2973	996
MG 4	1.7995	.2316	7.7694	.1145	.2644	805
MG 5	2.0564	.3287	6.2562	.0997	.2627	530
Normalized Model (f)						
MG 1	2.4121	.4152	5.8082	.2879	.2118	720
MG 2	3.6688	.2886	12.7109	.2200	.3375	1259
MG 3	3.4884	.3449	10.1120	.2426	.3054	996
MG 4	2.6936	.3230	8.3388	.2227	.2823	805
MG 5	2.9607	.4637	6.3837	.1986	.2677	530
Model (g)						
MG 1	1.5166	.1564	9.6935	.6135	.3402	720
MG 2	2.2175	.1209	18.3367	.6855	.4594	1259
MG 3	1.8078	.1411	12.8060	.6402	.3763	996
MG 4	1.6250	.1592	10.2022	.6453	.3387	805
MG 5	1.7314	.1876	9.2293	.4775	.3727	530
Model (h)						
MG 1	2.0713	.2156	9.6069	1.1651	.3375	720
MG 2	3.1563	.1718	18.3658	1.3844	.4600	1259
MG 3	2.3672	.1918	12.3400	1.1821	.3645	996
MG 4	2.1781	.2186	9.9628	1.2158	.3317	805
MG 5	2.3819	.2727	8.7327	1.0096	.3553	530

C = 2000
K = 0.20

LIST OF REFERENCES

1. Studies Prepared for the President's Commission on An All Volunteer Armed Force, V. 1 and V. 2, November 1970.

2. Gray, B. C., Supply of First-Term Military Enlistees; A Cross-Section Analysis in V. 1 of [1].

3. Fisher, A. C., The Supply of Enlisted Volunteers for Military Service, Ph. D. Thesis, Columbia University, 1968.

4. Fisher, A. C., "The Cost of the Draft and the Cost of Ending the Draft," American Economic Review, V. 59, P. 239-254, June 1969.

5. Nelson, G. R., Economic Analysis of First-Term Reenlistments in the Army in V. 1 of [1].

6. Wilburn, R. C., Impact of Income, the Draft and Other Factors On the Retention of Air Force Enlisted Men: A Cross Section Analysis in V. 1 of [1].

7. Hogg, R. V. and Craig, A. T., Introduction to Mathematical Statistics, 2nd. ed., Macmillan, 1965.

8. Nerlove, M., "A Note on Error Components Models," Econometrica, V. 39, p. 383-401, March 1971.

9. Wallace, T. D., and Hussain, A., "The Use of Error Components Models in Combining Cross Section With Time Series Data," Econometrica, V. 37, p. 55-72, January 1969.

10. Theil, H., Principles of Econometrics, Wiley, 1971.

11. Kuh, E., "The Validity of Cross-Sectionally Estimated Behavior Equations in Time Series Applications," Econometrica, V. 27, p.197-214, 1959.

12. Chetty, V. K., "Pooling of Time Series and Cross Section Data," Econometrica, V. 36, p. 279-290, April 1968.

13. Gray, B. C., ANRAS Working Paper No. 2, prepared for the Center for Naval Analyses, 29 June 1971.

INITIAL DISTRIBUTION LIST

		Number Copies
1.	Chief of Naval Personnel Pers-11b Department of the Navy Washington, D. C. 20370	1
2.	Naval Postgraduate School Department of Operations Research and Administrative Sciences (Code 55) Monterey, California 93940	1
3.	Defense Documentation Center Cameron Station Alexandria, Virginia	2
4.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
5.	Assoc. Professor R. S. Uhler, Code 55 Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	1
6.	Lt. D. A. Altergott SMC 2471 Naval Postgraduate School Monterey, California 93940	1

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE
A Study of Navy First-Term Reenlistments

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Master's Thesis; June 1972

5. AUTHOR(S) (First name, middle initial, last name)
Dennis Albert Altergott

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 1972	95	13

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Naval Postgraduate School Monterey, California 93940

13. ABSTRACT

It occassionally happens in economic analyses that the correctly specified model contains variables for which no observed data has been collected. When the data in a linear regression model are cross-sectional it is possible, under certain conditions on the nature of the variables, to estimate the independent effects of a specific set of explanatory variables on the dependent variable. A procedure for doing this is presented.

A commonly used model of reenlistment behavior, for which the data base is cross-sectional, satisfies the requisite conditions. This permits the estimation of the independent effect of the military wage on reenlistment rate, as an illustration of the proposed procedure.

DD FORM 1473 1 NOV 65 (PAGE 1)

S/N 0101-807-6811

Security Classification

A-31408

Security Classification

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
irst-term reenlistments						
inear regression analysis						

Thesis
A4254
c.1

133845

Altergott
A study of Navy
first-term reenlistments.