DIFFERENCES BETWEEN VETERAN AND NON-VETERAN OWNED BUSINESSES

by
David Rothenberg
INDUCTIVE INFERENCE, INC.
210 Riverside Drive
New York, New York 10025

This study was conducted under SBA Contract #7215-VA-83 to Dun & Bradstreet, Inc., "Developing Key Financial and Income Statements Data Base for Veteran Studies".

TABLE OF CONTENTS

ABSTRACT

The purpose of this study is to determine which financial and industry related characteristics, if any, distinguish veteran owned businesses from non-veteran owned and Vietnam veteran owned businesses. The analysis of data samples from Dun & Bradstreet's Market Identifiers and Financial Profiles did not reveal significant financial differences between veteran owned and non-veteran owned businesses. Small financial differences between Vietnam veteran owned and other businesses were observed that appear to be related to the former group's smaller size and more recent origin. Marked differences, however, were found between the distributions of each of the three groups of businesses across industries. Most of these differences cannot be accounted for by imbalances in the growth of new businesses in each industry at the inception times of the veteran owned and Vietnam veteran owned businesses. The characteristics that account for the differences are also reflected by the distribution of Vietnam veteran employment across industries. Details are presented that may be pertinent to public policy and to the deployment of veteran assistance programs.

I. PURPOSE OF STUDY

The purpose of this study is to determine those financial and industry related characteristics, if any, that distinguish veteran owned businesses from non-veteran owned businesses, and Vietnam veteran owned businesses from other veteran owned businesses. It is intended that the results of this study will provide insight into the effects of various aspects of public policy upon veterans and will help assess programs designed to assist veterans.

II. BACKGROUND AND SIGNIFICANCE OF PROBLEM

Although studies of the distribution of Vietnam veteran employment have been made (Ref. 1), little is known about the distribution of veteran owned businesses. More precisely targeted use of the resources of the variety of federal veteran assistance programs could result from knowledge of the characteristics of veteran owned businesses. Vietnam veterans and other veterans became entrepreneurs at different times, with different training, both in and out of the services, and with different resources at their disposal. Do these groups of veterans differ from other entrepreneurs at the same time? Do they differ with respect to the businesses they formed or with respect to their need for and use of capital? Are they in different industries than other businesses? Are they more or less liquid,

profitable, or managed in a different style? These questions are pertinent to the effective deployment of veteran training programs, small business loans and others aspects of public policy.

III. DATA SOURCES AND ADEQUACY

Financial statements for three random samples of firms were required. These were non-veteran owned businesses, Vietnam veteran owned (i.e., more than 50% owned) businesses and other veteran owned businesses. Each group was to contain 1750 firms.

To this end respondents to the SBA Ownership Characteristics Survey (OCS), which was based on a 3% sample of Master Establishment Records (MEL), was matched against Duns Financial Profiles (DFP) files in order to obtain financial statements wherever available.

Each business in the file of OCS respondents (i.e., 10% of the sample of MEL records) was then matched against the Duns Market Identifiers (DMI) files, which contain the Standard Industry Code (SIC), sales and number of employees for about four (4) million firms. Note that all firms for which financial statements are available (in the DFP file) are also in the DMI file (although the reverse is not true).

Examination of the portion of firms that was found on the DMI files revealed that the sample was too small. Another 3% sample was then drawn from the DMI files and the Dun and Bradstreet Information

Reports were searched to determine which of these businesses fell into each of the three ownership categories. There were still too few businesses in some of the categories.

The above sampling and searching procedure was therefore repeated until the following (still below target) number of firms in each of the three categories were obtained. The collection of these firms will be referred to as File A:

Non-veterans - 10,355 businesses
Other veterans - 3,430 businesses
Vietnam veterans - 1,112 businesses

Dun and Bradstreet (D & B) then matched the DFP files against the businesses in File A, and the following numbers of firms with financial statements in each of the categories resulted. This collection will be referred to as File B:

Non-veterans - 5,106 businesses
Other veterans - 1,550 businesses
Vietnam veterans - 461 businesses

Although this sample is more than adequate for conventional statistical analyses where a hypothesis is being tested, a different situation applies here. The data are, in fact, being "ransacked" in search of a hypothesis. That is, we are examining a large number of

characteristics of the firms in each of the categories in the hope that some one or combination of characteristics will distinguish at least two of the three categories.

The probability that, when testing many characteristics against the data, some will spuriously appear to distinguish the classes, increases with the number and (mutual) independence of these characterisitics. Fortunately, characteristics defined in terms of a large number of ratios of financial statement items are not independent of one another, and such independence decreases as the number of ratios increases.

The adequacy of the sample, in this case, depends upon the number of characteristics (and combinations of characteristics) considered, the mutual independence of these characteristics, and upon the number of firms with each characteristic that are in each of the three subsamples. The latter consideration makes it impossible to determine the adequacy of the samples in advance of the analysis to be performed.

It should also be noted that the above data gathering procedure does not guarantee a random sample from the population of all firms, although it is unclear what biases might have been introduced. Note, however, that 7.5% of File A are Vietnam veteran owned businesses, while 7.8% of the work force are Vietnam veterans (Ref. 1, page 279), which is not unreasonable (i.e., it implies that a Vietnam veteran is about as likely to start a business as anyone else in the work force).

IV. METHODOLOGY

The purpose of this study was to examine File B, above, and determine which, if any, characteristics could be defined in terms of the available financial and business information that would distinguish the non-veteran owned businesses, Vietnam veteran owned businesses, and other veteran owned businesses from one another.

A. Formation of Variables

It is difficult to form hypotheses about how the financial characteristics of veteran owned businesses and non-veteran owned businesses might significantly differ. Such differences as might be consequences of pre-induction characteristics, in-service training or inception dates of veteran owned businesses and remain evident after a lapse of many years, are likely to be extremely subtle (except, possibly, the differences between the more recently formed Vietnam veteran owned businesses and other businesses).

Using a conventional discriminant analysis requires a hypothesis about which financial variables and what algebraic function of these variables would capture the essential differences between the veteran owned, Vietnam veteran owned and non-veteran owned businesses. We could find no reasonable basis for such a hypothesis. A "stepwise" discriminant analysis, which is an automated "ransacking" of the data (see previous section), would bring the reliability of small differences into question.

FOR THESE REASONS,

(a) a very large collection of (not mutually independent) variations of standard financial variables were formed, and

(b) an unconventional discriminant analysis that is especially sensitive to small differences and is designed for preserving reliability when data are ransacked (see Appendix A), was used.

The reasons for considering the particular variables that are defined below will be discussed later. Note that the analytic technique is designed to deal with a very large number of variables that are not mutually independent. Also, in this analysis, redundancies resulting from the inclusion of non-independent variables increases, rather than reduces, the reliability of the results (see Appendix A).

A set of variables were formed from the financial and other information by the following techniques:

(1) A collection of financial quantities (for example, sales, total assets, etc.) and financial ratios that have appeared in the literature of financial analysis (Refs.3-10) and strategic planning were computed for each firm. These quantities will be referred to as raw variables.

(2) D & B provides, as part of its DFP records, "sel norms" for some of the above ratios. A "sel" is a collection of DFP firms defined by a four digit SIC, geographic region and size range of assets. These "sel norms" provide a median, upper quartile (boundary) and

lower quartile for each available ratio in each sel. For each business and for each ratio for which such norms were provided we defined the deviation of that ratio as:

$$[x - M_s(x)]/[M_s(x) - U_s(x)] \text{ when } x \geq M_s(x)$$

$$[x - M_s(x)]/[L_s(x) - M_s(x)] \text{ when } x < M_s(x)$$

where $M_s(x)$ denotes the median of ratio x for the sel in which the business lies, U_s denotes the corresponding upper quartile, and $L_s(x)$ denotes the corresponding lower quartile.

(3) For another collection of ratios [among the raw variables (1) above], D & B provides sel means. We defined normalized ratios (for each such ratio) as:

$$x/N_s(x)$$

where $N_s(x)$ denotes the mean of ratio x for the sel in which the business lies.

(4) "Pseudo-means" were approximated for the remainder of the raw variables by algebraic manipulations of the sel medians and means. Normalized variables for each of these variables were defined as in (3) above, where "$N_s(x)$" denotes the pseudo-mean.

(5) For each of the raw variables a growth was defined as

$$x_t/x_{t-1}$$

where x_t is the value of the variable in the most recent year of available information and x_{t-1} is the value of that variable in the previous year.

(6) For each raw variable a normalized growth was defined as:

$$[x_t/x_{t-1}]/[N_s(x_t)/N_s(x_{t-1})]$$

where $N_s(x_t)$ is the median of x_t, if such is given, else the mean or pseudo-mean (of the sel) of x_t. ($N_s(x_{t-1})$ is, of course, similarly defined for the previous year.)

(7) Corresponding to each of the sel medians, upper and lower quartiles and means, a corresponding quantity for each "sector" was obtained from D & B. A sector is a collection of businesses defined by a two digit SIC and geographic region. Pseudo means were calculated as before. Each of the deviations, normalized variables and normalized growths defined in (2) - (6) above were defined for sectors as well as for sels. These variables will be referred to as sector normalized variables, sector deviations and sector normalized growths.

(8) In addition, both the sel and sector medians (or means or pseudo-means), for the sel and sector in which each firm lies, are used as variables. These will be referred to as norms.

(9) In order to capture the effects of business size upon the financial ratios, additional variables were formed by dividing each sel norm by its corresponding sector norm.

The different types of variables formed above are obviously not independent of one another (i.e., uncorrelated). The reason for forming such an exhaustive collection of variables is the hope that some of the transformations (i.e., normalizations) of the raw ratios would characterize businesses regardless of their industry, size or geographic location. We might thereby avoid the fragmentation of the sample by characteristics that would otherwise have different significances within differing industries, size groups, or geographical regions.

Normalization by sectors as well as sels was done for two reasons: the first is that many sel norms are based upon a small number of firms and therefore vary grossly from year to year when the identities of these firms change. Another is that, since sectors make no distinctions by size of business, some of the sector normalized variables may act as crude measures of market share, investment intensity, etc. (e.g., sales of the firm divided by the sector norm of sales, and the sector norm of fixed assets divided by the sector norm of total assets, respectively).

Norms (and ratios of sel to sector norms) are included as variables in case differences in financial characteristics of the firm's industry memberships, rather than their individual financial characteristics, prove significant.

A list of all variables included in the analysis is given in Appendix B. These variables will be referred to as candidate variables.

B. Primitive Charateristics

A primitive characteristic is defined as a statement that the value of a particular candidate variable lies within a certain segment (i.e., interval) of its range, or does not lie within a given segment of its range; for example,

$$a \leq x < b, \quad x \geq c, \quad x < d$$

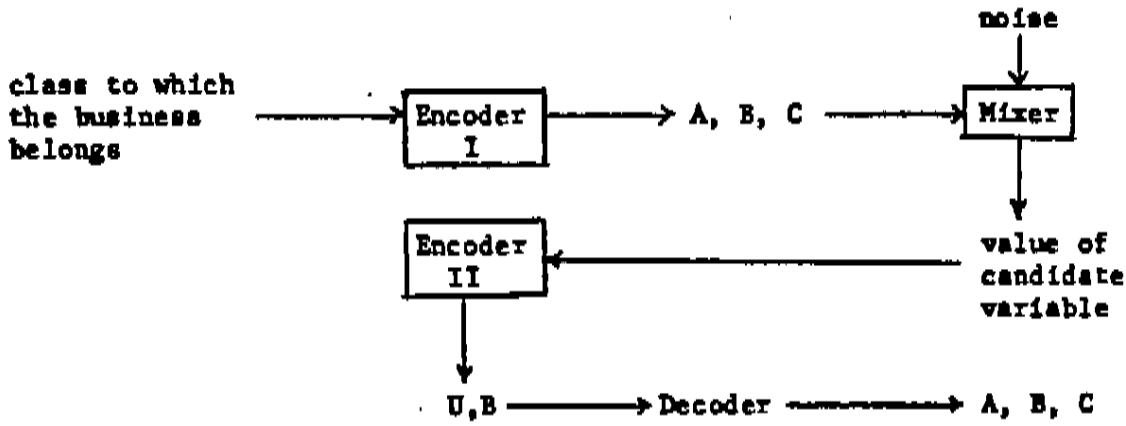
$$\text{not}(a \leq x < b), \quad \text{not}(x \geq c), \quad \text{not}(x < d)$$

where x is a candidate variable and a,b,c, and d are constants.

Primitive characteristics are constructed by the following technique to be called segmentation, which is performed separately for each candidate variable:

The combination (i.e., union) of the three subsamples (i.e., veteran owned businesses, Vietnam veteran owned businesses and other businesses) is first partitioned into two parts, one consisting of that portion for which the value of the candidate variable lies below a given boundary value, and the other consisting of the remainder. Each firm in the combined sample has a value of the candidate variable that lies either above or below this boundary value. Such a firm also lies in one of the three subsamples (which are drawn from three corresponding populations, each of which will be called a class). The boundary value is selected so the knowledge that a firm's candidate variable lies above (or below) each such boundary is maximally informative about the identity of the class to which it belongs.

More precisely, an information theoretic model (see the diagram below) is used. It is supposed that, for a given firm, the identity of the sample to which it belongs is encoded and transmitted. (The three classes will here be encoded as A, B and C.) This information is then mixed with "noise" so that the encoded information is replaced by the value of the candidate variable for that firm. This value is then replaced by "U" or "L"; which denotes that the value lies in the interval above, or the interval below the boundary value, respectively. This boundary value is chosen so that, given U (or L), A, B or C can be most often correctly determined. That is, $\{U,L\}$ gives maximal information, in the technical sense, about $\{A,B,C\}$.



In the above diagram Encoder II is determined by the selection of the partition boundary. A partition of the combined sample into many parts would merely increase the size of the alphabet generated by Encoder II.

The combined sample is, in fact, partitioned into many parts so as to maximize discrimination between the three classes. This analysis, in effect, selects the cell boundaries of a one dimensional contingency table. This is done so that the contingency table maximally captures the differences between the three classes.

The analysis operates by first partitioning the range of each candidate variable into two parts, then three, then four, etc. The successive partitions are restricted so that each is a subdivision of some interval resulting from the previous partition. The increase in information resulting from each successive partition is calculated. The partitioning stops when either this increase in information becomes negligible or when five partitions have been produced (see discussion preceding Appendix C).

The partitioning of each candidate variable is, therefore, a non-parametric univariate discriminant analysis that produces, on the basis of that variable, maximal separation between the three classes.

For better separation, two classes were first discriminated, one consisting of all veterans and the other of non-veterans. Then the two subclasses of the former class, one consisting of Vietnam veterans and the other of the remaining veterans, were discriminated.

C. Compound Characteristics

The above is the first step in a more complex discriminant analysis. Note that a statement that some candidate variable does or does not lie within one of the intervals produced by partitioning is one of the "primitive characteristics" defined above. The next step in this discriminant analysis is to form combinations (i.e., conjunctions) of primitive characteristics, such as:

$$a < x \leq b \quad \underline{\text{and}} \quad \text{not}(c \leq y < d) \quad \underline{\text{and}} \quad z < e$$

where x, y and z are candidate variables and a, b, c, d and e are constants (i.e., interval boundaries).

Such a combination of primitive characteristics will be called a compound characteristic. A pattern is defined as either a primitive or a compound characteristic. When two classes are discriminated, these patterns are formed such that every business in each class

satisfies at least one such predictor. (That is, the population is "covered.") We will refer to the set of businesses that satisfy a predictor as a cell. These cells, unlike the cells in a contingency table, overlap (i.e., a single business may lie in several distinct cells).

These cells are formed so that a business's membership in one of the classes being discriminated may be most effectively determined by its cell memberships. If the intersections of cells are also considered as cells, this analysis may then be considered to be a multidimensional extension of the above (segmentation) analysis.

D. Reliability

The collection of patterns that covers one of the classes (e.g., Vietnam veterans) "models" that class. Similarly, each pattern models a subset of the population. When a variety of possible models is tested against a single (all classes combined) data sample, results are notoriously unreliable. (See Ref.2 for a dramatic example). That is, because there are many possible models, one is likely to succeed in discriminating the samples from each class, but not the corresponding populations.

The analysis discussed above avoids this problem by using two techniques. One of these, "cross-validation," is standard. The other, a non-standard technique, to be described, proved more sensitive in the analysis performed in this study.

Cross-validation was here performed by dividing the data sample into two parts, one consisting of a random sample three quarters of the size of the original sample (to be called the analysis sample) and the remainder (to be called the test sample). Models were derived using only the data in the analysis sample. The results were then tested against the data in the test sample. The results of this test are "unbiased." This technique, however, depletes the sample used to derive the model by reserving a portion of the data for testing purposes (an exception is the "leaving one out method" used with linear models). Furthermore, the assumption underlying this technique is that the analysis and test samples are independent random samples. This, in practice, is often not the case because they both share biases that were inadvertently introduced in the selection of their parent sample.

The following describes the second of the two techniques which were used in this study to ensure reliability:

Consider first the segmentation portion of the above analysis, which divides the range of each candidate variable into non-overlapping intervals. Consider a particular candidate variable, X, and interval, I. Two primitive charateristics are defined by this interval:

(1) $X \in I$ and $X \in \bar{I}$

where $\bar{I}$ denotes the complement of I and $\in$ abbreviates "is an element of." Denote two classes to be discriminated, veteran owned and non-veteran owned businesses, by V and N, respectively. (Note that $N = \bar{V}$, the complement of V.) Let p(V), p(N) and p(I) denote the fraction of elements of the sample that are in class V, in class N and in interval I, respectively. Let p(V/I) denote the fraction of elements that are in I that are also in sample V. Then p(V/I)/p(V) measures how much more (or less) than the fraction of V in the sample lies in I. $p(V/\bar{I})/p(V)$ provides a similar measurement for the complement of I, $\bar{I}$, and $p(\bar{V}/I)/p(\bar{V})$ and $p(\bar{V}/\bar{I})/p(\bar{V})$ provides similar measurements with respect to the subsample, $N = \bar{V}$. Let p(V,I) denote the fraction of the combined sample that is in subsample V and in interval I, and similarly for $p(V,\bar{I})$, $p(\bar{V},I)$ and $p(\bar{V},\bar{I})$. Then the weighted geometric mean of the above four measures,

$$(2) \qquad p(V,I)\log[p(V/I)/p(V)] + p(V,\bar{I})\log[p(V/\bar{I})/p(V)] + p(\bar{V},I)\log[p(\bar{V}/I)/p(\bar{V})] + p(\bar{V},\bar{I})\log[p(\bar{V}/\bar{I})/p(\bar{V})]$$

is a measure of how well the primitive characteristics, (1) above, discriminate between sample V and sample N.

We therefore define the <u>score</u> of each of the primitive characteristics $X \in I$ and $X \in \bar{I}$ by this formula, (2), above.

Since the range of each candidate variable is partitioned into several intervals (say I_1, I_2..., I_n) we average over all of these, and (2) becomes:

(3) $$\sum_{j=1}^{n} (p(V,I_j)\log[p(V/I_j)/p(V)] + p(\bar{V},I_j)\log[p(\bar{V}/I_j)/p(\bar{V})])$$

which is defined as the information provided by the segmentation of cadidate variable, X. (This is the same as the information theoretic measure discussed earlier.)

Suppose a particular primitive characteristic, Q, has score, S. Consider all those primitive characteristics whose score is equal to or greater than S and that are satisfied by at least as many sample elements as satisfy Q. For any of these there is a calculable probability that its score is spuriously high, i.e., that it is purely accidental and reflects nothing about the populations from which the samples from each class were drawn. This probability, to be called the spuriousness of primitive characteristic, Q, depends upon its score, the fraction of the sample that satisfies it, the number of candidate variables considered, their mutual independence and the size of the samples from each class. An estimate can be obtained by the following simulation, which destroys the relation between the samples from the classes to be discriminated and the values of the primitive characteristics:

Randomly assign the elements of the combined sample to V and N (disregarding their original membership) so that these proportions, p(V) and p(N), remain unchanged. Call the result of this reassignment a synthetic sample.

Perform the same segmentation analysis that was performed on the original samples on this synthetic sample. Consider a particular primitive characteristic, Q, with score, S, that is satisfied by K sample elements. Count the number of (paired) primitive characteristics with a score no less than S and that are satisfied by at least K sample elements.

Form another, distinct, synthetic sample. Repeat the analysis and counting procedure many times and form the distribution of the number of primitive characteistics with scores no less than S that are satisfied by at least K sample elements.

Suppose that N primitive characteristics have a score, in the original sample, that is equal to or greater than S and are satisfied by at least K points. The spuriousness of primitive characteristic, Q, is then estimated by the fraction of synthetic samples in which N or more primitive characteristics meet these same conditions.

Obviously, the above information destroying procedure is expensive to simulate. Spuriousness is therefore estimated by algebraic means.

Suppose a primitive characteristic has a high score, s, in the original combined sample. If its spuriousness is acceptably low we would conclude that the primitive characteristic discriminates well between the classes, V and N.

Clearly, the score, s, is favorably biased (i.e., probably too high) because the primitive characteristics with the highest scores are selected from among many possibilities. "Bayesian" methods exist for removing this bias, but were not used in this study because of the marginality of the results obtained (to be discussed).

A similar, but more complex analysis, is done for the multidimensional analysis that forms compound characteristics:

All possible patterns (i.e., either primitive or compound characteristics) are organized into a sequence of complexity classes so that each contains more patterns than any of its predecessors. That is, one pattern is defined as "less complex" than another if it involves fewer primitive characteristics and has fewer variables in its component primitive characteristics and also is satisfied by a greater number of sample elements. Starting with the smallest of the classes, the spuriousness (defined only over patterns in the class) of the highest scoring patterns in that class is calculated. (Since patterns overlap, their scores are defined somewhat differently than those of primitive characteristics.) This score is then reduced, using "Bayesian" methods, in order to eliminate bias. Such adjusted scores of the highest scoring patterns generally increases and then

decreases as the complexity classes grow larger. The "best" patterns (i.e., those with the highest adjusted scores) lie at such turning points. For a more detailed discussion, see Appendix A.

V. PROCEDURE AND RESULTS

A. Analysis of Financial Data: Veteran versus Non-veteran owned Businesses.

The 2011 veteran owned businesses in File B were combined with a random selection of 2011 non-veteran owned businesses from File B (see discussion of data, above). The analysis sample consisted of three quarters of this combined sample, and the remainder formed the test sample. Both samples contained equal numbers of veteran owned businesses (to be abbreviated by "VOB") and non-veteran owned businesses (abbreviated by "NVOB").

The segmentation analysis was performed on the analysis sample and the results were validated by examination of the test sample. That is, the (favorably biased) scores of the primitive characteristics in the analysis sample were compared with their (unbiased) scores in the validation sample.

No clear and reliable results were obtained, except that more VOB's are, on the average, smaller. All observed differences between VOB's and NVOB's were marginal and few survived validation. A rough calculation of the spuriousness (see previous section) of the higher

scoring primitive characteristics indicated that most had significant likelihood of not (even marginally) distinguishing the classes. Furthermore, when primitive characteritics formed from normally (i.e., logically) correlated variables showed marginal differences between VOB's and NVOB's, these differences were often inconsistent. The possibility, however, remained that VOB's and NVOB's differ with respect to their industry memberships (to be discussed). The results of the above analysis of financial data are shown in Appendices C and D.

The multidimensional analysis described in the previous section was attempted, but (unsurprisingly) no positive results were obtained.

B. Analysis of Financial Data: Vietnam Veteran versus Other Veteran owned Businesses.

Because File B contained only 461 Vietnam veterans, the intended segmentation analysis was modified in order to eliminate the bias that results from the freedom to select interval boundaries (i.e., primitive characteristics). The interval boundaries that resulted from the above analysis (to distinguish VOB's from NVOB's) were used. The sample of veteran owned businesses was then distributed across these intervals, and the fraction of Vietnam veteran owned businesses (abbreviated by VVOB) in each such interval was calculated. The scores (as defined in the previous section) of each of the

corresponding primitive characteristics were then calculated, as was the information contained in the segmentations of their corresponding variables.

Two clear results were that VVOB's were, as expected, smaller and younger than other veteran owned businesses (OVOB's). Otherwise only very marginal differences were observed. That is, the number of VVOB's and OVOB's that satisfied a primitive characteristic typically differed by only a few percent (see Appendix E). Since many primitive characteristics were examined, but only those that (marginally) discriminated VVOB's from OVOB's are listed below, a favorable bias on their scores exists. A crude estimation of their spuriousness (as defined in the previous section, but adjusted because interval boundaries could not be freely selected) indicates that approximately 15% of the statements below are probably not valid (for the population of VOB's rather than the sample):

(1) VVOB's are marginally more liquid. (Quick ratio and current ratio were used as measures of liquidity.)

(2) VVOB's are in industries with marginally higher average liquidity.

(3) VVOB's have a marginally higher debt to equity ratio.

(4) VVOB's are in industries with a very slightly lower average debt to equity ratio.

(5) VVOB's have a marginally higher ratio of fixed assets to equity.

(6) VVOB's have a very slightly higher return on equity.

(7) VVOB's are in industries with a slightly higher average return on equity, but appear to have a slightly lower than average return within their industries.

(8) VVOB's have a very slightly lower return on sales.

(9) VVOB's are in industries with a very slightly higher average return on sales.

(10) VVOB's have a marginally lower sales turnover.

(11) VVOB's are in industries with a very slightly higher sales turnover.

(12) VVOB's have very slightly lower assets per employee.

(3), (4), (5) and (12) indicate that VVOB's, in addition to being smaller (in assets) than OVOB's, are, slightly higher in debt relative to equity.

(8), (9), (10), and (11) indicate that VVOB's have a slightly lower sales turnover, despite their (on the average) smaller assets per employee.

(6) indicates that VVOB's have a slightly above average return on equity, perhaps because of a lower equity to assets ratio. This is supported by the fact that VVOB's do not have above average return on assets. (3) and (5) also support this conjecture.

(1), (2) (6), and (7) indicate that VVOB's are differently distributed over industries than OVOB's.

A multivariate analysis produced no reliable patterns that were not subsumed by the above.

As might be expected, VVOB's appear to be smaller, younger, more in need of capital, and in different industries than other businesses.

The sample of VVOB's was apparently too small to significantly affect the previously described attempt to discriminate VOB's from NVOB's (even though the VVOB's were included among the VOB's).

The results of the analysis of financial data prompted an analysis of the distribution of File B across industries.

C. Analysis of Industry Membership

In this analysis industry was represented by a four digit Standard Industrial Classification (SIC). The sample of 2,011 VOB's was combined with a randomly selected equal number of NVOB's. The number of businesses (in this combined sample) in each four digit SIC and the

fraction of such businesses that were VOB's were determined. In order to ensure reliability, when fewer than twenty businesses were in the SIC, this fraction was not calculated. So that this information would not be completely disregarded, a similar analysis was performed for three digit SIC's, then for two digit SIC's, and finally for major industry categories.

The above analysis was then repeated to discriminate the 461 VVOB's and 1,550 OVOB's in File B.

The results of these analyses appeared to show significant differences, both between the VOB's and NVOB's and between the VVOB's and OVOB's. Therefore, in order to capture differences at a finer SIC level, similar analyses were performed to discriminate

(a) the 4,542 VOB's in File A from a randomly selected approximately equal number (4,553) of NVOB's in File A and

(b) the 1,112 VVOB's in File A from the 3,430 OVOB's in File A (see Appendices F and G, respectively).

The combined sample of 9,095 businesses in (a) above will be called sample A and the combined sample of 4,542 businesses in (b) will be called sample B. The principal results of the analyses of both these samples from File A are summarized below. Since File B is a subset of File A, the results of the prior identical analyses of File (not sample) B are not discussed.

The distributions of Vietnam veteran owned businesses (VVOB's) and all veteran owned businesses (VOB's) across different major industry categories (to be abbreviated by "MIC's") will be compared with one another and with the distributions of businesses and business growth in the years 1978-82. In order that the comparisons be transparent, the following transformations of relevant data are performed:

Let $p(VOB/MIC_i)_A$ denote the fraction of VOB's in Sample A that are in MIC i and $p(VOB)_A$ denote the fraction of VOB's in Sample A that are VOB's. When A is a random sample from the entire population, the subscript will sometimes be omitted. Then $p(VOB/MIC_i)/p(VOB)$ measures how much larger a fraction of VOB's are in MIC i than in all MIC's combined. This quantity will be denoted by $M(VOB,MIC_i)$ and will be called the marginal concentration of VOB's in MIC i . Note also that, if $p(MIC_i/VOB)_A$ denotes the fraction of VOB's in Sample A that are in MIC i and $p(MIC_i)_A$ denotes the fraction of all businesses in Sample A that are in MIC i,

$$p(MIC_i/VOB)_A/p(MIC_i)_A = p(VOB/MIC_i)_A/p(VOB)_A$$

The above notations will be used with similar meanings when "VOB" is replaced by "VVOB", and when Sample A is replaced by Sample B.

Since the sample of VOB's and NVOB's that were discriminated consisted of approximately equal samples of each, rather than a random selection of small businesses, we must transform $M(VOB,MIC_i)$ so as to apply to a random sample rather than the sample used. Denote the (mythical) random sample by R and the sample used by A. If we assume that the VOB's in A are randomly selected from

among the VOB's in R and that the NVOB's in A are randomly selected from among the NVOB's in R,

$$p(MIC_i/VOB)_R = p(MIC_i/VOB)_A \quad \text{and}$$

$$p(MIC_i/NVOB)_R = p(MIC_i/NVOB)_A$$

where the subscript, R or A, indicates the sample in which the quantity is computed. Then

$$p(MIC_i \text{ and } VOB)_R = p(MIC_i/VOB)_A \, p(VOB)_R \text{ and, since}$$

$$M(VOB,MIC_i) = p(MIC_i \text{ and } VOB)_R/(p(MIC_i)_R \, p(VOB)_R),$$

$p(VOB)_R$ is eliminated after substitutions and

$$M(VOB,MIC_i) = (p(VOB/MIC_i)_A/p(VOB)_A)(p(MIC_i)_A/p(MIC_i)_R$$

$p(MIC)_R$ in 1982 can be calculated from the chart on page 65 of Ref.1, and the analysis of sample A provides $p(VOB/MIC_i)_A$, $p(VOB)_A$ and $p(MIC_i)_A$ (See Appendix F).

Figure (1) shows the calculations of $M(VOB,MIC_i)$ for all major industry categories. (In Figures 1-5, the MIC's are abbreviated as follows: Agriculture, Forestry and Fishing by AFF; Construction by Constr.; Manufacturing by Manuf.; Transportation, Communication & Other Utilities by TC&U; Wholesale by Wholes.; and Finance, Insurance & Real Estate by FIRE).

Note again that the value of the marginal concentration of VOB's $M(VOB,MIC_i)$, indicates the degree to which MIC i has a greater or lesser fraction of VOB's than all MIC's combined. For convenience, the concentration of VOB's will be said to be above average when the marginal concentration of

VOB's exceeds one, and below average when it is below one.

FIGURE 1

CALCULATION OF THE MARGINAL CONCENTRATION OF VOB'S IN EACH MIC

MIC	$p(VOB/MIC)_A$	$\frac{p(VOB/MIC)_A}{p(VOB)_A}$	$p(MIC)_A$	1982 $p(MIC)_R$	$\frac{p(MIC)_A}{p(MIC)_R}$	M(VOB,MIC)
All	.4994	1.0000	1.0000	1.0000	1.0000	1.0000
AFF	.4979	.9970	.0269	.0283	.9293	.9265
Mining	.5441	1.0895	.0075	.0083	.9036	.9845
Constr.	.5595	1.1203	.1228	.1424	.8624	.9661
Manuf.	.4742	.9495	.1085	.0863	1.2572	1.1937
TCdU	.5051	1.0114	.0322	.0354	.9096	.9200
Wholes.	.5146	1.0304	.1434	.0998	1.4369	1.4806
Retail	.4719	.9449	.2736	.2899	.9438	.8918
FIRE	.4890	.9792	.0549	.0756	.7262	.7111
Services	.5045	1.0102	.2219	.2341	.9479	.9576

We received the data that was analyzed in December,1984 and therefore it contained, at the latest, 1983 statements. Many of the statements were for earlier years. Data is available that specifies growth in the number of enterprises in each MIC in 1978-80 and 1980-82 (Ref.1, Pg.64) and the number of enterprises in each MIC in 1982 (Ref.1, Pg.65). 83% of the businesses in sample A had been in business less than four years and 36% were less than two years old. 86% of the VOB's were less than four years old and 41% were less than two years old. Comparison of marginal concentrations of VOB's in each MIC with growths in number of enterprises in each MIC in the years 1978-1980 and 1980-1982 was therefore appropriate.

Although most of the businesses in sample A were started in 1978 or later, we used growths in number of enterprises rather than startups because

(a) The death rate continuously depleted the number of startups between 1978

and 1982, and therefore is relevant to the number of businesses extant in 1983, and

(b) Since growth equals births minus deaths, growths provide a more conservative measure for comparison with marginal concentrations of VOB's than startups.

Note that establishment births are, in effect, being used as a proxy for new business formations (see Ref.11 p. 123, footnote).

Let p(s-start 80-82) denote the fractional growth in number of enterprises between 1980 and 1982 and p(s-start 80-82/MIC$_i$) denote such fractional growth in number of enterprises in MIC i. ("s-start" may conveniently be considered as abbreviating "surviving startups".)

Define M(growth 80-82,MIC$_i$) as p(s-start 80-82/MIC$_i$)/p(s-start 80-82) and call it the concentration of growth in MIC i . It measures how much larger a fractional growth occurred in MIC i than in all MIC's combined and is therefore suitable for comparison with M(VOB,MIC$_i$). Similar notation will be used for the years 1978-80.

Figure 12) shows the calculations of M(growth 80-82,MIC) and M(growth 78-80,MIC) for all MIC's.

In order to facilitate comparison, we also define p(MIC$_i$) as the fraction of enterprises in MIC i and M(MIC$_i$) as p(MIC$_i$) divided by the average of p(MIC) over all MIC's. M(MIC$_i$), to be called the marginal concentration of all businesses in MIC i, therefore measures how much larger a fraction of

businesses are in MIC i than the average over MIC's.

FIGURE 2

CALCULATION OF THE MARGINAL CONCENTRATION OF GROWTHS IN EACH MIC

MIC	p(MIC) 1980	p(MIC) 1982	p(s-start,MIC) 1978-80	p(s-start,MIC) 1980-82	M(growth/MIC) 1978-80	M(growth/MIC) 1980-82
All	1.0000	1.0000	6.300	10.000	1.000	1.000
AFF	.0273	.0283	.099	.142	1.571	1.42
Mining	.0071	.0083	.131	.266	2.079	2.66
Constr.	.1442	.1424	.058	.087	.921	.87
Manuf.	.0886	.0863	.049	.071	.778	.71
TC&U	.0350	.0354	.087	.114	1.381	1.14
Wholes.	.1008	.0998	.068	.089	1.079	.89
Retail	.3015	.2899	.030	.058	.476	.58
FIRE	.0734	.0756	.048	.132	.762	1.32
Services	.2220	.2341	.095	.160	1.508	1.60

Similar measures were calculated for VVOB's as were calculated (above) for VOB's. A complication exists because the fraction of VVOB's was computed using sample B, which contains only VOB's, rather than Sample A, which contains NVOB's as well (See Appendix G). This fraction therefore is the fraction of VOB's that are VVOB's and are in MIC i to be denoted by

$$p(VVOB/MIC_i \& VOB)_B ,$$

rather than the fraction of VVOB's that are in MIC i.

We therefore calculate

$$p(VVOB/MIC_i)_B = p(VOB/MIC)_B \; p(VVOB/MIC_i \& VOB)_B$$

and then compute

$$M(VVOB,MIC_i) = p(VVOB/MIC_i)_B / p(VVOB)_B$$

by the same procedure that was used to calculate $M(VOB,MIC_i)$ above. This is shown in figure 3.

FIGURE 3

CALCULATION OF THE MARGINAL CONCENTRATION OF VVOB'S IN EACH MIC

MIC	$p(VOB/MIC)_B$	$p(VVOB/MIC\&VOB)_B$	$p(VVOB/MIC)_B$	$\frac{p(VVOB/MIC)_B}{p(VVOB)_B}$	$\frac{p(MIC)_B}{p(MIC)_R}$	$M(VVOB,MIC)$
All	.4994	.2448	.1223	1.0000	1.0000	1.0000
AFF	.4979	.3193	.1590	1.2999	.9293	1.2080
Mining	.5441	.1892	.1029	.8417	.9036	.7606
Constr.	.5595	.2624	.1468	1.2004	.8624	1.0353
Manuf.	.4742	.2415	.1145	.9364	1.2572	1.1772
TC&U	.5051	.2230	.1126	.9210	.9096	.8377
Wholes.	.5146	.1967	.1012	.8277	1.4369	1.1893
Retail	.4719	.2342	.1105	.9037	.9438	.8529
FIRE	.4890	.1680	.0822	.6717	.7262	.4878
Services	.5045	.2917	.1472	1.2033	.9479	1.1406

Note again that $M(VVOB,MIC_i)$ measures how much larger a fraction of VVOB's are in MIC i than in all MIC's combined. Note also that 81% of the VVOB's in the sample were less than two years old and 95% were under four years old. The distribution of VVOB's is therefore appropriately compared with distribution of business startups in 1980-82.

All of the quantities to be compared are shown in Figure 4, below.

FIGURE 4

COMPARISON OF THE MARGINAL CONCENTRATIONS OF VOB'S AND VVOB'S IN 1983 WITH CONCENTRATIONS OF GROWTHS IN 1978-82 AND MARGINAL CONCENTRATION OF BUSINESSES IN 1982

MIC	1982 M(MIC)	1978-80 M(growth,MIC)	1980-82 M(growth,MIC)	1982 M(VOB,MIC)	1982 M(VVOB,MIC)
AFF	.25	1.57	1.42	.93	1.21
Mining	.08	2.08	2.66	.98	.76
Constr.	1.28	.92	.87	.97	1.04
Manuf.	.78	.78	.71	1.19	1.78
TC&U	.32	1.38	1.14	.92	.84
Wholes.	.90	1.08	.89	1.48	1.19
Retail	2.61	.48	.58	.89	.85
FIRE	.68	.76	1.32	.71	.49
Services	2.10	1.51	1.60	.96	1.14

It appears clear, from figure 4, that veteran owned businesses had (in 1982) above average concentration in the wholesaling and manufacturing industries. Vietnam veteran owned businesses, in addition, had above average concentration in the agriculture, forestry and fishing and service industries.

VOB's were underrepresented in the financial, insurance and real estate, retailing and, less markedly, in the transportation, communication and other utilities and agriculture, forestry and fishing industries. VVOB's were far more severely underrepresented in the finance, insurance and real estate industries and were also underrepresented in mining, transportation, communication and other utilities and retailing.

Most noticeable is that

(a) The marginal concentration of VVOB's in finance, insurance and real estate is far less than that of VOB's in general, and both are

underrepresented in these industries.

(b) The marginal concentration of VVOB's in manufacturing industries is higher than that of VOB's in general, and both are overrepresented in these industries.

(c) VVOB's have a far greater marginal concentration in the agriculture, forestry and fishing industries than either VOB's or all businesses combined (i.e., $M(MIC_i)$).

(d) VVOB's have higher marginal concentration in the service industries than either VOB's or all businesses combined.

(e) VVOB's are significantly underrepresented in the mining industry, while VOB's in general are not.

Some of the above marginal concentrations may be connected with the fact that mining, manufacturing, transportation, communications and other utilities, and finance, insurance and real estate are large business dominated, while agriculture, construction, wholesale trade, retail trade and services are small business dominated (Ref.11 pg.124). Some marginal concentrations may also, in part, be traced to differential concentrations of growths (in number of enterprises) in each MIC between 1978 and 1982:

A disproportionately high concentration of growths in agriculture, forestry and fishing between 1980 and 1982 more than matches the high marginal concentration of VVOB's in those industries. (Note again that 81% of VVOB's ar

less than two years old.) On the other hand, an even higher concentration of growth in these industries occurred in 1978-80, but VOB's are slightly underrepresented in these industries.

A particularly high concentration of growths in the service industries occurred in 1980-82 which, again, might more than account for the high marginal concentration of VVOB's in these industries. An almost equally high concentration of growths in the service industries in 1978-80, however, is not reflected by an above average concentration of VOB's.

More striking are the contrasts between the marginal concentrations in various industrial groups of the VOB's and VVOB's and the corresponding concentrations of growths:

Extraordinarily high concentrations of growths in the mining industry in both 1978-80 and 80-82 are not reflected by above average concentrations of either VOB's or VVOB's. VVOB's, in fact, show a significantly low marginal concentration in this industry.

Low concentration of growth in manufacturing industries in both 1978-80 and 80-82 occurred, while a far above average concentration of VVOB's and an above average concentration of VOB's are in these industries.

High concentration of growth in transportation, communication and other utilities that occurred in both 1978-80 and 80-82 are not reflected by the marginal concentrations of either VOB's or VVOB's.

A significantly high marginal concentration of growth in the finance, insurance and real estate industries in 1980-82 is paired with a very low marginal concentration of VVOB's. This applies, in less extreme form, to VOB's.

The high marginal concentration of growth in the service industries in 1978-82 is not reflected by the marginal concentration of VOB's in these industries. The marginal concentration of VVOB's is greater than one, but is not comparable to the growth concentration.

The marginal concentrations of both VOB's and VVOB's in the retailing industries exceed the concentration of growth in these industries (although all these concentrations are below average).

Note, with respect to the last two statements, that the degree to which marginal concentrations of VOB's or VVOB's might be expected to respond to marginal concentrations of growths is unclear. (See, in this regard, the closing remarks in this section).

An examination of the marginal concentrations of all businesses in each major industry category (i.e., $M(MIC_i)$) in 1982 (as expected given the age of most VOB's and VVOB's) has no apparent bearing upon the marginal concentrations of VOB's or VVOB's, nor upon the above statements.

It should be noted that the above results are peculiar to the point in time examined, since the distribution of businesses over industries varies significantly over time. Also several recessions and recoveries have occurred

since the Vietnam war, each changing the composition of the population of businesses. This is particularly true of VOB's and VVOB's, which are smaller than average (see Ref.11 pg.135 & 192)

It is not easy to form hypotheses to account for the above observed differences. These also should be viewed with some measure of skepticism because the method by which the data samples were acquired does not guarantee a random sample (although it is unclear what kind of bias, if any, might have been introduced - see section III).

Differences in the distribution of Vietnam veteran owned businesses across industries may be due to characteristics of Vietnam veterans that are reflected in the distribution of their employment across industries. 1979 data is available about Vietnam veteran employment distribution (Ref.1, pg.279). In figure 5, below, $M(VVE,MIC_i)$, to be called the marginal concentration of Vietnam veteran employment in MIC i, denotes the ratio between the fraction of Vietnam veterans employed in MIC i in 1979 and the fraction of Vietnam veterans in all industries. $pe(MIC_i)$ will denote the fraction of the work force in MIC i.

Examination of figure 5 shows that the high marginal concentration of VVOB's in manufacturing industries is more than matched by the 1979 marginal concentration of Vietnam veteran employment in these industries.

The above average marginal concentration of VVOB's in the wholesaling industries is reflected, in magnified form, by the 1979 marginal concentration of Vietnam veteran employment in these industries.

The low marginal concentrations of VVOB's in the retailing and finance, real estate and insurance industries also applies to their employment in 1979 (but in less extreme form in the case of finance, real estate and insurance).

The low marginal concentration of VVOB's in mining, transportation, communications and other utilities contrasts sharply with the employment of Vietnam veterans in these industries in 1979.

The high marginal concentration of VVOB's in service industries contrasts with the employment of Vietnam veterans in 1979.

FIGURE 5

COMPARISON OF THE MARGINAL CONCENTRATIONS OF VIETNAM VETERANS EMPLOYMENT AND ALL EMPLOYMENT IN 1979 WITH THAT OF VVOB'S IN 1982

MIC	1979 pe(VVE/MIC)	1979 pe(MIC)	1979 Me(VVE/MIC)	1982 M(VVOB,MIC)
ALL	7.8	1.0000	1.000	1.000
AFF	unknown	unknown	unknown	1.21
Mining	.130	.0118	1.67	.76
Constr.	.092	.0670	1.18	1.04
Manuf.	.175	.2971	2.24	1.78
TC&U	.135	.0743	1.73	.84
Wholes.	.134	.0547	1.72	1.19
Retail	.051	.1994	.65	.85
FIRE	.057	.0736	.73	.49
Services	.050	.2220	.64	1.14

In summary, it appears that many (but not all) of the characteristics of the distribution of Vietnam veteran owned businesses over industries also apply to the distribution of Vietnam veteran employees. Nearly all characteristics of the distribution of Vietnam veteran owned businesses that cannot be accounted

for by the distribution of growths also apply to the employment of Vietnam veterans. This applies to the distributions of VVOB's in the following industrial groups:

finance, insurance and real estate (low marginal concentration)
wholesaling (high marginal concentration)
transportation, communication and other utilities (low marginal concentration)
manufacturing (high marginal concentration)
construction (high marginal concentration)

A notable exception is the mining industry where there is a low marginal concentration of VVOB's and a high concentration of both growth and Vietnam veteran employment.

It appears clear from the above that the peculiarities of the distribution of Vietnam veteran owned businesses across the major industry characteristics reflect differences between Vietnam veterans and both other entrepreneurs and other employees.

Vietnam veteran owned businesses are also differently concentrated than other veterans in specific industry groups, notably finance, insurance and real estate (lower marginal concentration), agriculture, forestry and fishing (higher marginal concentration), mining (lower marginal concentration), manufacturing (higher marginal concentration) and services (higher marginal concentration)

Veteran owned businesses also show differences (but less marked than Vietnam veteran owned businesses) from other businesses, notably in finance, insurance and real estate (lower marginal concentration), wholesaling and manufacturing (both higher marginal concentrations) and retailing (lower marginal concentration).

With notable exceptions, the differences in marginal concentrations between veteran owned businesses and other businesses also distinguishes other veteran owned businesses from other businesses, but to a far lesser degree. Few of the exceptions, however, can be accounted for by the more recent vintage of Vietnam veteran owned businesses (i.e., by comparison with the concentration of growth in 1980-1982).

Reasons for the differences between the distributions of Vietnam veteran owned businesses, other veteran owned businesses, and nonveteran owned businesses (as well as differences in the distribution of Vietnam veteran employment and the remainder of the work force) can be hypothesized. Potentially relevant factors include in-service training received by veterans, interruptions in education, capital resources, availability or unavailability of "GI bill" to finance higher education and ethnic, demographic or other pre-induction differences between the three groups.

The explanation of the differences that have been set forth is beyond the scope of this study, which was to have included only the search for financial differences between VOB's, VVOB's and NVOB's. Additional analyses would be required to investigate hypotheses about the reasons for the peculiarities of each of the three groups of businesses. However, in order to provide some

additional, possibly relevant information, we summarize the results of the analysis of the distributions over two and four digit SIC's.

FIGURE 6

MARGINAL CONCENTRATIONS OF VOB'S AND VVOB'S RELATIVE TO VOB'S IN TWO DIGIT SIC'S

TWO DIGIT SIC		#FIRMS	M(VOB,SIC)	#VOB	M(VVOB/SIC)
01	Agricultural Production - Crops	88	1.0910	48	.5956
02	Agricultural Production - Livestock	40	.7000		
07	Agricultural Services	106	1.0378	55	1.9310
13	Oil & Gas Extraction	52	1.0384	27	.7565
23	Apparel & Other Finished Fabric Products	37	.8108		
25	Furniture & Fixtures	21	.6666		
28	Chemicals & Allied Products	42	.7142		
33	Primary Metal Industries	24	.7500		
35	Machinery Except Electrical	179	1.0614	95	1.2471
38	Measur, Analyz & Control Photographic Equip	41	.8292		
39	Miscellaneous Manufacturing Industries	49	.9796	24	1.3615
41	Local, Surburban, Interurban Transit	24	.6666		
47	Transportation Services	53	.9056	24	.5106
51	Wholesale Trade - Nondurable Goods	416	1.0048	209	.6646
55	Automotive Dealers & Gas Service Station	406	1.2266	249	.6891
56	Apparel & Accessory Stores	209	.6986	73	.6156
58	Eating & Drinking Places	316	.8988	142	1.2083
60	Banking	32	.2500		
64	Insurance Agents, Brokers & Services	120	1.2166	133	.5037
67	Holding & Other Investment Companies	24	.8334		
72	Personal Services	221	1.0136	112	1.2402
75	Automotive Repair, Services, Garages	213	1.2018	128	1.1810
76	Miscellaneous Repairs & Services	258	1.3256	171	1.4093
79	Amusement & Recreational Services	81	.8888	36	1.2484
80	Health Services	198	.9394	93	1.4056
81	Legal Services	51	1.0980	28	.7296
82	Educational Services	31	.5806		
83	Social Services	37	.3784		
86	Nonprofit Membership Organizations	40	.2500		

Figure 6 lists those two digit SIC's in which either $M(VOB,SIC_i)_A$ or $M(VVOB,SIC_i)_B$, where B is the sample of VOB's, either exceeds 1.2 or is less than .83. (These limits correspond to equal quantities of "information").

Also shown is the number of businesses and the number of VOB's in each such SIC, so that reliability judgements about the margins can be made. The Values of the margins are omitted when the corresponding samples (in the SIC) contain fewer than twenty businesses.

Some suggestive observations can be made, such as the low marginal concentrations of VOB's in banking, educational services, social services, nonprofit organizations, and their high marginal concentrations in automotive dealerships and gasoline service stations, miscellaneous repairs and services automotive repair, services and garages, and insurance and brokerage services.

VVOB's have high marginal concentrations in agricultural services, miscellaneous repairs and services, health services, miscellaneous manufacturing industries, personal services, amusement and recreational services, machinery other than electrical, and eating and drinking places They have low marginal concentrations in insurance and brokerage services, transportation services, agricultural production (crops), apparel and accessory stores, and wholesale trade in nondurable goods.

A higher level of detail can be achieved by looking at four digit SIC's. Figure 7 (below) is identical to Figure 6, except that four digit SIC's are considered:

FIGURE 7

MARGINAL CONCENTRATIONS OF VOB'S AND VVOB'S RELATIVE TO VOB'S IN FOUR DIGIT SIC'S

FOUR DIGIT SIC	#FIRMS	M(VOB,SIC)	#VOB	M(VVOB SIC)
0181 Ornamental Floriculture & Nursery Prods	27	1.2592		
1311 Crude Petroleum & Natural Gas	26	1.2308		
1542 General Contractors - Non residential bldg	67	1.0150	34	.7210
1623 Water,Sewer,Pipeline,Communic,Power Constr	28	1.2858		
1721 Painting, Paper Hanging & Decorating	37	1.1352	21	.7794
1731 Electrical Work	142	1.2254	87	1.1270
1752 Floorlaying & Other Floor Work, nec.	21	1.3334		
1771 Concrete Work	21	.6666		
1799 Special Trade Contractors, nec.	44	1.0000	22	1.2998
2751 Commercial Printing, Letterpress & Screen	39	1.3334	26	1.0997
2752 Commercial Printing, Lithographic	74	1.0270	38	.6450
3599 Machinery Except Electrical, nec.	85	1.1530	49	1.4171
4722 Arrangement of Passenger Transportation	41	.8292		
5013 Automotive Parts & Supplies	99	1.0708	53	1.4644
5023 Home Furnishings	24	.7500		
5031 Lumber, Plywood & Millwork	22	.7272		
5039 Construction Materials, nec.	43	1.0698	23	.1777
5065 Electronic Parts & Equipment	40	1.1500	23	.7104
5072 Hardware	23	1.2174		
5074 Plumbing & Heating Equipment & Supplies	34	1.2352	21	.7782
5083 Farm & Garden Machinery & Equipment	69	.8696	30	.4085
5084 Industrial Machinery & Equipment	129	1.0542	68	.5408
5085 Industrial Supplies	29	1.4482	21	.7782
5099 Durable Goods, nec.	34	.8236		
5153 Grain	24	.8334		
5171 Petroleum Bulk Stations & Terminals	25	1.2000		
5191 Farm Supplies	55	.9090	25	.6536
5211 Lumber & Other Bldng Material Dealers	76	.9736	37	1.3248
5399 Misc. General Merchandise Stores	60	1.1234	34	1.4416
5411 Grocery Stores	178	1.0338	92	1.2431
5462 Retail Bakers, Baking & Selling	21	1.2380		
5511 Motor Vehicle Dealers - New & Used	92	1.1740	54	.2271
5521 Motor Vehicle Dealers - Used	36	1.2220	22	.7426
5531 Auto & Home Supply Stores	84	1.3096	55	.8170
5541 Gasoline Service Stations	133	1.2030	30	.8681
5551 Boat Dealers	22	1.6362		
5611 Men's & Boy's Clothing & Furnishings Store	45	.9778	22	.1859
5611 Women's Ready to Wear Store	48	.5416		
5661 Shoe Stores	46	.7826		
5699 Misc. Apparel & Accessory Stores	27	.6666		
5712 Furniture Stores	86	.8838	38	.7524
5732 Radio & Television Stores	53	1.1320	30	1.7700
5812 Eating Places	229	.8734	100	1.2255
5912 Drug Stores & Proprietary Stores	138	.8260	57	.6450
5921 Liquor Stores	59	.8136	24	.6810

FIGURE 7 (cont'd)

5931	Used Merchandise Stores	28	.7142		
5941	Sporting Goods Stores & Bicycle Shops	90	1.0222	46	1.6871
5943	Stationary Stores	22	.6364		
5944	Jewelry Stores	60	.8666	26	.6283
5947	Gift, Novelty & Souvenier Shops	60	.6666		
5992	Florists	52	1.6538		
5999	Misc. Retail, nec.	130	.8770	57	1.2900
6411	Insurance Agents, Brokers & Services	120	1.2166	73	.5037
6512	Operators of Nonresidential Buildings	59	.8814	26	.4714
7011	Hotels, Motels & Tourist Courts	93	.7956	37	.6626
7215	Coin Operated Laundries & Dry Cleaning	24	1.2500		
7221	Photographic Studios, Portrait	32	1.2500		
7231	Beauty Shops	35	.5142		
7261	Funeral Services & Crematories	40	1.0500	21	.7782
7311	Advertising Agencies	34	1.2352	21	1.3615
7531	Top & Body Repair Shops	39	1.2820	25	1.1438
7538	General Automotive Repair Shops	96	1.1876	55	1.3615
7622	Radio & Television Repair Shops	45	1.5556	35	1.1671
7629	Electrical & Electronic Repair Shops, nec.	33	1.2122		
7699	Repair Shops & Related Services, nec.	102	1.2942	66	1.5474
7999	Amusement & Recreation Services, nec.	28	.7858		
8011	Offices of Physicians	75	1.0934	41	1.5940
8021	Offices of Dentists	39	1.0770	21	1.5564
8111	Legal Services	51	1.0980	28	.7296
8911	Engineering, Architect & Survey Services	136	1.1618	79	.6205
8931	Accounting Services	71	1.0140	36	1.7022

A great many two digit SIC's were omitted from Figure 6 because they contained fewer than 20 firms, and far more four digit SIC's are omitted from Figure 7.

Nonetheless, some interesting observations can be made, such as the extremely low marginal concentrations of VVOB's as operators of non-residential buildings, owners of hotels, motels and tourist courts, liquor stores, wholesalers of construction materials, farm and garden machinery, industrial machinery and equipment, and men's and boy's furnishing stores.

VVOB's have high marginal concentrations in radio and television stores, sporting goods and bicycle shops, repair shops and related services, offices of physicians,dentists, miscellaneous general merchandising stores and accounting

services.

VOB's have a different and less pronounced pattern of concentration in four digit SIC's than VOB's. This can be seen by examining Figure 7.

The above is suggestive of possible requirements for capital and training and possible consequences of in-service training. More comprehensive results can be obtained by subjecting all of file A (see Section III) to the analysis that was here performed on samples A and B from this file. Estimation of the spuriousness of the marginal concentrations would be necessary because those SIC's where these concentrations are more extreme are selected for attention from among a very large number of four digit SIC's. Such an analysis could be useful in targeting veteran's training programs and loans.

We close with a note to the effect that, by comparing marginal concentrations rather than distributions, similarities and differences are more clearly revealed. For example, we examined (qualitatively rather than statistically) the hypothesis that the distributions of VOB's across MIC's is proportional to the distribution of growths in number of enterprises, i.e.

$$p(VOB/MIC_i) = K\ p(s\text{-}start/MIC_i)$$

Why did we not simply compare p(VOB/MIC) with $p(s\text{-}start/MIC_i)$? The magnitudes of $p(VOB/MIC_i)$ and $p(s\text{-}start/MIC_i)$ are not similar and require an adjustment by the unknown, K. By the definitions of $M(VOB/MIC_i)$ and $M(growth/MIC_i)$

$$\frac{M(VOB,MIC_i)}{M(growth,MIC)} = \frac{p(VOB/MIC)}{p(s\text{-}start/MIC)} \frac{p(s\text{-}start)}{p(VOB)}$$

Hence p(s-start)/p(VOB) may be considered as an estimate of K, and the quantities to be compared are similar in magnitude.

VI. SUMMARY RESTATEMENT OF CONCLUSIONS

While 7.8% of the work force are Vietnam veterans, 7.5% of the businesses examined were Vietnam veteran owned. If the sample is representative, then a Vietnam veteran is about as likely to own a business as anyone else in the work force.

Except that veteran owned businesses are, on the average, smaller, no clear and reliable financial differences between veteran owned and nonveteran owned businesses were detected.

Vietnam veteran owned businesses appear to be smaller, younger and more in need of capital than other businesses, including other veteran owned businesses. Other differences are very slight (see part V, section B).

Vietnam veteran owned businesses (VVOB's), veteran owned businesses (VOB's) and businesses in general differ significantly in their distribution across industries in 1982.

The following tables show the business categories in which VOB's and VVOB's are overrepresented and underrepresented (i.e., have high or low marginal

concentrations respectively). Note again that the marginal concentration of VOB's in an industry is defined as the ratio between the fraction of VOB's in that industry and the fraction of all businesses that are VOB's. The table entries are in descending order of the degree of such overrepresentation or underrepresentation:

VOB's overrepresented	VOB's underrepresented
wholesaling	finance, insurance and real estate
manufacturing	retailing
	transportation, communications and other utilities
	agriculture, forestry and fishing

VVOB's overrepresented	VOB's underrepresented
manufacturing	finance, insurance and real estate
agriculture, forestry and fishing	mining
wholesaling	transportation, communications and other utilities
services	retailing

The above may, in part, be related to the fact that mining, manufacturing, transportation, communications and utilities and finance, insurance and real estate are large business dominated while agriculture, construction, wholesale trade, retail trade and services are small business dominated.

VVOB's and VOB's also differ in that:

(a) The concentration of VVOB's in finance, insurance and real estate is far less than that of VOB's, who have significantly low marginal concentration in these industries.

(b) The concentration of VVOB's in manufacturing industries is

higher than that of VOB's, who have high marginal concentration in these industries.

(c) VVOB's have a notably high marginal concentration in the agriculture, forestry and fishing industries.

(d) VVOB's have a high marginal concentration in the service industries.

(e) VVOB's have a very small marginal concentration in the mining industry.

Few of the differences between VOB's, VVOB's and businesses in general appear to be accounted for by growths in number of enterprises in each industry category during the inceptions of the VOB's and VVOB's. In particular,

(a) Extraordinarily high concentrations of growth in the mining industry in 1978-82 are not reflected by above average marginal concentrations of either VOB's or VVOB's. VOB's show a significantly low marginal concentration in this industry.

(b) Low concentrations of growth in manufacturing in 1978-82 contrast with the far above average concentration of VVOB's and above average concentration of VOB's in these industries.

(c) High concentrations of growth in 1978-82 in transportation, communication and other utilities are not reflected by the marginal

concentrations of either VOB's or VVOB's.

(d) Significantly high concentration of growths in 1978-82 in the finance, insurance and real estate industries were not reflected by the concentration of VOB's.

(e) The high concentrations of growth in the service industries in 1978-82 are not reflected by the concentration of VOB's in these industries (although the marginal response of VVOB's is greater than one).

Neither the marginal concentrations of VOB's nor VVOB's are accounted for by the distribution of businesses across industry categories.

Many of the peculiarities of the marginal concentrations of VVOB's in different industries also apply (usually in less extreme form) to VOB's. These peculiarities are, for the most part, paralleled by differences in employment in these industries between Vietnam veterans and the remainder of the work force. This is particularly true of characteristics of the distribution of VVOB's across industries that contrast with what might be expected given the distribution of business growths. It appears clear that the peculiarities of the distribution of VVOB's across the major industries reflect differences between Vietnam veterans and both other entrepreneurs and other employees.

The examination of differences between the marginal concentrations of the three classes of businesses at the two and four digit SIC level also showed differences that are suggestive of possible specific requirements for capital

and training. (See section V).

Determination of the reasons for the differences between the distributions over industries of Vietnam veteran owned businesses, veteran owned businesses and other businesses requires additional study. A great many factors may be relevant, such as in service training received by veterans, interruptions in education, capital resources, availability or unavailability of "GI bill" to finance higher education and ethnic, demographic or other pre-induction differences between the three groups. The marginal concentrations of the three groups in two and four digit SIC's is suggestive in this regard. Many two and four digit SIC's, however, contained too few elements of the samples that were analyzed to yield reliable information. This difficulty can be largely overcome by performing a similar analysis using the entire file of businesses obtained in this study. Although the results could not, taken alone, determine the reasons for differences between the industry memberships of the three groups, it would pinpoint the nature and degree of these differences in a manner that might be relevant to public policy and to the deployment of financial resources and training programs.

REFERENCES

1. THE STATE OF SMALL BUSINESS: A Report of the President, U.S. Government Printing Office, March 1984.

2. Frank, R.E., Massy, W.F. and Morrison, D.G., "Bias in Multiple Discriminant Analysis", JOURNAL OF MARKETING RESEARCH, Vol. 2, pp. 250-258, August 1965.

3. Altman, E.I., "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy", JOURNAL OF FINANCE, Vol. 23, No. 4, pp. 589-609, 1968.

4. Beaver, W.H., "Financial Ratios as Predictors of Failure", Empirical Research in Accounting: Selected Studies 1966, JOURNAL OF ACCOUNTING RESEARCH, Vol. 12, No. 1, pp. 1-25, 1966.

5. Chen, K.H. and Shimerda, T.A., "An Empirical Analysis of Useful Financial Ratios", FINANCIAL MANAGEMENT, Vol. 10, pp. 51-60, 1981

6. Deakin, E.B., "Distributions of Financial Accounting Ratios: Some Empirical Evidence", THE ACCOUNTING REVIEW, Vol. 51, No. 1, pp. 167-179, 1976.

7. Edminster, R.O., "An Empirical Test of Financial Ratio Analysis For Small Business Failure Prediction", JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS, Vol. 7, No. 2, pp. 1477-1493, 1972.

8. Libby, R., "Accounting Ratios and the Prediction of Failure: Some Behavior Evidence", JOURNAL OF ACCOUNTING RESEARCH, Vol. 13, pp. 150-161, 1975.

9. Pinches, G.E., and Mingo, K.A., "A Multivariate Analysis of Industrial Bond Ratings", JOURNAL OF FINANCE, Vol. 28, pp. 1-18, 1973.

10. Wilcox, J.W., "A Simple Theory of Financial Ratios as Predictors of Failure", JOURNAL OF ACCOUNTING RESEARCH, Vol. 9, No. 2, pp. 389-395, 1971.

11. THE STATE OF SMALL BUSINESS: A Report of the President, U.S. Government Printing Office, May 1985.