1975

The impact of negative income taxes upon the labor supply of farm operators

Wendell Eugene Primus
Iowa State University

76-9195

PRIMUS, Wendell Eugene, 1946-
THE IMPACT OF NEGATIVE INCOME TAXES UPON
THE LABOR SUPPLY OF FARM OPERATORS.

Iowa State University, Ph.D., 1975
Economics, general

Xerox University Microfilms, Ann Arbor, Michigan 48106

The impact of negative income taxes upon the

labor supply of farm operators

by

Wendell Eugene Primus

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of

The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major: Economics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1975

TABLE OF CONTENTS

CHAPTER I. INTRODUCTION

The Current System of Income Maintenance

Projected figures for 1975 show that income security payments will total $141.7 billion accounting for 45 percent of all federal public expenditures [19]. Primarily these payments go to families who do not have the ability to earn an adequate income, i.e. retired individuals, fatherless families with children, and the mentally and physically handicapped.

The two major forms of assistance are social insurance including Social Security, Unemployment Insurance and other social insurance programs, and need-based benefits including Aid to Families with Dependent Children (AFDC), Food Stamps, Medicaid, Housing, Education, and similar social services.

These need-based programs have been criticized for a variety of undesirable effects. Many public welfare programs have resulted in inequities where nonworkers and those who fail to fulfill their family responsibilities are better off than workers and couples with children. For example, cash benefits available to certain families with unemployed fathers (AFDC-UF) exceed the maximum Unemployment Insurance (UI) benefits in some states [5]. Also since every UI recipient is not eligible for maximum payments, the number of cases for which welfare is better than UI is significant.

Other criticisms of the current system are the notch effect and high negative tax rates. The notch effect is a situation where a small change in income greatly affects the amount of benefits. For example, in 24 states the last dollar that removes a family from eligibility for cash welfare also terminates its right to Medicaid, valued at several hundred dollars per family [89]. The negative tax rate is defined as the reduction in governmental benefits for each additional dollar earned. Even though individual programs have low negative tax rates, the integration of the various programs inevitably produces a high negative tax rate. The higher this tax rate, the greater the disincentive to work. For example, the AFDC program has a negative tax rate of 67 percent. For each dollar the recipient earns, she can expect a net gain of 33 cents. However the Food Stamp program, taking note of her 33 cent gain, will raise stamp prices 10 cents per extra dollar earned, and Public Housing will raise rent by 8 cents per extra dollar earned. Thus the negative tax rate rises to 85 cents or a net gain of only 15 cents. Consequently, it is not reasonable to expect persons to work for a small net gain, especially if it is unpleasant work [89].

Also within the current system, there are monetary incentives for families to break up. For example, in July 1974, a family head in North Dakota with three children earning $2.00 per hour could have earnings of $278 per month. With these earnings a family would be eligible for a $100 Food Stamp bonus and no Medicaid payments. If the father left home, however, the mother and three children could receive $300 in AFDC benefits, $67 in Food Stamps, and $70 in Medicaid for a total of $437. With the additional $278 in earnings by the father, there would be a net gain of $337 in the separated family [89].

Other problems also exist. The disparity between state benefit levels is large. For fatherless families, AFDC and Food Stamps range from a low of \$197 per month in Mississippi to a high of \$448 in New York and \$463 in Alaska. Fraud, inefficient management, and duplicity between different government agencies compound problems within the current system.

One solution to the inadequacies of the current welfare system is an impersonal, mathematically determined benefit based on income, family size, and resource levels. This system has the distinct advantage of not categorizing any family by such conditions as employability, type and costs of housing, and medical payments.

While several different kinds of formulas exist which include wage and earnings subsidies, this discussion will be limited to the negative income tax. The formula for this scheme is as follows:

$$\text{Payment} = GL - t \times Y \text{ if } 0 < Y < \frac{GL}{t}$$

$$= 0 \text{ if } Y \geq \frac{GL}{t}$$

$$= GL \text{ if } Y \leq 0$$

where Y = income

t = negative tax rate

GL = guarantee level or the benefit level if income equals zero.

Note that as income increases, the amount of benefit declines by the negative tax or reduction rate. GL is usually a simple function of family size. As family size increases, GL increases, but at a decreasing rate. The additional benefit is smaller for the sixth child than the second child, and after the eighth or ninth child, the additional benefit might be zero.

Resource levels are usually incorporated into the formula in two ways. The first is that if resources are greater than some specified level, the family is ineligible. The second is that, after certain dollar level exclusions for home and business property, the resource level is multiplied by a low percentage, say 5 to 10 percent, and added to income. In this manner notch effects are avoided and incentives to change assets from one form to another form are minimized. The prime rationale for this tax on resources is that low income individuals should be encouraged to liquidate a portion of their assets to cover fluctuations in their current income stream [6].

Plan of the thesis

Chapter two presents some theoretical expectations of what the potential work effort response of farm operators will be to the negative income tax (NIT). It begins by reviewing some of the recent literature on labor supply and its application to the problem of work disincentives of government tax and transfer policies. A formal statement of the classical theory of work-leisure choice is extended to include that of self-employed individuals. Chapter three discusses several dependent variables used to measure labor supply, estimation methodology, and independent variables used in the study. Chapter four presents the results of the empirical investigation of what the labor supply response of self-employed farm operators is to a negative income tax scheme. Chapter five is a digression where one problem, that of response errors, is investigated in detail. The differences between initial survey responses and an extensive editing procedure are discussed, and finally the difference in results between the two different data sets are contrasted and evaluated. Chapter six concludes with a discussion of the policy implications and shortcomings of the approach.

Description of the data

The data used were collected by University of Wisconsin personnel in conjunction with the rural negative income tax experiment. This major social experiment was launched in July 1969 by the Office of Economic Opportunity approximately eighteen months after its predecessor, the Urban experiment in New Jersey and Pennsylvania.

The sites for the rural experiment were Dauphin County, North Carolina, and Calhoun and Pocahontas Counties in Iowa. Selecting the poor from two geographic areas rather than using a nationwide sample eliminated the high administrative costs associated with the latter. The selection of these areas was based on the following four prerequisites:

1) The area must have a reasonable degree of agricultural diversity rather than most farmers specializing in the production of one commodity. This would allow for observations of changes in enterprise mix and potential elimination of secondary activities (e.g., livestock feeding).

2) Specific counties within the area should not be located within 25 miles of a city with over 50,000 persons.

3) Mountainous and swampy areas were to be avoided where the consequent low population density would render the prospective sample population too inaccessible.

4) Areas containing a few counties with a moderately-sized city between them (about 10,000 persons) were given preference to those with no

nearby city at all. It was anticipated that the existence of some manufacturing and/or marketing activities within reasonable reach of the rural residents would offer the opportunity to observe potential changes in their employment patterns.

The sample was chosen by a two-stage process. Initially, a short screening interview was administered to a random sample of all households in the three county area. From these, families whose income was less than 1.6 times the poverty line were chosen and given a more extensive baseline survey. On the basis of this information, a sample was selected whose income was less than 1.5 times the poverty line, where income was defined by the rules of operation. These families were then randomly assigned to either the control or experimental group and enrolled in the experiment. During enrollment, for those in the experimental group, the rules of operation were explained in laymen's terms, namely, that as income goes up, payments go down and vice versa; that as family size increases, payment level increases and vice versa; that the payment check could be spent in any way desired; and that the family must supply copies of its W2 form and file honest monthly reports of income and expenses.

The experiment lasted three years and one month with an additional eight month follow-up period. During this time period, a quarterly interview was administered every three months. It was from these surveys that the basic data for this study was obtained.

Different types of information were collected at different intervals. For example income and expense data and family composition data were collected once every three months; asset data, inventories, and physical units of production were collected once per year; and nutrition information was collected twice during the experiment. The amount of data collected was immense and it covered almost every area of human endeavor.

The payments process was administered separately from the quarterly surveys. Payments were based on the monthly income report form in the hope that information on quarterly surveys would not be tainted by reporting effects. Experimental payment families had an obvious incentive to report as many expenses and as little income as possible on these monthly income report forms.

Almost all information was obtained directly from the respondent. Some outside information was also available, such as Internal Revenue Service (IRS) forms and Social Security records. The former were collected from the respondent, while the latter were obtained directly from the Social Security Office in Baltimore, Maryland.

This describes the major features of the experiment and its data collection apparatus. For further details see [24]. See also Appendix A for a description of the sample and Appendix B for a description of data editing procedures.

CHAPTER II. LABOR SUPPLY

Review of the Literature

Early labor supply theory

The beginning of modern day econometrics is often traced to the famous studies by Ernst Engel in estimating the effect of income on family expenditure patterns. Questions of what effect price had, or more specifically whether a tax on earnings, particularly a progressive tax, would have an effect on labor supply were soon raised.

Lionel Robbins' classical article on labor supply provided the framework for analyzing the effects a tax on income would have. By transforming labor supply into a study of the demand for leisure, he was able to define an income and substitution effect, namely as the wage rate increased, the price of leisure increased, which *ceteris paribus* meant demand for leisure was lowered. However, at the same time, the amount of goods one could buy with a given expenditure of work increased. Thus with leisure being a normal good, this rise in income should increase the purchase of leisure. Thus theoretically the answer was indeterminate.

Nonexperimental empirical studies

Proposals in the late 1960's for alternative welfare systems have stimulated much work in estimating income and substitution effects. If the effect on labor supply of an income compensated wage rate change is small, very little attention need be devoted to selection of a correct tax rate. And, if the income effect of high guarantee rates is small, the selection of a correct guarantee can depend entirely upon considerations of cost. On the other hand, if empirical work indicates differently, careful attention must be given to the choice of a proper tax and guarantee rate.

The Greenberg-Koster's study reported in Cain and Watts [16], was the first of many studies which were done on nonexperimental data to simulate the effects a nationwide program would have. Their basic labor-supply model is presented in terms of the economic theory of choice. Their study focuses on male headed families with income less than $15,000 based on the 1967 Survey of Economic Opportunities' data base.

A unique feature of their study is an attempt to control for difference in preference structures. They argue that persons with a relatively strong preference for asset accumulation will tend to work more, consume less, and have larger than average nonemployment income. This in turn will tend to depress labor supply through the income effect. On the other hand this is offset by the strong preference for asset accumulation. Consequently, the labor supply effect of a negative income tax program could be misstated if no attempt were made to control for these differences.

A study by Hall, also reported in Cain and Watts [16], is characterized by the development of a "potential wage" variable in the first stage estimation which is used as a price variable in further analysis. He also makes extensive use of "form-free" regression models that utilize binary variables. The former was developed primarily because of two problems: a) measurement error in wage rate and b) the fact that many housewives, etc., do not have an observation on wage rate.

Orley Ashenfelter and James Heckman [3] describe the formulation of the theoretical restrictions on the labor supply function of both the husband and wife simultaneously. These restrictions are that (1) own substitution effects must be positive, (2) cross substitution effects must be equal and (3) the husband's own substitution effect times the wife's own substitution effect must be greater than the multiplication of these cross substitution effects. These restrictions are tested and, when accepted, are imposed on the data, resulting in more consistent estimators.

Experimental studies

The forerunner to the Rural Income Maintenance Experiment (RIME) was the urban experiment conducted in four sites in New Jersey and Pennsylvania [79, 95, 97, 23]. Utilizing four different guarantee levels and three different tax rates, payments were administered to a sample of approximately 600 experimentals for a period of three years. Information was collected on both the control and experimental group through the use of quarterly interviews. Using labor supply models based on the theory of economic choice and pooled regression analysis, estimates of the behavioral response were made. These results were mixed, particularly with respect to race. For white male headed families, the overall results indicated a small insignificant effect on the labor supply of the male heads. As could be expected, the results for spouses and other secondary earners indicated a rather large (around 35 percent) significant disincentive. The number of true observations upon which this conclusion was based is small. Many versions of the dependent variable were used, including hours of work per week, number of earners in a family, wage income, and labor force participation status. All were generally consistent.

Several problems with the results exist. The major problem is the positive incentives, contrary to economic theory, which were indicated for the black male headed families. In addition, positive effects were found on wage rates for male heads. While this is plausible, given that experimental families can engage in longer job search activities because of the payments and consequently obtain a higher wage, there is no indication within the data that a longer job search was in fact undertaken. This provides little explanation for the high wage rate, unless it is connected with the intensity of job search activities. Several minor inconsistencies also exist in the behavioral response among different tax and guarantee rates. Occasionally lower disincentives are associated with higher tax and guarantee levels.

Because this was the initial experiment of its kind, several problems were encountered in the data collection methodology. The experiment suffered from a high attrition rate and a large loss in use of usable observations. Furthermore, original data items were often inconsistent or changed on the basis of little information, with no apparent way existing to determine how these changes affected results.

The Rural Income Maintenance Experiment was undertaken to determine the behavioral response to different tax and guarantee rates for the self-employed. The analysis will differ from the urban experiment in several important ways which may affect the response. These differences can be summarized as follows:

1) As hours decline or increase, the value of the marginal product of labor is no longer constant.

2) Self-employed, particularly farmers, may be able to change work effort more easily, i.e. there is no institutional restraint such as a 40 hour work week.

3) Elements of risk, cash flow, and investment, all enter the picture when capital is combined with labor in the productive process.

Theory

Several authors (Lee, 1965 [56]; Huffman, 1973 [43]) have attempted to explain the allocating of resources between farm and nonfarm use particularly with respect to labor resources. However, with the exception of Meyer and Saupe (1970) [66], very little literature exists on the behavior of the self-employed under an income maintenance scheme.

To explain the behavior of a farm family firm, the following model will be developed. Our farm family is presumed to behave as one individual maximizing a utility function, U, which is composed of income per time period, Y, and the number of hours of leisure, L.

Income is generated by three different means, all of which involve an allocation of the individual's scarce time.[1] These are crop growing enterprises, livestock growing enterprises, and wage work. Unearned income,

[1] Overhead hours exist for self-employed enterprises. These involve such activities as bookkeeping, filling out tax forms, and trips to town. Conceptually it will be assumed that these hours can be divided between crop and livestock activities. Marginal decisions regarding the allocation of time are then handled straightforwardly by the model developed in the remainder of the chapter.

which appears in many labor supply models, is deleted from this analysis because of its small magnitude in the sample under study (see Appendix A). Income from the two farm enterprises will be defined as price per unit of output times a production function explaining the output. Each of these production functions will have four arguments. The first is a fixed amount of capital that is unlikely to change over the course of the experiment. This would include land, buildings, machinery or other such investments. The second argument is a variable amount of capital (intermediate inputs) representing the decision a farmer must make each year on how much fertilizer to use, spray to apply, gasoline to buy, additional labor to hire, etc. The third argument represents the amount of time the farm operator devotes to the enterprise, and the fourth argument represents entrepreneurial skills. The latter is assumed to be constant throughout the duration of the experiment for any particular farmer. The opportunity cost of money invested in capital is defined as the cost or price of capital (r) multiplied by total capital. This opportunity cost plus the cost of the intermediate goods subtracted from price times output will yield labor income for each of the two farm enterprises.

Our individual maximizes:

$$U = U(Y, T - h_w - h_c - h_1) \tag{1}$$

where:

U = a utility function constrained by the availability of T (total time)

Y = income (to be defined more fully shortly)

h_w = hours employed in the nonfarm market at a fixed wage rate

h_c = hours engaged in crop growing enterprises

h_1 = hours engaged in livestock growing enterprises

T = total time available to an individual in a given time period.

Income, Y, can be earned as follows:

$$Y = P_c F(\overline{K}_c, K_c, h_c, E_c) + P_1 G(\overline{K}_1, K_1, h_1, E_1) + wh_w - r(\overline{K}_c + \overline{K}_1) - (1 + r)(K_c + K_1) \quad (2)$$

P_c, P_1 = the average price of output from crop (livestock) growing enterprises

F, G = a production function for crop (livestock) output, respectively

$\overline{K}_c$, $\overline{K}_1$ = fixed capital stock used in crop (livestock) enterprises, respectively

K_c, K_1 = the amount of variable capital used in growing crop (livestock) output

E_c, E_1 = entrepreneurial and management skills used in raising crops (livestock)

w = fixed wage rate received in a job outside the farming operation

r = opportunity cost of capital.

A few comments are necessary before proceeding. It is presumed the farmer is operating in a competitive industry and consequently P_c, P_1, and r are determined exogenously. The production functions, F and G, are assumed to behave normally in that all first-order partial derivatives are positive while the second-order partial derivatives are negative. This reflects that while in the economic region of production, an increase in the amount of any input will increase output. However, further equal increases in the amount of inputs holding all other inputs constant, will not increase output as much, due to diminishing marginal returns.

It will be assumed throughout this analysis that wage and farm work are equivalent with respect to nonpecuniary benefits. Aside from economic returns, there is no incentive to remain in one occupation versus the other.

The time constraint facing a farmer is the following:

$$L = T - h_w - h_c - h_1 \tag{3}$$

or the amount of leisure is equal to total time minus hours devoted to wage work, crop, and livestock growing enterprises, respectively.

Equations 1, 2, and 3 can be combined to yield the following constrained utility maximizing equation:

$$\begin{aligned} \max U = U(Y, T - h_w - h_c - h_1) &+ \lambda[P_c F(\bar{K}_c, K_c, h_c, E_c) \\ &+ P_1 G(\bar{K}_1, K_1, h_1, E_1) + wh_w - r\ (\bar{K}_c + \bar{K}_1 + K_c + K_1) \\ &- (K_c + K_1) - Y] \end{aligned} \tag{4}$$

From the first-order utility maximizing conditions we can derive the determinants of total labor supply, the allocation of capital between livestock and crops, and the allocation of work hours between crop, livestock, and the outside labor market.

$$\frac{\partial U}{\partial h_c} = \frac{-\partial U}{\partial (T - h_c - h_1 - h_w)} + \lambda P_c f_{h_c} = 0 \tag{5}$$

$$\frac{\partial U}{\partial h_1} = \frac{-\partial U}{\partial (T - h_c - h_1 - h_w)} + \lambda P_1 g_{h_1} = 0 \tag{6}$$

$$\frac{\partial U}{\partial h_w} = \frac{-\partial U}{\partial (T - h_c - h_1 - h_w)} + \lambda w = 0 \tag{7}$$

$$\frac{\partial U}{\partial K_c} = \lambda P_c f_{K_c} - \lambda r - \lambda = 0 \tag{8}$$

$$\frac{\partial U}{\partial K_1} = \lambda P_1 g_{K_1} - \lambda r - \lambda = 0 \tag{9}$$

$$\frac{\partial U}{\partial Y} = \frac{\partial U}{\partial Y} - \lambda = 0 \tag{10}$$

$$\begin{aligned} \frac{\partial U}{\partial \lambda} &= P_c F(\overline{K}_c, K_c, h_c, E_c) + P_1 G(\overline{K}_1, K_1, h_1, E_1) \\ &+ wh_w - r(\overline{K}_c + \overline{K}_1 + K_1 + K_c) - (K_c + K_1) - Y = 0 \end{aligned} \tag{11}$$

Assuming a "well-behaved" utility function so that the second-order conditions insure a utility maximum, the first-order conditions yield the following:

$$MRS_{L \cdot Y} = P_c f_{h_c} = P_1 g_{h_1} = w \tag{12}$$

and

$$P_c f_{K_c} = P_1 g_{K_1} = r + 1 \tag{13}$$

$MRS_{L \cdot Y}$ is the marginal rate of substitution of income for leisure while $P_c f_{h_c}$ and $P_1 g_{h_1}$ are the value of the marginal product of labor in crop and livestock enterprises, respectively. An individual will work as long as the value of his marginal product (VMP) or wage rate is greater than $MRS_{L \cdot Y}$. The allocation of work between the three enterprises is based simply upon where the highest value for his labor can be obtained. In the economic region of production, the VMP, holding capital constant, in both crops and livestock enterprises will decline as labor is increased.

The following implications result directly from Equations 12 and 13 above. If VMP_c (value of the marginal product of labor in crops) $> w > VMP_l$, this would imply an individual growing crops until $VMP_c = MRS_{L \cdot Y}$. If at this point $w > MRS_{L \cdot Y}$, additional hours would be added to the wage job and crop hours until $w = VMP_c = MRS_{L \cdot Y}$. Livestock enterprises would only be entered if there were difficulty in obtaining wage employment.

If the VMP_l from both crops and livestock is less than w for all levels of h_c and h_l, an individual should leave farming and work at a job. The only exception to this would be if sufficient wage hours could not be obtained in off-farm work. In this case $w > MRS_{L \cdot Y} = VMP_c = VMP_l$.

If $VMP_c > VMP_l > w$, an individual would allocate additional time until $VMP_c = VMP_l$. If at this point $VMP_l > MRS_{L \cdot Y}$, additional hours would be worked in both crop and livestock enterprises, thus maintaining their equality. Only if $VMP_l = VMP_c = w > MRS_{L \cdot Y}$ would wage work be done. Enough wage hours would be worked until the equality in Equation 12 is reached. Also as more hours are worked, Y increases, and the rate of substitution between leisure and income decreases. Stated differently, for each additional hour worked, the rate at which one is willing to trade one hour of leisure for income increases.

Equation 13 tells us that variable capital will be hired until the marginal product of capital in crop and livestock operations equals the cost of capital plus its opportunity cost, r. Note also the simultaneity between Equations 12 and 13. As more hours are devoted to crops, the marginal product of capital in crops increases and vice versa.

Introducing the negative income tax changes Equation 2 above. The payments formula is as follows:

Payment = GL when $Y < 0$

Payment = $GL - t \times Y$ when $0 \leq Y \leq GL/t$, and

Payment = 0 when $Y > GL/t$

where:

GL = guarantee level or amount of payment when Y = 0 and

t = negative tax or reduction rate.

Defining $Y^* = GL + (1-t)\,Y$ and replacing Y^* for Y in the constrained utility maximizing Equation 4, the first-order conditions become:

$$MRS_{L \cdot Y} = (1-t)P_c f_{h_c} = (1-t)P_1 g_{h_1} = (1-t)\,w \qquad (14)$$

and

$$(1-t)P_c f_{K_c} = (1-t)P_1 g_{K_1} = (1-t)(r+1) \qquad (15)$$

The amount of adjustment is primarily determined by the slope of the marginal product curves. These are illustrated graphically in Figure 2.1 for the economic region of production.

Prior to the experiment, let the income-leisure trade-off be determined as $MRS_{L \cdot Y_o}$.[1] Utility is maximized by doing no wage work, working h_{1_o} hours in livestock enterprises and h_{c_o} hours in crop enterprises. After the experiment is enacted, letting t = .50 and $MRS_{L \cdot Y_o}$ remain unchanged, Equation 14 becomes $VMP_c = VMP_1 = w = 2MRS_{L \cdot Y}(MRS_{L \cdot Y_1})$. In this example,

---

[1]$MRS_{L \cdot Y_o}$ is determined by the tangency of the utility function or indifference curve with the budget constraint. This budget constraint is an ordered (highest to lowest) VMP of an additional hour in one of the three different enterprises. Figure 2.1 is concerned primarily with showing the relative trade-offs between the various enterprises.

hours devoted to crops have decreased from h_{c_o} to h_{c_1}, and livestock hours declined to h_{l_1} from h_{l_o}. Thus original work effort has declined from $h_{c_o} + h_{l_o}$ to $h_{c_1} + h_{l_1}$. In our example, no wage work was done before or after the experiment because w was less than $MRS_{L \cdot Y_o}$.

It should be noted that $MRS_{L \cdot Y}$ will become smaller as hours of leisure are substituted for income. Consequently, $2MRS_{L \cdot Y}$ will occur at a lower level than $2MRS_{L \cdot Y_o}$. If $MRS_{L \cdot Y_1}$ had occurred at a higher level, the livestock operation may have been completely abandoned. Thus it is clear that the impact upon farm operators from the assumptions of this model will be to reduce hours worked. However, the amount of reduction, all else being equal, should be less than that of an urban wage earner. The amount of reduction and where it occurs is determined chiefly by the slope of the VMP curves. Because those are more steeply sloped than the VMP curve (i.e., wage rate) of the wage earner, the amount of reduction for the farmer should be less.

Turning to Equation 15 briefly, which determines how variable capital is allocated, the fact that (1-t) multiplies each term will mean the experiment will have no effect on capital allocation, ceteris paribus. This may not be true if the stream of benefits and costs are not proportioned equally between experimental and post experimental time periods. Equations 14 and 15 are really a simultaneous system with the level of variable capital expenditures determining the VMP curves for labor which influence the amount of labor which in turn influence the VMP curves for capital.

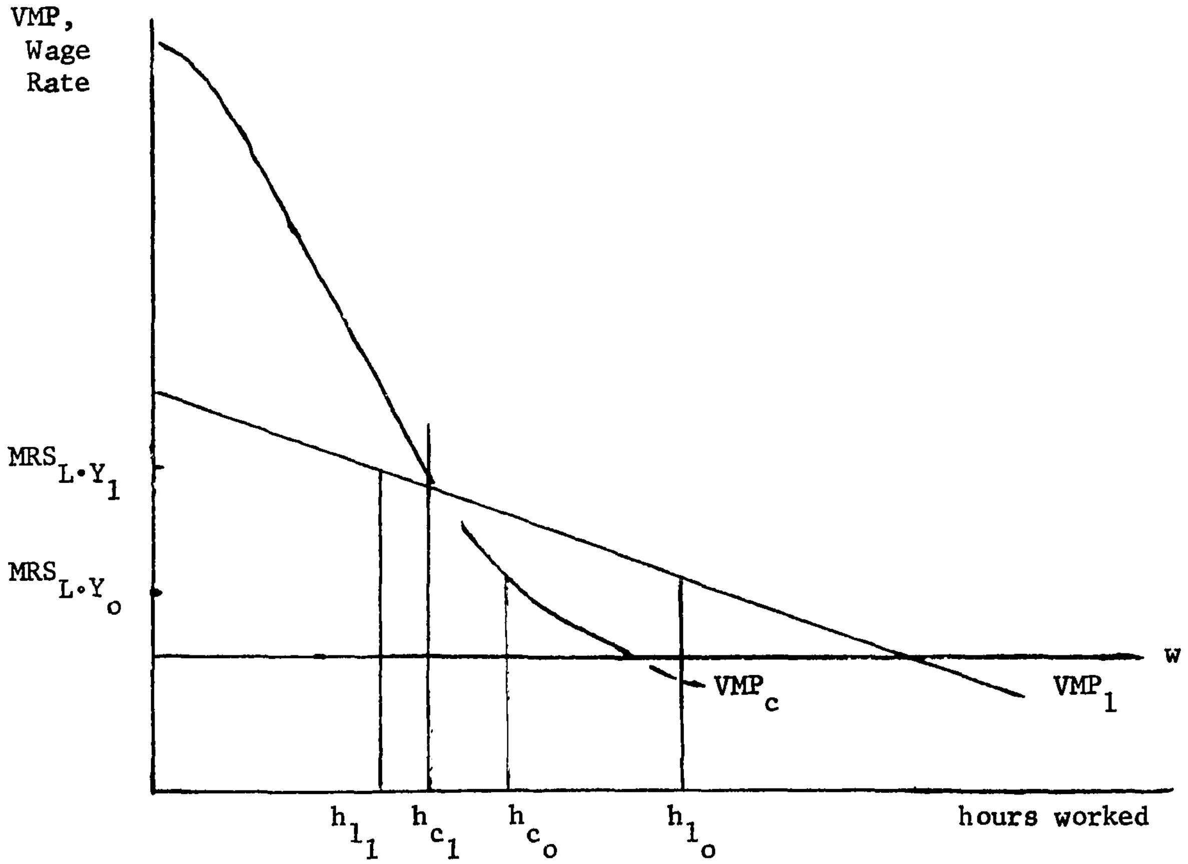

Figure 2.1. The effect of a negative income tax on the allocation of work between livestock, crop and nonfarm activities

It should also be clear that as t increases, the amount of the reduction in hours worked increases. Dividing Equation 13 by (1-t) gives values of 1.42 $MRS_{L\cdot Y}$, 2.0 $MRS_{L\cdot Y}$, and 3.33 $MRS_{L\cdot Y}$ for t of .3, .5, and .7 respectively. A quick examination of Figure 2.1 reveals that with downward sloping VMP curves, the reductions will be greater with higher tax rates.

Also, as GL increases, the amount of reduction should increase. This is true because the additional income should decrease the rate of substitution between leisure and income, everything else constant.

Given the limited duration of the experiment,[1] two factors acting together will determine where the labor response will occur. These two factors are large, physical fixed capital stocks that are highly illiquid and inflexible and relatively fixed coefficient production functions. Response to the experimental stimulus will be negligible when both factors are present. In Iowa, for example, for a given set of farm machinery and managerial skills, the range of feasible labor amounts per corn acre in the relevant range of the production function is small. Essentially the cropland must be readied for seed, planted, and harvested. Any amount less than this will yield zero output. Harrowing the land ten times as opposed to once increases output very little. Some adjustment is possible in

---

[1]Different expectations about the permanence of the program will imply differential behavioral response to the program. If all families view the program as a temporary phenomenon rather than a permanent program, the labor supply response will be biased downwards in comparison with the enactment of a permanent nationwide NIT program. For a mathematical statement and estimation of this bias, the reader is referred to [23, Volume II, Chapter 3]. The arguments in this paragraph are advanced with the assumption that most families view the experiment as being temporary.

cultivating, spraying, or fertilizer applications. However, if the individual farmer perceives his labor VMP curve as steep or that it is extremely important that he cultivate three times, he will behave as indicated.

Basically, with fixed managerial skills, the farmer will apply fixed combinations of labor and variable capital to each unit of fixed capital. Units of this limited fixed capital will not go unused because of its cost and the fact that landlords would still expect a rental payment. The likelihood of temporarily leasing the land is reduced because of the small chance of re-obtaining the land after three years.

However, if either one of the two factors is missing, the labor supply response to the experiment may be large. Small fixed capital enterprises can be abandoned without much monetary loss. Large enterprises with wide ranges of feasible labor inputs could also be affected by the experiment. This leads to the testable hypothesis that those enterprises with relatively high fixed capital endowments and fixed coefficient production functions should have the least reduction in labor input.

As the analysis proceeds from theory to empirical results, several assumptions should be highlighted. First, there is no difference between price expectations (i.e., P_c and P_1) on the part of control versus experimentals. Second, there is no difference in the cost of obtaining capital (i.e., that both groups face the same r). Third, there is no change during the experiment that influences F and G, and fourth, the fixed capital components remain fixed.

CHAPTER III. DEFINITION OF VARIABLES

Choice of Dependent Variable

For the wage earner three different dependent variables, each with several variations, could be examined to determine the impact of a negative income tax program on work effort. These three variables are labor force participation, earnings, and hours worked. For the self-employed there are also many variables, and the results may differ between the variables. These variables are net farm income, gross farm income, and various measures of hours worked.[1] In this section each variable will be carefully defined, followed by arguments citing the strengths and weaknesses of each variable.

Net farm income

Net income refers to gross income minus total expenses during a designated time period, usually a year. Gross income is defined as all income during a time period less the purchase amounts paid for cattle, hogs, or sheep which are sold during the designated time period. This income may occur from livestock sold, crops sold, acreage diversion payment, gas tax refunds, and all other sources of farm income.

---

[1]Most of these measures are concerned with the labor supplied by the entire farm family in the farming enterprise. This study has not focused on the substitution of operator labor for that of his spouse or dependents. Many other factors besides the experiment are more important in determining this substitution. Substitution between hired labor and family labor is covered in this study. One other obvious and interesting question not addressed in this study is the change in off-farm hours induced by the experiment. This area was investigated intensively by another member of the research team working on the RIME experiment [50]. Essentially this study showed a large negative wage work reduction for farm operators in the experimental group relative to the control group. The

Expenses include fertilizer, crop insurance, interest, depreciation, and other similar expense items. Farmers had the option of electing straight line or accelerated depreciation,[1] consistent with their reporting to the Internal Revenue Service (IRS). In all cases the net and gross farm incomes were to be on a cash basis as opposed to the accrual method.[2] For the expenses to be deductible they had to be paid in the designated time period. For example, if the farmer just incurred the expenses by charging them at the local store and not paying for them, they presumably were not reported to us. The prepayment of expenses is immediately deductible even though the goods may be delivered at a much later date. Thus, net income refers to the amount of income received during the time period minus expenses paid during the same time period.

For farmers renting on a cash basis, the rental payment would simply be an expense item. For farmers renting on a share basis, the definition of gross income does not include income paid to the landlord, nor are expenses paid by the landlord included in deriving net income.

---

probability of the farm operator participating in the wage market was greater for the experimental group relative to the control group in Iowa. The reverse was true in North Carolina. The average number of hours worked for those farm operators who participated in the wage market decreased rather drastically in both regions. In the middle year of the experiment, the percentage disincentives were 48 and 28 percent respectively for Iowa and North Carolina. The wives' wage work was mixed but also showed rather large decreases in Iowa. The measure of farm work used in this study was recall hours, and this variable showed rather large positive incentives for the experimental group relative to the control group.

[1]Farmers were asked to report their depreciation expenses once a year on a special form. They were encouraged to report the same depreciation amount as given to the Internal Revenue Service.

[2]Theoretically the accrual method is a better measure of income because it includes the changes in inventory between the beginning and end of an accounting period. However for administrative reasons, the cash basis was chosen over accrual as the accounting method to be used in the experiment.

From a policy relevant viewpoint, net farm income may be the appropriate variable. This measures a family's ability to attain a decent standard of living, and if through the experiment this variable has a treatment effect,[1] it would have substantial cost implications for a national negative income tax program. However, gross farm income and net farm income suffer due to timing implications. A farmer, to a large extent, controls the timing of income and expenses. An experimental farmer relative to a control farmer has a strong incentive to build inventories of grain and livestock and pay expenses immediately, while delaying the receipt of income. In this way his payment would be maximized. A farmer with access to credit would not have much difficulty in paying expenses immediately and delaying income and consequently would show a larger disincentive for income than what actually occurred.

To the extent that control and experimental families have different sizes and kinds of farming operations, a change in the relative farm produce prices would affect the amount of income. Consequently, this would affect the amount of predicted incentive or disincentive. For example, if experimental families on the average had relatively more crop production and less livestock initially, a simple analysis might erroneously suggest a small incentive, given that crop prices rose more than livestock prices. A change induced by the experiment (i.e., a livestock reduction) might be muted. Or if an experimental farmer reduced livestock operations significantly and switched from corn to soybeans,

---

[1]Treatment effect is the generic term for the experimental effects which occur because of the difference in tax rates and guarantee levels between the control and experimental (payment) families in the experiment.

net farm income may not be as drastically affected as the actual quantity of labor supplied.

Reporting of income is probably affected by the experiment. The amount of payments the families received is a direct function of the amount of income that the respondent reports. Consequently, an experimental individual has every incentive to report all the expenses he can. Furthermore, he may even change his behavior and incur expenses within the duration of the experiment that otherwise might be delayed. Examples of this would be major overhauls of tractors or building upkeep. Adjusting the period of analysis would be an attempt to overcome this problem.

A control family, in contrast, has a tendency to round numbers, devotes less time to reporting, and does not have a built-in incentive to report all expenses. Each quarter a question was asked regarding where farm information originated: 1) memory and/or 2) farm account records. There was a significant difference between control and experimentals with experimentals reporting more frequently from farm account records.[1]

The above problems are in addition to the normal data collection problems incurred by collecting information via personal interviews. These include timing (i.e., was the expense reported in the prior period, especially if it occurred close to the boundary between the two periods) and other forms of response errors.

---

[1]For a more complete discussion of this point, the reader is referred to Chapter 5 of this dissertation.

Gross farm income

This variable has been defined previously and its advantages and disadvantages are similar to net farm income. To some extent, reporting problems may be less than those for net farm income. Income transactions are usually bigger, less frequent and consequently likely to be reported more accurately than expenses, resulting in a better measure of gross income than net income. Gross income is a poorer measure of cost implications, disposable income, or a standard of living from a policy standpoint.

Recall hours

During the interviews which were conducted once every three months, a question was asked regarding the number of hours worked by the respondent on his farm or business the previous week. This is entirely a recall question. No effort was made to have the respondents keep track of the number of hours worked on a day-to-day basis. By multiplying by 13 and summing over the four quarters in a year, a yearly estimate of hours worked could be made.

The chief problem with this variable is the Hawthorne[1] type effect. While a strict intrepretation of the Hawthorne effect is a change in behavior resulting from observation, this will be extended to include changes in reporting behavior upon which the analysis is based. For example, imagine that you have an eighth grade education and are receiving $1,500 a year from "heaven." On each opportunity that you see the

---

[1]Hawthorne effect refers to the general effect where experimental groups alter their behavior because they realize they are being watched or monitored.

benefactor, there is probably a tendency to convey an image of a hard-working and industrious farmer. There might even be a feeling that the payments are somehow tied to their work effort, despite statements from project administrators to the contrary.[1]

The control families, however, have no such incentives. They do not receive much financial reward (i.e., $100 per year for a family with a head and a spouse) and by the fourth or fifth interview there is an effort to get the interview over with quickly. Consequently, estimates of the response based upon this variable may be incorrect.

The interviews were conducted during the months of March, June, September and December. Because of the seasonal nature of farming, multiplying these four one week snapshots by 13 may not give a correct yearly estimate of hours worked.

Furthermore, inclement weather the week before or during the interview may affect the number of hours worked. If one is willing to assume that these time effects are distributed randomly with respect to the treatment variable, they should not affect the resultant direction of the conclusion. Data from other studies for similar farming operations could probably reduce the influence of these last two factors, but the Hawthorne effect problem would still exist.

---

[1] During August 1970, or approximately nine months after the study began, the respondents were tested on their knowledge of the rules of the experiment. At the time of enrollment and in this following test, the families were told how the check was calculated and that there were no strings attached on how this money could be used. See [25, Chapter 4] for a more complete analysis.

Budgeted hours

Each year the respondents were asked the number of acres grown in each crop and the number of different kinds of livestock that were sold. By obtaining a coefficient from other published research and extension farm planning manuals (Acker, et al., 1968-69 [1]; James, 1968 [46], Missouri Farm Planning Handbook, 1972 [70]; Van Arsdale, 1962, 1965 [92]; Willet and Saupe, 1973 [98]), an estimate of the number of hours a particular farming enterprise required was constructed. The table below lists the coefficients used for each major enterprise (the same coefficients were used for both North Carolina and Iowa).[1]

Crop	Hr/Acre	Livestock	Hr/Animal
Corn	5.5	Feeder pigs	.7
Wheat	4.0	Market hogs (birth to market)	2.5
Tobacco	292.0	Market hogs (40 lbs to market)	1.8
Soybeans	5.0	Feeder calves (implies a stock cow herd)	13.0* # stock cows
Oats	5.5		
Hay	7.0	Fatten cattle (500 lbs to market)	4.6
Beans/Peas	45.0	Chickens	.0125
Blueberries	547.0	Sheep	1.8

[1]This statement is misleading in that for those enterprises grown in both states such as corn, soybeans, and some livestock activities, the percentage of total labor hours these common activities constitute in North Carolina for this study is small.

Crop	Hr/Acre
Pickles	150.0
Sweet Potatoes	102.0
Squash/Eggplant	157.0
Peppers	138.0
Milo/Sorghum	6.2
Diverted Acres	1.0
Watermelon	96.0
Sunflowers	4.0

Livestock	Hr/Animal
Eggs (Dozen)	
<4000	.13/doz.
≥4000	.04/doz.
Goats	2.0
Contract pigs (respondent labor only)	.25
Feeders (small calves to feeders)	5.0

	# of cows	Total Hours
Dairy cows	1	240
	2-20	240+85* cows
	21-40	840+55* cows

The chief advantage of using this particular variable is that it is less likely to be affected by the Hawthorne type effect. It also suffers least from other reporting problems. The variables required are numbers which farmers are more likely to remember; the number of acres of row crop or the number of cattle in the feed lot is a status symbol. Furthermore, it is these numbers that are basic to all farm management decisions. For example, the number of acres in corn and soybeans would serve as a guide to how much fertilizer and seed to buy.

There are four distinct disadvantages in using budgeted hours.

1. The main disadvantage in using this variable is that the per unit labor requirement is assumed to be the same for all farmers despite differences in amount and size of farm machinery, farming methods employed, and the level of farm operator skills.

2. There is farm overhead labor not directly related to any productive enterprise, such as yard, ground and buildings maintenance, repairs of unspecialized machinery (e.g., general purpose tractor), and gathering of marketing or technological information. This suggests that accurate recall will exceed accurate budgeted labor estimates, but they should be highly correlated.

3. Some problems in constructing the coefficient for each operation were encountered. The categories for which budgeted hours were available from outside sources did not match categories from the interview. For example, from the interview the number of feeder pigs and market hogs sold, the dollar value of the sales, and the cost of any purchased animals are known. The labor coefficients that can be obtained from previous studies are for hog operations that a) raise pigs from birth to market, b) raise pigs from 40 pounds to market, or c) raise feeder pigs to 40 pounds.

Some assumptions are required to convert the information from the interview into the number of different pigs falling under each of the three categories for each farmer. For the most part, the assumptions seem realistic and are applied to both control and experimental families indiscriminately.

4. Another criticism of using budgeted hours as a measure of operator labor is that it represents the labor required by the entire farm operation -- the operator, his family and hired labor. Consequently it would be possible for an experimental farmer to increase his production level slightly but substantially increase the amount of labor hired, thus having

the net effect of reducing his own labor input. This would then be erroneously misconstrued as an increase in labor input when it clearly was not.

Scaled hours

As noted above, budgeted hours suffers because the coefficients do not reflect differences in farm sizes or methods. Scaled hours does adjust for farm size by making the simplistic assumption that size of operation perfectly predicts farming methods employed. For example, if a farmer has 300 acres of corn, he probably has large tractors with a five to six bottom plow, while a farmer who farms only 60 acres of row crop probably has small tractors with a two to three bottom plow.

The coefficients were constructed in the following manner. All livestock hour requirements were assumed to be explained by a second-degree polynomial of the form $ax^2 + bx + c$ where x is the number of animals sold. Two end points which covered the range of livestock activities and one point in the center were chosen and hours from previous studies [1, 46, 70, 92, 98] assigned to these quantities. The values of a, b, and c were then computed. The values and function for each livestock operation are shown in Table 3.1.

All other livestock operations were assumed to be linear and hence were the same as budgeted hours.

For crop operations a question asked in the post interview about machinery size was used to determine which coefficients to use, while for tobacco a simple function of yield and acres determined the number of hours. Using information from the post experimental period interview

Table 3.1. Description of functions used in generating scaled hours

	Livestock		
	Low	Middle	High
Fat cattle			
Hrs/head	12	8	4
Quantity	1	60	400
Function	$-.0116x^2 + 8.64x + 3.37$		
Stock cows			
Hrs/head	40	15	10
Quantity	1	22	60
Function	$-.1136x^2 + 16.42x + 23.69$		
Dairy cows			
Hrs/head	240	115	100
Quantity	1	15	40
Function	$-.3864x^2 + 112.25x + 128.14$		
Eggs			
Hrs/head	100	1,120	4,375
Quantity	1,000	16,000	125,000
Function	$-.000000308x^2 + .0732x + 27.11$		
Ewes			
Hrs/head	9	5	4
Quantity	1	35	70
Function	$-.0273x^2 + 5.865x + 3.1623$		
Hogs (feeder pigs)			
Total hours	13	200	591
Quantity	8	150	500
Function	$-.0004x^2 + 1.38x + 1.99$		
Market hogs (birth to market)			
Total hours	45	529	1,620
Quantity	8	184	720
Function	$-.001x^2 + 2.94x + 21.54$		

could cause problems if the variables used had been affected by the treatment. In this case, if investment in machinery were related to treatment parameters, an analysis of the hours variable could be misleading. However, as will be shown later, there is no evidence of this. The coefficients for the Iowa grain crops are shown below.

Hours per acre

	Corn	Soybeans	Oats	Hay	Diverted acres
4 Row	5.7	4.7	5.5	7.0	1.0
6 Row	4.7	3.8	5.5	7.0	1.0
8 Row	3.8	2.8	5.5	7.0	1.0

While the scaled hours variable may be an improvement over straight budgeted hours, it makes some very arbitrary assumptions which should not be overlooked. For example, if a farmer has a reasonably large farm and then makes a reduction in farm size, the method would understate the amount of work reduction. This is true primarily because the method and capital employed would still be the same, and yet in making a reduction in the number of animals, scaled hours automatically presumes an increase in per unit labor requirements. On the other hand, the amount of hours is understated when a farm increases its level of production without changing method or amount of farm capital employed.

For both budgeted and scaled hours, it was possible to construct crop and livestock hour variables. Also by making assumptions about wage rates and the amount of labor connected with machine hire and custom work, a measure of scaled hours can be constructed which represents total farm family labor. Results will be presented for each of the major classes of dependent variables described above.

Estimation Methodology

Because the study was conducted over a three year period, a natural problem of pooling time series and cross section data arose. Consequently the prime tool for these analyses was a time series - cross section pooling regression program developed from the procedures in Nerlove (1971)[73].[1] The sample used was for farm families with a male head less than or equal to 69 years of age at the beginning of the experiment, who farmed each year, had constant marital status, and had at least 400 hours of work in one of the years between 1969 and 1972. Eight cases were eliminated because of missing information and faulty data.

If the experiment induced many individuals to leave farming, the above subpopulation definition would bias the treatment effect upward. However, if a few individuals left farming for reasons other than the experiment, and these were concentrated slightly in the experimental group, and no variable represented the reasons, then the treatment effect would be biased downward. The latter is the primary reason for the population chosen. Separate analysis is needed to study movements in and out of farming.

Because of different production functions and methods of production, the Iowa and North Carolina samples were split. Tobacco growing is

---

[1]During most of the preliminary work on this chapter, ordinary least squares were used. There is little difference in overall results between the two estimation techniques.

primarily a labor intensive operation while corn-growing in Iowa is a highly capital intensive operation. This is reflected readily in the tables in Appendix A which illustrate sample averages for a number of important variables by year and treatment in both Iowa and North Carolina.

Each year of the experiment for each farmer constitutes one observation. This has the advantage of increasing degrees of freedom and allowing the behavioral response to the program over time to be traced. The number of individual farmers by site is shown in Table 3.2.

Table 3.2. Number of farmers by plan by region

Tax Rate	30	50	50	50	70	Total	Control	Total
% Guarantee	75	50	75	100	75	Exp.		
Iowa	13	9	13	10	5	50	54	104
North Carolina	9	11	15	11	6	52	42	94

One of the methods for analyzing the experimental effect would be to construct a demand-supply model. A properly constructed model could estimate correctly the treatment coefficients taking into account the simultaneity problems inherent in Equations 14 and 15. Such a model, however, needs adequate measures of the VMP of labor and capital, respectively in crop and livestock operations. Given the income data, the ex post nature of the data, and poor measures of capital, no attempt was made to estimate the structural equations. The prime focus of the study is to estimate accurately the treatment parameters in a reduced form model. The following is a description of the major independent variables used in the analysis.

Major Independent Variables

These variables fall naturally into four groups - control, experimental, interaction between control and experimental variables, and measurement variables. Control variables are included to insure that differences between the control and experimental samples, which occurred as a by-product or accident of random sampling, are not included in the coefficients on the treatment variables. They are needed to yield consistent experimental coefficients given the stratified design of the sample. Also, these variables are often important determinants of the dependent variables. Without their inclusion, the equations would suffer from specification error resulting in biased experimental variable coefficients. Experimental variables are the different formulations of the tax and guarantee parameters. Interactions are needed between the two groups of variables to test for differences between subpopulations (e.g., race) or as further refinements to determine where the treatment response took place. Measurement variables alleviate misinterpreting results when major explanatory and dependent variables are measured with error. The following is a description of the major independent variables used in the analysis, discussed in order by group.

Control variables

Pre-farm size These variables are the 1969 values for some variant (total, crop, livestock) of scaled or budgeted hours. They have been entered as linear, quadratic, or spline functions. Their primary role is to control for pre-experimental work effort, motivation, family needs, and size and kind of farming operation.

When scaled or budgeted hours is the dependent variable, these variables are a pseudo lagged dependent[1] variable and essentially convert everything into a "change" model. Consequently, when age or family size are entered into the equation, their interpretation must be to explain the change between 1969 and the other years. This lagged dependent variable should reflect the variation in hours worked in 1969 due to age, family size, and other such demographic variables. In the linear model, the coefficient should be 1.0 or slightly smaller and be highly significant.

Age Age is entered primarily as a control variable for the stage in life cycle. Theoretical expectations and evidence from other studies indicate that work effort is related to age.

This variable has been entered in several forms including the following:

1. Age
2. Age, Age^2

[1]The word pseudo is used because the lagged dependent variable is always the 1969 variable even when the observation pertains to calendar year 1971 or 1972. A true lagged dependent variable would associate 1970 with 1971 and 1971 with 1972.

3. Age, 1 if over 55

4. Age

Age55 (0 if ≤ 54)
(age if > 54)

<u>Wage</u> Throughout the farm analysis, participation in the wage labor market is assumed to be an exogenous variable.[1] The purpose behind entering this variable into the farm analysis is to control for changes in the level of participation in this other market. Changes in wage work probably affect the amount of farm work. Consequently, omitting the variable could lead to biased and inefficient treatment coefficients (i.e., standard problems associated with incorrect specification).

An interaction with the experimental variable is avoided so incorrect inferences will not be made. The purpose of this study is to estimate farm work response to the experiment. Thus, forcing the effect of a change in off-farm work on farm work to be identical for control and experimental families simplifies the resulting analysis. For example, assume a reduction in wage and farm work occurs for the experimental group. A wage-treatment interaction would capture some of the reduction in farm work thus making an interpretation of the other treatment variables more difficult.

Several formulations of this variable have also been tried, including the following:

1) Current year's head's wage and business recall hours

---

[1]For an analysis where this assumption is relaxed, the reader is referred to [50].

2) Change in head's wage and business recall hours. The change is calculated as deviation from 1969 head's wage and business recall hours.

3) 1 and 2 combined.

Education This variable is included as a proxy for a farmer's management ability. The variables used are the number of years of formal education completed and the Ammons and Ammons Quick Test.[1]

The effect of education on work effort or income is indeterminate. A higher education or management level might be associated with a greater ability to manipulate the payments system. For example, a greater ability to time income and expense streams so as to maximize payments would show a greater income disincentive and a smaller work disincentive. One could also argue that more education would result in a higher degree of farm mechanization. Consequently, an operator would have to work less for a similarly sized farm operation. In other words, a higher education may provide a way for the farm operator to become more efficient. On the other hand, education may serve as a proxy for an incentive variable. Those individuals with higher education might have more incentive to farm bigger operations. If there is no other control for degree of management, education would probably have a positive influence on hours worked. However, given the change model, there seems little justification for an education variable except as a control for the random selection process.

[1]The Ammons and Ammons quick test score is based on a word recognition test. The scale range is 0-50.

Family composition and size As family size increases, it is expected that the number of hours that the family works will increase. The larger the family, the greater is the need for a larger income. However, dependents aged 14 and above may supply labor to the farm and, hence, the head may be able to reduce his work effort. Again given the "change" model, there seems little justification for a family unit size variable.

Family composition is assumed to be exogeneous and given the importance and amount of labor needed to grow tobacco, several variables representing size and composition were added to the model in North Carolina to insure treatment coefficients were not drastically affected. These variables included the number of members in given age groups and a dummy variable for whether a spouse is present.

1969 normal income This variable was constructed by the field staff from the screening and pre-enrollment interviews. This variable was the basic stratification variable in assigning the family to a cell.

Farm asset variables In the models presented here, the role of assets is difficult to define. There are several aspects that enter into the picture. Total farm debt divided by total assets reflects a farmer's ability to obtain credit. One can argue that as this ratio increases, the farmer has a certain inflexibility about reducing his work effort. Many authors have pointed out that for wage earners, the 40 hour work week reduces freedom of choice involving the number of hours worked. This lack of substitutability will reduce the impact of a short-run experimental program. Similarly, large debts will reduce the flexibility of altering work patterns for a farm operator.

The amount and number of assets also reflects a farmer's past work behavior or incentive along with several other phenomena including inheritances. Consequently, the major role played by assets is as a control variable. If by accident of random sampling the number of assets were unevenly distributed between control and experimental, the coefficients on the experimental variables could be distorted.

Treatment variables As noted earlier in the labor supply chapter, increasing the tax or guarantee level should result in a larger reduction of hours worked. Consequently, besides a simple control-experimental dummy (C/E), other variables such as tax rate or guarantee percentage may be important in explaining the experimental response. Entering one variable for tax rate (4 values: 0, 30, 50, 70) or one variable for the guarantee percentage (4 values: 0, 50, 75, 100) places a linear restriction on the response. In essence it restricts or forces the middle plan (i.e., 50 percent tax rate, 75 percent guarantee) to lie between and equidistant from the other two plans. A more unrestricted formulation is to enter each plan as a separate dummy variable. However, even this is a highly restricted model. It presumes that the treatment effect should be constant for all levels of farm size, debt ratios, or age. These interactions, which may be important explanatory variables of where the experimental response took place, are described below.

Treatment interactions To remove the restriction that the treatment response be identical for all levels of farm size, an interaction between farm size and C/E is required. Economic theory sheds little light on whether there should be a differential effect or what direction it would go with respect to farm size. Quadratic and spline interactions were also tried. The former was chosen over the latter because the adjustment process is forced to be continuous, and it forestalls justifying arbitrary spline functions.

Because of institutional reasons alluded to earlier, the amount of debt could limit a farm operator's choice. This was empirically investigated by including a C/E interaction with the debt ratio. The expected sign is positive, i.e., as the debt ratio increases, the amount of work effort should increase. Or if there is a disincentive, the amount of the disincentive should be less.

The last primary interaction concerns age. Stage in life cycle could alter response. In particular, as one approaches retirement age, the response might be greater. This was investigated by including one of the following formulations:

1) Age x C/E
2) Age if over 54 x C/E,
   otherwise 0
3) 1 and 2 combined

Measurement error variables These variables probably are poorly named, but hopefully will explain some of the variation in the experimental group response. A simple look at the distribution of scaled hours worked for 1970-1972 versus 1969 hours worked for the experimental group shows a considerable variation. Some experimental families increased their hours substantially while other families' hours declined substantially. There are several possible explanations, some of which fall under the general heading of measurement error problems. For example, as argued earlier, experimental farmers who can alter timing of income and expenses are not affected in the same way as farmers who cannot alter their timing. The following four subtopics fall under this generic classification.

The first subtopic involves incorrect selection of several high-income farmers. From the pre-enrollment interview, an estimate of normal income determined eligibility. If errors were made in this determination, however, and a large percentage of the farmers had normal incomes well above the breakeven level, the experiment probably would have little or no effect upon work effort. Thus the average treatment effect would be biased (probably downward), if inference were made to a population of low income farmers. This selection problem was particularly acute in Iowa where, for example, several well-to-do and highly efficient farmers became eligible because of hail damage in 1969 which temporarily reduced their cash income. A variable which indicated whether a family was over the breakeven level in 1969 based on ex post information rather than expectations in August to

October of 1969 on the pre-enrollment interview was added to the model to partially correct for these measurement problems.

A second subtopic relates to the timing of income and expenses. Particularly in year three, farmers who could control the timing of their income and expenses could have substantial payments without any reduction in work effort. Consequently, their marginal tax rate has been altered and their behavioral response will be different. An attempt is made to capture this effect in several ways. The amount of net equity and the ratio of total debt to total assets reflects the amount of credit a farmer has. This, in turn, reflects a farmer's ability to change the timing of income and expenses. A farmer with a good line of credit does not have a need to sell in a certain time period and also can incur expenses before receiving receipts. An effort was made to classify income which could be delayed. Milk and livestock sales are more inflexible than crop sales. However, the use of this variable was rejected because of simultaneity problems.

A third subtopic is the reporting behavior of farmers. Farmers who intentionally or unintentionally misreport income to the payments process via their monthly income report forms might behave differently. Upon the edited estimates of quarterly[1] income, an estimate of the amount of payments farms should receive can be estimated.[2] Based upon this, some

---

[1]For a complete description of data editing methods see Appendix B. For a discussion of changes which arose throughout the analysis, the reader is referred to Chapter 5.

[2]The same mathematical formula as was used in the actual payments was applied to information collected by the surveys. All components of income and expenses were included. The survey's estimate can also differ because of the accounting period. Actual payments used a one or three month averaging period with a 12 month carryover procedure. Surveys did not

farmers received one thousand dollars more in payments than they should have received.[1]

Theoretically, these experimental farmers are facing a different marginal tax rate than their assigned tax rate. A change in their true income does not affect their payment as the subscribed mathematical formula underlying the NIT transfer scheme would indicate. Consequently, their behavioral response will be altered as compared to other farmers in the experimental group. The income effect would still be present, but the substitution effect would dissipate. The strength of the substitution effect is determined primarily by the amount of misreporting and closeness to the breakeven level. Other studies, mentioned in the introduction, have differed upon the relative weight of income versus substitution effects. If we assume these weights are approximately 1/3 and 2/3, respectively, and if there is an average disincentive of 9 percent, farmers whose assigned tax rate was completely altered (i.e., became zero or like control) would have only a 3 percent effect. Furthermore if this income were viewed as completely transitory, the effect would be zero. As a partial control for this phenomenon, a variable (PN) was added which is actual minus predicted payments divided by the mean of the two variables. Since it is hypothesized that income level and work effort do not

---

collect information monthly, so the accounting period was simply a year with no carryover scheme. This could not have, particularly in the last two years of the experiment, accounted for some of the huge differences in actual versus predicted payments.

[1]Predicted payments were on the average $845 less in North Carolina and $676 less than actual payments in Iowa. The standard deviation of the differences was $845 and $1030 respectively.

explain the misreporting behavior, this variable should not be correlated with the error terms, and its corresponding undesirable effects upon the estimaters from a multiple regression analysis should not be present.[1]

A fourth subtopic is the often weak connection between a farmer's income and his work effort. Farmers experiencing damage from hail or livestock disease problems may alter their behavior in response to these stimuli. Because these events may not be randomly distributed between control and experimental groups, and because of the small sample size, they might give a distorted view of what is happening.

Each quarter a farmer was asked in an open ended question if he were satisfied with the production of each crop and livestock activity. If not, he was asked why not. Reasons indicating random weather or disease effects were noted and coded as dummy variables.

---

[1]The correlation coefficients of the dependent variables with PN were always less than 0.20. Attempts to explain the variation in PN with a variety of variables always netted an R^2 less than .25.

CHAPTER IV. RESULTS

To summarize the results takes ingenuity and perseverance. Advantages of using a reduced form estimation procedure are the relative straightforwardness of interpreting the regression results and the use of a similar model for all dependent variables. Without sacrificing any distinctly different results, only findings related to net and gross farm income and scaled crop, scaled livestock, total scaled, adjusted scaled, head and total recall hours will be presented. Budgeted hours results are practically identical to scaled hours results. Adjusted scaled hours are total scaled hours minus hired labor expense divided by two minus hired machine expense divided by ten plus custom work done divided by ten. The divisions convert dollar amounts into hours. Total recall hours is a summation of head hours plus .9 times spouse recall hours plus .8 times dependent recall hours.[1]

The results for each region will be presented separately. After a basic model is described, several simple parameterizations of the results will be presented. From these simple models a flavor of the overall results can be obtained. A selected complex model will be presented along with partial F-tests for classes of independent variables and predicted incentives between control and experimental groups will be estimated.

---

[1]These numbers were deemed reasonable by farm extension specialists at Iowa State University.

Iowa

The basic model chosen for presentation was estimated by the time series-cross section pooling regression program in the following form:

$$Dep = b_o + b_1\ HR69 + b_2\ AGE + b_3\ AGE55 + b_4\ EDUC + b_5\ DR + b_6\ FE + b_7\ \Delta HDWBH + b_8\ D71 + b_9\ D72$$

where the variables are defined in the following way:

HR69 - 1969 total scaled hours. When the dependent variable is scaled crop or scaled livestock hours, the associated independent variable is 1969 scaled crop or scaled livestock hours respectively.

AGE - age of the farm operator

AGE55 - age of the farm operator if greater than 54; otherwise zero

EDUC - number of years of formal schooling of the farm operator

DR - total farm debts/total farm assets

FE - net farm equity

ΔHDWBH - the change in amount of off-farm work done by the farm operator

D71 - 1 if 1971; 0 otherwise

D72 - 1 if 1972; 0 otherwise

In Table 4.1, several models representing simple treatment parameterizations are presented. Model I adds to the basic model a simple dummy variable (C/E) where one represents experimental and zero represents control. The coefficient represents an average difference between control and experimental groups controlling for the variables in the basic model.

If the coefficient is negative, the experimental group declined relative to the control group for the dependent variable in question. The F-value determines if this difference is statistically significant. The reverse is true for a positive coefficient.

Model II is slightly more complicated in that both C/E and C/E interacted with 1969 total scaled hours are added to the basic model. This removes the restriction that the difference between the control and experimental groups must be equal for all levels of the 1969 scaled hours variable. It allows the treatment effect to be different between large and small sized farms. For example, if the treatment effect is always a ten percent reduction with respect to 1969 total scaled hours, the interaction would allow this percentage to be estimated. Model I only allows a constant absolute difference in hours worked or income earned. The F-value represents whether the two treatment variables jointly and significantly add to the explanation of the dependent variables. The C/E coefficient and interaction may now be of different signs and it may no longer be readily apparent what the percentage incentive or disincentive is. One estimate of the percentage can be ascertained by evaluating the model at different levels of the HR69 variable. In Table 4.1 this percentage incentive has been calculated for the mean value of HR69.

Model III is again slightly more complicated than Model II. Two variables, HR69 and $HR69^2$ interacted with C/E have been added to Model II. Essentially, the response can be curvilinear rather than the linear restriction imposed by Model II. The sign on the treatment variables may be different so the direction of the effect can not be directly seen.

Table 4.1. Iowa treatment coefficient estimates, F-values and predicted incentives for models with simple treatment parameterizations for selected dependent variables

	Dependent Variables							
	Net Farm Income	Gross Farm Income	Head Recall Hours	Total Recall Hours	Total Scaled Hours	Adjusted Scaled Hours	Scaled Crop Hours	Scaled Livestock Hours
Model I								
C/E	-381	-1389	23	97	-168	-222	-23	-150
F	.40	.95	.29	.39	6.37[a]	7.84	.41	5.35[a]
Model II								
C/E	-1772	1088	246	263	9.08	-31.6	94	-131
C/E*HR69	.70	-1.25	-.11	-.08	-.089	-.10	-.11	-.021
F	.78	.81	.31	.32	4.01[a]	4.60[a]	.99	2.71[c]
% Incentive	-8.8	-7.0	1.0	3.9	-8.0	-11.5	-2.1	-15.8
Model III								
C/E	-4761	6561	351	314	136	-165	-125	-70
C/E*HR69	3.88	-6.51	-.18	-.10	-.21	.048	.24	-.191
$C/E*HR69^2$	-.00066	.00110	.00001	.00000	.00002	-.00003	-.00011	.00006
F	1.07	.70	.28	.32	2.52[c]	2.94[a]	.28	2.02[d]
% Incentive	7.5	-10.3	1.8	4.9	-8.5	-9.7	.1	-19.7

[a]Significant at the 5 percent level.

[b]Significant at the 1 percent level.

[c]Significant at the 10 percent level.

[d]Significant at the 20 percent level.

This is resolved by evaluating the model at several levels of HR69. Only the mean value of HR69 is reported in Table 4.1 along with the overall F-value on the three treatment variables.

Examining the treatment coefficients of Model I as presented in Table 4.1, net farm income, gross farm income, head recall, total recall, and scaled crop hours are all insignificant. Scaled livestock hours is negative and significant which in turn makes total scaled and adjusted scaled hours negative and significant. The average experimental effect was -152 for scaled livestock hours, -173 for total scaled hours, and -230 for adjusted scaled hours. Since adjusted scaled hours is more negative than total scaled hours, it indicates experimentals hired more labor or did less custom work.

In Model I, both income variables are negative but insignificant. This agrees with expectations in that the least profitable enterprise should be curtailed. This will lead to an hours reduction but less of an income disincentive.

These basic results hold up in Models II and III. Scaled livestock hours continues to be negative and significant making the measures of total scaled hours negative and significant. The income variables are insignificant although the percentage incentives are sometimes positive, indicating that experimental families are earning relatively more than control families with similar characteristics. This is somewhat surprising given that experimental families have a built-in incentive to underreport income, overreport expenses and consequently show less net farm income than their control counterparts. This, however, is not true for all levels of HR69; it is not significant, and therefore can be discounted.

The percentage incentives are calculated by subtracting the control group y-hat ($\hat{C}$) from the experimental group y-hat ($\hat{P}$), dividing by $\hat{C}$ and multiplying by 100. This percentage can be interpreted as the average percentage amount the experimental changes relative to the control group holding all other variables constant. As previously mentioned the y-hats are calculated by using the mean values for all independent values. In Models II and III these percentages are -16.6 and -20.4 percent for scaled livestock hours, -8.3 and -8.9 percent for total scaled hours, and -12.3 and -10.7 percent for adjusted scaled hours respectively.

Both measures of recall hours (total and head) are positive and insignificant. This is contradictory to the results obtained from scaled hours and raises the question of which variable is to be believed.

Proving unequivocally which variable is correct is a highly elusive goal. A measurement of hours worked for the farmer is the intersection between his labor demand and labor supply curves. It cannot be argued that recall hours represents the supply curve, and budgeted hours represents a labor demand curve. Both variables are attempts to measure the intersection of the two curves by different methods. To establish which is the better variable, one must determine which measurement technique is best. To argue that economic theory should determine which is correct because one agrees better with the theory is not an independent test of the theory and not a particularly useful way of doing empirical research. The variable which is judged best by independent criteria, statistical methodology, or whatever, will hopefully correspond best to theoretical predictions.

It probably is a valid criticism to argue that budgeted hours does not measure the same thing as recall hours. Budgeted hours contains no measurement of overhead hours. It only measures time spent in direct productive activities. Overhead activities include time spent bookkeeping, maintaining buildings, repairing fences and machinery, and driving to town on errands. If it is assumed that these activities are proportional to productive hours, then the estimate of scaled hours understates rather than overstates the amount of the disincentive.

Remember that recall hours are based solely upon a respondent's declaration of the number of hours worked the previous week, while scaled hours depends directly upon numbers of acres of corn, acres of soybeans, hogs sold, cattle sold, etc. These numbers are more likely to be remembered because they influence all subsequent management decisions. This is not to say that there were no errors in the reporting of year end numbers; but if and when corrections to these quantities were needed, the changes could be documented.[1]

A further argument can be advanced. Presume that one has a number of different productive activities which need to be summarized as one. One immediately thinks of constructing a linear combination of these different

---

[1]For a more complete discussion of this point, the reader is referred to Chapter V of this dissertation.

activities. It is contended here that those weights should represent labor coefficients from other studies. However, because individually most productive enterprises decrease in quantity with respect to treatment variables, all reasonable coefficient sets would show a negative treatment effect. Consequently, if recall hours and budgeted hours are to be reconciled, three possibilities arise:

1) A Hawthorne effect does exist and is responsible for the difference.

2) While the experimental families have reduced their farm work effort in terms of growing corn or raising pigs, they have increased the number of overhead hours. While this may be true, it certainly is not policy relevant that the farmer must now take five trips to town instead of the previous two to accomplish the same mission.

3) An accident of random sampling resulted in experimental families using more labor intensive techniques of production. This hypothesis is essentially negated, however, when various functional forms of HR69 are entered in the model.

Furthermore, evidence from the quarterly interviews would indicate similar investments in machinery and equipment. Table 4.2 shows for experimental and control groups respectively the overall means of initial machinery and equipment stock plus the amount of investment each year. The simple test of means is not significant for either the initial stock of machinery or equipment or the subsequent investment by year in machinery and equipment.

Thus, depending upon the context within which the treatment effects are evaluated, differing conclusions can be reached. However, the result which seems policy relevant in terms of farm goods produced would indicate a disincentive of approximately 8 to 12 percent.

Table 4.2. Overall means of initial machinery and equipment stock and investment by year, treatment, and region[a]

	Initial Machinery Equipment Stock	Gross Investment 1970	Gross Investment 1971	Gross Investment 1972
North Carolina				
Control	564	244	408	620
Experimental	554	134	143	449
t-value	.04	1.24	1.65	.52
Iowa				
Control	10,199	1,473	993	2,615
Experimental	10,943	1,157	1,404	2,102
t-value	-.74	.66	-1.10	.79

[a]All simple t-values are insignificant.

These initial results can be expanded by determining if the treatment effect is related to age, to guarantee or tax rates, to measurement variables that were previously discussed, to year or to amount of farm debt. To test these hypotheses, partial F-values were calculated and are presented in Table 4.3. Two equations from which some of the F-values were derived are shown in Tables C.1 and C.2 in Appendix C for each dependent variable.

The results in Table 4.3 will be discussed line by line. To the basic model, five treatment variables are added and the joint F-statistics calculated. Similar to the simple models previously discussed, the

Table 4.3. Iowa partial F-statistics for selected dependent and independent variables

Groups of Independent Variables	Dependent Variables: Net Farm Income	Gross Farm Income	Head Recall Hours	Total Recall Hours	Total Scaled Hours	Adjusted Scaled Hours	Scaled Crop Hours	Scaled Livestock Hours
Due to HR69*C/E, HR69^2*C/E, D70*C/E, D71*C/E, D72*C/E	.79	.58	.57	1.08	1.81[a]	2.20[b]	.30	1.69[a]
PN	28.73[c]	21.25[c]	.44	.81	1.85[a]	2.22[a]	.03	2.22[a]
DR*C/E	.0009	2.82[b]	.006	.07	4.45[d]	.36	8.94[c]	.64
AGE*C/E, AGE55*C/E	2.49[b]	2.37[b]	3.81[d]	4.85[c]	2.51[b]	.81	2.52[b]	.97
All of above except PN	2.16[d]	1.71[b]	1.34	1.93[b]	2.26[d]	1.77[b]	1.53[a]	1.50[a]
HR69*C/E, HR69^2*C/E, C/E	1.03	.70	.28	.32	2.52[b]	2.94[d]	.28	2.02[a]
D71*C/E, D72*C/E[e]	.42	.38	1.02	2.22[a]	.75	1.09	.34	1.19
T30, T70[f]	.52	3.98[d]	.14	.08	.19	.22	.54	.08
G50, G100[f]	.56	2.80[b]	.56	.52	1.35	1.29	1.39	1.25
T30, T70, G50, G100[f]	.21	4.08[c]	.76	.87	.98	.91	2.04[b]	.78
Family Comp. Variables[g]	.71	1.01	1.24	1.30	2.04[b]	2.21[d]	3.34[c]	.80

[a]Significant at the 20 percent level.
[b]Significant at the 10 percent level.
[c]Significant at the 1 percent level.
[d]Significant at the 5 percent level.
[e]D70*C/E, D71*C/E and D72*C/E are year dummies interacted with the simple treatment variable C/E.
[f]T30, T70, G50, G100 are all simple dummy variables representing each of the different experimental plans. The 50% tax and 75% guarantee plan has been excluded.
[g]See footnote on page 61.

treatment variables are significant in the scaled livestock, total scaled, and adjusted scaled equations.

While the main treatment variables were not significant for net and gross farm income in Iowa, a "measurement" variable (PN) is highly significant. Recall that this variable is constructed by subtracting predicted negative income tax payments from survey data from actual negative income tax payments and dividing by the mean of the two variables in the numerator. PN has a mean of .72 and a standard derivation of 1.07 for Iowa experimentals. As expected, the sign is positive in all equations where the variable is significant. If the family received overpayments, the effect of the negative income tax transfer scheme was less (more incentive, less disincentive) when compared to other experimentals with similar characteristics. This variable was significant in all equations except the measures of recall hours and scaled crop hours.

When the debt ratio interacted with treatment was added to the model, it was negative and significant in the gross farm and total scaled hours equations. This is contrary to expectations in that experimental farmers with high debt ratios would be expected to have less flexibility in reducing their work effort. However, relative to their control counterparts they are better off because of the payments, and consequently the result is plausible.

When AGE and AGE55 are interacted with treatment, the partial F is significant in five out of the eight equations. Yet, the results are mixed in that the signs are often different between the variables. However, with the exception of gross farm income, AGE*T is always negative in the five equations and larger than AGE55 if the latter is of opposite sign. Therefore, the conclusion that the experiment affected disincentives more as age increased should probably be accepted.

The next line is the partial F-statistic on all treatment variables represented by the lines above with the exception of the measurement variable (PN) which is included in the basic set of variables. The treatment variables are significant in six out of the eight equations.

A new equation is represented by the next line. The basic model is identical to that of line one. The only difference is that treatment is entered as one variable rather than as three variables, each being a different year dummy, interacted with treatment. Consequently these results are almost identical to the first line, except generally showing a higher significance level.

The test of the hypothesis that each year was similar is shown by the partial F-test on the next line in Table 4.3. The alternative hypothesis is that the treatment effect was different in one of the years of the experiment. The latter actually has much appeal. One could argue convincingly that because the experiment started in late 1969, the farming operation for 1970 was already decided. Furthermore decisions probably would not be changed immediately until the payment checks came for several months as promised. The data do not support this view, however. Given that the major reduction is in the livestock operation and that this adjustment can be made easily and quickly, the overall effect is the same for the three years. It is quite possible that some families overreacted, cut their livestock production drastically the first year, found they had too much time on their hands, and thus increased their production in subsequent years. This was balanced by some farmers who cut production slowly every year.

The treatment parameters, tax rate and percentage guarantee, were insignificant in almost all formulations except gross farm income and scaled crop hours, and were often of incorrect sign. Theoretically, the higher the tax rate the larger the reduction in hours, holding everything else constant. Consequently when the tax parameter is unconstrained, the coefficient on the 70 percent plan should be more negative than the 50 percent which in turn is more negative than the 30 percent plan. Part of the time these expectations held, but often they did not. The prime rationale is the small sample. As shown in Table 3.2, the number of observations in each cell is quite small. Near the bottom of Table 4.3, the partial F-tests are shown for the addition of the four plan dummies to the basic model and the five treatment variables. In addition the F-statistic is shown for the addition of tax dummies when the guarantee dummies have been added to the above model and vice versa.

The major variable influenced by plan effects is gross farm income. The coefficients [see Table C.2 in Appendix C] on the guarantee dummies are -126 and -6183 respectively for G50 and G100, while they are -1099 and 8400 for T30 and T70. Only G100 and T70 are significant. The sign on G100 agrees with expectations while that on T70 is of the wrong size and quite large. However, this coefficient is essentially determined by only five observations, thus making it almost impossible to make any conclusions about plan effects.

The family composition[1] variables were significant in the scaled hours equations and were for the most part of the correct sign.

Other variables described earlier such as 1969 normal income and other "measurement" variables had little influence upon the significance and estimates of the treatment coefficients presented above.

Because of the many interactions with treatment variables, it is difficult to determine the net or overall treatment effect from the selected models in Tables C.1 and C.2 in Appendix C. These results have been summarized in Table 4.4 by estimating a predicted y-hat for 1971 for both control and experimentals at selected values of HR69. In calculating the y-hats all other variables' overall means were used. By subtracting the control group, a percentage point estimate of the disincentive is obtained. Algebraically this may be represented as $(\hat{P} - \hat{C})/\hat{C}*100$.

One must use caution in interpreting and using these statistics. Values of 1000 and 2600 are a long distance from the mean and the variance of those estimates is extremely large. They should not be interpreted as average results but only a prediction for families with selected characteristics.

The following highlights from Table 4.4 deserve special mention. First, net and gross farm income are generally positive but go from negative to positive in a rather drastic fashion as PN increases.

---

[1]The family composition variables are:
- # of males, ages 13-15
- # of females, ages 13-15
- # of males, ages 16-20
- # of females, ages 16-20
- # of males ages 21-60
- # of females, ages 21-60
- 1 if spouse present, 0 otherwise.

Table 4.4. Iowa predicted incentives for selected dependent variables from the model in Table C.1.

	Selected HR69 Values				
	1000	1400	1800	2200	2600
Net Farm Income					
Control	3538	3944	4248	4450	4550
Experimental (PN=0)	1955	2993	3781	4320	4610
% Incentive	-44.7	-24.1	-16.0	-2.9	1.3
Experimental (PN=.72)	3151	4189	4978	5517	5807
% Incentive	-10.9	6.2	17.2	24.0	27.6
Experimental (PN=1.79)	4929	5967	6756	7295	7585
% Incentive	39.3	51.3	59.0	63.9	66.7
Gross Farm Income					
Control	10860	15165	18812	21802	24133
Experimental (PN=0)	13823	16204	18306	20129	21672
% Incentive	27.3	6.9	-2.7	-7.7	-10.2
Experimental (PN=.72)	15489	17871	19972	21795	23338
% Incentive	42.6	17.8	6.2	.03	-3.3
Experimental (PN=1.79)	17966	20347	22448	24271	25814
% Incentive	65.4	34.2	19.3	11.3	7.0
Scaled Adjusted Hours					
Control	1131	1479	1817	2147	2467
Experimental (PN=0)	956	1268	1567	1852	2125
% Incentive	-15.5	-14.3	-13.8	-13.7	-13.9
Experimental (PN=.72)	988	1300	1599	1884	2156
% Incentive	-12.6	-12.1	-12.0	-12.2	-12.6
Experimental (PN=1.79)	1036	1348	1646	1932	2204
% Incentive	-8.4	-8.9	-9.4	-10.0	-10.7
Scaled Hours					
Control	1141	1559	1956	2330	2684
Experimental (PN=0)	1142	1472	1793	2105	2408
% Incentive	.1	-5.6	-8.3	-9.7	-10.3
Experimental (PN=.72)	1170	1501	1822	2134	2437
% Incentive	2.5	-3.7	-6.9	-8.4	-9.2
Experimental (PN=1.79)	1213	1543	1864	2176	2479
% Incentive	6.3	-1.0	-4.7	-6.6	-7.6

Table 4.4. Continued

	Selected HR69 Values				
	1000	1400	1800	2200	2600
Total Recall Hours					
Control	1956	2299	2587	2823	3005
Experimental (PN=.72)	2024	2266	2479	2665	2823
% Incentive	3.5	-1.4	-4.2	-5.6	-6.1
Head Recall Hours					
Control	1726	2071	2359	2589	2761
Experimental (PN=.72)	1872	2087	2273	2429	2554
% Incentive	8.5	.8	-3.6	-6.2	-7.5
	600	800	1000	1200	1400
Scaled Crop Hours					
Control	623	843	1050	1245	1426
Experimental (PN=0)	703	907	1100	1281	1452
% Incentive	12.8	7.6	4.8	2.9	1.8
Experimental (PN=.72)	705	909	1102	1284	1454
% Incentive	13.2	7.8	5.0	3.1	2.0
Experimental (PN=1.79)	709	913	1106	1287	1458
% Incentive	13.8	8.3	5.3	3.4	2.2
	200	600	1000	1400	
Scaled Livestock Hours					
Control	375	739	1090	1428	
Experimental (PN=0)	188	481	789	1110	
% Incentive	-49.9	-34.9	-27.6	-22.3	
Experimental (PN=.34)	217	510	818	1138	
% Incentive	-42.1	-31.0	-25.0	-20.3	
Experimental (PN=1.17)	259	553	861	1181	
% Incentive	-30.9	-25.2	-21.0	-17.3	

For example, using a value of 1800 for HR69, when PN = 0 there is an 11 percent disincentive in net farm income; this becomes 17.2 percent when PN = .72 and 59 percent when PN = 1.79. Zero and 1.79 are approximately one standard deviation away from the mean of .72. The same phenomenon holds true for gross farm income. At PN = 0, the percentage disincentive is -2.7 percent; at PN = .72 it is 6.2 percent, and at PN = 1.79 the percentage is 19.3.

Second, the livestock hours variable is negative on the order of 20 to 40 percent and crop hours is positive, approximately 5 to 14 percent, thus making total scaled hours negative around 4 to 8 percent. Scaled adjusted hours is slightly more negative again due to the increased use of hired help by the experimental farmers.

Third, the percentage incentive increases as PN increases for scaled adjusted, total scaled, and scaled livestock hours, and also slightly for scaled crop hours. For example, at PN = 0, scaled hours exhibits -8.3 percent disincentive when HR69 = 1800. This becomes -6.9 percent at PN = .72 and -4.7 percent at PN = 1.79.

North Carolina

In many respects, the results between the two regions are quite similar. The results for North Carolina will be discussed in a similar fashion to that of Iowa. However, the results for scaled livestock hours will not be discussed except through total and adjusted scaled hours variables. Livestock hours are a very minor enterprise in North Carolina and the analysis is dominated by a few observations.[1] The basic model for

[1]The average number of scaled livestock hours in North Carolina is only 152 hours. Only 12 farmers had over 500 livestock hours.

North Carolina is identical to that of Iowa except for the addition of a RACE variable.

The simple parameterizations of the treatment variables is shown in Table 4.5. For Model I, the coefficients are all negative except recall hours. This result is identical to that of Iowa. The variables which have been significantly affected have changed however. The treatment coefficient is now significant in the gross farm income and scaled crop hours equations, continues to be significant for total scaled hours, and is no longer significant for adjusted scaled hours. The latter is puzzling and will be discussed at length shortly.

From Model I, the average disincentive for scaled crop hours is 290 hours, 83 hours for adjusted scaled hours, and 266 hours for total scaled hours. Head and total recall hours show a positive incentive of 141 and 148 hours respectively. The income variables are -181 for net farm income and -1425 for gross farm income.

The results from Model I are essentially identical to Models II and III. The partial F-values retain their same level of significance, and a point estimate of treatment effect reverses sign only in Model III for the adjusted scaled hours variable where the treatment variables are insignificant anyway. These point estimates, as previously discussed, are constructed by calculating predictions based on sample means.

The results for adjusted scaled hours are inconsistent with total scaled hours. This results primarily because the controls purchased relatively more hired labor than experimental individuals. This is even more unusual when you consider that the experimental farmer was essentially

Table 4.5. North Carolina treatment coefficient estimates, F-values and predicted incentives for models with simple treatment parameterizations for selected dependent variables

	Net Farm Income	Gross Farm Income	Head Recall Hours	Total Recall Hours	Total Scaled Hours	Adjusted Scaled Hours	Scaled Crop Hours
Model I							
C/E	-189	-1425	141	148	-266	-83	-290
F	.32	3.17[a]	1.12	.83	3.76[a]	.72	4.84[b]
Model II							
C/E	-591	588	238	117	-13	-139	65
HR69*C/E	.26	-1.32	-.06	.02	-.17	.04	-.25
F	.47	2.97[a]	.67	.42	2.62[a]	.43	4.32[b]
% Incentive	-7.4	-13.8	9.0	7.2	-12.7	-5.9	-15.1
Model III							
C/E	-398	-1104	353	129	-686	-454	-295
C/E*HR69	.033	1.10	-.22	.015	.80	.49	.33
C/E*HR69^2	.00005	-.00060	.00004	-.0	-.00024	-.00011	-.00016
F	.28	2.37[a]	.51	.29	3.91[c]	1.11	4.16[c]
% Incentive	-6.7	-8.1	6.5	7.1	-1.3	2.2	-6.9

[a]Significant at the 10 percent level.

[b]Significant at the 5 percent level.

[c]Significant at the 1 percent level.

being subsidized for this hired labor and had every reason to report the amount paid fully.

The only rationale that could explain the inconsistency in results are the following:

a) On similar government forms (IRS), many families report paying money to teenage children for work done on the farm. Through a process of trial and error, the experimental families learned that when filling out the monthly income and expense forms upon which the payment is based, hired labor expenses for dependents should not be reported. Thus their added knowledge from the payments process, relative to control, essentially made them report hired labor expense for teenagers less frequently on the quarterly interviews.

b) Several unidentifiable "outliers" are responsible for the inconsistency.

c) Reporting of hired labor expense was notoriously poorer for control than for experimentals. During the editing process, a wrong fudge factor was used in correcting the misreporting of hired labor expense, which led to the above erroneous results.[1]

All of these explanations are plausible and upon investigation would tend to remove the inconsistency, but not without being very arbitrary. In the final analysis, the outcome should depend upon which variable is most likely to have been reported accurately. On this basis, it seems hard to refute the evidence offered by the consistency of scaled crop with net and gross farm income.

[1] This was mostly corrected in the data used for this analysis.

Building upon the models discussed above, more complicated models were estimated similar to that in Iowa. The coefficients from two selected models are shown in Tables C.3 and C.4 in Appendix C. Partial F-statistics from these equations are shown in Table 4.6.

The five treatment variables shown on the first line made a significant difference in explaining the variance of the dependent variable in five of the seven equations, namely, gross farm income, total recall, head recall, total scaled, and scaled crop hours equations. In comparison to Models I to III above, the treatment variables have increased in significance in all equations except the gross farm income equation.

The "measurement" variable (PN) plays a highly significant role in all equations except recall hours. It is always positive when significant and, as will be seen shortly, influences disincentive predictions greatly.

The debt ratio interacted with treatment is negative and weakly significant in five out of the seven equations. Although this was unexpected, it agrees with the Iowa results.

The age variables interacted with treatment are insignificant in all equations as shown by the partial F-statistic on the next line. Race interacted with treatment was also insignificant in all equations except scaled crop hours, and then only weakly significant. Black families showed less of a disincentive for scaled crop hours than white families in comparison with control families of similar characteristics and of the same race.

The next line shows the joint significance of all treatment variables and treatment interactions except PN which was included in the reduced model. The treatment variables were significant in the gross farm income,

Table 4.6. North Carolina partial F-statistics for selected dependent and independent variables

Group of Independent Variables	Dependent Variables: Net Farm Income	Gross Farm Income	Total Recall Hours	Head Recall Hours	Adjusted Scaled Hours	Total Scaled Hours	Scaled Crop Hours
HR69*C/E, $HR69^2$ x C/E D70*C/E, D71*C/E, D72*C/E	.43	1.49[a]	1.53[a]	1.76[a]	.82	2.38[b]	2.43[b]
PN	19.53[c]	3.80[d]	.18	.64	2.36[b]	6.50[b]	6.3[b]
DR*C/E	.17	.0004	1.99[a]	1.77[a]	1.77[a]	2.04[a]	3.42[d]
AGE*C/E, AGE55*C/E	.89	1.06	.08	.11	.95	.22	.36
RACE*C/E	.23	.59	.07	.81	.29	.48	1.69[a]
All of the above except PN	.96	1.49[a]	1.10	1.34	1.12	2.12[b]	2.59[c]
HR69*C/E, $HR69^2$*C/E, C/E	.28	2.37[d]	.29	.51	1.11	3.91[c]	4.01[c]
D71*C/E, D72*C/E[e]	.66	.18	3.38[b]	3.62[b]	.39	.09	.06
T30, T70[f]	1.69[a]	3.95[b]	2.61[d]	3.35[b]	6.11[c]	4.66[b]	3.43[b]
G50, G100[f]	3.28[b]	2.90[d]	1.71[a]	1.63[a]	7.45[c]	6.81[c]	5.01[c]
T30, T70, G50, G100[f]	2.21[d]	2.56[b]	1.84[a]	1.86[a]	6.16[c]	4.79[c]	3.33[a]
Family Comp. Variables[g]	.78	.46	1.31	1.22	1.46[a]	1.15	1.11

[a]Significant at the 20 percent level.

[b]Significant at the 5 percent level.

[c]Significant at the 1 percent level.

[d]Significant at the 10 percent level.

[e]D70*C/E, D71*C/E and D72*C/E are year dummies interacted with the simple treatment variable C/E.

[f]T30, T70, G50 and G100 are all simple dummy variables representing each of the different experimental plans. The 50 percent tax and 75 percent guarantee plans have been excluded.

[g]See footnote on page 61.

total scaled, and scaled crop equations. This is entirely consistent with the results from Models I to III and those shown by the next line, namely HR69*C/E, $HR69^2$*C/E and C/E.

A test to determine if the treatment effect differed between years is shown on the next line. As in Iowa, the null hypothesis was accepted for all equations except recall hours, meaning that the experiment probably did not have a differential impact with respect to year.

The next three lines show the significance of the tax and guarantee effects. A joint F-test to determine if at least one of the five plans is different from the mean treatment effect is performed. The reduced model includes the variables in the basic model plus those represented by line one. The reduced model for the F-test on tax variables includes the guarantee dummies and vice versa for the F-test on the guarantee variables. The results here are vastly different from Iowa's results. The variables are significant, often at the one and five percent level, in every equation.

As previously discussed, the G50 and T30 coefficients should be positive and the G100 and T70 coefficients should be negative. For the 28 coefficients (7 equations x 4 variables), 18 are of the wrong sign and most of these are significant as shown below.

Significance Level

Sign	>20%	10-20%	5-10%	1-5%	1%	Total
Correct	7	2	0	1	0	10
Incorrect	4	0	3	4	7	18

Furthermore only one of the ten correct signs is less than the ten percent level of significance, while fourteen of the eighteen incorrect signs are significant at the ten percent level or less. The inclusion of the measurement variable PN only alters the above table slightly.

Explaining these results is difficult. To conclude that theory is incorrect is also to contradict a large wealth of empirical evidence which supports that theory. A more plausible approach is to argue that the number of observations is too small, that the link between income and hours is indirect for self-employed families, that other measurement errors are present, and that therefore, tax and guarantee effects are inconclusive.

Family composition variables were insignificant in all equations but one. Other measurement variables mentioned earlier were insignificant and had little effect upon the treatment coefficients.

Again, similar to Table 4.4 for Iowa, a predicted disincentive table has been constructed for North Carolina for a family with selected characteristics at different values of HR69. These are shown in Table 4.7.

Net farm income moves from negative values (disincentives) to positive values (incentives) as PN increases. For example, when HR69 equals 1400: at PN = 0, the predictived incentive is -15.5 percent; at PN = .91 it becomes 9.8 percent and at PN = 1.79, it is 34.3 percent. Because the coefficient on PN is positive in all equations, this same phenomenon holds true for all dependent variables.

Consistently evaluating the response at HR69 equalling 1400: at PN = 0, gross farm income has a predicted incentive of -17 percent which increases to -1.5 percent at PN = 1.79; adjusted scaled hours increases

Table 4.7. North Carolina predicted incentives for selected dependent variables from the model in Table C.3.

	Selected HR69 Values				
	600	1000	1400	1800	2200
Net Farm Income					
Control	1531	2093	2550	2903	3152
Experimental (PN=0)	1305	1757	2155	2500	2792
% Incentive	-14.8	-16.1	-15.5	-13.9	-11.4
Experimental (PN=.91)	1951	2402	2800	3145	3437
% Incentive	27.4	14.8	9.8	8.3	9.0
Experimental (PN=1.79)	2575	3026	3424	3769	4061
% Incentive	68.2	44.6	34.3	30.0	28.8
Gross Farm Income					
Control	7288	8530	9858	11271	12768
Experimental (PN=0)	5823	7062	8181	9178	10054
% Incentive	-20.1	-17.2	-17.0	-18.6	-21.3
Experimental (PN=.91)	6600	7839	8957	9954	10831
% Incentive	-9.4	-8.1	-9.1	-11.7	-15.2
Experimental (PN=1.79)	7351	8590	9708	10705	11582
% Incentive	.9	.7	-1.5	-5.0	-9.3
Adjusted Scaled Hours					
Control	1017	1192	1362	1528	1690
Experimental (PN=0)	789	1075	1329	1553	1746
% Incentive	-22.4	9.8	-2.4	1.6	3.3
Experimental (PN=.91)	895	1181	1435	1659	1852
% Incentive	-12.0	-.9	5.4	8.6	9.6
Experimental (PN=1.79)	998	1283	1538	1761	1954
% Incentive	-1.9	7.6	12.9	15.2	15.6
Total Scaled Hours					
Control	1338	1651	1969	2293	2621
Experimental (PN=0)	904	1379	1797	2158	2462
% Incentive	-32.4	-16.5	-8.7	-5.9	-6.1
Experimental (PN=.91)	1051	1526	1944	2305	2610
% Incentive	-21.4	-7.6	-1.3	.5	-.4
Experimental (PN=1.79)	1194	1668	2086	2448	2752
% Incentive	-10.8	1.0	5.9	6.8	5.0

Table 4.7. Continued

	Selected HR69 Values				
	600	1000	1400	1800	2200
Total Recall Hours					
Control	1764	2009	2209	2365	2476
Experimental (PN=.91)	1721	1936	2132	2310	2469
% Incentive	-2.4	-3.6	-3.5	-2.3	-.3
Head Recall Hours					
Control	1213	1451	1651	1812	1935
Experimental (PN=.91)	1464	1586	1706	1823	1938
% Incentive	20.7	9.3	3.3	.6	.2
Scaled Crop Hours					
Control	1196	1584	1956	2314	2655
Experimental (PN=0)	894	1305	1674	2000	2285
% Incentive	-25.3	-17.6	-14.4	-13.6	-13.9
Experimental (PN=.91)	1042	1453	1821	2148	2432
% Incentive	-12.9	-8.3	-6.9	-7.2	-8.4
Experimental (PN=1.79)	1184	1595	1964	2291	2575
% Incentive	-1.0	.7	.4	-1.0	-3.0

from -2.4 to 12.9 percent; total scaled hours grows from -8.7 to 5.9 percent; and scaled crop hours goes from -14.4 to .4 percent.

As argued earlier, the coefficient on PN should be positive; namely, as PN increases, the substitution effect is eroded but the income effect remained, implying that there should still be disincentive on hours and income, but that it should be less than when PN = 0. What explains the large positive incentives as PN increases?

Temporarily assume that payments information is more accurate than survey information, or, to understand the argument more clearly, assume that payments information is perfectly accurate. Then, when survey income is greater than payment information, PN will be positive. Under this assumption, PN primarily reflects editing and response errors in the survey information. Almost by definition when income is added incorrectly, the response will be positive for those families. This is a possible explanation for the sign and significance of PN.

On the other hand, if survey information is more accurate (e.g., because it is less likely to be affected by reporting effects of the experiment), the following is true. PN identifies characteristics of experimental families which are associated with a high variability of income within a year and an ability for families to effectively use the payments system to their advantage. Presumably these same characteristics exist within the control families. However, it is impossible to identify these families, and consequently all of the treatment comparisons with respect to PN are made to an average control family. Thus, the amount of incentive is overstated when PN is high and understated when PN is low.

It could be that if the comparisons between the correct experimental and control families were made, the incentive would be negative throughout, but significantly less negative when PN is large.

For total scaled and scaled crop hours the disincentives are primarily negative, but the estimates exhibit much variance among the different levels of HR69.

Conclusion

The hypotheses introduced in the theory section have been tested with mixed results. The operations with small amounts of fixed capital and relatively fixed coefficient production functions have experienced the least disincentive. Observe the performance of livestock operations in Iowa and crop operations in North Carolina versus that of the capital intensive crop operations in Iowa. In terms of scaled hours, the former had significant disincentives of 5 to 30 percent while the latter had small insignificant positive incentives.

The major hypothesis advanced is that work effort should decline with an introduction of a negative income tax scheme. The answer to this hypothesis unfortunately depends upon the variable being analyzed. If recall hours are chosen, one would conclude that the experiment had no effect upon hours worked. On the other hand if scaled hours or some weighted combination of acres and livestock sold is used, the evidence is quite strong that the treatment did affect hours of work in a significant negative direction. Probably the weakest aspect of this evidence is the inconsistency in North Carolina between adjusted scaled hours (removing the effect of machine hire, hired labor, and addign custom work) and total scaled hours. These inconsistencies must be evaluated carefully, and their resolution is chiefly dependent upon the data collection accuracy of the variables in question. Upon these grounds, this author concludes that scaled hours is a "better" measure of work effort and thus the experiment induced a decline in work effort.

These results are roughly consistent with net and gross farm income. Particularly this is true when a measurement variable (actual negative income payments minus predicted negative income tax payments based upon edited quarterly data) is added to the model. Those families where actual payments exceeded predicted payments by a large amount are affected less by the program parameters (tax and guarantee rates) than families where actual payments equals predicted payments. Consequently there is more disincentive in families where the latter occurred in terms of hours and income.

The results on the program parameters are disappointing and inconclusive. In the few cases where the plan dummies significantly improved the regression relationship, the parameters were often of the wrong sign and internally inconsistent. Probably the chief reason for this inconclusiveness is the lack of observations.

With respect to year effects, one must always accept the conclusion that the effect in each year was identical. While a strong argument can be made for the effect to have grown over time, the data lend very little evidence to support that contention. Recall that the response of the experimental families was <u>varied</u>. Some families experienced no reduction in hours, some experienced increases, some experienced a large decline in all three years, versus others where there was a decline in only one of the three.

The race variable was insignificant in North Carolina, and most of the other results (i.e. effect of age and education) agreed with a priori expectations.

While there remain several disturbing aspects of the data, it seems clear that farmers reduced their crop hours significantly in North Carolina and livestock hours in Iowa. In each of these cases, the overall effect on total hours is negative, significant, and on the order of 4 to 10 percent.

CHAPTER V. IMPACT OF DATA ERRORS UPON TREATMENT ESTIMATES OF THE FARM POPULATION

Introduction

A survey operation must attempt to insure that data collected are accurate and provide unbiased answers to the major objectives of the study. In the Rural Income Maintenance Experiment (RIME), this implies that the data collection process should not be biased[1] with respect to the treatment variables (tax rate, guarantee level, or being in the experimental group versus the control group). Furthermore the data gathered on farm income, expenses, assets, inventory, and physical production should be free from large errors so the precision of treatment estimates will be enhanced.

A brief outline of the farm data would include the following:

a) Cash income and expenses--obtained by item (e.g. swine sales, fertilizer expenses, machinery repair expenses) every three months on the quarterly interview.

b) Income and expenses--as itemized on the Internal Revenue Service (IRS) Schedule F. While this is available for a majority of farmers, nearly all farmers reported a nonitemized total of gross and net income as given to IRS.

[1]The effect of biases can enter at several states during the experiment. These would include biases arising from refusing the screening interview, the pre-enrollment interview, enrollment, or by attriting sometime during the experiment. None of the above are examined in this chapter. Only the bias arising from response errors is discussed.

c) Assets, inventory, and related debts--by category (e.g., land and buildings, machinery, equipment, livestock by type, unsold grain) were obtained once a year. New purchases and sale of assets were also obtained. Some information on assets is also available from the depreciation schedules provided by many farmers on their IRS Schedule F. In addition, a questionnaire section on the use of credit facilities provides yet another source of information.

d) Hours worked on the farm--every three months the farmer was asked to recall the number of hours he spent working on the farm during each of the previous seven days.

e) Physical production--at the end of each year the farmer was asked to indicate total acreage planted in each different crop, yield for each crop, livestock sales by type of animal, and milk and egg production.

f) Miscellaneous--each year the farmer's landlord-tenant contractual relationships were noted. Also, a continuous off-farm wage work record was obtained for the farmer and all other members of his household over fifteen.

Reports from the field (interviewers) indicated several difficulties in data collection. They included the following:

1. Complex income flows, especially in Iowa. The average farmer had about twenty different transactions to be recalled every three months. The omission or double-counting of a particular transaction was common.

2. Low literacy and no record keeping, especially in North Carolina. Data collection problems were especially severe when the farmer was a

sharecropper whose landlord handled the books (if the landlord himself kept any).

3. A quite hazy distinction between farm activity where output was destined for home consumption versus purely commercial activity. This problem was common in North Carolina.

4. Murky patterns of asset ownership--just who in the family had what legal title to what portion of the land--and uninformed estimates by the farmers of their assets' current market value.

5. A time slice problem--due to the decision to collect information on a three month basis, all transactions which occurred near the artificial three month boundaries were likely to be double reported or not reported at all. Consequently, a load of 25 cattle, which is relatively small by Iowa standards, sold around the first of March and not reported in either quarter could have resulted in a $7500 underestimate of net farm income. For a particular family these transactions would not balance. Perhaps for the entire population the treatment parameter may not be affected.[1]

A glance at the data often revealed glaring inconsistencies such as feed expenditures exceeding livestock sales with no appreciable increase in inventory value. It should be noted that many apparent inconsistencies occurred and were not necessarily the result of response errors. For example, feed expenditures without livestock sales might be perfectly

[1] If income is the dependent variable, if no measurement problems exist in the independent variables, and if no correlation exists between the treatment parameter and measurement error, the treatment parameter should be unbiased. If income is an independent variable, the coefficient will be based towards zero.

reasonable for someone who had just entered the livestock business or had retired from it with carryover feed bills.

Given the wealth and detail of information collected, the availability of outside information and the possibility of reporting incorrect conclusions, an intensive editing of the data was undertaken. Essentially two data bases were created. The original (ORIG) data base was information as reported on the quarterly interview with "obvious" errors and most of the data processing errors corrected. For example, if a corn yield of 1000 bushels per acre were reported, 100 was assumed or the value replaced by standard statistical techniques for eliminating missing data. No corrections were made to data that were reasonable in a univariate framework. These data may contain obvious inconsistencies between data items. This data base compares favorably with procedures and editing techniques used on national surveys like the Current Population Survey (CPS).

An edited (EDIT) data base was created by removing a large percentage of inconsistencies. Out of a class of all possible data changes, the change requiring the fewest updates which made the entire picture "consistent" was typically chosen. Outside information, e.g., farm budget and extension bulletins, farm practices from previous years, and value judgments concerning the quality and direction of the different data sources, were all used in arriving at what data was to be changed on the edited data base.[1] An example of the latter would be that income information as reported to IRS is more often understated than overstated.

---

[1]See Appendix B for a complete discussion of the distinction between the original and edited data bases.

With the goal of defining the approximate magnitude of data errors and their impact upon treatment estimates, the following topics will be examined in sequence.

a) A theoretical discussion of the impact of data errors upon treatment coefficients.

b) A test of the hypothesis that the data collection apparatus was unbiased with respect to control/experimentals.

c) A description of the changes between the original and edited data bases.

d) A rationale for accepting the edited data base as most accurate.

e) An attempted explanation of the changes between the original and edited data bases.

f) The effect of the changes on results.

The population analyzed is defined as all farm families with constant marital status, with the head less than 69 years of age at the beginning of the experiment, with positive budgeted hours of farm work in each of the years 1969 to 1972, and with a total of 400 budgeted farm hours in at least one of those years. Budgeted farm hours, a constructed variable, is a weighted combination of crop acres and of livestock production. The weights, derived from State Extension Surveys, are the average number of hours per year required to tend an acre of a particular crop or raise a particular type of livestock on a farm with average mechanization.

Theory

This section traces the effect data errors will have on treatment estimates beginning with a simple two variable model. Assuming that y and x are related by

$$y = \alpha + \beta x + e$$

where

y is a dependent variable of interest, particularly income or hours

x is an independent variable which controls for nonrandom sampling present in the experimental design and

e is a random disturbance term.

While this may not be an interesting model for this analysis because it ignores the treatment effect, it can be highly instructional. This theory is well-known and has been derived many times before. Consider Y and X the observed variables where both are measured with error. Let small x, y denote the true values of the variables. Then

$$X = x + u$$

$$Y = y + v$$

If the true values are related by the relationship

$y = \alpha + \beta x + e$ then by substitution

$$Y - v = y$$

$$X - u = x$$

and

$$Y - v = \alpha + \beta(X-u) + e$$

$$Y = \alpha + \beta X - u + v + e$$

or

$$Y = \alpha + \beta X + w$$

Where $w = v - \beta u$.

Since w is not independent of X because it includes the term $-\beta u$, ordinary least squares will produce biased estimates of α and β even if the sample is infinite and even if the mean values of the errors are zero. The estimates of βn for a sample of size n based on the observed variables is

$$\beta n = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n} (X_i - \overline{X})^2}$$

If the errors terms are introduced and rearranged the following expression is obtained:

$$\beta n = \frac{\Sigma(x + u - (\overline{x} + \overline{u}))(y + v - (\overline{y} + \overline{v}))}{\Sigma(x + u - (\overline{x} + \overline{u}))^2}$$

Letting $x' = x - \overline{x}$

$y' = y - \overline{y}$

$u' = u - \overline{u}$

$v' = v - \overline{v}$

The expression is

$$\beta n = \frac{\Sigma x'y' + u'y' + \Sigma v'x' + \Sigma v'u'}{\Sigma x'^2 + 2\Sigma u'x' + u'^2}$$

The true β is simply $\frac{\Sigma x'y'}{\Sigma x'^2}$. Thus the bias in the estimated β as compared to the true β depends upon the following:

1. u'^2 - if u has a large variance (not unusual considering some

empirical evidence to be presented later), βn will be biased downward.

2. $\Sigma u'x'$ - if u is positively correlated with X, e.g., in the case when high asset families report assets less accurately while low income families report assets more accurately, βn will be biased downward.

If u is negatively correlated with X, e.g., when high income people report income accurately because of records while low income individuals do not report accurately, βn will be biased upward.

3. $\Sigma u'v'$ - if u and v are positively correlated, e.g., when X is the 1969 value of the dependent variable y, βn will be biased upward. This is true in cases where respondents consistently overestimate or underestimate income for consecutive years. If the converse is true then βn will be biased downward.

4. $\Sigma v'x'$ and $\Sigma u'v'$ - this will cause the coefficient to be biased if the true value of one variable is correlated with the error term in the other. Again this may be likely to occur when x is a lagged independent variable of y.

Consequently, if one can do validity studies to predict the components of the above expression, then one could accurately predict βn.

For similar reasons, $Var(\beta)$ is also distorted indeterminately by the different components of error.

Turning to a more interesting model,

$$Y = \alpha + \beta X + \gamma t + e$$

Y, X are previously defined with both containing measurement error, while t is a treatment variable measured without error. Let the experimental group be represented by a one and the control group with a zero.

An unbiased estimate of γ is required to accurately estimate the effect of the experiment. Letting

$$t' = t - \bar{t}$$

$$y' = y - \bar{y}$$

$$x' = x - \bar{x}$$

$$v' = v - \bar{v}$$

$$u' = u - \bar{u}$$

Then

$$\gamma \text{ (without errors)} = \frac{\Sigma y'x' \; \Sigma x't' - \Sigma y't' \; \Sigma x'^2}{(\Sigma x't')^2 - \Sigma t'^2 \; \Sigma x'^2}$$

$$\begin{aligned} \gamma \text{ (with errors)} = {} & (\Sigma y'x' + \Sigma x'v' + \Sigma y'u' + \Sigma u'v') \\ & (\Sigma x't' + \Sigma t'u') - (y't' + \Sigma v't') \\ & (\Sigma x'^2 + 2\Sigma u'x' + \Sigma u'^2)/(\Sigma x't')^2 \\ & + 2 \; \Sigma t'u \; \Sigma t'x' + (\Sigma t'u')^2 - \Sigma t'^2 \\ & (\Sigma x'^2 + 2 \; \Sigma u'x' + \Sigma u'^2). \end{aligned}$$

If the following assumptions are made

1. X is measured without error
2. Y and X are independent
3. Y and t are negatively correlated, (i.e., experimental families tend to understate income)

then the intuitive notion that the treatment coefficient is overstated is readily shown.

Other assumptions about the behavior of u and v can be traced without much difficulty in this three variable world. In a multiple variate world the expressions become complicated very quickly. One must resort to matrix algebra and strict assumptions to gain any further insights into the problem.

Because these assumptions were hard to meet, and because in many cases using a priori information it seemed possible to identify where errors existed in the data, the edited data base was created in accordance with the following criteria:

1. Where inconsistencies were noted among the six different sources of farm data listed previously, preference would be given to data generated from written sources. Thus if a farmer gave a yield estimate for his soybeans that was at wide variance with his sales as recorded from his books, the latter probably would be accepted and the yield estimate edited to make it consistent.

2. If no information were from written sources, preference was given to the remembered acres and yield estimates over remembered cash sales.

3. It was assumed that omission of income was a more likely error than the false inclusion of income. Therefore, if four of the six sources failed to mention hog sales, but one of the four included a mention of, say, a feeder pig inventory, or feed purchase, it was assumed that the farmer probably had some swine sales, and the highest reasonable estimate for those sales was edited in.

4. Most additions of income or expenses to the verbatum record of the farmer had to have evidence for their existence in at least two

different sources. Thus, in the above example, if mention of swine had been made only once, it may have been edited out rather than editing it in on all the other sources.

Magnitude of Data Collection System Bias

Three approaches are possible for testing the data collection system. First, a direct measure of certain data quality variables may be possible. One of the variables that exemplifies this technique is recording how respondents answer the various income and asset questions. If the information comes from records or other written documentation, the information may be of a different quality than that coming from recall. Many respondents kept a monthly index of receipts, or did all transactions by check; or in one case, an accounting system kept track of all farm business dealings. These records do not completely eliminate mistakes but are much better than recall responses.

A second approach is to verify the information being collected on interviews with an outside source such as Internal Revenue Service (IRS) data. Although information from IRS forms was not obtained directly from the agency, interviewers were able to make copies of the forms sent to IRS by obtaining them from the families. This does allow the possibility that the forms given to RIME were different than the actual ones submitted to IRS. However, the probability of this seems small.

IRS information undoubtably is also biased in the sense that income probably is underreported and expenses overreported. However, one has to assume perfect information on the part of respondents to argue that the introduction of the negative income tax alters the reporting behavior of experimentals vis-a-vis control to the IRS. Actually the smart individual would have overreported income to IRS and underreported income to RIME. By filing amended returns to IRS at the end of the experiment, he could have recouped his additional tax payments from both RIME and IRS.

A more valid criticism of this technique is the refusal bias problem. The families which did not volunteer their IRS forms may be the families with the highest or lowest level of misreporting. No easy method is available for circumventing this problem.

Approximately the same percentage of control and experimental families let interviewers peruse their IRS information. In Table 5.1 the number of families who volunteered their itemized schedule F's as well as the number who reported just their total gross income, depreciation, interest, and net farm income is shown. This information is significant because of the use of IRS information in the formation of the second data base. It is gratifying to note that no significant control/experimental differential in reporting behavior is present in either Iowa or North Carolina.[1]

A third approach is to estimate a relationship between variables that is known a priori and which is not subject to a treatment effect. One has to proceed carefully here. Some relationships, which on first glance are not suspected of containing a treatment effect, in fact might. For example, gross farm income in Iowa can be predicted by number of acres of corn, soybeans, oats, and hay, diverted acres, and number of market hogs,

[1]The chi square test of independence yielded insignificant values of 1.15 and 1.70 for Iowa and North Carolina respectively.

Table 5.1. Number of farmers for whom tax information is available who are less than 69 years of age, have constant marital status, and farm in each of the years from 1969 to 1972

	Iowa		North Carolina	
	Control	Experimentals	Control	Experimentals
Itemized Schedule	132	127	67	81
Percent	(81.5)	(84.7)	(50.8)	(49.1)
Abbreviated Schedule	10	10	19	33
Percent	(6.2)	(6.6)	(14.4)	(20.0)
No Tax Information Available	20	13	46	51
Percent	(12.3)	(8.7)	(34.8)	(30.9)

other hogs, market steers, other cattle, and other livestock sold, with an appropriate adjustment for tenure status. The above question is not interesting to an economist, for it is practically an accounting equation. One could argue that a control/experimental dummy inserted into the above relationship should carry a coefficient not significantly different from zero. Another more preferred way of testing for a control/experimental difference would be to estimate the relationship separately for control and experimentals, then perform a Chow test to determine if the two sets of coefficients are significantly different from each other.

However, a counter argument exists which suggests that the experimental group may hold grain inventories for a longer period, obtain a higher price, and consequently might make the coefficient on number of acres different between the two. Furthermore, because of a reduction in labor supplied to the farm firm due to the treatment effect, animals might be sold earlier at lighter weights. This again would cause the coefficients between control and experimentals to be theoretically different. All three methods will be explained further beginning with a direct measure of the data collection system.

Reporting from records

During each quarter after the initial quarterly interview, the interviewer recorded whether information about farm income and expenses came from written documentation (scored as one) or memory (scored as zero). These were aggregated for each year into a score from zero to four with zero representing all information coming from memory and four meaning all information came from records.

A similar variable was created by taking into account the average amount of income and expenses reported each quarter by region. Therefore, rather than taking equal weights for each quarter, the 0, 1 variable was weighted by the combined amount of income and expenses typically reported that quarter relative to the total amount of income and expenses for the entire year. In North Carolina, for example, most of the farm income and expenses occurred during the third and fourth quarters. Consequently, if these quarters came from records and the other two did not, then the latter variable properly accounts for that distribution of the use of records being less serious than if reporting from records were randomly distributed throughout the year.

Table 5.2 shows the distribution of the raw unweighted reporting scores by region. It also shows the means for both the raw and weighted reporting scores. Each farmer is represented in the table three times, once for each year. Two facts are immediately obvious from the table. One is the amount of difference between Iowa and North Carolina. The North Carolina means range from .86 to 2.22 while in Iowa the range is 3.20 to 3.91. This means that in North Carolina the average farmer used records one or two times out of four for each year while in Iowa the average farmer used records more than three out of four times he was interviewed. This difference is highly significant and probably can be attributed to a lower level of education in North Carolina and a much smaller farming enterprise.

The second obvious fact in the table is the difference between control and experimentals in each region. For both measures the difference is slightly over one in North Carolina and slightly under one in Iowa. This implies that on the average, experimental families used records approximately one additional quarter each year compared to control families.

The simple chi square test of independence which tests the null hypothesis that the distribution of raw reporting scores is identical between the North Carolina control group and the North Carolina payment group had to be rejected at the one percent level. The same was true for the Iowa population.

Using the weighted score as the dependent variable, regression equations were estimated to determine whether the control/experimental difference was related to other variables like time or income. Variables were also introduced into the relationship to control for initial sampling differences between control and experimentals. These variables were

Table 5.2. Raw reporting scores for farmers who are less than 69 years of age, have constant marital status, and farm in each of the years from 1969 to 1972[a]

	Mean Weighted Raw Score	Mean Raw Score	Raw Score 0	1	2	3	4	Total
North Carolina Control	1.00	.86	70 (53)	33 (25)	15 (11)	6 (5)	8 (6)	132 (100)
North Carolina Experimentals	2.22	1.95	37 (22)	41 (25)	19 (11)	29 (18)	39 (24)	165 (100)
Iowa Control	3.24	3.20	7 (4)	14 (9)	20 (12)	20 (12)	101 (63)	162 (100)
Iowa Experimentals	3.91	3.91	0 (0)	0 (0)	2 (1)	10 (7)	138 (92)	150 (100)

[a]The number in parentheses indicates the percentage each raw score is of the total.

expected to increase the explanatory power of the equation. A brief description of the major independent variables and their expected signs is presented below:

Year dummies - because of a learning curve and a concerted effort by project administrators, both experimental and control scores should increase over time.

Year * treatment (C/E) dummies - these variables test for the difference by year between control and experimentals. The field staff reported that they were noticing increased payment cooperation relative to control in using records. This is quite believable because of the negative income tax payments being given to the experimental group. Consequently our expectation is for a positive sign, hopefully not significant.

Spouse present dummy - in many farm families the wife is responsible for record keeping, and bachelors are notoriously poor recordkeepers, as well as housekeepers. Although there are not many single males in the sample, having a wife probably will result in more information coming from records.

Farm size, age, education and quick test score[1] - these reflect the management ability of the farm operator. Consequently the better the manager, the higher will be the probability of the farmer using, and consequently reporting from, records. Farm size reflects or is an output of past management decisions while age, education, and quick test represent current management skill and human capital endowment.

Race and the race * treatment interaction were added to the equation for North Carolina to test for black-white reporting differences. To the extent that race for a given educational attainment represents less management skills (because of poorer schooling and less chance to learn from experience and agricultural extension services), a negative coefficient is expected to occur. The sign of the race-treatment interaction is not

[1]The Ammons and Ammons quick test score is based on a word recognition test. The scale range is 0-50.

known a priori. If negative, it implies a smaller difference between black experimental and black control families than between white experimental and white control families. If positive, the opposite result is indicated.

Pre-experimental net farm income was included to control for initial income differences between experimental and control. The higher the income, the more likely the farmer is a good manager and consequently keeps good records.

The regression equations as reported in Table 5.3 were estimated by ordinary least squares. With the exception of the North Carolina 1972 year dummy variable, all of the coefficients had the expected sign and nearly all were significant.

Income and farm size were positively correlated with reporting behavior. Also education, age, and quick test as measures or proxies for management skills were positively and significantly associated with reporting behavior.

All the treatment variables except the one for 1972 in Iowa were significant at the one percent level. In 1970 the Iowa experimental families reported from records 1.27 quarters more than their control counterparts. As shown by the year treatment interactions, this difference declined to .24 in 1972. In North Carolina, the 1970 difference between control and experimentals is 1.60 quarters for white families and .98 for black families. However in both 1971 and 1972 the difference grew rather than declined.

Table 5.3. Regression equations of reporting scores for farmers who are less than 69 years of age, have constant marital status, and farm in each of the years from 1969-1970

Dependent Variable: Weighted Reporting Score

Variable	Iowa Coefficient	Iowa t-value	North Carolina Coefficient	North Carolina t-value
Constant	.03	.04	-2.59	-3.18[a]
1971 Dum	.848	5.63[a]	.126	.457
1972 Dum	1.10	7.06[a]	-.0456	-.165
1970 Dum * C/E	1.27	8.25[a]	1.60	5.23[a]
1971 Dum * C/E	.562	3.66[a]	1.63	5.35[a]
1972 Dum * C/E	.241	1.56	1.71	5.60[a]
Ŝp Pres Dum	.323	1.14	.740	1.65[b]
Ave Frm Size Spline[c]	.000203	3.21[a]	.000148	1.34
Lrg Frm Size Spline[d]	.000123	3.22[a]	.000130	1.90[b]
Educ	.0619	1.95[b]	.0706	2.31[e]
Age	.0130	2.23[e]	.0388	4.17[a]
Quick Test	.0147	1.16	.00924	1.13
Net Frm Inc	.0000158	1.46	.0000684	1.78[b]
Race			-.544	-2.21[e]
Race * TR			-.620	-2.00[e]
	$\bar{R}^2$ = .31		$\bar{R}^2$ = .34	
	N = 312		N = 297	

[a]Significant at the 1 percent level.

[b]Significant at the 10 percent level.

[c]Equal to total scaled hours for farms between 1600-2199 hours in 1969, 0 otherwise. Total scaled hours are similar to total budgeted hours in that there are weighted combinations of physical units of productions. The coefficients have been arbitrarily adjusted for mechanization. See Chapter IV for a complete description.

[d]Equal to total scaled hours for farms greater than 2199 hours in 1969, 0 otherwise.

[e]Significant at the 5 percent level.

While these results are not conclusive proof that the data have a treatment bias, the possibility that this happened is enhanced. This evidence provides no answers to what the nature of the bias is, i.e., do experimental families report more accurately or less accurately, and what is the ultimate effect upon labor supply predictions.

Comparison with IRS

The second method of verifying the data collection process is comparison with an outside source that is less biased with respect to the treatment parameters. A detailed comparison with IRS is possible for the 407 observations who gave RIME their itemized schedule F (see Table 5.1). A simple regression equation can be estimated for each income/expense category. The quarterly interview value is regressed on a similarly defined IRS value, a control/payment dummy (C/E) and the control/experimental dummy interacted with the IRS value (C/E * IRS). If there is no treatment bias in collecting information and no data errors, all coefficients on the IRS variable should equal 1.0, and variables involving the control/experimental dummy should not be significantly different from zero.

In Table 5.4, simple means and the results of the regressions described above are presented by income and expense category. In column 1, the original data base mean is shown along with the standard deviation. The IRS mean and standard deviation are shown in column 2. Column 3 contains the coefficient on the IRS variable along with the standard error of estimate for the simple relationship between IRS and the original data base value. Column 4 contains statistics obtained from estimating original

Table 5.4. Regression equations of itemized income and expenditures on the original data base for farmers who are less than 69 years of age, have constant marital status, and farm in each of the years from 1969 to 1973[a]

Variable	Orig. Mean	IRS Mean	IRS Coeff. (S.E. of Est.)	Diff. (F Value)
Tobacco Sales	2270 (4149)	2323 (4149)	.98 (909)	-57.75 (7.64)[b]
Grain Sales	5581 (6656)	5951 (6850)	.93 (1814)	-114.23 (4.34)[c]
Acreage Diverted Payments	880 (981)	953 (998)	.87 (463)	-60.19 (7.94)[b]
Other Crop Sales	105 (396)	37.7 (259)	.80 (338)	3.46 (.03)
Cattle Sales	1555 (3486)	1599 (3201)	.94 (1754)	-48.85 (.94)
Hog Sales	3646 (5960)	4180 (6953)	.80 (2085)	274.26 (27.43)[b]
All Other Income	1183 (2109)	1381 (2328)	.81 (951)	-119.37 (10.38)[b]
Labor Expense	747 (1129)	809 (1130)	.90 (495)	52.48 (1.02)
Fertilizer, Seed, Herbicide	1605 (1769)	1653 (1672)	.89 (968)	23.68 (9.62)[b]
Machinery Repair, Supplies	1067 (1263)	1198 (925)	.75 (1055)	79.91 (.96)
Feed, Vet, Breeding Fees	2767 (4645)	2975 (4915)	.92 (1106)	59.97 (5.43)[b]
Gas, Fuel, Utilities	1227 (726)	1260 (730)	.78 (450)	-2.73 (.02)
All Other Expenses	1856 (2276)	2747 (3340)	.68 (1614)	417.75 (12.33)[b]
Total Gross Income	15219 (10655)	16426 (11460)	.87 (3647)	-143.06 (.18)
Total Gross Expenses	9269 (7895)	10642 (8443)	.87 (2911)	703.68 (6.79)[b]
Net Farm Income	5950 (5083)	5784 (4589)	.84 (3315)	-849.70 (4.98)[b]

[a]One observation was deleted because of data problems.
[b]Significant at the 1 percent level
[c]Significant at the 5 percent level.

data base values as a function of IRS, C/E, C/E * IRS. The top number is the predicted difference by which experimentals exceeds control evaluated at the IRS mean. A negative number would indicate that control exceeds experimentals. The number in parentheses is the partial F-value testing the significance of the treatment variables.

Notice, first, that the original mean is lower than the IRS mean in fourteen out of the sixteen categories.[1] Only in the variables other crop sales and net farm income, are the original means higher than the IRS mean. This implies that transactions are typically forgotten rather than incorrectly reported. This rationale could also explain why net farm income is higher on the original data base. Since relatively more transactions are made on the expense ledger, and assuming the errors of omission are inversely related to size of the transaction, gross expenses would be biased downward more than gross income, leading to a net income overstatement.

The second observation is that the standard errors of estimates in column 3 are relatively high. They typically are 25 to 50 percent of the standard deviation of the dependent variable and in one case 83 percent. For identically defined variables, this seems unusually high. While categorization errors still exist, the categories were chosen to minimize this kind of error. For example, fertilizer, seed, and herbicide were put together, as well as machinery repair and supplies. In most cases the categories are well delimited.

The third observation is that all the coefficients on the IRS variable are less than one. This agrees with statistical theory which says

[1]This is slightly misleading in that three of the sixteen categories are a function of the other thirteen categories.

that if two variables are measured with error, the coefficient will be biased downward from the true coefficient, which in this case should be 1.0. Furthermore, the amount of bias, given certain assumptions about the measurement errors, is related to the size of the measurement error in the independent variable, which in this case is the IRS variable. A priori expectations about the difficulty of measuring the different categories are borne out in most cases. For example, all other expenses is furthest from 1.0 while tobacco sales, a big item with few transactions is closest to 1.0. However, this does not explain why the coefficient on acreage diverted payments should be lower than the coefficient on grain or cattle sales.

The final point is that income items tend to have negative signs for the predicted difference, while expense categories tend to have positive signs. Recall that a positive sign implies that the experimental group exceeds the control group, while the reverse is true for a negative sign. Overall gross income is $143 less for the experimental than for the control group, while expenses are $704 more for the same level of the IRS variable. This leads to net farm income being $850 more for the control group. There is no good explanation for this result other than a treatment bias in reporting. Obviously the predicted difference between control and experimentals should be zero and insignificant rather than negative and significantly related to treatment.

If IRS is assumed to be correct and all the farmers are below breakeven, the reporting of income via the quarterly interview would cost taxpayers $425 per year with a 50 percent tax rate. The effect reporting of income has on the costs of a national NIT program could be substantial.

Accounting equations

Another method of determining the adequacy of the data collection system is estimating accounting equations. These equations should be identical for control and experimentals, but must be carefully constructed to avoid including an inadvertant treatment effect. The equations should have a high r^2, and more importantly the treatment variables should not be significantly different from zero.

Essentially no treatment effect was found in the accounting relationships on the original data base, and consequently the equations will not be reported here. However, a typical simple accounting relationship will be described. In North Carolina, tobacco sales are estimated as a function of the following:

TOBLB - Number of tobacco acres grown by respondent times his reported yield times a tenure adjustment. The tenure adjustment is based on each parcel of land and is one if the land is owned or rented for cash. If land is rented on a share basis, the adjustment is equal to the percent of income which the respondent receives. This adjustment is necessary because the tobacco sales reported on the quarterly interviews are sales which the respondent receives. By definition, he should not be reporting the amount of sales his landlord receives on a share basis.

C/E - This variable was entered as a dummy with one representing the treatment group.

TOBLB*C/E - Again this variable should not be significantly different from zero.

For Iowa the entire grain sales amount is predicted. The primary difference between the two regions is the relative importance of several crops in Iowa versus just one in North Carolina, the importance of livestock production and the subsequent feeding of these animals with home grown feedstuffs. Consequently, the equation was changed to reflect these basic differences.

Equations estimating livestock sales for the two regions were estimated in a similar way.

Magnitude of Differences Between Original and Edited Data Bases

The magnitude of changes between the data bases was immense. In Tables 5.5 and 5.6 for North Carolina and Iowa, respectively, the means and standard deviations of the important analytic variables are presented for the original data base (ORIG) and the edited data base (EDIT). Several other variable means and standard deviations are also presented. These include the following:

a) Edited minus original data base (Diff).

b) Absolute value of (a) (ABS). The number of changes could cancel each other out so that Diff equals zero. ABS presents a truer measure of the changes which took place.

c) Original data base divided by the edited data base (RATIO). The variable was constrained to be between -2.00 and 2.00. Presenting an unconstrained mean and having a change of net farm income from, say, -10 to 1,830 would present misleading statistics.

These tables demonstrate vividly that the changes between the original and edited data bases were enormous. In Iowa, for example, (Table 5.6), total family income has an average difference of 849 with a standard deviation of 4,640. The absolute value of these differences is 2,633 with a standard deviation of 3,921. This means net farm income was changed by a total of $821,496 for the 104 Iowa farmers over the three year period. This does not count changes to the data base which offset each other in a summary variable like net farm income.

The changes were most dramatic in income measures. This probably is due to two reasons. One is that income tended to have many cross-checks whereas asset values and debts did not. Adjustments were not made to the edited data base unless there were inconsistencies and an indication of where the inconsistency lay. Consequently, income tended to be changed often relative to asset data. The second reason is that compared to number of acres of corn or number of cattle sold, income and expense transactions happened frequently and probably were more often forgotten.

The physical production numbers (acres grown of each crop or number of livestock sold) which were asked near the end of each year did not change much. Probably this is true because the size of an individual operation is a status symbol and is a basis for most operational decisions made by the farmer. Consequently the hours variables which are constructed from these production numbers did not change substantially.

Table 5.5. Important analytical variable means for original and edited data bases in North Carolina

Variable	ORIG Mean	ORIG St. Dev.	EDIT Mean	EDIT St. Dev.	DIFF Mean	DIFF St. Dev.	ABS Mean	ABS St. Dev.	RATIO Mean	RATIO St. Dev.
TOT FAM INC[a]	4390	4045	4915	3759	524	2094	1312	1712	.91	.58
GROSS FARM INC[b]	7564	6863	8559	7100	995	2008	1157	1919	.87	.22
NET FARM INC[c]	2109	2881	2587	2213	478	2063	1261	1700	.80	.77
CROP SCAL[d]	1673	1367	1709	1382	36	361	88	352	.95	.24
LIV SCAL[e]	189	555	249	644	60	283	71	280	.46	.51
TOT SCAL[f]	1862	1425	1957	1451	96	455	156	438	.95	.21
TOT FARM VAL[g]	11324	12155	11573	11860	249	4094	1804	3682	.91	.35
FARM DEBT[h]	3218	5196	3325	5066	107	1475	438	1412	.58	.58
NET EQUITY,[i] LAND AND BUILDINGS	7772	8907	7790	8168	16.8	3856	1751	3434	.91	.51

[a] Total family income including unearned income, net farm and business income, plus wage income of all family members older than 15 years of age.

[b] Gross farm income--total amount of sales of farm produce, commodity loan payments, crop insurance proceeds less the amount paid for animals purchased that were sold during the year.

[c] Net farm income--gross farm income minus gross farm expenses including interest and depreciation.

[d] Scaled crop hours--weighted combination of acres raised. Previously defined in Chapter IV.

[e] Scaled livestock hours--weighted combination of animals and produce sold. Previously defined in Chapter IV.

[f] Total scaled hours--scaled crop plus scaled livestock hours.

[g] Total farm value--respondent's estimate of the total market value of his farm assets including land, machinery and equipment, livestock on hand, plus inventories of feed and gasoline.

[h] Total farm debt--total indebtedness of the farm operator on his farm operation.

[i] Net equity--total market (respondent estimated) value of farm land and buildings minus debt.

Table 5.6. Important analytical variable means for original and edited data bases in Iowa[a]

Variable	ORIG		EDIT		DIFF		ABS		RATIO	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
TOT FAM INC	5126	5863	6020	5373	894	4640	2633	3921	.83	.84
GROSS FARM INC	19013	11420	20860	12131	1846	4946	2576	4607	.92	.17
NET FARM INC	3808	5175	4649	4522	861	4629	2618	3911	.74	.90
CROP SCAL	1111	432	1112	432	2	16	2	16	1.00	.02
LIV SCAL	744	714	835	760	91	333	122	323	.80	.37
TOT SCAL	1855	849	1947	889	93	334	124	323	.96	.11
TOT FARM VAL	32079	29695	34280	29604	2201	5292	3401	4611	.91	.19
FARM DEBT	11484	17813	11534	17563	50	3981	1029	3845	.85	.44
NET EQUITY, LAND AND BUILD-INGS	13399	13336	14689	13474	1290	3909	1867	3668	.85	.90[b]

[a]See Table 5.5 for a definition of the variables.

[b]This high value is caused by the fact that many ratios are defined to be zero.

While absolute value differences are typically smaller for North Carolina than for Iowa, the percentage change is not much different. The reason Iowa's differences are higher probably is because the scale of operation in North Carolina is smaller and revolves around one major commodity--tobacco. Iowa is more diversified with large crop and livestock operations. Also it is common for an Iowa farmer to have several crop and several livestock operations.

Tables 5.7 and 5.8 contain a frequency distribution of the changes between the edited and original data bases for the variable, total family income, differentiated by year and control/experimentals. Even though the distribution is roughly centered, a large number (45 in Iowa, 10 in North Carolina) had changes greater than $5000. Also note that distributions between control and experimentals are different. In Iowa, for example, 51 families, or 31 percent of the control families, had changes of less than a -$1,000 compared to 17 percent for experimental families. However, the number of changes above 1000 were 27 percent for controls as compared to 37 percent for experimentals. Correspondingly, for North Carolina the percentage of control families with changes less than -1,000 was 19 percent versus 12 for experimental families, while the percentage of families above 1,000 was identical at 25 percent.

In Iowa the simple chi square test[1] yielded a value of 18.11 which is significant at the one percent level. This implies that the distribution is different. In North Carolina the chi square value is 5.33 which is insignificant.

[1]The categories < -5000 and -4900 to -2501 were added together prior to doing the chi square test.

Table 5.7. Frequency distribution of differences between original and edited data base for total family income in North Carolina

	Edited - Original							
	<-5000	-4900 to -2501	-2500 to -1000	-999 to +999	1000 to 2500	2501 to 5000	>5000	Total
CONTROL								
1970	1	1	6	19	8	6	3	44
1971	1	3	4	24	8	3	1	44
1972	2	2	7	21	11	3	0	44
TOTAL	4	6	17	64	27	12	4	132
EXPERIMENTAL								
1970	0	1	3	35	11	5	0	55
1971	0	3	4	39	6	2	1	55
1972	0	0	8	30	12	2	3	55
TOTAL	0	4	15	104	29	9	4	165

Table 5.8. Frequency distribution of differences between original and edited data base for total family income in Iowa

	Edited - Original							
	<-5000	-4900 to -2501	-2500 to -1000	-999 to +999	1000 to 2500	2501 to 5000	>5000	Total
CONTROL								
1970	2	7	12	20	7	1	5	54
1971	4	7	8	20	3	4	8	54
1972	4	7	10	17	9	1	6	54
TOTAL	10	21	30	57	19	6	19	162
EXPERIMENTAL								
1970	1	2	7	24	3	10	3	50
1971	0	2	4	27	5	6	6	50
1972	0	1	8	29	3	3	6	50
TOTAL	1	5	19	70	11	29	15	150

One may wonder if the extent of changes does not nullify any resulting analysis using edited data. The reader must ultimately pass judgment on this issue, but before doing this he should read Volume I, particularly Chapters 7 to 9 of the Final Rural Income Maintenance Report [24]. Field staff were carefully selected and trained, all interviews were pretested, and generally the methodology used in this data collection operation was the latest and best in terms of conventional practice.

The reasons for these differences have been alluded to in the introduction. Other surveys have encountered similar difficulties. For example, welfare income typically has been understated by 40 percent where population totals are estimated from stratified samples. The 1966-67 Survey of Economic Opportunity (SEO), when compared by researchers at the Urban Institute with the actual annualized caseload count, produced a 43 percent undercount. Approximately 1,527,000 AFDC cases were actually present in the sample universe while only 877,000 cases were identified as such by the SEO data [63].

A government sponsored survey in 1970 produced a 41 percent undercount of the state's actual welfare count. A weighted 2125 ANFC families were identified in the survey, as opposed to the actual total of 3590 cases on the rolls as of September 5, 1970 [63].

The current Population Survey (CPS) for 1972 produced a 34.5 percent undercount of welfare benefits [59].

The information available on the reporting of self employed income is similar. Some recent comparisons between Internal Revenue Service and USDA estimates of farm income have been made with wide differences. While conceptual differences in the definition of USDA and IRS farm income do exist, these do not appear to explain the differences [81].

For farm income in 1972, the CPS produced an undercount of 42.3 percent, while for nonfarm self-employment the undercount was 13.0 percent. These undercount percentages were derived by comparing CPS with benchmark income statistics from other sources [59].

Consequently, it is not surprising that the wide differences did occur in the rural negative income tax sample. These data problems underscore what will be the major conclusion of this chapter, that the accounting and reporting system used by any future income maintenance program will have more serious cost implications than corresponding labor supply effects. One must not only consider the overpayments resulting from families who understate their income, but also the underpayments resulting from low income families overstating their income, which is equally as serious given the legislative mandate of any income maintenance program.

Determining Which Data Base is Best

This section is modeled closely after the second section. Determining which data base is best can only be done by a comparison with outside sources or the estimation of known data relationships. An objective evaluation of which data base is best is practically impossible. If the known data relationships or the outside sources were used in creating the edited data base, the edited data base should always be better than the original data base when using these two criteria for evaluation purposes. Such is the situation here. A third source of data or known data relationships not used in the editing process may provide a somewhat objective evaluation. However these do not exist; and even if they did, one might question why they would not be used as well in the editing process.

In the actual editing process, conflicts arose between the different data sources. Furthermore changes were not made to strictly conform to Internal Revenue Service data or with accounting relationships. Consequently this section primarily explores the remaining differences between the edited data base and IRS data. The primary justification or rationale for accepting the edited data base as best is the methodology or procedure under which the edited data base was created. This is described further in Appendix B. The best data base is one which provides an unbiased and efficient estimator of the treatment coefficients.

The table presented earlier for the original data base (Table 5.4) is presented for the edited data base in Table 5.9. The edited means are higher than the IRS means in eight out of sixteen categories as compared with the previous fourteen out of sixteen. Gross income is now \$300 higher than the IRS mean as compared to \$1,200 lower on the original data base. Expenses are less by \$600 as contrasted to the original data base being \$1,400 less. Net income is \$900 more than IRS on the edited data base. One of the prime reasons net income is above and expenses below IRS is the refusal to allow certain car expenses.

The coefficient on the IRS variable increased in all sixteen categories and moved closer to 1.0 in fifteen categories. The lone exception was tobacco sales. The standard error of estimates declined in fourteen

Table 5.9. Regression equations of itemized income and expenditures on the edited data base for farmers who are less than 69 years of age, have constant marital status, and farm in each of the years from 1969 to 1973[a]

Variable	Edited Mean	IRS Mean	IRS Coeff. (S.E. of Est.)	Diff. (F Value)
Tobacco Sales	2430 (4330)	2323 (4149)	1.03 (576)	-49.44 (2.98)
Grain Sales	6143 (7241)	5951 (6850)	1.02 (2007)	112.93 (1.46)
Acreage Diverted Payments	920 (994)	953 (998)	.97 (239)	6.20 (.30)
Other Crop Sales	126 (456)	37.7 (259)	1.06 (364)	12.95 (.49)
Cattle Sales	1710 (3379)	1599 (3201)	.99 (1143)	75.02 (5.00)[b]
Hog Sales	4045 (6694)	4180 (6953)	.92 (1865)	230.06 (12.49)[b]
All Other Income	1376 (2366)	1381 (2328)	.96 (769)	-63.64 (.80)
Labor Expense	784 (1107)	809 (1130)	.95 (293)	31.09 (1.34)
Fertilizer, Seed, Herbicide	1754 (1747)	1653 (1672)	.93 (794)	-155.86 (2.12)
Machinery Repair, Supplies	1131 (958)	1198 (925)	.81 (601)	-20.38 (.47)
Feed, Vet, Breeding Fees	2977 (4830)	2975 (4915)	.97 (609)	-17.92 (1.21)
Gas, Fuel, Utilities	1232 (697)	1260 (730)	.86 (307)	3.50 (1.95)
All Other Expenses	2184 (2817)	2747 (3340)	.72 (1446)	313.29 (19.98)[b]
Total Gross Income	16749 (11482)	16426 (11460)	.97 (2949)	318.30 (4.89)[b]
Total Gross Expenses	10061 (8166)	10642 (8443)	.93 (2113)	249.03 (1.21)
Net Farm Income	6688 (5016)	5784 (4589)	.95 (2470)	-36.94 (8.60)[b]

[a] One observation was deleted due to data problems.
[b] Significant at the 1 percent level.

equations, went up slightly in one, and increased substantially in only one equation. In some cases the decline was large, particularly in expense categories.

In Table 5.4, ten equations had a significant treatment effect. On the edited data base this declined to five equations. More importantly the $850 control/experimental differential for net farm income declined to $37. This suggests that if IRS is unbiased with respect to treatment, the edited data base has been adjusted correctly for income and expenses and is less prone to contain a bias between control and experimentals.

Another method of determining which data base is best is estimating accounting relationships which should be identical for control and experimentals. As alluded to in the introduction, the equation must be carefully constructed to avoid including an inadvertant treatment effect. Further evidence of which data base is best can be obtained by looking at the total variation explained by the accounting equation and by having zero treatment coefficients. Removing outliers, as was done in the transition from the original to the edited data base, should increase the explanatory power of any accounting relationship.

The same accounting relations as described previously in section 3 were re-estimated for the edited data base. In all equations, the R^2 increased and no treatment effects were encountered. This evidence combined with the procedure used to create the edited data base suggests that the edited data base is preferred to the original data base.

An Explanation of the Changes Between the Original and Edited Data Bases

The primary purpose of this section is to determine if the changes between the original and edited data bases were systematically related to any family or farm characteristics, particularly treatment variables. The effect of reporting from records was found to have a profound impact on the extent of changes between the two data bases.

There is scant previous literature on what variables should explain nonresponse errors or what families are most likely to report inconsistencies in the data.

As an economist, one might turn first to a benefit/cost approach. In this case, there is no benefit or loss from reporting inaccurately on the interview. The survey operation and payment operation were entirely independent. Each month the respondent filled out a form upon which payments were based. Information that was reported to the payments department was not adjusted or made to conform with information on the interview. Even in cases of wide discrepancy between the two sources of information, no audits were performed. Probably, however, experimental families did not realize the distinction, and if they had something to hide, they may have reported identically to both payment and survey departments.

One can postulate the following hypotheses to explain the incidence and magnitude of errors:

a) Errors would be related to an individual's ability to report accurately. This could be measured by an individual's education, quick test score,[1] management ability, and age. Higher management abilities, education levels, etc., would be associated with lower error levels.

b) Errors should be related to the difficulty of reporting for a particular farm operation. For example, holding everything else constant, the farm operation having no share or tenancy arrangements should report better than a farm operation with several tenancy arrangements. Other ways of measuring this concept are number of different crop operations, number of different livestock operations, percent of cropping operation in corn and soybeans (tobacco in North Carolina), percent of livestock operation in one enterprise, number of different landlords, and 1.0 minus the fraction of land rented on a share basis. The farmer who has several landlords but rents all land on a cash basis has only one or two simple transactions to remember, while those renting on a share basis have many transactions to recall. The higher the percentage in the major crop or livestock operation, the more accurate the reporting should be.

c) The effect of farm size upon reporting accuracy is probably indeterminate. Controlling for everything else, e.g., management ability, the farmer with the higher level of business activity would be more likely to have a greater error level. However, size of farm may also serve as a proxy for management ability. The fact that a farmer has a larger than average business activity may also indicate a higher management ability. Thus if management ability is inappropriately measured, size of farm could have a negative sign.

---

[1]The Ammons and Ammons quick test score is based on a word recognition test. The scale range is 0-50.

d) Error level should be related to whether information comes from records. As the number of quarters for which records were used increases, the accuracy of reporting should increase.

e) The level of error should decrease with time. As the respondents learn what is expected of them, they respond (perhaps by using records) by reporting more accurately.

f) The treatment effect on accuracy of reporting is indeterminate. Because experimental families on the average were receiving $1400 per year from the experiment, their degree of cooperation or willingness to provide information probably would enhance that fact alone. Evidence presented earlier showing experimentals reporting more frequently from records supports this viewpoint.

On the other hand, one could argue that experimental families may have attempted to maximize the size of their payment check and reported the same information to both payments and surveys. Relative to the control group, this might result in less income and more expenses being reported.

Separating or disentangling these two effects will be achieved in the following manner. Essentially the model will be estimated in two forms. One form is the simple difference between the edited and original data bases. This variable presumes the errors are intentional. As explained in section 2 the level of reporting is related to the treatment parameters. If the treatment parameter is significant, after controlling for the level of reporting, it suggests that some experimental families manipulated the payment system to their advantage.

The other form of the model is the absolute difference between the edited and original data bases. If the treatment variables are significant and negative, this suggests that being in the experimental group reduced the error level or amount of changes between the two data bases. If most survey response errors are unintentional, the test of hypotheses a through e above should be stronger or more evident by the use of this model.

In Table 5.10 for North Carolina and in Table 5.11 for Iowa, the partial F-statistics are shown for testing hypotheses a through f above on two selected independent variables. These variables are total family income and scaled crop hours in North Carolina and total family income and scaled livestock hours in Iowa. These two variables were chosen to be representative for a large number of dependent variables. One variable was chosen from income and one from hours with the hours variable which was most sensitive to the treatment being the prime candidate.[1] This was done to determine if the editing procedure had an impact upon disincentive predictions. The full equations are reported in Appendix D.

In North Carolina, with scaled crop hours primarily a function of acres, the only variable which explains any portion of the change between the data bases is farm size. However, for total family income, several things are evident. Difficulty of reporting and measures of an

---

[1] Many important analytical variables in several forms (ABS, DIFF, RATIO) were estimated as a simple function of treatment parameters. From these equations, the selected variables (total family income, scaled crop hours in North Carolina and scaled livestock hours in Iowa) were chosen. In the other variables, the differences in the data bases seemed insensitive to treatment parameters. More work would need to be done before that could be concluded definitely, however.

individual's ability to report are significant at the one percent level for the absolute error model and are insignificant for the difference model. This result follows directly if one assumes the errors are unintentional.

Table 5.10. North Caroline partial F-statistics for determining the significance of various groups of independent variables for explaining the changes between the original and edited data bases[a]

	Dependent Variables			
	Diff Total Family Income	ABS Total Family Income	Diff Scaled Crop Hours	ABS Scaled Crop Hours
1 Due to Individual's Reporting Ability	.58	5.21[b]	.24	.49
2 Difficulty of Farm Enterprise	2.07	4.35[b]	.38	1.29
3 Farm Size	7.34[b]	11.46[b]	5.30[b]	4.50[c]
4 Reporting from Records	7.98[b]	10.38[b]	.81	.05
5 Time	2.49	.86	.81	1.50
6 Treatment	1.83	.23	.90	1.88
7 Time * C/E	.63	2.97	1.14	2.97
8 DIS, DIS * C/E	5.00[b]	7.22[b]	.43	3.37[c]

[a]The numbers in front of each line refer to the variables which comprise each group. These variables are described in Appendix D. Each group of variables was entered into the equation in the order shown above.

[b]Significant at the 1 percent level.

[c]Significant at the 5 percent level.

Table 5.11. Iowa partial F-statistics for determining the significance of various groups of independent variables for explaining the changes between the original and edited data bases[a]

	Dependent Variables			
	Diff Total Family Income	ABS Total Family Income	Diff Scaled Crop Hours	ABS Scaled Crop Hours
1 Due to Individual's Reporting Ability	.71	2.42[b]	2.03	2.03
2 Difficulty of Farm Enterprise	.64	2.52[b]	3.38[c]	4.17[b]
3 Farm Size	20.81[c]	21.74[c]	13.84[c]	25.79[c]
4 Reporting from Records	.66	4.96[b]	.45	8.11[c]
5 Time	.70	1.22	.73	.56
6 Treatment	9.56[c]	8.68[c]	7.62[c]	6.16[c]
7 Time * C/E	.10	.58	.46	.39
8 DIS, DIS * C/E	4.83[c]	3.07[b]	4.00[b]	3.38[b]

[a]The numbers in front of each line refer to the variables which comprise each group. These variables are described in Appendix D. Each group of variables was entered into the equation in the order shown above.

[b]Significant at the 5 percent level.

[c]Significant at the 1 percent level.

There is no evidence of a learning curve or a treatment effect. Reporting from records is highly significant in both equations and the sign (shown in Appendix D, Table D.1) indicates that errors are reduced when records are used in reporting. Errors increase with the level of farm size, holding everything else constant. The farm size variable is highly significant in both models. It probably reflects a number of phenomena, one being that the degree of cooperation from respondents typically fell if much effort was expended in reporting information.

Size of farm is also an indicator of the number of transactions and, with more transactions, the higher probability of an error.

For Iowa, some results are similar and some are quite different. Because livestock operations are more varied and complex than crop operations in North Carolina, the differences between the data bases are significantly related to several variables as opposed to just one in North Carolina. Along with farm size, the complexity of the farming enterprise, reporting from records, and the treatment variable are significant. More will be explained about the treatment effect later.

As in North Carolina, for total family income, difficulty of reporting and measures of an individual's ability to report are significant for the absolute error model but are insignificant for the difference model. The difference between regions is that treatment is significant while reporting from records is insignificant in the difference model.

The significance of the treatment effect needs to be analyzed carefully. Ideally one would hope that any changes to the data would not be related to treatment variables, for that raises the question of which data base to believe. One could argue that if errors were being eliminated, primarily the precision of the results would be affected and not the level of the predicted incentive or disincentive.

In Iowa this ideal obviously did not happen. However, if the original data collection system was biased with respect to the treatment parameters particularly in the reporting of income, then corrections to the system will be correlated with treatment parameters. This is the prime rationale for the significant treatment effects. It also implies that experimental farmers were intentionally underreporting to the quarterly interviews in concordance with their reporting on the payments system. This question is addressed elsewhere.[1]

Thus far, the last line of Tables 5.10 and 5.11 has been conspicuously avoided. DIS measures total scaled hours for each year minus 1969 total scaled hours. A negative sign would indicate a decline in farm hours from the 1969 level. As an afterthought, a measure of the change in labor supply was inserted into the equation explaining the difference in total family income. This variable was insignificant, but when interacted with the control/experimental variable it was significant in seven of the eight equations. This phenomenon, although not originally hypothesized, was so strongly related to treatment parameters, that it could not be ignored. The explanation is subtle but should give the reader an insight into the editing process. DIS is based upon physical units of production, i.e., acres of corn, number of cattle sold, etc. Consequently, whenever an inconsistency arose, for example between number of cattle sold, cattle sales data, IRS sales, inventory numbers, or feed expenditures, all data were consulted. When control families were involved, more often than not the inconsistency did not lie in the first two values but in the overall relationship between the five different pieces of data. Consequently, the consistency was most often resolved (if changes in income or hours were to occur at all) by changing number of cattle sold and cattle sales in the same direction--either increasing or decreasing both values.

---

[1]See [25] for a more complete discussion of the differences between payments data and data from the quarterly interviews.

However, when experimental families were involved, the inconsistency frequently did lie between the first two numbers. Consequently, the discrepancy only could be resolved by changing either one of the two variables or by changing both in opposite directions. This explains the sign and significance of the DIS and DIS * C/E variables in Tables 5.10 and 5.11.

A careful analysis of the models without one variable, number of times records were used, reveals an interesting paradox for Iowa. For the difference model, the treatment variable is positive and large suggesting that the average difference for experimentals is larger than that for controls. However, the absolute model indicates that the average control change was larger than the average experimental change. This apparent contradiction can be explained quite easily. The average experimental change was typically small (relative to control) but usually in one direction; the changes to the control families were large but in both directions, making the difference lie closer to zero.

The changes between the data bases are explainable by some family and farm characteristics. Reporting from records reduced the number of changes. There was no evidence of a learning curve, but there was evidence of treatment effects in Iowa. This undoubtedly will have implications for the next section which examines the effect editing had upon treatment coefficients.

The Effect of the Changes Upon Results

The main purpose of this excercise is to determine the effect which editing had upon experimental results. While all the point estimates of the various treatment effects are changed, none are changed drastically with the exception of the Iowa net farm and gross farm income equations and the North Carolina net farm income equation. This was expected given the results of the previous section, namely, that the changes between the original and edited data bases in Iowa were related primarily to treatment and predominately for income variables. In addition the treatment variable in the North Carolina adjusted scaled hours equation became significant.

The following is a brief summary of the major independent variables and the reason for their inclusion in the model.[1] Algebraically the models estimated for Iowa can be expressed as follows:

$$Y = b_0 + b_1(\text{HR69}) + b_2(\text{HR69}^2) + b_3(\text{AGE}) + b_4(\text{EDUC}) + b_5(\text{NET EQUITY}_{69}) + b_6(\text{DEBT RATIO}_{69}) + b_7(\text{AGE55}) + b_8(\text{YR}) + b_9(\Delta\text{OFF-FARM HOURS})$$

b_0, b_1, b_2, b_7, and b_8 are allowed to be different between control and experimentals. The other coefficients are constrained to be identical between control and experimental families.

[1]For a further detailed discussion of these variables, the reader is referred to Chapter IV. While the model is slightly different than the one in Chapter IV, the difference does not affect the analysis of this chapter or its conclusions.

A brief description of the variables follows:

HR69 - Total scaled hours in 1969 which controls for initial differences between the control and experimental groups. When the dependent variable is crop or livestock scaled hours, this variable is the 1969 version of crop and livestock scaled hours respectively.

$HR69^2$ - Allows the growth in income or hours between 1969 and the current year to be curvilinear.

AGE, EDUC, NET $EQUITY_{69}$, DEBT $RATIO_{69}$ - these variables control for sampling variation and also allow growth in the dependent variable to vary linearly with different levels of the variables in question. The debt ratio is the amount of total farm debt divided by the total asset level.

AGE55 - A spline function for age which is assumed to have a kink at age 55. This variable is zero for all ages less than 55 and is equal to the age of the respondent for all years above 54. Taking the age variables as a set, the Δ in the dependent variable is assumed to be linearly related to age, with a different linear function for those under and over age 55. However the formulation restricts the overall function to be continuous at age 55. The treatment is presumed to affect those in the over age 55 group differentially with respect to age.

YR - composed of two dummy variables, one for 1971 and one for 1972.

Δ OFF-FARM HOURS - defined to be the level of wage and business hours for each year minus the 1969 level of wage and business hours.[1] This variable is assumed to be exogenously determined.

The model for North Carolina is slightly different. It includes a race variable plus six variables representing family composition. The latter are included because of the importance of labor, especially unpaid family labor, in the growing of tobacco. The race coefficient is allowed to be different between control and experimentals while the other coefficients are constrained to be identical.

The population is identical to that of earlier sections in this chapter with the exception of five families being eliminated from the North Carolina sample. The reduced form is used to derive treatment coefficient estimates.

The full regression equations are presented in Appendix D, Tables D.3 and D.4. It is interesting to note the difference in disincentives predictions that the two data bases generate based upon the above model. This is clearly shown in Table 5.12 for North Carolina and Table 5.13 for Iowa.

---

[1]Business hours for 1969 was not asked in the interview. Consequently 1970 business hours was used for both 1969 and 1970. This fudge affects few observations.

Table 5.12. North Carolina predicted incentives for the original and edited data bases for selected dependent variables from the regression models in Appendix D, Table D.3.

	Selected Y_{69} Values					
Dependent Variable	800	1200	1600	2000	2400	2800
Total Scaled Hours						
EDIT DB	-25.8	-11.1	-5.1	-4.6	-7.6	-12.8
ORIG DB	-32.8	-18.6	-11.6	-9.2	-9.7	-12.1
DIFF	7.0	7.5	6.5	4.6	2.1	-.7
Scaled Adjusted Hours						
EDIT DB	-17.0	-4.0	2.8	5.1	4.4	1.4
ORIG DB	-11.2	.6	7.3	10.5	11.2	10.2
DIFF	-5.8	-3.4	-4.5	-5.4	-6.8	-8.8
Net Farm						
EDIT DB	3.0	.7	.3	.8	2.0	4.0
ORIG DB	3.5	15.7	20.7	22.5	22.4	20.7
DIFF	-.5	-15.0	-20.4	-21.7	-20.4	-16.7
Gross Farm						
EDIT DB	-7.2	-8.1	-10.1	-12.8	-15.8	-19.0
ORIG DB	-10.3	-9.5	-10.8	-13.4	-16.8	-20.8
DIFF	3.1	1.4	.7	.6	1.0	1.8
	400	800	1200	1600	2000	2400
Crop						
EDIT DB	-35.6	-18.9	-12.7	-11.6	-13.1	-16.2
ORIG DB	-40.4	-25.4	-18.9	-15.9	-16.8	-18.6
DIFF	4.8	6.5	6.2	4.3	3.7	2.4

Table 5.13. Iowa predicted incentives for the original and edited data bases for selected dependent variables from the regression models in Appendix D, Table D.4.

Dependent Variable	Selected Y_{69} Values					
	800	1200	1600	2000	2400	2800
<u>Total Scaled Hours</u>						
EDIT DB	3.2	-5.0	-8.6	-10.2	-10.7	-10.7
ORIG DB	-9.3	-7.0	-6.9	-7.8	-9.3	-11.2
DIFF	12.5	2.0	-1.7	-2.4	-1.4	.5
<u>Net Farm</u>						
EDIT DB	-28.7	-4.7	11.2	21.7	28.1	31.0
ORIG DB	-49.2	-9.7	15.6	21.9	11.5	-9.7
DIFF	20.5	5.0	-4.4	-.2	16.6	21.3
<u>Gross Farm</u>						
EDIT DB	54.7	18.5	2.9	-4.9	-8.8	-10.4
ORIG DB	74.7	23.4	2.0	-9.1	-15.3	-18.7
DIFF	-20.0	-4.9	.9	4.2	6.5	8.3
	400	800	1200	1600	2000	2400
<u>Crop</u>						
EDIT DB	6.1	4.1	1.6	-1.1	-4.0	-7.0
ORIG DB	8.0	4.8	1.6	-1.7	-5.1	-8.6
DIFF	-1.9	-.7	0	.6	1.1	1.6
	0	200	400	600	800	1000
<u>Livestock</u>						
EDIT DB	-70.3	-44.8	-35.7	-30.5	-26.9	-24.1
ORIG DB	-71.1	-46.0	-34.8	-28.8	-25.4	-23.3
DIFF	.8	1.2	-.9	-1.7	-1.5	-.8

For each of the five variables[1] for which editing was done, a predicted incentive percentage is presented for both the original and edited data bases for various levels of 1969 total scaled hours. This incentive percentage is defined as (P-C)/C * 100.0. A negative sign indicates a disincentive. The original data base percentage incentive is subtracted from the edited data base percentage incentive to obtain a difference in predictions between the two data bases. A positive sign indicates that the edited data base is predicting a higher incentive, while a negative sign indicates a lower incentive or higher disincentive.

Turning to North Carolina first, Table 5.12, several things are immediately noticeable. The edited data base predicts a higher incentive for crop hours which in turn influences or leads to a higher incentive for total scaled hours. However, the edited data base predicts a lower disincentive for net farm income. Why this inconsistency? If one inspects the results from the original data base, some highly unusual results abound. For example, there is a fairly large incentive for net farm income and yet experimental farmers are working fewer hours. This is even harder to explain in light of the substantial gross farm income disincentive. The edited data base does much to dispel these inconsistencies because the net farm income incentives were reduced to small percentages, while the hours incentives were increased.[2]

---

[1]Iowa scaled adjusted hours and total scaled hours yield identical results so results for the former are not presented. Livestock hours is dominated by just a few operators in North Carolina and consequently the results are not presented.

[2]While doing the editing process, it was not known whether the family was control or experimental and no thought was given to whether the changes would be easier or harder to justify.

The last inconsistency is between scaled adjusted hours and total scaled hours. The former is an adjustment to total scaled hours based on hired labor, machine hire, and custom work done. Because of the experiment, one would theoretically expect a larger disincentive with scaled adjusted hours than with total scaled hours alone. Such was not the case however. Original scaled adjusted hours shows a positive incentive while original scaled hours shows a substantial disincentive with a difference between the two variables of around 20 percent. While not completely reconciling the two measures, the inconsistency is reduced to around 10 percent in the edited data base. The basic pattern of results for North Carolina is the same between the two data bases.

In Iowa, this is even more true. For all hour variables, for all percentages except one (lowest level of total scaled hours), the changes are small and are all less than 2.5 percent. The income variables have changed somewhat, however. For gross farm income, the original income incentive went from 75 percent to -19 percent. This has been scaled down to a range from 55 percent to -10 percent. Net income underwent a similar change. The original percentages were negative for low values of 1969 total scaled hours, became positive, and then returned to negative at high levels of 1969 total scaled hours. The edited data base has changed this pattern. At low levels the incentive percentages begin at negative values and then become increasingly more positive.

There is one other aspect of the results between the edited and original data bases which deserves mention. That is the effect upon precision or accuracy of the results. These are shown for North Carolina in Table 5.14 and for Iowa in Table 5.15.

Table 5.14. North Carolina partial F-statistics and related regression statistics for the original and edited data bases from the regression equations in Appendix D, Table D.3.

	Dependent Variables				
	Net Farm Income	Gross Farm Income	Total Scaled Hours	Adjusted Total Scaled Hours	Scaled Crop Hours
Orig Eq. S.E.[a]	2738	5255	810	646	786
Edit Eq. S.E.[a]	1938	5052	778	563	754
Treatment Variables					
ORIG	.95	3.10[b]	3.72[b]	1.32	3.66[b]
EDIT	.83	3.91[c]	9.28[c]	2.78[b]	8.85[c]
Year * TRT					
ORIG	.48	2.1	.52	.41	.55
EDIT	.36	.09	.04	.26	.04
Age * TRT					
ORIG	1.36	3.10	.04	.62	.10
EDIT	4.44[b]	4.39[b]	.73	3.15	.61
Race * TRT					
ORIG	.96	1.49	3.71	.13	2.64
EDIT	.62	.97	1.98	1.45	1.68
Overall R^2					
ORIG	.17	.44	.61	.47	.62
EDIT	.28	.50	.66	.56	.68

[a]Standard error of estimate.

[b]Significant at the 5 percent level.

[c]Significant at the 1 percent level.

Table 5.15. Iowa partial F-statistics and related regression statistics for the original and edited data bases from the regression equations in Appendix D, Table D.4.

	Dependent Variables				
	Net Farm Income	Gross Farm Income	Total Scaled Hours	Scaled Crop Hours	Scaled Livestock Hours
Orig. Eq. S.E.[a]	4865	8471	494	207	468
Edit Eq. S.E.[a]	4152	8217	416	203	398
<u>Treatment Variables</u>					
ORIG	6.59[b]	5.06[b]	6.38[b]	.94	7.35[b]
EDIT	1.58	1.48	4.82[b]	.64	3.99[b]
<u>Year * TRT</u>					
ORIG	1.40	.50	.24	.17	.17
EDIT	.33	.12	.41	.15	.59
<u>Age * TRT</u>					
ORIG	1.42	2.89	1.90	6.49[c]	.15
EDIT	2.41	4.98[c]	3.08	5.67[c]	.13
<u>Overall R^2</u>					
ORIG	.21	.50	.69	.79	.61
EDIT	.21	.57	.80	.79	.75

[a]Standard error of estimate.

[b]Significant at the 1 percent level.

[c]Significant at the 5 percent level.

Turning to Iowa first, three important distinctions between the two data bases should be noted. As expected, the explanatory power of the equation without treatment variables increased for the edited data base in all equations. This is evidenced by the lower standard error of estimate as shown in line 1 of Table 5.15. Also the R^2 of all models has increased or stayed the same when comparing the edited data base with the original.

The second observation is that one would have concluded that the treatment influenced income variables if only results from the first data base were available. This is rejected for the edited data base. The earlier result was significant at the one percent level, while the edited version has a very low significance level.

The third result is that the precision of results with respect to year and age are roughly equivalent between the two data bases. Age became a slightly more significant determinant in the gross farm income equation. Overall, the significance of the treatment variables typically declined from the original to the edited data base.

In Table 5.14 the reader will notice two similarities and one dissimilarity between the North Carolina and the Iowa results. As in Iowa, the standard error of estimates declined in all equations and the R^2 increased when comparing the edited data base with the original data base. Also as for Iowa, the year interacted with treatment and age interacted with treatment variables were roughly equivalent. As in Iowa the significance of age interacted with treatment increased.

The basic difference between the regions is that the significance level of the treatment variables increased in North Carolina while the significance level declined in Iowa.

Conclusion

Several conclusions can be drawn from the evidence presented in this chapter. The major result is that an income maintenance scheme which must administer a program to the self-employed will entail more cost from income reporting problems than cost from disincentives in labor supply.

For example the program may have induced a $900 income under-reporting problem in Iowa which translates approximately into an average $300 over-payment. This assumes some families are over the breakeven level. However, an average estimate of net farm income disincentive is at most $100 (from Chapter IV), which translates roughly into $50 of extra cost.

The second major conclusion is that the changes between the data bases were large. This entire analysis raises doubts about any survey's ability to correctly recall or construct income in a detailed, disaggregated manner. The profession ought to tackle with diligence the problems of measurement, especially in the areas of income and asset information. While that is hardly a new conclusion, one cannot help but wonder about the validity of our current income statistics from the census and the CPS.

Another conclusion is that the direction and magnitude of predicted incentive percentages were not significantly changed for hours variables. Income results were changed, particularly in Iowa. Based on the original data base, a treatment effect for income would have resulted. The edited data base contains no such effect.

This raises the question of whether the entire edited data base concept was worthwhile. The answer is unequivocally yes. Without the edited data base, researchers would have compared IRS information with the original data base, and also could have shown by the analysis in section two of this chapter that the data collection system was potentially biased with respect to the treatment parameters. This would have produced some very nagging questions and one would have continually wondered about the validity of the experiment's results. While the edited data base does not erase all those questions, it does lend considerable credibility and validity to the study. The cost of this effort was small (less than two percent) compared to the cost of the entire study.

CHAPTER VI. CONCLUSION

Summary

The prime focus of this analysis has been to estimate the farm work response of farm operators. Several dependent variables were defined, each with different policy implications. Income reflects a family's ability tc earn income and obtain a decent standard of living. With this variable, one should be able to estimate program costs. It is expected that a farm operator would reduce the enterprises where the marginal value of his time is the lowest. Theoretically it would be possible to have a 15 to 20 percent decrease in labor supply which would have only a small impact upon income and the cost implications of a nationwide income maintenance program.

The measurement of labor supply was undertaken in two directions. The first was to have the farmer recall his hours of farm work for one week at four different points throughout a year. The second was to construct a linear combination of crops grown and livestock and produce sold where the coefficients are obtained from outside sources. They reflect the number of hours required to grow one unit of the enterprise in question on a farm with average mechanization. This assumption was relaxed by assuming the scale of operation perfectly predicts farming methods employed.

Using reduced form equations, a time series - cross section estimation routine based upon Nerlove [73], and controlling for important farm variables in the pre-enrollment year, treatment coefficients were estimated. The two formulations of hours worked produced different results.

The recall hours variable was usually positive and statistically insignificant while constructed hours was primarily negative and statistically significant. This was due to adjustments to livestock operations in Iowa and crop operations in North Carolina. These operations are flexible, require short planning periods and can be adjusted easily. This is contrasted with crop operations in Iowa where no reduction in hours occurred, since institutionally land tenure arrangements do not allow for much flexibility.

For North Carolina, in Model II the treatment is parameterized as a simple dummy variable, and this dummy is interacted with hours worked in 1969, the pre-experimental year. The constructed hours has a disincentive of 10 to 15 percent, net farm income a disincentive of 7 percent, gross farm income a disincentive of 14 percent, and recall hours an incentive of 7 to 9 percent.

In Iowa the constructed hours has a disincentive of 8 to 17 percent, net farm income a disincentive of 10 percent, gross farm income a disincentive of 7 percent, while recall hours is positive from 1 to 4 percent. The resolution of what hours variable should be used primarily depends upon the underlying data collection methodology. Since acres and number of livestock are more accurately reported than number of hours worked the previous week, this author favors the results from constructed hours.

These results are generally consistent with theoretical predictions. However, the effect of tax and guarantee rates upon labor supply are mixed, often significant, and of the wrong sign. The primary reason for this is a lack of observations and a possible misspecification of the model.

During the data collection phase, much information about the farmer was accumulated. This included inventories of livestock at the beginning and end of each year, detailed itemization of income and expenses, acres devoted to each cropping enterprise, number of livestock sold, plus the schedule F as reported to the Internal Revenue Service. It soon became evident that there were many discrepancies in responses. These were highlighted by the inconsistency with an outside source (Schedule F) as well as internal inconsistencies within the interview (e.g., cattle in the beginning inventory, no record of sales or death, and no cattle in the ending inventory). Furthermore, it was noted that there was a correlation between treatment parameters and whether the family used records in giving information to the interviewer. Therefore an intensive data editing task was initiated.

Employing a priori information, making judgments about the quality and direction of the various data sources, and changing as few responses as possible, an internally consistent picture of the farm was generated. The original information, devoid of processing errors, was also retained for comparison purposes and insurance. The changes between the data bases were fairly drastic. Net farm income was changed by an average of $1261 in North Carolina and $2618 in Iowa. The overall consistency with the outside source was improved. More importantly, relatively little change was detected in the significance and value of treatment coefficients. The major exceptions to this are the income variables in Iowa. From the original data base one would have concluded that net farm income was significantly reduced for the experimental group relative to the control families. While the edited data base is still negative, the significance level is very low. Generally most other point estimates of disincentives were smaller in the edited data base relative to the original data base.

Explaining the variation in differences between the data bases as a function of family or farm characteristics was generally poor in terms of R^2. It was easier to explain the absolute difference rather than the simple difference between the two data bases. Generally farm size, reporting from records, difficulty of farm enterprise, and an individual's ability to report were significant in explaining the changes. No learning curve was detected in either site. In Iowa, the treatment parameters were significant in explaining the changes. Furthermore there is some evidence which indicates experimental families attempted to maximize their payments through inaccurate reporting, particularly during the final year of the experiment.

Consequently, families who intentionally or unintentionally misreported income had a different marginal tax than that assigned by the experiment. Their behavior is different from other experimental families who did not misreport and can be partially explained by inclusion of a variable named PN. This variable is constructed as actual payments minus predicted payments divided by their average. As expected, when PN is large, disincentive effects are less.

Issues and problems for further research

The analysis has been concerned primarily with estimating labor response of farm operators in two areas. These results should be extended and a simulation study undertaken to analyze the impact of these results upon the supply and prices of the entire agricultural market. By estimating which farmers are eligible, determining their current impact upon the total market, and using the information from this study (i.e., a presumed reduction in livestock operations in Iowa and tobacco in North Carolina), the change in consumer prices could be estimated. While it may be apparent that the short run impact would be to raise prices, it is not clear that this is necessarily the long run effect. Perhaps resources freed up by low-income farmers would be used more efficiently by higher income individuals, resulting in a different long run effect.

The farm model developed here was essentially static and did not fully take into account the simultaneous decision-making aspects of the farmer in his farm enterprises. For example, the farmer simultaneously makes an investment/consumption decision at each point in time, a work/leisure decision, as well as numerous decisions about the use of intermediate goods and the hiring of farm labor. Perhaps the data in this experiment can be used to model the entire family farm firm.

The three year design of the experiment and the effect this has upon labor supply predictions under a permanent national program should be carefully investigated. The theory would suggest that farm operators should have less labor supply effects than their urban wage earner counterparts working in an environment where the marginal worth of their time is equal to their wage rate. However, the empirical results from the income maintenance experiments do not suggest this. As shown by this analysis, greater disincentives were uncovered in enterprises where the farmer has flexibility in changing the scale of operation and his labor input quickly. Thus the discrepancy between these results can be resolved by (1) suggesting that farmers and wage earners are entirely different entities because basic motivational patterns are dissimilar; (2) arguing that previous wage analysis has incorrectly modeled the institutional constraints of a 40 hour work week (i.e., changing labor supply implies changing jobs or reducing work hours to zero); or (3) believing that recall hours is a more accurate measure than constructed hours.

This analysis has highlighted the differences between different sources of data and the extent and variance of misreporting. If a national program were instituted, careful attention should be given to the proper method of collecting income data. Given the cost implications of misreporting, administrative experimentation should be undertaken to determine how income questions should be formulated, the frequency of reporting, and the extent and magnitude of audits.

Direct comparison of survey income data and payments data should be done. This could provide more direct evidence as to the extent of misreporting and with which farm and family characteristics misreporting occurred. This analysis suggests that the misreporting was large and that it affected the behavior of farm families. Response errors in the survey data were analyzed but not the differences between the payment and survey data.

The analysis would also suggest that the prime consideration for determining actual tax and guarantee rates should not be based upon labor supply implications but upon cost implications. Evidence from other income maintenance experiments probably should be consulted before this conclusion is entirely accepted, given the paucity of observation in certain plans in this analysis.

Policy recommendations

The implications regarding the labor supply response are not new. Essentially they suggest a disincentive or a reduction in hours worked of from 5 to 15 percent for a tax rate of 50 percent and a guarantee rate of 75 percent. This should not constitute a reason for failure to enact a national income maintenance program. Presumably there are less labor supply effects under this program than under the current welfare system, where tax rates can exceed 100 percent.

The prime policy recommendation of this study for a national program is that misreporting has more serious cost implications than does work disincentives. For example, the program may have induced a $900 average income under-reporting problem in Iowa, which translates approximately into an average $300 over-payment. However, an average estimate of net farm income is at most $100, which translates into roughly $50 of extra cost.

It also casts considerable doubt on the validity of poverty statistics based upon one time surveys. The difference between the original versus the edited information was substantial. This also may be reflected in the low participation rates for Food Stamps. The number of families that truly are below a certain income cut-off may be lower than what these statistics would indicate, which automatically would lower the participation rates.

BIBLIOGRAPHY

1. Acker, Darrell, et al. "Wisconsin Farm Enterprise Budgets, No. 1, Dairy Cows and Replacements; No. 2, Swine; No. 3, Sheep; No. 4, Beef Cow Herds; No. 6, Cattle Feeding; No. 7, Field and Forage Crops." Department of Agricultural Economics and University Extension, University of Wisconsin, Madison, 1968-1969.

2. Aigner, Dennis. "An Appropriate Econometric Framework for Estimating a Labor-Supply Function from the SEO File." Institute for Research on Poverty Discussion Papers, University of Wisconsin, Madison, 1973.

3. Ashenfelter, Orley and Heckman, James. "The Estimation of Income and Substitution Effects in a Model of Family Labor Supply." Econometrica 42 (January 1974):73-83.

4. Background Papers. The President's Commission on Income Maintenance Programs, 0-376-440. Washington, D.C.: U.S. Government Printing Office, 1970.

5. Barth, Michael C.; Carcagno, George J.; and Palmer, John L. Toward an Effective Income Support System: Problems, Prospects, and Choices. Institute for Research on Poverty, University of Wisconsin, Madison, 1974.

6. Bawden, D. Lee. Administrative Guidelines for Income Maintenance Programs Covering the Self-Employed. U.S. Congress. Joint Economic Committee Studies in Public Welfare, Paper No. 5 Part 3, Implications of the Income Maintenance Experiments 98-110, 93d Congress, 2d Session. Washington, D.C.: U.S. Government Printing Office, 1973.

7. Bawden, D. Lee. "Income Maintenance and the Rural Poor: An Experimental Approach." American Journal of Agricultural Economics 52 (1970):438-441.

8. Bawden, D. Lee. "Implications of a Negative Income Tax for Rural People." American Journal of Agricultural Economics 53 (1971): 754-760.

9. Becker, Gary S. Human Capital. New York, N.Y.: Columbia University Press, 1954.

10. Becker, Gary S. "A Theory of the Allocation of Time." The Economic Journal 75 (1965):493-517.

11. Becker, Gary S. "The Allocation of Time and Goods Over Time." Working paper in the Department of Economics, University of Chicago, June 1967.

12. Borus, Michael E. "Response Error and Questioning Technique in Surveys of Earnings Information." Journal of American Statistical Association 64 (June 1970):566-675.

13. Boskin, Michael J. "The Negative Income Tax and the Supply of Work Effort." National Tax Journal 20 (December 1967):353-367.

14. Boskin, Michael J. "The Economics of the Labor Supply." Research Center in Economic Growth Memorandum No. 100 (revised). Stanford University, November 1970.

15. Cain, Glen. Married Women in the Labor Force: An Economic Analysis. Chicago, Ill.: University of Chicago Press, 1966.

16. Cain, Glen G. and Watts, Harold W., eds. Income Maintenance and Labor Supply: Econometric Studies. Chicago, Ill.: Rand McNally College Publishing Co., c1973.

17. "Comparative Costs and Returns for Livestock." Department of Agricultural Economics, University of Illinois, Champaign-Urbana, 1962.

18. Conlisk, John. "Simple Dynamic Effects in Work-Measure Choice: A Skeptical Comment on the Static Theory." Journal of Human Resources 3 (Summer 1968):324-326.

19. Economic Report of the President. Washington, D.C.: U.S. Government Printing Office, January 1975.

20. Evans, L.T. "Relative Economic Efficiency of Farms." In Final Report of the Rural Income Maintenance Experiment, Vol. 4, Institute for Research on Poverty, University of Wisconsin, Madison, Wisconsin, (forthcoming).

21. Ferber, Robert; Forsythe, John; Guthrie, Harold W.; and Maynes, E. Scott. "Validation of Consumer Financial Characteristics: Common Stock." Journal of the American Statistical Association 64 (June 1969):415-432.

22. Ferber, Robert; Forsythe, John; Guthrie, Harold W.; and Maynes, E. Scott. "Validation of a National Survey of Consumer Financial Characteristics: Savings Accounts." Review of Economics and Statistics 51 (November 1969):436-444.

23. Final Report of the New Jersey Graduated Work Incentives Experiment. Vol. I, II. Institute for Research on Poverty, University of Wisconsin, Madison, Wisconsin and Mathematica, Inc., Princeton, N. J., December 1973.

24. _Final Report of the Rural Income Maintenance Experiment_. Vol. I. Institute for Research on Poverty, University of Wisconsin, Madison, Wisconsin, (forthcoming).

25. _Final Report of the Rural Income Maintenance Experiment_. Vol. II. Institute for Research on Poverty, University of Wisconsin, Madison, Wisconsin, (forthcoming).

26. Finegan, T. A. "Hours of Work in the United States: A Cross-Sectional Analysis." _Journal of Political Economy_ 70, 5 (October 1962):452-470.

27. Friedman, Milton. _A Theory of the Consumption Function_. Princeton, N.J.: Princeton University Press, 1957.

28. Fuller, W. A. "Properties of Estimates in the Errors in Variables Model." Paper presented at the annual meeting of the Econometrics Society, 1971. (mimeographed).

29. Gallaway, Lowell. "Negative Income Tax Rates and the Elimination of Poverty." _National Tax Journal_ 19 (September 1966):298-307.

30. Gallaway, Lowell. "Negative Income Tax Rates and the Elimination of Poverty: Reply." _National Tax Journal_ 20 (September 1967): 338-343.

31. Garfinkel, Irwin. "The Effects of Welfare Programs on Experimental Responses." _Journal of Human Resources_ 9 (Fall 1974):504-529.

32. Garfinkel, Irwin. "On Estimating the Labor Supply Effects of a Negative Income Tax." Institute for Research on Poverty Discussion Paper 101-71, University of Wisconsin, Madison, November 1971.

33. Goldberger, Arthur S. _Econometric Theory_. New York, N.Y.: John Wiley and Sons, Inc., 1964.

34. Goode, Richard. "The Income Tax and the Supply of Labor." _Journal of Political Economy_ 57 (October 1949):428-437.

35. Greene, Christopher. "Negative Taxes and Monetary Incentives to Work: The Static Theory." _Journal of Human Resources_ 3 (Summer 1968):280-288.

36. Griliches, Zvi. "Estimates of the Aggregate Agricultural Production Function from Cross-Sectional Data." _Journal of Farm Economics_ 45 (May 1963):419-428.

37. Hall, Robert E. "Labor Supply and the Negative Income Tax Experiment." Prepared for the Brookings Panel on Social Experimentation, Department of Economics, Massachusetts Institute of Technology, Cambridge, Mass., April 1974.

38. Hannan, M. T. and Young, A. A. "Estimation in Panel Models: Results on the Pooling of Cross-Sections and Time Series." Stanford University, Palo Alto, California, unpublished paper. Circa 1974.

39. Heady, Earl O. "Production Functions from a Random Sample of Farms." <u>Journal of Farm Economics</u> 28 (1946):989-1004.

40. Heady, Earl O. "Resource Returns and Productivity Coefficients in Selected Farming Areas." Journal paper J-1762, Iowa Agricultural Experiment Station, Iowa State University, Ames, Iowa, 1954.

41. Henderson, James M. and Quandt, Richard E. <u>Microeconomic Theory</u>. New York, N.Y.: McGraw-Hill Book Co., Inc., 1958.

42. Hollister, Robinson G.; Lefcowitz, Myron J.; and Orr, Larry L. (Principal Investigators). "A Comprehensive Analysis of the Issues for Research, Demonstration in Income Maintenance." A report of the Institute for Research on Poverty to the Department of Health, Education and Welfare, 18. University of Wisconsin, Madison, 1969.

43. Huffman, Wallace E. "A Cross-Sectional Analysis of Off-Farm Work of Farm Operators and Their Wives." Oklahoma State University, Stillwater, Oklahoma, December 1973.

44. Huffman, Wallace E. "Farm Household Labor Supply and Home Goods Production." Unpublished paper, Iowa State University, Ames, Iowa, 1975.

45. Huffman, Wallace E. "The Productive Value of Human Time in U.S. Agriculture." Unpublished paper, Iowa State University, Ames, Iowa, 1975.

46. James, Sydney C., ed. "Midwest Farm Planning Manual." Ames, Iowa: Iowa State University Press, 1968.

47. Johnston, J. <u>Econometric Methods</u>, 2nd ed. New York, N.Y.: McGraw-Hill Book Company, c1963.

48. Johnston, W. E. and Tolley, G. S. "The Supply of Farm Operators." <u>Econometrica</u> 36 (April 1968):365-382.

49. Kalachek, E. D. and Raines, F. Q. "Labor Supply of Lower Income Workers." In Technical Studies, pp. 159-185. Edited by The President's Commission on Income Maintenance Programs. Washington, D.C.: U.S. Government Printing Office, 1970.

50. Kerachsky, Stuart. "On Farm-Off Farm Work Decisions." In Final Report of the Rural Income Maintenance Experiment, Vol. 4, Institute for Research on Poverty, University of Wisconsin, Madison, Wisconsin, (forthcoming).

51. Kesselman, Jonathan. "Labor-Supply Effects of Income, Income-Work, and Wage Subsidies." Journal of Human Resources 4 (Summer 1969): 275-292.

52. Kmenta, Jan. Elements of Econometrics. New York, N.Y.: The Macmillan Co., c1971.

53. Kosters, Marvin. "Income and Substitution Effects in a Family Labor Supply Model." Rand Corporation Paper P-3339, December 1966.

54. Lampman, Robert. Ends and Means of Reducing Income Poverty. Institute for Research on Poverty Monograph Series. Chicago, Ill.: Markham Publishing Co., c1971.

55. Lansing, John B. and Morgan, James N. Economic Survey Methods. Survey Research Center, University of Michigan, Ann Arbor, Michigan, 1971.

56. Lee, J. E. "Allocating Farm Resources Between Farm and Nonfarm Uses." Journal of Farm Economics 47 (February 1965):83-92.

57. Leuthold, Jane H. "An Empirical Study of Formula Income Transfers and the Work Decision of the Poor." Journal of Human Resources 3 (Summer 1968):312-323.

58. Leuthold, Jane H. "Formula Income Transfers and the Work Decision of the Poor: An Application of a Model of Work-Leisure Choice." Unpublished Ph.D. dissertation, University of Wisconsin, Madison, 1968.

59. McClung, Nelson. "Editing Census Survey Tape Files for Income and Wealth." Unpublished paper, Washington, D.C., August 1972.

60. Madansky, Albert. "The Fitting of Straight Lines When Both Variables are Subject to Error." Journal of the American Statistical Association 54 (March 1959):173-205.

61. Maddala, G. S. "The Use of Variance Components Models in Pooling Cross-Section and Time Series Data." Econometrica 39 (March 1971):341-358.

62. Malinvand, E. Statistical Methods of Econometrics. Amsterdam: North-Holland Publishing Co., 1966.

63. Mathematica, Inc. "Report on the Baseline Survey and Cost Projection." State of Vermont Family Assistance Program Planning Papers Volume V, Prepared Jointly by State of Vermont Family Assistance Planning Unit and Mathematica, Inc., January 1971.

64. Maynes, E. Scott. "The Anatomy of Response Errors: Consumer Saving." Journal of Marketing Research 2 (November 1965):378-387.

65. Merrill, William C. and Fox, Karl A. Introduction to Economic Statistics. New York, N.Y.: John Wiley and Sons, Inc., c1970.

66. Meyer, Charles W. and Saupe, William E. "Farm Operators under the Negative Income Tax." American Journal of Agricultural Economics 52 (1970):255-262.

67. Miller, Herman P. "An Appraisal of the 1950 Census Income Data." Journal of the American Statistical Association 48 (1953): 28-43.

68. Miller, Herman P. Income Distribution in the United States. Washington, D.C.: U.S. Government Printing Office, 1966.

69. Mincer, Jacob. "Labor Force Participation of Married Women." Aspects of Labor Economics, pp. 63-97. Edited by B. G. Lewis. Washington, D.C.: National Bureau of Economic Research, 1962.

70. "Missouri Farm Planning Handbook." College of Agriculture Manual 75, FM 7000, University of Missouri, Columbia, 1972.

71. Morgan, James N., et al. Income and Welfare in the United States. New York, N.Y.: McGraw-Hill Book Co., Inc., 1962.

72. Musgrave, Richard A. The Theory of Public Finance. New York, N.Y.: McGraw-Hill Book Co., Inc., 1959.

73. Nerlove, Marc. "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections." Econometrica 39 (March 1971):359-382.

74. Nerlove, Marc. "A Note on Error Components Models." Econometrica 39 (March 1971):383-396.

75. Nicholson, Walter. "The Income Data Series in the Graduated Work Incentive Experiment: An Analysis of Their Differences." In Final Report of the New Jersey Graduated Work Incentive Experiment, Institute for Research on Poverty, University of Wisconsin, Madison, Wisconsin, and Mathematica, Inc., Princeton, N.J., December 1973.

76. Ostle, Bernard. Statistics in Research, 2nd ed. Ames, Iowa: The Iowa State University Press, c1963.

77. Perlman, Richard. "A Negative Tax Plan for Maintaining Work Incentives." Journal of Human Resources 3 (Summer 1968):289-299.

78. Rao, Potluri. "Some Notes on the Errors-in-Variables Model." The American Statistician 27 (December 1973):217-218.

79. Rees, Albert. "An Overview of the Labor-Supply Results." Journal of Human Resources 9 (Spring 1974):158-180.

80. Reese, R. B. "Estimating Nonparticipation in the Food Stamp Program." Unpublished paper, Economic Research Service, USDA, Washington, D.C., 1972.

81. Reinsel, Edward I. "Farm and Off-Farm Income Reported on Federal Tax Returns." Economic Research Service, U.S. Department of Agriculture, Washington, D.C., Circa 1969.

82. Rosen, Sherwin and Welch, Finis. "Labor Supply and Income Redistribution." Review of Economics and Statistics 53 (August 1971): 278-282.

83. Rosen, Sherwin and Welch, Finis. "A Note on the Estimation of Labor Supply Effects and Income Redistribution Policy." Journal of Human Resources 7 (Winter 1972):104-109.

84. Sirken, Monroe C.; Maynes, E. Scott; and Frechtling, John A. "The Survey of Consumer Finances and the Census Quality Check." In National Bureau of Economic Research, An Appraisal of the 1950 Census Income Data Volume 23 of Studies in Income and Wealth. Princeton, N.J.: Princeton University Press, 1958.

85. Stone, R. "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand." Economic Journal 64 (September 1954):511-527.

86. Stuart, Walter J. "Computer Editing of Survey Data--Five Years of Experience in BLS Manpower Surveys." American Statistical Association Journal 61 (June 1966):375-383.

87. Tintner, Gerhard and Brownlee, O. H. "Production Functions Reviewed from Farm Records." Journal of Farm Economics 26 (1944): 566-571.

88. U.S. Bureau of the Census. Evaluation and Research Program of the U.S. Censuses of Population and Housing: Record Check Study of Accuracy of Income Reporting, Series ER 60, 8. Washington, D.C.: U.S. Government Printing Office, July 1970.

89. U.S. Congress. Joint Economic Committee Income Security for Americans: Recommendations of the Public Welfare Study, 93rd Congress, 2nd Session. Washington, D.C.: U.S. Government Printing Office, 1974.

90. U.S. Congress. Joint Economic Committee Studies in Public Welfare Paper No. 14, Public Welfare and Work Incentives: Theory and Practice, 93rd Congress, 2nd Session. Washington, D.C.: U.S. Government Printing Office, 1974.

91. U.S. Department of Labor, Bureau of Labor Statistics. A Micro Model of Labor Supply. BLS Staff Paper 4. Washington, D.C.: U.S. Government Printing Office, 1970.

92. Van Arsdale, R. FM 20 and AE 4112, Department of Agricultural Economics, University of Illinois, Urbana, 1962 and 1965.

93. Warren, R. D.; White, J. K; and Fuller, W. A. "An Errors-in-Variables Analysis of Marginal Role Performance." Journal of the American Statistical Association 69 (December 1974): 886-893.

94. Watts, Harold W. "Adjusted and Extended Preliminary Results from the Urban Graduated Work Incentive Experiment." The Institute for Research on Poverty Discussion Paper No. 69-70 (revised 1970).

95. Watts, Harold W. "Graduated Work Incentives: An Experiment in Negative Taxation." American Economic Review 59 (May 1969): 463-472.

96. Watts, Harold W. "The Iso-Prop Index: An Approach to the Determination of Differential Poverty Income Thresholds." Journal of Human Resources 2 (Winter 1967):3-18.

97. Watts, Harold W. "Mid-Experiment Report on Basic Labor-Supply Response." Institute for Research on Poverty Discussion Paper No. 98-71, University of Wisconsin, Madison, May 1971.

98. Willet, G. C. and Saupe, W. E. "Production Practices of Wisconsin Dairy Farmers." Research Report R2479, College of Agricultural and Life Sciences, University of Wisconsin, Madison, 1973.

99. Zellner, A.; Kmenta, J.; and Dreze, J. "Specification and Estimation of Cobb-Douglas Production Function Models." Econometrica 34 (October 1966):784-795.

ACKNOWLEDGEMENTS

Many individuals have assisted me in the preparation of this dissertation. The intellectual stimulation and patience of Charles Meyer enriched my entire graduate education and his direction was essential to the successful completion of this manuscript.

I have also depended upon the advice and profited from the constructive criticism of the RIME staff, namely Lewis Evans, Stuart Kerachsky, Bill Harrar and William Saupe. A major debt is owed to D. Lee Bawden, principle investigator of the experiment and whose guidance was instrumental in the formulation of this study.

A special thanks is also due to Michael Boehlje and Wallace Huffman who read the final draft and provided several valuable comments. Wayne Fuller, who also read a final draft, provided valuable assistance as a sounding board on data editing techniques and their justification during the early stages of the study.

I would also like to acknowledge the support given to me by staff of the Institute for Research on Poverty. They are too numerous to mention but performed well the tedious task of data collection and processing.

A special word of thanks is also due Betty Ingham who deciphered my writing and provided valuable clerical assistance in completing this thesis.

I cannot let go unmentioned a major debt to my wife, Virginia R. Primus, whose editorial assistance and confidence made this thesis possible.

APPENDIX A. DESCRIPTION OF THE DATA

Table A.1. Iowa sample averages for selected characteristics of farms included in the rural income maintenance experiment by year

	Experimentals				Controls			
	1969	1970	1971	1972	1969	1970	1971	1972
Sources of income:								
Wage income	443	879	750	972	521	757	940	1095
Net business income	6	46	60	28	110	11	235	195
Net farm income - surveys	2573	3185	3945	6188	2564	3531	3984	7131
Net farm income - IRS	4590	2774	2879	4933	4245	2725	3134	6214
Unearned income[a]	70	-582	-602	-649	8	-554	-579	-502
Unusual income[b]	0	73	537	648	0	32	299	452
Total income	3092	3601	4690	7187	3202	3777	4880	8372
NIT payments - actual	N.A.	1397	1552	1652	N.A.	0	0	0
NIT payments - predicted	N.A.	1024	957	593	N.A.	0	0	0
% with wage income	0.42	0.68	0.72	0.72	0.46	0.69	0.76	0.76
Farm summary:								
Total tillable acres	217	229	228	217	202	213	216	207
Diverted acres	28.3	25.4	47.6	35.8	25.9	24.6	44.3	36.7
% with diverted acres	0.88	0.84	0.94	0.96	0.85	0.85	0.98	1.0
Total crop sales	N.A.	8346	10097	11914	N.A.	6846	8960	9369
Total livestock sales	N.A.	7223	6934	9047	N.A.	11269	10661	15224
Fertilizer, sod, insecticide	N.A.	2082	2328	2455	N.A.	2006	2099	2245
Fuel, repair, machine hire	N.A.	2787	2682	3041	N.A.	2573	2795	2936
Feed, vet., breeding fees	N.A.	361	442	439	N.A.	306	326	334

[a]Includes welfare income, income from assets, and retirement benefits minus childcare, alimony, child support, and medical expenses.

[b]Includes one-time, lump sum payments (e.g., large gifts, inheritances, and windfall income).

Table A.1. Continued

	Experimentals				Controls			
	1969	1970	1971	1972	1969	1970	1971	1972
Assets:								
Farm land and bldg. - value	5460	5500	5558	5490	9759	9907	13370	16051
Farm land and bldg.-debt	2043	1963	1931	1951	4341	4565	6852	8920
% owning land	0.12	0.12	0.14	0.14	0.20	0.20	0.26	0.30
Farm machinery and eqt. - value	10727	11209	12015	11889	11971	12555	13711	13879
Farm machinery and eqt.-- debt	2936	3203	3601	2720	2777	2330	2930	2615
Other farm assets - value	8764	10966	11144	13565	14815	17923	16564	21723
Other farm assets - debt	2309	20093	21019	22493	3775	28613	29468	32937
Total farm value	24950	27675	28717	30944	36545	40386	43646	51654
Total farm net equity	17662	20093	21019	22493	25653	28613	29468	32937
Liquid assets	2890	2741	2594	3174	2366	2181	2423	2222
Total other net equity	0	92	104	97	518	535	601	712
New purchases - mach. and eqt.	N.A.	2275	2689	4028	N.A.	2127	1955	4042
Hours:								
Head								
Scaled crop	1110	1143	1147	1090	1060	1104	1115	1077
Scaled livestock	821	732	678	698	974	991	991	891
% with livestock	0.88	0.86	0.80	0.82	0.93	0.96	0.94	0.96
Total scaled	1931	1875	1825	1788	2033	2095	2107	1968
Farm recall	N.A.	2617	2351	2369	N.A.	2605	2304	2461
Wage	135	160	132	127	129	162	171	178
Business	N.A.	1	1	1	N.A.	2	4	1
Spouse								
Unpaid farm or bus.	N.A.	215	163	189	N.A.	98	156	129
Wage and business	21	53	74	122	59	84	116	141

Table A.1. Continued

	Experimentals				Controls			
	1969	1970	1971	1972	1969	1970	1971	1972
Other members								
Wage	86	165	68	76	34	96	133	188
Unpaid	N.A.	0	94	97	N.A.	0	61	217
Age of head	42.6	43.6	44.6	45.6	43.5	44.5	45.5	46.5
Family size	4.6	4.6	4.5	4.6	4.7	4.7	4.8	4.7

Table A.2. North Carolina sample averages for selected characteristics of farm included in the rural income maintenance experiment by year

	Experimentals				Controls			
	1969	1970	1971	1972	1969	1970	1971	1972
Sources of income:								
Wage income	1251	1759	1976	1977	2073	2680	3151	3948
Net business income	4	9	32	49	113	45	-11	131
Net farm income - surveys	1212	2427	1941	2820	1005	2523	2429	2892
Net farm income - IRS	1751	1552	1397	2054	1479	1303	952	2517
Unearned income[a]	118	-282	-194	-145	84	-257	-317	-269
Unusual income[b]	0	8	603	232	0	0	3	0
Total income	2584	3921	4357	4933	3275	4991	5254	6702
NIT payments - actual	N.A.	1867	1578	1725	N.A.	0		0
NIT payments - predicted	N.A.	919	967	767	N.A.	0	0	0
% with wage income	0.69	0.85	0.81	0.73	0.81	0.91	0.83	0.83
Farm summary:								
Tobacco acres	4.42	4.85	4.37	4.71	5.14	6.38	5.60	5.62
Total crop sales	N.A.	5292	5349	6555	N.A.	7109	7274	8280
Total livestock sales	N.A.	1298	1369	1328	N.A.	1549	1313	1598
Fertilizer, seed, insecticide	N.A.	659	577	483	N.A.	625	624	349
Fuel, repair, machine hire	N.A.	1009	1093	1160	N.A.	1311	1179	1258
Feed, vet., breeding fees	N.A.	625	624	349	N.A.	659	577	483
Hired labor	N.A.	1185	1061	1154	N.A.	1694	1450	1706

[a]Includes welfare income, income from assets, and retirement benefits minus childcare, alimony child support, and medical expenses.

[b]Includes one-time, lump sum payments (e.g., large gifts, inheritances, and windfall income).

Table A.2. Continued

	Experimentals				Controls			
	1969	1970	1971	1972	1969	1970	1971	1972
Assets:								
Farm land and bldg. - value	6274	6500	7267	7479	7098	7296	7956	8444
Farm land and bldg.-- debt	1449	1273	1246	1368	2231	2200	2609	2581
% owning land	0.54	0.54	0.58	0.60	0.55	0.57	0.57	0.57
Farm machinery and eqt. - value	1588	2414	2468	2507	1992	2524	2820	3242
Farm machinery and eqt. - debt	679	684	643	513	612	927	903	1008
Other farm assets - value	828	983	1612	1448	572	688	546	578
Other farm assets - debt	72	61	20	1016	211	42	25	734
Total farm value	8690	9897	11347	11434	9662	10509	11322	12263
Total farm net equity	6490	7879	9439	8537	6608	7339	7785	7941
Liquid assets	194	330	195	889	459	316	249	294
Total other net equity	244	594	640	758	907	795	1059	1255
New purchases - mach. and eqt.	N.A.	367	386	821	N.A.	868	1108	1123
Hours:								
Head								
Scaled crop	1355	1523	1472	1538	1457	1950	1861	1812
Scaled livestock	129	168	173	175	102	129	133	123
% with livestock	0.58	0.62	0.62	0.60	0.55	0.62	0.57	0.62
Total scaled	1484	1691	1645	1713	1559	2079	1994	1934
Farm recall	N.A.	1792	1623	1629	N.A.	1816	1597	1210
Wage	333	317	316	242	464	370	461	631
Business	N.A.	1	1	2	N.A.	3	4.8	11

Table A.2. Continued

	Experimentals				Controls			
	1969	1970	1971	1972	1969	1970	1971	1972
Spouse								
Unpaid farm or bus.	N.A.	360	412	389	N.A.	389	417	291
Wage and business	202	293	309	290	417	397	478	543
Other members								
Wage	71	197	194	314	99	342	356	405
Unpaid	N.A.	0	74.8	52.8	N.A.	0	95	137
Age of head	48.3	49.3	50.3	51.3	48.4	49.4	50.4	51.4
Family size	4.32	4.35	4.19	4.21	4.55	4.19	3.95	3.88

APPENDIX B. DESCRIPTION OF EDITING TECHNIQUES

General Description of Coding Techniques

Introduction

The achievement of a high quality data base has been the single most important criterion governing the behavior of the coding and data processing offices. The construction of a data base where the coefficients on the treatment variables would be least biased by processing and coding techniques was a guiding principle. Because the staff lacked omniscience, this ideal was not perfectly achieved; and the purpose of this appendix is to provide an overview of the techniques and procedures used in the data processing office for assuring data quality. To put the rest of this appendix in proper focus, a list of the initial premises or tenets of the data manager (the author) is presented.

First, given the quantity of data to be collected from the families and the unavoidably high prevalence of response errors in the data, the data entry procedure was not to be checked on a step by step basis. In other words, respondents' errors would not and could not be removed fully prior to coding. There seemed little benefit in devising a complex checking mechanism to insure that the data base agreed perfectly with the interviews given the large number of interviewer and respondent errors that would go undetected. Rather, it was thought that the coded data should be placed onto the data base with a minimum of errors, and then the whole process from the field to a printed copy of constructed analytical variables constructed from the data base would be checked. There is little merit in tracing down errors between interviews if there is no way of dealing with them effectively once they are discovered. For example, presume that in an earlier part of the year, a respondent claimed that he owns acres and yet in the asset evaluation section there is no mention of any farm land. Why spend the time detecting the inconsistency if no resolution to the problem exists?

A second tenet of the data processor was to keep boring, tedious clerical work to a minimum, and wherever possible let the computer do the work. The physical transcription of the data is necessarily a rather boring process. To assign someone to copy the telephone book perfectly brings no lasting reward. If one decides to recheck this transcription process, not too many errors would be found because the checker would become lazy and, consequently, would not find as many errors as really existed. A corollary to keeping boring work to a minimum is that human beings make mistakes, and the greater number of clerical hours needed, the larger will be the number of errors.

A third requirement was a strict insistence that no missing information in critical data fields be recorded onto the data base. It was a policy throughout the data collection apparatus that all information related to income, assets, and family composition be declared not missing. If the respondent could not give an estimate of his machinery value or remember the number of hours he worked last week, the respondent was forced to make a guess or present other information that allowed a good approximation to be estimated for the desired answer. For example, if machinery value was unavailable, the field would obtain numbers, kind and age of the different pieces of machinery and an estimation would be derived in the office. Schedule F's at the end of the year occasionally helped remove missing information. If that failed in the field, the office would fabricate the answer based on past information. These fudges, which occurred infrequently, were recorded and are available to researchers, both with this project and for future use. The rationale for this procedure was that the researcher could not be preoccupied with missing data techniques on each of 300 variables. Missing value techniques work well on a large number of observations and few variables and not vice versa. An educated guess for variables like income or assets based upon related variables supplied by the respondent may be better than any of the statistical techniques for massaging or correcting missing data, especially considering the wealth of information that is available.

A fourth demand was that everyone in the organization at all levels of operation should have a good understanding of the goals of the research project and a broad picture of how these goals would be achieved. This was done primarily to motivate coders and programmers into a higher sense of work value so, hopefully, they in turn would do a better job. This led to a rejection of the editor concept of coding an interview where first an editor would read over the entire interview and make certain value judgments and a coder would merely transcribe the data during a second operation. On this project transcription of the data, which is done by the coder, is handled by the same individual who edits the interview. Editing implies checking to insure that the respondent answered the questions according to what the question intended. For example, several times farmers would lump business with farm work, saying their net farm income was $1100, where in reality they had a net farm income of $1300 and a trucking business on the side that had a net income of -$200. Those would be two separate enterprises and, for our research purposes, should be treated separately even though for accounting reasons, they were lumped together by the respondent.

The respondent was urged by the interviewing staff to differentiate between the two businesses. Having two businesses causes difficulty in collecting the information, but separating the two according to type of expense was properly the job of the editors based on notes from the interviewer. Other functions would be to question outliers, check consistency of information between interviews, and code open-ended questions.

Chronological development of coding procedures

With this viewpoint in mind, a brief chronological order of the coding operation will be explained. The data as it initially came from the field had been checked to insure that all data entries on the interview were asked correctly, that the interview itself contained no missing information, and that the interviewers had followed the proper skip logic. See [24, Chapter 7] for more details. An appropriate coding instruction manual was defined, along with some editing rules, and the coder-editor (one person) was to look for inconsistencies in the data as the information was transcribed. For example, if in the earlier example assets were collected in one section and acres owned were in another section of the same interview, the editor insured that both values were positive or both values were zero.

Initially, a large amount of faith was placed in the syntaxing operation. Syntaxing implies that as the coded cards were processed onto the main data base the cards were checked by computer programs and had to obey logical skip patterns and range checks. For example, if a respondent had two children, the syntax program insured that two children's data, (e.g. age, sex, education, etc.), would be entered onto the data base. If the respondent was not a wage earner, the wage earner section would get skipped and everything about the wage earning section should be zero or designated as nonapplicable. If the respondent rented the house, the rental arrangement section would be answered and the home owner section skipped. The logical pattern of the interview was followed in the coding operation, and a program insured that it was perfectly obeyed. If it wasn't, it was brought to the attention of an editor and corrected before placing the data onto the data base. Range checks allowed only values within a certain range to be placed on the data base. For example, if a question could be answered with only a yes or a no, the program assured that no other answer except a yes or a no was recorded. The fallacy with this approach was that very little checking was done across interviews; also there were a number of ranges such as number of hours worked last week, which could vary essentially from zero or one hour per week all the way to 99. Since this was a two digit field essentially no checking occurred. The latter case is an extreme example, but there were several instances where a housewife tended the laundromat next door and assured us on several occasions that she worked from 7 a.m. to 10 p.m. seven days a week. That yielded a very high number of hours worked per week at a very low wage rate.

Development of initial edit sheets

The syntax program could not be expanded to perform these edit checks across interviews. Separate and independently written edit programs were a better approach to the problem. When editors were doing their job well and comparing interviews across time, many problems were appearing which could not be handled easily. There were many gross inconsistencies between interviews that were inexplicable. A policy was established that whenever possible a problem resolver was written and sent back to the field for resolution, either with the next quarterly interview or on a special call back basis. Over 600 of these were sent back, with most of the discrepancies concerning income, assets, and family composition. Because of this increasing awareness of sloppy reporting, a new procedure was initiated whereby pre-printed forms from previous interviews would be sent to the field as the current year's information was gathered. For example, the loan sections from pre-enrollment and third quarter were pre-printed so that when the interviewer asked about loan sections on the seventh quarter, the previous year's information was close at hand. If a new loan was mentioned that was over one year old, this was politely brought to the respondent's attention and resolved. If one of the huge loans on an earlier year no longer appeared, the respondent was probed and asked specifically about the loan in question. This resulted in a large number of corrections to the data base.

A serious mistake was made probably in that this was not done more frequently. In this setting the initial edit sheets were created with the wage edit sheet being done first. The wage rate of the head, hours of the head, and wage income of the head for each quarter were printed on top of each other. The primary guideline used in editing this information was the wage rate between quarters. This procedure involved going back to the interviews a large number of times, and where there were real problems, there was no way of resolving them. For example, if the respondent said his wage rate was one dollar for one quarter, three dollars the next, and two dollars the next, there was no good nor effective way of resolving these apparent discrepancies. Furthermore, the interviewer was going back to the files a large number of times because relevant information had not been printed on the edit sheet. These trips back to the files could have been eliminated if it had been noted that a true job change had taken place, or that the respondent had been laid off, was sick, or had worked a considerable amount of overtime in a particular quarter.

Description of General Editing Techniques

Despite the quality control measures described earlier both in the field and in the coding office, it became increasingly clear that there were large inconsistencies in the information for one family over time and between different data sources. It also became clear that these discrepancies existed on a large number of families. Different data sources were the Social Security information as reported to the Baltimore, Maryland office and the W-2 and other Internal Revenue Service information supplied by the respondent. Presumably this is identical to information supplied to IRS directly from the respondent's employer. Inconsistencies are defined as situations in which certain pieces of evidence in the interview support one estimate of a respondent's earnings, while other data items support a different (i.e., 20 percent variation in estimate or in some cases a \$1000 or more difference) estimate of earnings. Confronted with these realities and fully conscious that the resolution of them could affect experimental results, the following possibilities were examined. The first was resolving no data discrepancies except those arising from processing errors and allowing statistical techniques to arrive at an appropriate estimate of a respondent's earnings.

The chief arguments for this approach were that no human bias in making editorial decisions would be injected into the data and that a large number of statisticians and econometricians would favor this technique.

The chief argument against this technique was that much information was lost by using statistical techniques. For example, in a family where the entire income came from wages, and the IRS estimate and the Social Security estimates were roughly equivalent and yet survey data was \$1000 lower caused by a mysterious reporting of a one dollar lower wage rate and a loss of two weeks work in one out of the four quarters, the weight of evidence favors the estimate of income that is the higher of the two. Probably no other data collection process in the United States has ever had access to the amount, kind, and detail of data that this experiment enjoyed. The older techniques may not be optimal when considering this vast detail of information.

Consequently, a second possibility was to select a definition of income which by some criteria was judged best. However, one serious criticism of this approach could be raised. In the analysis phase, many more variables besides income (e.g. hours, wage rates) would be examined. If the IRS version of wage income was deemed to be correct yet the quarterly estimate of hours times wage rate was not changed to coincide with this estimate, analysis that did not focus on income may be led astray.

A third possibility was to select on the basis of some arbitrary rule those cases that were the most inconsistent. On these cases income, hours, wage rates and other detail could be changed to agree with the best estimate of wage income. The chief criticism of this method is what role the arbitrary rule has in subsequent analysis. For example, presume that payment reporting is unbiased with a small variation from the truth, while control reporting is essentially unbiased but with a wide variation. If the rule detected only those X percent above the truth, and 3X percent below the truth, the result of the editing would change control estimates downward relative to changes made to payment families. This would ultimately cause a negative bias in treatment coefficients estimated by ordinary least squares (i.e., it would understate any experimental effect that might exist).

When considering the merits of the above three alternatives a fourth possibility emerged by selecting the best features of these alternatives. One set of data referred to hereafter as the original data base, was made to coincide exactly with what the interview said, i.e., all processing errors would be removed. Also some obvious interviewer errors were removed. A second set of data, hereafter referred to as the edited data base, was constructed from the original data base by removing all data inconsistencies. Data inconsistencies were resolved in a way that required the least number of changes in the data, adjusting for the quality and assumed bias of the source of each data item. For example, W-2 forms were rarely changed in a downward direction, however W-2 forms could be increased because of in-kind payments and also because respondents may have had wage income from employers who did not have to file W-2's. Whenever appropriate and where respondent cooperation allowed, a contact was made with the respondent to allow large discrepancies to be resolved. Essentially this solution allowed for an evaluation of the impact of editing at some later date. Furthermore, independent researchers who might be skeptical of the value judgments introduced into the data by RIME staff, could rely upon statistical techniques and work primarily with the original data base.

Given the above rationalization there existed several ways of implementing editing. The method chosen involved printing information from the data base in a neatly labeled and logical sequence. In addition several obvious errors were detected by the computer and printed for correction. This later technique, however, was not used on a wide scale primarily because of the small number of families. Initially it was thought to be infeasible and economically inefficient to program a large number of edit checks given the number of families in our sample. A trained staff of seven to eight individuals went through roughly a ten page printed format for each family. Upon encountering an inconsistency or two different estimates of the same variable they first checked to determine whether the error was due to processing. These processing errors were corrected both on the original and the edited data base. If the error turned out to be a true logical inconsistency, all possible data or evidence was brought to bear on the inconsistent data relationship. The editors would suggest what corrections they thought best to make. These were then reviewed by

Bill Harrar or Wendell Primus, and the change okayed by these individuals would be placed on the edited data base. Roughly seven man years were involved in the editing process.

At the time editing began, the coding staff was quite small. New staff were hired, rather than coders being promoted into that position. These newly hired individuals went through an extensive training period. One day was involved in acquainting them with the project, the purpose of the experiment, the design of the interview, and quality control procedures. The second and third days were spent explaining the design and use of the coding instruction manual, the organization of the data base, and teaching how to code interviews for one family on a typical quarterly interview. The next several days were spent practice coding under close supervision, and learning about information collected in each of the different quarters and their individual idiosyncrasies. After these individuals had mastered the coding process, they were lectured on the source of errors in surveys.

Computer printouts of previously coded information were presented and explained. From these printouts certain error checks were defined, and an idea given on how to detect errors on the printout. Administrative procedures were also discussed some time during the training period which took place over a three week period.

In essence, the administrative procedures adopted were quite simple. The following items deserve some mention:

1) Along with the editing, certain information was coded. Primarily, information that had been intentionally skipped in the ninth quarter interview, identification of which data contained the

best estimate of income, and certain other miscellaneous information was coded.

2) The families were divided by region with Bill Harrar making all edited data base decisions for North Carolina and Wendell Primus making all decisions relating to Iowa. The families were further subdivided into farmers, wage earners, and one other category containing attritions, other adults and new filers. This allowed some specialization by editors and permitted the wage analysis to commence at an earlier date.

3) Because many editors were not familiar with farming, several interviewers in Fort Dodge were selected to do the Iowa Farm Edit. Bill Harrar assumed the burden of the North Carolina farm edit with clerical assistance.

4) The editing was done in two phases. The first two years of experimental data were done first, and it was at this time that information from quarter nine was coded. Some previous edit work had been done prior to this time, as alluded to in the introduction. This provided information on how to proceed in these two final edits. Information from these earlier edits enabled the computer printout to be designed efficiently, and guided certain other administrative decisions. The last edit was of a slightly different nature. It involved checking to insure that corrections in the earlier edits had been done correctly while primarily editing third year information. After this edit and resulting corrections, printout was obtained for every family to insure

that corrections made during the final (third year) edit phase were done accurately.

5) The editing of one family was done by one editor. However, the farm section was edited independently of the other sections. While not allowing for specialization, one editor became completely familiar with a family's profile and could make appropriate cross checks between the different sections.

The rest of this appendix will present one sample family, an Iowa farmer. Only the computer printouts for the farm section and asset section will be presented. Along with these computer printouts a brief description of the printout, editing guidelines used by the editors, abbreviations used in the printout, and examples of how inconsistencies in the data and between the different sources of data were resolved are provided.

A Specific Example: The Farm Section

Description

The farm section is comprised of five different sections. The first section prints in matrix form detailed income and expense items by quarter for both the first and second data bases. It also shows itemized schedule F's when available.

At the end of this section is a summary listing showing total gross income, expenses, depreciation, interest, and net farm income by year from both the survey data and IRS. Section two shows accounting equations which predict sales of grain from a farmer's acres, yield per acre, and an average imputed price. On the right hand side of the equation is the actual grain sold from a year's crop, inventory on hand, and the amount fed to livestock. Section three is a similar section for livestock. At the end of each year, the farmer reported the number of each type of animal sold. This times an average price was compared to the actual sales during the year. Near the bottom of this page are the inventories on hand for each type of livestock at quarters one, five and nine. Section four is an accounting equation estimating, usually from acres and a predicted amount per acre, the expenses incurred throughout the year. This is then compared with the actual expenses. The final section entitled Farm Operation Edit Sheet shows rental arrangements by land parcel for each year. In addition, it shows in summary form the number of acres grown each year and the number of livestock sold by type each year.

Editing guidelines

1) Net farm income between IRS and quarterly should be within ten percent or $300, whichever is greater adjusting for some conceptual differences between the two estimates. This check was used primarily as a signal that something might be astray. Income and expense items were not necessarily changed to agree with the check.
2) Total predicted sales should be within total production by ten percent or $500, whichever is greater.
3) Total predicted livestock sales should be within ten percent or $500 of actual livestock sales, whichever is greater.
4) Large changes in rental arrangements between years were investigated.

5) Shifts in production of more than 20 percent in dollar terms between years were investigated.

6) The edit checks that were possible on the farm were innumerable given the detail of our data. The main checks are described above. Essentially these involved resolving differences between IRS and survey data, checking the consistency of the data via accounting equations, and examining large changes in behavior over time.

Example

Presume that IRS gross farm income is $4000 more than the quarterly interviews. It might be added parenthetically that amounts of this magnitude were not unusual. First, an examination would be conducted to determine which income type caused the bulk of the discrepancy. Typically this might be more than one type, but assume for this example that most of the discrepancy was due to a difference in cattle sales. Next the editor insured that there were no coding mistakes in recording cattle sales. If no coding mistakes were found, all the evidence relating to cattle sales would be accumulated.

This would involve looking at feed expenditures, the dollar value of cattle purchases, the number sold, an indication whether this was more, less or equivalent to the previous year's sales for the year in question, and sales during the preceding and following years.

The estimate that would be recorded on the edited data base would be the one that was most consistent with the other evidence. Certain value judgments about the different sources of data were also presumed. For example, values given to IRS on the income side were generally judged to be lower bounds. Outside of clerical errors, not many reasons exist why a farmer would overreport income to IRS. The number of cattle sold per year was judged to be fairly accurate, particularly if a farmer sold only one load of cattle during the year. This is contrasted with an estimate of hogs sold, where typically many different sized bunches could be sold during a year's time. In this case the number reported at the end of the year was presumed to be less accurate.

Another example of an inconsistency is the following. Presume IRS corn sales are \$10,000, quarterly sales are \$13,000, and predicted sales are \$20,000. In this case, coding or processing errors would be investigated first. Assuming coding errors did not exist, investigation was made to determine whether misclassification of soybean or other grain sales had occurred. Also information was gathered that determined whether the grain could still be on hand or had been fed to livestock. This was determined by comparing cattle sales minus cattle purchases with feed costs. A record of sales after the inventory date of March 1 provides some information about inventory value. In many instances discrepancies were resolved by increasing the inventory on hand. If all of these areas, along with an investigation of possible crop damage, produced nothing, the sale values would be changed to approximately \$18,000. The above situation happened infrequently. Usually a discrepancy could be resolved by some combination of the activities mentioned above.

Abbreviations

Q1(69), Q1(70):	because quarters overlapped year boundaries, an effort was made to split the quarter into the appropriate year
Q1(69):	refers to the income and expenses allocated to 1969.
IRS:	Internal Revenue Service
SCH_F:	itemized schedule of income and expenses used by farmers in reporting to IRS
IRS_D:	a coder check to determine whether Schedule F was coded properly
PSALES:	predicted sales
INV:	inventory as of March 1 for each year in question
STOCK ADJ:	value of corn that was fed to livestock; unless sold, all hay and oats were assumed fed to livestock
SHR:	share
PRFT:	profit
TOB:	tobacco

Table B.1. Income and expense summary for household X: first data base

Description	Q1(69)	Q1(70)	Q2	Q3	Q4&Q5	Total	SCH F	Q5(71)	Q6	Q7	Q8	Q9(71)	Total	SCH F
<u>Income</u>														
Cattle (including dairy), horses	0	963	0	0	1381	2344	2344	0	650	0	0	180	830	3406
Swine (hogs)	529	0	0	1755	336	2091	1755	954	2155	1952	977	0	6038	6654
Corn, soybean, other grain, gener.	3364	5027	159	1401	7087	13674	9409	0	536	548	2641	0	3725	3725
Custom work	302	0	0	0	281	281	325	201	0	0	300	440	941	941
Dividends (mainly patronage)	0	0	0	0	0	0	0	32	0	0	0	0	32	32
Farm commodity loan	0	0	0	0	2018	2018	4700	2373	0	0	230	2208	4811	6600
Acre div. payments, land conser.	0	0	0	1308	0	1308	1308	0	0	1396	33	0	1429	1658
Gas tax refunds (state and fed.)	129	132	108	216	18	474	504	0	213	131	39	0	383	345
Other	0	0	0	0	0	0	0	0	0	77	0	0	77	877
Crop, livestock insur. claims	0	0	0	0	0	0	0	119	0	0	0	0	119	0
Totals less live-stock purchased	4324	6122	267	4680	11121	22190	20345	3679	3504	4104	4220	2828	18335	24238

First quarter total = 10446 Fifth quarter total = 11917

Table B.1. Continued

Description	Q1(69)	Q1(70)	Q2	Q3	Q4&Q5	Total	SCH F	Q5(71)	Q6	Q7	Q8	Q9(71)	Total	SCH F
<u>Expense</u>														
Labor hired. Soc. Sec. hired help	0	0	75	0	137	212	333	0	0	311	242	52	605	605
Repairs, maint.	0	158	425	0	590	1173	1876	512	188	111	436	230	1477	1477
Interest	68	70	0	0	0	70	1101	0	318	0	4	983	1305	1308
Rent-Farm/bus. pasture, other	615	0	0	0	615	615	615	0	0	0	0	615	615	615
Feed purchased	101	370	286	8	1585	2249	2286	1065	684	2032	261	769	4811	4811
Seed, plants, insecticides, etc.	0	0	208	282	0	490	280	0	74	0	0	532	606	606
Fertilizers, lime misc.	404	1074	0	0	1231	2305	1892	621	0	0	0	0	621	2075
Machine hire	72	440	111	148	0	699	509	0	106	6	0	284	396	665
Supplies purch.	0	144	75	0	383	602	515	0	57	44	127	49	277	254
Vet.-livestock medicine	0	9	10	112	60	191	401	102	43	34	110	396	675	675
Gas, fuel, oil	114	932	343	98	766	2139	1872	613	167	9	574	652	2015	2015
Taxes (state and fed.)	0	0	0	0	30	30	0	307	64	0	0	0	371	0
Insurance (farm not personal)	0	0	138	0	27	165	348	0	0	186	0	0	186	136
Utilities (farm share only	22	56	77	81	126	340	292	24	115	60	66	29	294	349
Freight, truck and pickup costs	0	0	0	0	318	318	404	0	65	114	179	0	358	0
Tax-not state, fed., soc. sec.	0	0	30	0	0	30	60	0	6	0	0	0	6	6
Car, legal, bank other bus. costs	0	0	60	4	294	358	1050	0	9	59	34	222	324	1519

Table B.1. Continued

Description	Q1(69)	Q1(70)	Q2	Q3	Q4&Q5	Total	SCH F	Q5(71)	Q6	Q7	Q8	Q9(71)	Total	SCH F
Poultry, sheep, swine purch.	0	0	0	0	0	0	0	0	50	0	0	0	50	0
Depreciation	0	0	0	0	0	0	1508	0	0	0	0	0	0	2512
Totals less livestock purch.	1328	3183	1838	733	6132	11886	12733	2937	1514	2966	2029	3820	13266	15809

First quarter total = 4410 Fifth quarter total = 6726

	1970 Quarterly	1970 IRS	1971 Quarterly	1971 IRS	1971 IRS_D
Gross farm income less livestock purchased	22190	20811	18335	24267	24238
Farm expenses except interest and depreciation	11886	12849	13266	15808	15808
Depreciation	1507	1507	2512	2512	2512
Interest	1101	1101	1308	1308	1308
Net farm income	7696	5354	1249	4639	4610

Table B.2. Income and expense summary for household X: second data base

Description	Q1(69)	Q1(70)	Q2	Q3	Q4&Q5	Total	SCH F	Q5(71)	Q6	Q7	Q8	Q9(71)	Total	SCH F
<u>Income</u>														
Cattle (including dairy), horses	0	963	0	0	1381	2344	2344	0	650	2451	0	180	3281	3406
Swine (hogs)	529	0	0	1755	336	2091	1755	954	2155	1952	1593	0	6654	6654
Corn, soybean, other grain, gener.	3364	5027	159	1401	5456	12043	9409	0	536	548	2641	0	3725	3725
Custom work	302	0	0	0	325	325	325	201	0	0	300	440	941	941
Dividends (mainly patronage)	0	0	0	0	0	0	0	32	0	0	0	0	32	32
Farm commodity loan	0	0	0	0	2018	2018	4700	2373	0	1789	230	2208	6600	6600
Acre div. payments, land conser.	0	0	0	1308	0	1308	1308	0	0	1396	262	0	1658	1658
Gas tax refund (state and fed.)	129	132	108	216	48	504	504	0	213	131	39	0	383	345
Other	0	0	0	0	129	129	0	0	0	77	0	0	77	877
Crop, livestock insur. claims	0	0	0	0	138	138	0	119	0	0	0	0	119	0
Totals less livestock purch.	4324	6122	267	4210	9831	20430	20345	3679	3504	8344	5065	2828	23420	24238

First quarter total = 10446 Fifth quarter total = 11917

Table B.2. Continued

Description	Q1(69)	Q1(70)	Q2	Q3	Q4&Q5	Total	SCH F	Q5(71)	Q6	Q7	Q8	Q9(71)	Total	SCH F
<u>Expense</u>														
Labor hired, Soc. Sec., hired help	0	0	75	0	137	212	333	0	0	311	242	52	605	605
Repairs, main-tenance	0	158	425	600	590	1773	1876	512	188	111	436	230	1477	1477
Interest	68	70	0	0	0	70	1101	0	318	0	4	983	1305	1308
Rent-farm/bus., pasture, other	615	0	0	0	615	615	615	0	0	0	0	615	615	615
Feed purchased	101	370	286	8	1585	2249	2286	1065	684	2032	261	769	4811	4811
Seed, plants, insecticides, etc.	0	0	208	282	0	490	280	0	74	0	0	532	606	606
Fertilizers, lime, misc.	404	1074	0	0	1231	2305	1892	621	0	0	0	1454	2075	2075
Machine hire	72	440	111	148	0	699	509	0	106	275	0	284	665	665
Supplies purch.	0	144	75	0	383	602	515	0	57	44	127	49	277	254
Veterinary - livestock medi-cine	0	9	10	112	60	191	401	102	43	34	110	386	675	675
Gas, fuel, oil	114	932	343	98	765	2139	1872	613	167	9	574	652	2015	2015
Taxes (state and fed.)	0	0	0	0	30	30	0	307	64	0	0	0	371	0
Insurance (farm-not personal)	0	0	138	0	27	165	348	0	0	186	0	0	186	186
Utilities (farm share only)	22	56	77	81	126	340	292	24	115	60	66	29	294	349
Freight, truck and pickup costs	0	0	0	0	318	318	404	0	65	114	179	0	358	0
Tax-not state, fed., Soc. Sec.	0	0	30	0	0	30	60	0	6	0	0	0	6	6

Table B.2. Continued

	Q1(69)	Q1(70)	Q2	Q3	Q4&Q5	Total	SCH F	Q5(71)	Q6	Q7	Q8	Q9(71)	Total	SCH F
Car, legal, bank, other bus. costs	0	0	60	4	294	358	1050	0	9	59	34	222	324	1519
Poultry, sheep, swine purch.	0	0	0	470	0	470	0	0	50	0	0	0	50	0
Depreciation	0	0	0	0	0	0	1508	0	0	0	0	0	0	2512
Totals less livestock purch.	1328	3183	1838	1333	6182	12486	12733	2937	1514	3235	2029	5274	14989	15808

First quarter total = 4410 Fifth quarter total = 6726

	1970 Quarterly	1970 IRS	1971 Quarterly	1971 IRS	IRS_D
Gross farm income less livestock purchased	20430	20811	23420	24267	24236
Farm expenses except interest and depreciation	12486	12849	14989	15808	15303
Depreciation	1507	1507	2512	2512	2512
Interest	1101	1101	1308	1308	1308
Net farm income	5336	5354	4611	4639	4610

Table B.3. Income and expense summary for household X

Description	Q9(1972)	Q10	Q11	Q12	Total	SCH F	1971 (Total)	SCH F
<u>Income</u>								
Cattle (including dairy), horses	0	0	0	3604	3604	3604	3281	3406
Swine (hogs)	278	2507	1624	1352	5761	5304	6654	6654
Corn, soybean, other grain, gener.	0	1054	93	2307	3454	3454	3725	3725
Custom work	0	0	0	462	462	462	941	941
Dividends (mainly patronage)	0	0	0	0	0	20	32	32
Farm commodity loan	5134	0	0	5310	10444	10135	6600	6600
Acre div. payments, land conser.	0	0	1744	0	1744	1744	1653	1658
Gas tax refund (state and fed.)	76	39	39	248	402	464	383	345
Other	0	0	0	0	0	0	77	877
Crop, livestock insur. claims	0	0	0	0	0	0	119	0
Totals less livestock purch.	5488	3200	3200	13203	25091	25187	23420	24238
<u>Expense</u>								
Labor hired, Soc. Sec., hired help	18	137	125	177	457	492	605	605
Repairs, maintenance	179	366	250	581	1376	1464	1477	1477
Interest	86	9	51	967	1113	1113	1305	1308
Rent-farm/bus., pasture other	0	0	0	615	615	615	615	615
Feed purchased	691	560	691	1036	2978	2976	4811	4811
Seed, plants, insecticides, etc.	0	0	0	445	445	445	606	606
Fertilizers, lime, misc.	1454	0	0	50	1504	1501	2075	2075
Machine hire	0	0	0	301	301	301	665	665
Supplies purchased	68	118	107	225	518	454	277	254

Table B.3. Continued

Description	Q9(1972)	Q10	Q11	Q12	Total	SCH F	1971 (Total)	SCH F
Veterinary - livestock medicine	0	33	45	0	78	78	675	675
Gasoline, fuel, oil	534	172	450	1557	2713	2914	0	0
Storage, warehousing	0	0	0	118	118	118	0	0
Taxes (state and federal)	0	0	0	0	0	0	371	0
Insurance (farm - not personal)	0	234	0	200	434	0	186	136
Utilities (farm share only)	86	121	88	138	433	434	294	349
Freight, truck and pickup costs	110	79	69	126	384	383	358	0
Tax - not state, federal, Soc. Sec.	0	0	31	0	31	31	6	6
Car, legal, bank, other bus. costs	119	132	257	336	844	1097	324	1519
Poultry, swine, sheep purchased	0	400	300	80	780	0	50	0
Depreciation	0	0	0	0	0	2569	0	2512
Totals less livestock purch.	3259	1952	2113	5905	13229	13303	14989	15808

	1970		1971		1972		
	Quarterly	IRS	Quarterly	IRS	Quarterly	IRS	IRS_D
Gross farm income less livestock purchased	20430	20811	23420	24267	25091	25187	25187
Farm expenses except interest and depreciation	12486	12849	14989	15808	13229	13372	13303
Depreciation	1507	1507	2512	2512	2569	2569	2569
Interest	1101	1101	1308	1308	1113	1113	1113
Net farm income	5336	5354	4611	4639	8180	8133	8202

Table B.4. Iowa crop production for household X - 1970, 1971 and 1972

Grain	Acres	Yield/ Acre	Total Yield	Price	PSALES	Qtrs.[a]	Inv.	Stock Adj	Total Prod.
1970									
Corn	77.00	87.0	6699.0	1.17	7837.82	7803	0	591	8394
Oats	5.50	45.0	247.5	0.61	150.97	0	0	0	0
Soybeans	85.19	29.0	2470.5	2.55	6299.80	2044	2850	0	4894
Total					14288.59				13288
1971									
Corn	78.10	90.0	7029.0	1.21	8505.08	7572	0	1905	9477
Soybeans	90.69	32.0	2902.1	2.90	8416.03	2641	3680	0	6321
Total					16921.10				15798
1972									
Corn	78.10	92.0	7185.2	1.35	9700.01	12378	0	1955	14333
Soybeans	87.45	37.0	3235.6	3.29	10645.28	1879	4174	0	6053
Total					20345.29				20386

[a]For 1970, Quarters 4, 5; For 1971 Quarters 8, 9; and 1972, Quarters 12, 13.

Table B.5. Livestock production module for household X

Type	Number Sold/ Produced	Ave. Price	Pred. Total	Sales During Year
1970				
Cattle (feeder)	19	148.049	2812	
Total cattle			2812	2344
Hogs (feeder)	88	19.449	1711	
Total hogs			1711	2091
Total livestock			4524	4435
1971				
Cattle (feeder)	20	156.149	3122	
Total cattle			3122	3281
Hogs (market)	135	39.599	5345	
Hogs (feeder)	108	12.319	1330	
Total hogs			6676	6654
Total livestock			9799	9935
1972				
Cattle (feeder)	32	194.849	6235	
Total cattle			6235	3604
Hogs (feeder)	160	22.699	3631	
Hogs (breeding)	1	88.000	88	0
Total hogs			3719	5761
Total livestock			9955	9365

Table B.5. Continued

Type	Q1		Q5		Q9	
	Number	Price/Unit	Number	Price/Unit	Number	Price/Unit
Bull	1	300	2	200	1	300
Ram	1	20	0		0	
Ewes	11	19	0		0	
Stock cows	29	200	31	250	36	300
Young breeding st	7	160	12	110	2	130
Boar	0		1	100	1	160
Sows	0		25	100	25	130
Feeder pigs	0		56	12	0	
Feeder pigs	0		41	8	0	
Feeder pigs	0		50	30	0	

Table B.6. Livestock and crop expense module for household X

Type	Acres or Amt.	Price	Total	Income/ Expenses
1970				
Acre diversion payment	19	67.000	1325	1308
Rental expense	40	15.000	600	615
Fertilizer - corn	77	18.169	1399	
Fertilizer - beans	85	8.619	734	
Total fertilizer			2133	2305
Seed	167	10.000	1676	490
Fuel	304	5.500	1676	2139
Repairs	15000	0.081	1229	1773
Depreciation	15000	0.149	2249	1507
Feed - feeder cattle	19	122.029	2318	
Feed - feeder pigs	88	11.000	968	
Total feed			3286	2249
Veterinary	4435	0.029	133	191
Total expenses			12986	11269
1971				
Acre diversion payment	18	67.000	1206	1658
Rental expense	40	15.000	600	615
Fertilizer - corn	71	18.169	1290	
Fertilizer - beans	82	8.619	710	
Total fertilizer			2000	2075
Seed	153	10.000	1533	606
Fuel	306	5.500	1687	2015
Repairs	17625	0.081	1445	1477
Depreciation	17625	0.149	2643	2512
Feed - feeder cattle	20	124.029	2480	
Feed - market hogs	135	29.639	4001	
Feed - feeder pigs	108	11.000	1188	
Total feed			7669	4811

Table B.6. Continued

Type	Acres or Amt.	Price	Total	Income/ Expenses
Veterinary	9935	0.029	298	675
Total expenses			17879	14786
1972				
Acre diversion payment	20	67.000	1373	1744
Rental expense	40	15.000	600	615
Fertilizer - corn	71	18.169	1290	
Fertilizer - beans	79	8.619	685	
Total fertilizer			1975	1504
Seed	150	10.000	1505	445
Fuel	301	5.500	1655	2713
Repairs	17950	0.081	1471	1376
Depreciation	17950	0.149	2692	2569
Feed - feeder cattle	32	129.739	4151	
Feed - feeder pigs	160	11.000	1760	
Total feed			5911	2978
Veterinary	9365	0.029	280	78
Total expenses			16092	12278

Table B.7. Farm operation edit sheet - family X

Parcel Num.	Term	# of Acres	Type	# of Acres	Amount
			Rental Land Agreements		
1970					
1	Rented in	160.00	Fixed cash amt. - no tob.	40.00	15 $ per acre
			Share of crop or profit	120.00	50 % of crop/prft
2	Rented in	240.00	Share of crop or profit	240.00	50 % of crop/prft
1971					
1	Rented in	240.00	Share of crop or profit	200.00	50 % of crop/prft
			Fixed cash amt.-no tob.	40.00	15 $ per acre
2	Rented in	160.00	Share of crop or profit	160.00	50 % of crop/prft
1972					
1	Rented in	240.00	Fixed cash amt.-no tob.	40.00	15 $ per acre
			Share of crop or profit	200.00	50 % of crop/prft
2	Rented in	160.00	Share of crop or profit	160.00	50 % of crop/prft
1973					
1	Rented in	160.00	Share of crop or profit	160.00	50 % of crop/prft
2	Rented in	240.00	Share of crop or profit	240.00	50 % of crop/prft

Description of Crops

Crop	73 Acr	M/L	72 Acr	71 Acr	70 Acr	69 Acr	Ave. Yld/73	Ave. Yld/72	Ave. Yld/71	Ave. Yld/70	Deg. of Satis.-72
Corn	142.00	same	142.00	142.00	140.00	140.00		92.0 bu	90.0 bu	87.0 bu	Satisfied
Soybeans	190.00	more	159.00	164.89	154.89	134.89		37.0 bu	32.0 bu	29.0 bu	Satisfied
	8.00	less									
Bldgs, etc.	60.00		18.00	21.09							
Pasture			40.00								
Oats					10.00	30.00				45.0 bu	
Feedgrn Pr			41.00	36.00	36.00	36.00					

Table B.7. Continued

Kind	# Sold/73	M/L	# Sold/72	# Sold/71	# Sold/70	# Sold/69	Degree of Satisfaction/72
			Livestock Production				
Feeder pigs	720	More	160	108	88		Satisfied
Feeder calves	22	Same	32	20	19	20	Satisfied
# litters farrowed	70	Err					
Breeding hogs			1				Satisfied
Market hogs (fattened)				135			
Sheep (lambs)						5	
Wool						45	

A Specific Example: The Asset Section

Description

Fairly self-explanatory. The top section shows value and debt for major classifications of assets at each of the different time periods asked. The middle section shows changes to these classifications such as new acquisitions, property traded or sold, and depreciation. The last section shows a detailed debt picture by purpose and loaning institution. It should be noted that the debts under purpose are exactly equivalent to the ones under loaning institution.

Editing guidelines

1) Net equity should be within ten percent or $1000, whichever is greater, of the previous time period's net equity unless mitigating circumstances prevailed, e.g., inheritances, extremely large or negative incomes, etc.
2) Home asset information was checked for consistency with the housing section.
3) Land ownership was checked for consistency with the last part of the farm section, in particular, the Farm Edit Operations section.
4) Within each broad category of assets, and dependent upon the asset classification, all large changes were checked.

Example

Presume the following for net equities at each of the different time periods:

Pre-enrollment	\$12,500
1st Quarterly	20,000
5th Quarterly	23,000
9th Quarterly	15,000
13th Quarterly	30,000

In a situation like this, an initial tendency would be to increase net equity for the ninth quarterly and pre-enrollment interviews and perhaps lower net equity at the thirteenth quarterly interview. During thirteenth quarter however, especially in livestock inventories, there was a tremendous increase in prices, which would account for inventories increasing. All relevant information to the ninth quarter would be gathered. Assume that the following facts become known: depreciation increases tremendously between fifth and ninth, yet there are no machine acquisitions nor increases in machine inventory shown; presume further that machine hire expense has decreased and that liquid assets present in fifth quarter no longer exist; also that there is a machine debt in eleventh quarter that is inconsistent with information at either the ninth or thirteenth quarterly interview. In this case, machine inventory would be increased by a multiple of the increase in depreciation. If the discrepancy were not in machinery inventories the following would be investigated: unsealed grain inventories would be checked for consistency with the crop production section described earlier; the pre-enrollment values would be

Table B.8. Asset information for household X

	Pre-enrollment		First		Fifth		Ninth		Thirteenth	
	Value	Debt	Value	Debt	Value	Debt	Value	Debt	Value	Debt
Land & buildings									2500	2500
Machinery	14000		15000	2500	17625	5300	17950	2000	15381	2000
Livestock & farm inv.	10000	10000	7438	2500	14570	5400	14770	11020	17010	11701
Business inventory					468		87		30	
Liquid	9600		9425		9300		12100		16300	
Unsealed grain	1500		2850	50	2850	50	3680		4174	1139
Totals	35100	10000	34713	5050	44813	10750	48587	13020	55395	17340
Net equity	25100		29663		34063		35567		38055	

Assets information was checked and agrees with 5th quarter totals

	Depreciation	Investment	De-Investment		
	Depreciation	New Acquisition	Traded or Sold	Inherited	Total
Quarter 9 to Quarter 13					
Farm machinery & equipment	2569				-2569
Farm livestock & inventory			725		- 725
Farm land and buildings		2500			2500
Quarter 5 to Quarter 9					
Farm machinery & equipment	2512	3271	434		325
Farm livestock & inventory			900		- 900
Quarter 1 to Quarter 5					
Farm livestock & inventory			470		- 470
Farm machinery & equipment	1507	4210			2703

Table B.8. Continued

	Loan and Debt Information			
	Pre-enrollment	Third Quarter	Seventh Quarter	Eleventh Quarter
Purpose				
Farm machinery or equipment	6000	0	0	0
General farm operating expense	2480	14040	14000	16680
Car (private) & repair	500	0	0	0
Medical expenses	45	0	0	0
Loaning Institution				
Commercial private bank	6545	14040	14000	3240
Loan or finance company	0	0	0	13440
Stores or business firms	2480	0	0	0

checked to see which classifications were different from first quarter. If the differences were in items such as unsealed grain or livestock, no changes would be made since the pre-enrollment interview was administered at a different time of the year compared to quarters one, five, nine, and thirteen. The pre-enrollment interview was obtained prior to crop harvesting time. If the differences were in land values, the value of land would be increased so that it approximately corresponded to the value reported at the other quarters.

Conclusion

This has been a rather lengthy appendix. Undoubtedly despite these efforts there still exists in the data a large number of processing and respondent errors. Hopefully the impact of misreporting and processing errors in conclusions has been substantially reduced by the procedures outlined here.

On the positive side, the editing process, in the opinion of the author, allows for more confidence that the results are valid. All value judgments and imputations of RIME researchers have been opened to scrutiny. This permits the investigation of editing bias. At a minimum, it may provide some information about the mean and variance of response errors. Combining this information with several statistical assumptions, one could determine the range of plausible coefficient estimates due to response errors. Furthermore, the approach taken here may be the only method for obtaining this critical information about the possible variance of response errors. In this modern day of computer technology, there is no reason not to adopt, in certain instances, clerically assigned or computer algorithmically generated multiple data values for the same variable.

APPENDIX C. PRESENTATION OF THE FULL REGRESSION MODELS REPORTED IN CHAPTER FOUR

Table C.1. Iowa coefficient estimates for a selected model for selected dependent variables[a]

Independent Variables	Net Farm Income	Gross Farm Income	Total Recall Hours	Head Recall Hours	Adjusted Scaled Hours	Total Scaled Hours	Scaled Crop Hours	Scaled Livestock Hours
	Dependent Variables							
HR69[b]	1.781 (37.0)	15.70 (0.1)	1.258 (1.4)	1.295 (0.4)	.9373 (0.1)	1.208 (0.0)	1.324 (0.0)	.9421 (0.0)
$HR69^2$	-.003 (42.0)	-.0021 (2.5)	-.0002 (10.0)	-.0002 (4.3)	-.00003 (60.1)	-.00007 (13.0)	-.0002 (23.6)	-.00004 (56.1)
AGE	-28.50 (69.0)	-410.6 (1.3)	39.02 (3.5)	19.95 (21.5)	2.202 (82.3)	-3.445 (66.9)	-6.993 (8.6)	5.403 (51.7)
EDUC	-245.0 (15.8)	-58.29 (88.4)	30.21 (49.8)	12.07 (75.6)	-59.69 (1.3)	-36.12 (6.5)	5.048 (60.9)	-26.56 (17.1)
DR	-15.76 (47.6)	162.4 (0.2)	10.90 (5.6)	9.109 (6.7)	1.135 (71.0)	4.731 (5.8)	3.521 (1.1)	1.782 (47.7)
FE	.0216 (30.2)	.2075 (0.0)	.0095 (7.8)	.0049 (29.3)	-.0019 (50.9)	-.0015 (53.7)	.00009 (94.1)	-.0019 (43.0)
ΔHDWBH	-.7538 (47.4)	-1.206 (58.7)	-.3885 (9.8)	-.3243 (10.0)	-.3896 (0.3)	-.3356 (0.3)	-.1063 (6.8)	-.2786 (1.3)
AGE55	-10.94 (67.5)	86.66 (15.1)	-9.660 (15.1)	-4.428 (45.0)	-3.482 (33.4)	.2658 (92.8)	2.259 (12.7)	-2.523 (38.9)

[a]Variable definitions are included in the text. Number in () is the significance level expressed as a percent.

[b]For scaled crop and livestock hours, HR69 is the 1969 value of the dependent variable.

Table C.1. Continued

Independent Variables	Net Farm Income	Gross Farm Income	Total Recall Hours	Head Recall Hours	Adjusted Scaled Hours	Total Scaled Hours	Scaled Crop Hours	Scaled Livestock Hours
	Dependent Variables							
D71	461.8 (49.0)	1558. (12.3)	-195.7 (4.0)	-296.5 (0.0)	4.011 (94.5)	15.56 (79.0)	12.69 (63.6)	3.433 (94.9)
D72	3612. (0.0)	7067. (0.0)	64.66 (49.4)	-138.3 (6.5)	-142.4 (1.6)	-122.1 (3.7)	-25.13 (35.0)	-96.20 (7.2)
HR69*C/E	2.684 (30.3)	-7.651 (20.3)	-.4454 (50.5)	-.5304 (36.3)	-.0584 (87.1)	-.3113 (28.9)	-.1069 (77.1)	-.2418 (33.8)
$HR69^2$*C/E	-.0005 (37.3)	.0012 (32.0)	.00008 (54.5)	.00009 (45.7)	-.00001 (85.6)	.00004 (50.8)	.00002 (89.6)	.00008 (34.9)
D70*C/E	2078. (69.5)	10819. (37.5)	3912. (0.4)	3141. (0.8)	410.8 (57.3)	705.4 (23.8)	141.6 (66.1)	509.0 (30.1)
D71*C/E	2359. (65.7)	10716. (38.0)	3860. (0.5)	3163. (0.8)	318.6 (66.2)	631.1 (29.1)	129.9 (68.8)	444.4 (36.6)
D72*C/E	707.8 (89.4)	8616. (48.0)	3621. (0.8)	3008. (1.2)	417.2 (56.7)	711.8 (23.4)	108.8 (73.7)	544.0 (26.8)
AGE55*C/E	-4.148 (90.8)	-132.5 (11.1)	18.62 (4.5)	12.28 (12.9)	-1.567 (75.2)	-5.008 (11.7)	-3.899 (5.9)	.8440 (83.4)
PN	1662. (0.0)	2314. (0.0)	47.24 (32.1)	28.02 (45.9)	44.24 (12.9)	39.66 (16.0)	3.296 (80.2)	40.11 (12.5)
DR*C/E	-6.048 (82.9)	-106.3 (10.1)	-4.790 (50.7)	-2.605 (67.9)	-2.585 (50.4)	-6.767 (3.3)	-4.765 (0.4)	-2.903 (35.6)

Table C.1. Continued

Independent Variables	Net Farm Income	Gross Farm Income	Total Recall Hours	Head Recall Hours	Adjusted Scaled Hours	Total Scaled Hours	Scaled Crop Hours	Scaled Livestock Hours
	Dependent Variables							
AGE*C/E	-130.4 (15.3)	35.82 (86.5)	-72.36 (0.2)	-54.81 (0.8)	-7.629 (54.4)	-3.720 (71.7)	3.018 (56.0)	-11.01 (29.1)
CONSTANT	5655. (25.1)	6867. (54.4)	-1543. (22.2)	-485.6 (65.9)	777.8 (25.1)	437.9 (42.9)	40.47 (89.9)	213.5 (67.6)
R^2	.28	.51	.22	.26	.58	.70	.68	.62
OVERALL F	6.0	16.0	4.5	5.4	21.3	36.5	32.7	24.7

Table C.2. Iowa coefficient estimates for a selected model with treatment parameters for selected dependent variables[a]

Independent Variables	Net Farm Income	Gross Farm Income	Total Recall Hours	Head Recall Hours	Adjusted Scaled Hours	Total Scaled Hours	Scaled Crop Hours	Scaled Livestock Hours
	Dependent Variables							
HR69[b]	.9205 (63.7)	13.75 (0.2)	1.044 (3.9)	1.088 (1.3)	.8662 (0.1)	1.093 (0.0)	.9956 (0.2)	.9003 (0.0)
$HR69^2$	-.0002 (65.3)	-.0018 (3.9)	-.0001 (23.7)	-.0001 (11.9)	-.00002 (72.8)	-.00005 (23.7)	-.00003 (79.4)	-.00003 (67.0)
AGE	-80.41 (10.5)	-397.3 (0.0)	-4.932 (70.1)	-12.89 (24.7)	-1.338 (83.9)	-7.472 (17.7)	-7.679 (0.8)	-1.139 (83.6)
EDUC	-259.1 (15.2)	-222.8 (58.2)	24.86 (59.7)	11.43 (77.8)	-65.05 (0.7)	-44.58 (2.8)	-1.076 (91.9)	-34.08 (8.1)
DR	-18.94 (18.2)	94.68 (0.3)	5.471 (13.9)	5.333 (9.5)	-.3872 (83.7)	.7794 (62.3)	.5139 (54.3)	.1156 (94.1)
FE	.0207 (33.2)	.2267 (0.0)	.0073 (18.5)	.0032 (50.2)	-.0017 (55.5)1	-.0010 (68.8)	.00007 (95.4)	-.0018 (45.0)
ΔHDWBH	-.6381 (56.2)	-1.040 (64.5)	-.3723 (12.5)	-.3148 (11.9)	-.3529 (0.8)	-.3022 (1.0)	-.1004 (9.4)	-.2466 (2.8)
D71	460.5 (51.8)	1556. (13.4)	-195.9 (3.9)	-296.6 (0.0)	3.594 (95.1)	15.18 (79.6)	12.63 (63.6)	3.068 (95.4)

[a]Variable definitions are included in the text. Number in () is the significance level expressed as a percent.

[b]For scaled crop and livestock hours, HR69 is the 1969 value of the dependent variable.

Table C.2. Continued

Independent Variables	Net Farm Income	Gross Farm Income	Total Recall Hours	Head Recall Hours	Adjusted Scaled Hours	Total Scaled Hours	Scaled Crop Hours	Scaled Livestock Hours
	Dependent Variables							
D72	3610.	7064.	64.41	-138.4	-143.0	-122.6	-25.22	-96.69
	(0.0)	(0.0)	(49.7)	(6.5)	(1.6)	(3.8)	(34.5)	(7.2)
AGE 55	-13.61	12.80	1.362	2.961	-4.662	-2.298	.4738	-2.230
	(46.3)	(75.8)	(77.7)	(47.7)	(5.9)	(26.7)	(66.3)	(26.7)
HR69*C/E	3.885	-6.193	-.1708	-.2460	.0827	-.1998	.1918	-.1755
	(12.2)	(27.1)	(79.3)	(66.2)	(80.4)	(47.6)	(59.8)	(46.8)
$HR69^2$*C/E	-.0007	.0012	.00002	.00003	-.00004	.00002	-.0001	.00006
	(19.5)	(29.2)	(88.6)	(80.2)	(58.7)	(66.2)	(51.2)	(45.2)
C/E	-4004.	6712.	418.9	375.5	-85.96	232.7	-41.85	-8.063
	(17.5)	(30.3)	(57.7)	(56.3)	(82.4)	(47.5)	(85.1)	(96.0)
D71*C/E	281.3	-101.7	-51.71	22.34	-90.88	-73.09	-11.40	-63.46
	(78.4)	(94.6)	(70.6)	(83.6)	(28.7)	(38.9)	(76.7)	(41.3)
D72*C/E	-627.7	-1169.	-269.5	-119.5	27.73	25.57	-31.01	54.39
	(54.2)	(43.6)	(5.0)	(26.9)	(74.5)	(76.3)	(42.2)	(48.3)
G50	-1219.	-126.5	336.3	301.7	-268.8	-177.7	19.56	-200.5
	(34.0)	(96.5)	(31.1)	(29.3)	(11.4)	(21.3)	(79.6)	(15.0)
G100	-1043.	-6183.	118.1	115.2	-153.0	-213.2	-101.3	-167.2
	(41.4)	(3.1)	(72.2)	(68.9)	(36.7)	(13.6)	(17.6)	(24.3)
T30	-951.9	-1100.	-109.7	-97.98	-58.96	-72.24	-61.84	-11.72
	(40.9)	(67.0)	(71.4)	(70.5)	(70.0)	(57.5)	(36.5)	(92.6)

Table C.2. Continued

Independent Variables	Net Farm Income	Gross Farm Income	Total Recall Hours	Head Recall Hours	Adjusted Scaled Hours	Total Scaled Hours	Scaled Crop Hours	Scaled Livestock Hours
	Dependent Variables							
T70	382.4	8400.	3.957	66.80	71.96	2.938	-71.37	53.89
	(80.4)	(1.6)	(99.2)	(84.7)	(72.6)	(98.6)	(43.1)	(75.0)
CONSTANT	9415.	13426.	803.2	1298.	1148.	1010.	441.6	661.5
	(1.9)	(13.4)	(43.8)	(14.7)	(3.1)	(2.4)	(9.5)	(9.6)
R^2	.20	.50	.20	.24	.59	.70	.66	.62
OVERALL F	4.0	15.3	3.8	4.9	21.8	35.3	29.7	25.1

Table C.3. North Carolina coefficient estimates for a selected model for selected dependent variables[a]

Independent Variables	Net Farm Income	Gross Farm Income	Total Recall Hours	Head Recall Hours	Adjusted Scaled Hours	Total Scaled Hours	Scaled Crop Hours
	Dependent Variables						
HR69[b]	1.929	2.681	.8330	.7880	.4595	.7596	1.047
	(2.3)	(19.3)	(5.1)	(2.6)	(6.6)	(2.8)	(0.1)
$HR69^2$	-.0003	.0003	-.0001	-.0001	-.00001	.00002	-.00005
	(11.2)	(59.4)	(17.9)	(15.8)	(81.6)	(85.6)	(53.3)
AGE	-8.307	36.68	-1.431	-5.938	-8.049	-15.43	-15.61
	(85.4)	(73.8)	(95.0)	(75.2)	(54.6)	(40.2)	(37.9)
EDUC	97.38	441.0	-9.454	-24.53	15.89	43.10	39.36
	(14.3)	(0.7)	(77.7)	(37.3)	(41.6)	(11.1)	(12.8)
DR	3.259	-8.501	6.549	4.777	2.748	5.994	7.137
	(76.7)	(75.0)	(23.5)	(29.4)	(39.6)	(18.0)	(8.9)
FE	-.0080	.1189	.0349	.0281	.0022	-.0021	-.0130
	(74.8)	(5.0)	(0.6)	(0.7)	(76.0)	(83.5)	(18.7)
ΔHDWBH	-.8011	-1.851	-.3023	-.3819	-.2141	-.3510	-.3400
	(0.2)	(0.6)	(0.9)	(0.0)	(0.6)	(0.1)	(0.1)
AGE55	12.46	9.038	.8215	2.199	.1550	1.383	1.845
	(34.8)	(77.9)	(90.2)	(68.9)	(96.8)	(79.7)	(72.0)

[a]Variable definitions are included in the text. Number in the () is the significance level expressed as a percent.

[b]For scaled crop, HR69 is the 1969 value of the dependent variable.

Table C.3. Continued

Independent Variables	Net Farm Income	Gross Farm Income	Total Recall Hours	Head Recall Hours	Adjusted Scaled Hours	Total Scaled Hours	Scaled Crop Hours
	Dependent Variables						
D71	-53.22 (85.5)	335.8 (67.8)	-88.86 (47.2)	-181.5 (6.0)	18.65 (83.6)	-60.83 (59.3)	-65.65 (55.5)
D72	584.7 (5.0)	1805.0 (3.0)	-503.0 (0.0)	-503.2 (0.0)	-59.44 (52.0)	-50.54 (66.5)	-47.36 (67.8)
RACE	-594.3 (29.5)	-5392.0 (0.2)	-234.9 (41.0)	-121.2 (60.7)	-326.6 (5.1)	-344.9 (13.5)	-325.2 (13.3)
HR69*C/E	-.5348 (63.9)	1.022 (71.2)	-.2021 (72.3)	-.4708 (31.9)	.4090 (22.3)	.7112 (12.5)	.1906 (63.6)
$HR69^2$*C/E	.0002 (56.2)	-.0006 (34.0)	.0001 (56.4)	.0001 (32.5)	.00008 (31.2)	.0002 (8.8)	-.00008 (41.2)
D70*C/E	-342.1 (91.2)	2467.0 (74.2)	913.1 (55.6)	1272.0 (32.1)	-688.1 (44.9)	-1377.0 (27.3)	-834.4 (47.4)
D71*C/E	-525.8 (86.5)	2233.0 (76.6)	930.2 (54.8)	1269.0 (32.2)	-660.5 (46.7)	-1306.0 (29.8)	-763.2 (51.2)
D72*C/E	-496.6 (87.2)1	1992.0 (79.1)	1295.0 (40.3)	1577.0 (21.9)	-597.2 (51.1)	-1309.0 (29.7)	-775.2 (50.6)
AGE55*C/E	-17.85 (32.5)	-18.04 (68.2)	3.214 (72.4)	1.287 (86.4)	-6.263 (24.1)	-5.575 (44.9)	-4.982 (47.8)
PN	709.1 (0.0)	853.4 (5.0)	-28.05 (69.4)	-43.70 (43.7)	116.6 (2.0)	161.9 (1.2)	162.1 (1.0)

Table C.3. Continued

Independent Variables	Net Farm Income	Gross Farm Income	Total Recall Hours	Head Recall Hours	Adjusted Scaled Hours	Total Scaled Hours	Scaled Crop Hours
	Dependent Variables						
DR*C/E	-7.400 (59.3)	9.084 (78.7)	-9.496 (17.2)	-8.221 (15.2)	-5.000 (22.0)	-7.384 (19.0)	-9.780 (6.5)
AGE*C/E	14.47 (80.4)	-83.67 (55.4)	-12.15 (67.8)	-10.43 (66.6)	6.589 (70.0)	13.52 (56.7)	11.99 (59.6)
RACE*C/E	-349.1 (63.0	1353.0 (44.2)	-99.59 (78.4)	-270.5 (36.8)	115.1 (58.9)	203.7 (48.9)	175.2 (52.2)
CONSTANT	294.6 (91.0)	-111.9 (98.6)	1160.0 (37.5)	1140.0 (29.2)	954.9 (21.3)	1311.0 (21.6)	1099.0 (27.6)
R^2	.25	.38	.23	.30	.40	.49	.52
OVERALL F	4.1	7.6	3.7	5.3	8.3	12.1	13.5

Table C.4. North Carolina coefficient estimates for a selected model with treatment parameters for selected dependent variables[a]

Independent Variables	Dependent Variables: Net Farm Income	Gross Farm Income	Total Recall Hours	Head Recall Hours	Adjusted Scaled Hours	Total Scaled Hours	Scaled Crop Hours
HR69[b]	1.704 (3.1)	2.741 (14.7)	.6333 (10.7)	.5712 (8.1)	.4078 (5.6)	.6954 (2.1)	.9841 (0.0)
$HR69^2$	-.0003 (11.9)	.0002 (66.5)	-.00007 (44.2)	-.00006 (45.2)	.000006 (90.1)	.00005 (45.6)	-.000003 (96.0)
AGE	-4.042 (89.3)	-33.03 (64.5)	-9.357 (53.1)	-14.79 (23.6)	-6.107 (45.2)	-8.898 (43.6)	-8.793 (42.8)
EDUC	108.7 (10.3)	390.3 (1.5)	-13.45 (68.5)	-32.47 (24.1)	14.14 (43.3)	40.86 (10.8)	35.59 (15.1)
DR	.2472 (96.96)	-.3550 (98.2)	.5040 (87.6)	-.1357 (96.0)	.0046 (99.8)	.9597 (69.9)	.7476 (75.4)
FE	-.0112 (65.2)	.1298 (2.9)	.0359 (0.4)	.0298 (0.4)	.0029 (66.4)	-.0009 (92.4)	-.0114 (22.8)
ΔHDWBH	-.7876 (0.3)	-2.031 (0.2)	-.3177 (0.5)	-.4039 (0.0)	-.2152 (0.4)	-.3424 (0.1)	-.3187 (0.1)
AGE55	6.514 (47.2)	5.355 (80.5)	4.197 (35.3)	4.749 (20.8)	-1.748 (47.6)	.0299 (99.3)	.5589 (86.8)

[a]Variable definitions are included in the text. Number in () is the significance level expressed as a percent.

[b]For Scaled Crop, HR69 is the 1969 value of the dependent variable.

Table C.4. Continued

Independent Variables	Net Farm Income	Gross Farm Income	Total Recall Hours	Head Recall Hours	Adjusted Scaled Hours	Total Scaled Hours	Scaled Crop Hours
	Dependent Variables						
D71	-20.71 (94.5)	393.2 (62.7)	-88.77 (46.9)	-181.6 (5.8)	24.31 (79.0)	-53.92 (64.1)	-59.87 (59.7)
D72	581.1 (6.1)	1853.0 (2.5)	-498.9 (0.0)	-497.2 (0.0)	-59.13 (52.6)	-52.86 (65.5)	-52.98 (64.7)
RACE	-899.4 (1.9)	-4012.0 (0.0)	-370.5 (5.0)	-325.8 (3.9)	-321.9 (0.2)	-311.9 (3.1)	-274.2 (4.5)
HR69*C/E	-.6440 (54.9)	.0141 (99.6)	-.2237 (67.6)	-.3357 (45.2)	.2794 (33.8)	.5085 (21.5)	.1505 (68.7)
$HR69^2$*C/E	.0002 (40.5)	-.0004 (54.9)	.00005 (68.5)	.00006 (57.3)	-.00007 (34.1)	-.0002 (7.5)	-.0001 (22.3)
C/E	791.4 (43.2)	2204.0 (36.6)	455.4 (36.0)	657.2 (11.3)	-59.64 (82.8)	-142.2 (71.2)	159.1 (64.1)
D71*C/E	-466.1 (25.1)	-591.7 (58.6)	26.90 (87.0)	12.51 (92.2)	-19.20 (87.6)	7.192 (96.3)	8.283 (95.6)
D72*C/E	-246.7 (55.2)	-652.9 (55.5)	380.9 (2.5)	303.9 (2.1)	74.59 (55.1)	49.11 (75.7)	44.20 (77.6)
G50	-1303.0 (3.9)	-3623.0 (1.7)	-532.1 (9.1)	-462.3 (7.8)	-553.4 (0.1)	-798.5 (0.1)	-744.2 (0.2)
G100	342.8 (58.5)	-1190.0 (43.0)	.6983 (99.8)	-276.5 (29.0)	104.1 (54.1)	31.33 (89.6)	-78.39 (73.4)

Table C.4. Continued

Independent Variables	Net Farm Income	Gross Farm Income	Total Recall Hours	Head Recall Hours	Adjusted Scaled Hours	Total Scaled Hours	Scaled Crop Hours
	Dependent Variables						
T30	-1215.0 (6.8)	-4441.0 (0.6)	-749.9 (2.4)	-709.1 (1.0)	-601.6 (0.1)	-760.0 (0.3)	-741.2 (0.3)
T70	-396.6 (60.9)	-2298.0 (21.8)	-227.2 (55.8)	-221.8 (49.2)	-438.8 (3.8)	-448.1 (13.1)	-401.7 (16.6)
CONSTANT	608.9 (74.9)	3587.0 (43.2)	1800.0 (5.8)	1919.0 (1.6)	989.3 (5.6)	1102.0 (13.0)	887.2 (20.1)
R^2	.22	.40	.25	.31	.46	.54	.55
OVERALL F	3.6	8.6	4.2	5.8	11.0	15.1	16.0

APPENDIX D. PRESENTATION OF THE FULL REGRESSION MODELS REPORTED IN CHAPTER FIVE

Table D.1. North Carolina regression equations explaining the changes between the original and edited data bases[a]

Independent Variables	DIFF Total Family Income	ABS Total Family Income	DIFF Scaled Crop Hours	ABS Scaled Crop Hours
	Dependent Variables			
Constant	405.2	664.9	95.05	-157.2
1 Age	.9541 (.1)	-3.843 (-.3)	.6788 (.2)	2.273 (.8)
1 Educ	45.18 (.9)	69.55 (1.8)[b]	-3.662 (-.4)	8.494 (1.0)
1 Quick Test Score	-16.34 (-.8)	6.556 (.4)	-2.884 (-.7)	-3.684 (-1.0)
1 Tobacco Yield	.0222 (.1)	-.2894 (-.8)	-.0324 (-.4)	.0127 (.2)
2 Number Different Crop Operations	38.11 (.4)	68.92 (1.0)	-19.07 (-1.1)	-13.26 (-.8)
2 Number Different Livestock Operations	122.0 (.8)	-43.04 (-.4)	24.10 (.8)	-23.76 (-.9)
2 % of Labor in Major Crop Enterprise	-1.300 (-.3)	2.773 (.8)	-.9009 (-1.0)	-.2379 (-.3)

[a]Numbers in front of each line indicate what variables were grouped together for the partial F-statistics reported in Table 5.12.

[b]Significant at the 10 percent level.

Table D.1. Continued

Independent Variables	Dependent Variables DIFF Total Family Income	ABS Total Family Income	DIFF Scaled Crop Hours	ABS Scaled Crop Hours
2 % of Labor in Major Livestock Enterprise	-.3086	5.750	-.7852	.2648
	(-.1)	(2.1)[c]	(-1.2)	(.4)
2 Number of Different Landlords	180.5	389.9	-5.136	20.82
	(1.3)	(3.7)[d]	(-.2)	(.8)
2 % of Land Owned or Rented-Cash Basis	-1.171	-3.535	.7547	.1550
	(-.3)	(-1.3)	(1.0)	(.2)
3 Total Family Income Edited DB[e]	.0574	.00061	.0927	.1052
	(1.0)	(.01)	(3.2)[d]	(3.8)[d]
3 Total Farm Value	.0554	.0529	-.0021	.0011
	(3.7)[d]	(4.7)[d]	(-.8)	(.4)
4 Number of Quarters From Records	-357.2	-271.5	24.09	7.6834
	(-3.6)[d]	(-3.6)[d]	(1.3)	(.4)
5 D71	-817.3	-509.5	48.40	1.388
	(-1.9)[b]	(-1.6)[b]	(.6)	(1.0)
5 D72	-848.8	-843.5	127.0	183.5
	(-1.9)[b]	(-2.5)[d]	(1.6)[b]	(2.4)[d]

[c]Significant at the 5 percent level.

[d]Significant at the 1 percent level.

[e]Total scaled hours, Edit DB was substituted for the variable in the scaled hours equations.

Table D.1. Continued

Independent Variables	DIFF Total Family Income	ABS Total Family Income	DIFF Scaled Crop Hours	ABS Scaled Crop Hours
	Dependent Variables			
6 C/E	40.14	-420.9	31.71	55.25
	(.1)	(1.1)	(.2)	(.6)
6 C/E * Total Family Income[e]	-.0441	-.0476	-.0342	-.0677
	(-.6)	(-.9)	(-.9)	(-1.8)[b]
7 D71*C/E	246.0	384.6	10.82	19.67
	(.4)	(.9)	(.1)	(.2)
7 D72*C/E	795.6	1150.0	-127.1	-192.7
	(1.4)	(2.6)[d]	(-1.2)	(-2.0)[c]
8 DIS[f]	-.4353	-.1799	-.0314	-.0826
	(-2.7)[d]	(-1.5)	(-.9)	(-2.6)[d]
8 DIS*C/E[f]	.7658	.7081	.0350	.0946
	(3.0)[d]	(3.6)[d]	(.7)	(2.0)[c]
R^2	.1808	.2954	.0715	.1256
$\hat{\sigma}_u$	1966.0	1491.0	361.0	341.5

[f]DIS is total scaled hours for the current year minus 1969 total scaled hours.

Table D.2. Iowa regression equations explaining the changes between the original and edited data bases[a]

Independent Variables	Dependent Variables DIFF Total Family Income	ABS Total Family Income	DIFF Scaled Live-stock hours	ABS Scaled Live-stock hours
Constant	3537.0	116.9	295.6	-28.97
1 Age	2.627	3.790	-2.454	-1.298
	(.1)	(.1)	(-1.0)	(-.6)
1 Educ	97.13	91.35	-29.98	-18.59
	(.6)	(.7)	(-2.4)[b]	(-1.6)
1 Quick Test Score	-79.55	-24.58	1.87	3.339
	(-1.2)	(-.5)	(.4)	(.8)
1 Corn Yield	-25.35	-4.233	-.2427	.4548
	(-1.8)[c]	(-.4)	(-.2)	(.5)
2 # Different Crop Operations	-132.21	123.7	27.78	8.673
	(-.4)	(.5)	(1.2)	(.4)

[a]Numbers in front of each line indicate what variables were grouped together for the partial F-statistics reported in Table 5.13.

[b]Significant at the 5 percent level.

[c]Significant at the 10 percent level.

Table D.2. Continued

Independent Variables	DIFF Total Family Income	ABS Total Family Income	DIFF Scaled Live-stock hours	ABS Scaled Live-stock hours
	Dependent Variables			
2 # Different Livestock Operations	7.537 (.04)	-140.2 (-.10)	2.986 (.2)	8.896 (.7)
2 % of Labor in Major Crop Enterprise	15.3 (.7)	12.39 (.7)	1.569 (1.0)	2.748 (1.9)[c]
2 % of Labor in Major Livestock Enterprise	10.60 (1.2)	12.72 (1.8)[c]	-.8846 (-1.4)	-.6878 (-1.2)
2 # of Different Landlords	-278.3 (-.8)	382.3 (1.3)	-25.97 (-1.0)	-.7767 (-.03)
2 % of Land Owned or Rented-Cash Basis	3.743 (.4)	5.456 (.8)	.0408 (.1)	.7006 (1.2)
3 Total Family Income Edited DB	.1811 (2.5)[b]	.0265 (.4)	.1005 (2.4)[b]	.1788 (4.6)[d]
3 Total Farm Value	.0131 (1.2)	.0383 (4.5)[d]	.00045 (.6)	-.00018 (-.3)
4 # Quarters from Records	-581.5 (-1.9)[b]	-585.7 (-2.3)[c]	-22.54 (-1.0)	-72.57 (-3.5)[d]
5 D71	902.6 (.9)	829.3 (1.0)	-43.57 (-.6)	43.42 (.7)
5 D72	356.5 (.3)	163.7 (.1)	-43.34 (-.5)	52.05 (.6)

[d]Significant at the 1 percent level

Table D.2. Continued

Independent Variables	DIFF Total Family Income	ABS Total Family Income	DIFF Scaled Livestock hours	ABS Scaled Livestock hours
	Dependent Variables			
6 C/E	325.8 (.3)	-947.7 (-1.1)	-37.39 (-.5)	58.29 (.8)
6 C/E * Total Family Income	.2522 (2.5)[b]	.3215 (3.6)[d]	.1261 (2.6)[d]	.0711 (1.6)[c]
7 D71*C/E	-377.5 (-.3)	-1006.0 (-1.0)	22.32 (.3)	-38.24 (-.5)
7 D72*C/E	-215.7 (-.2)	-1036.0 (-1.0)	102.0 (1.2)	44.06 (.5)
8 DIS	1.653 (1.8)[c]	-.2169 (-.0)	-.0987 (-1.5)	-.0590 (-1.0)
8 DIS*C/E	-3.435 (-3.0)[d]	1.179 (-1.2)	.2210 (2.6)[d]	.1726 (2.2)[b]
R^2	.2257	.2759	.2344	.3041
$\hat{\sigma}_u$	4228.0	3455.0	301.7	279.0

Table D.3. North Carolina coefficient estimates for a selected model explaining work disincentives for selected dependent variables contrasting the original and edited data bases

	Dependent Variables				
	Orig Net Farm Income	Edit Net Farm Income	Orig Gross Farm Income	Edit Scaled Hours	Orig Total Scaled Hours
Constant	-1510.0	576.5	3277.0	4456.0	1823.0
HR69	1.657 (1.95)[a]	1.886 (3.13)[b]	2.694 (1.68)[a]	2.757 (1.79)[a]	.7177 (2.91)[b]
$HR69^2$	-.00025 (-1.32)	-.0003 (-2.14)[c]	.00019 (.53)	.00025 (.69)	.00002 (.33)
AGE	2.892 (.09)	-17.90 (-.79)	-57.10 (-.94)	-77.05 (-1.33)	-22.02 (-2.35)[c]
EDUC	109.6 (1.47)	106.8 (2.04)[c]	434.2 (3.09)[b]	481.2 (3.59)[b]	33.53 (1.55)
DEBT RATIO	-2.958 (-.44)	-2.587 (-.52)	-10.67 (-.84)	-5.444 (-.43)	-1.864 (-.95)
NET FARM EQUITY	-.0257 (-1.27)	-.0117 (-.64)	.0494 (1.28)	.1120 (2.39)[c]	-.0058 (-.98)
Δ OFF-FARM HRS	-.6427 (-1.90)[a]	-.6485 (-2.45)[c]	-1.682 (-2.64)[b]	-2.058 (-3.03)[b]	-.1785 (-1.82)[a]

[a]Significant at the 10 percent level.

[b]Significant at the 1 percent level.

[c]Significant at the 5 percent level.

Table D.3. Continued

	Dependent Variables				
	Orig Net Farm Income	Edit Net Farm Income	Orig Gross Farm Income	Edit Scaled Hours	Orig Total Scaled Hours
D71	760.0 (1.26)	-8.209 (-.02)	882.3 (.78)	332.6 (.31)	-134.0 (-.77)
D72	1158.0 (1.90)[a]	577.1 (1.34)	2227.0 (1.93)[c]	1908.0 (1.73)[a]	-213.7 (-1.21)
# Males 13-15	344.0 (.76)	-78.12 (-.24)	-570.8 (-.67)	-877.4 (-1.07)	265.1 ((2.02)[c]
# Females 13-15	180.0 (.37)	390.4 (1.15)	392.8 (.43)	371.7 (.43)	205.2 (1.46)
# Males 16-20	-79.38 (-.25)	244.4 (1.08)	378.5 (.62)	465.3 (.80)	182.4 (1.95)[a]
# Females 16-20	606.7 (1.72)[a]	301.4 (1.21)	1212.0 (1.82)[a]	1057.0 (1.66)[a]	140.7 (1.37)
# Males 21-60	-8.313 (-.01)	-3.198 (-.01)	-530.5 (-.42)	-266.7 (-.22)	-18.24 (-.09)
# Females 21-60	-64.58 (-.14)	-92.07 (-.29)	133.1 (.16)	77.82 (.10)	-42.95 (-.33)
RACE	-5.367 (-.01)	-623.3 (-1.49)	-5105.0 (4.65)[b]	-4652.0 (-4.35)[b]	-579.4 (-3.43)[b]
AGE55	206.8 (1.76)[a]	206.2 (2.44)[c]	494.5 (2.22)[c]	499.0 (2.31)[c]	55.10 (1.61)

Table D.3. Continued

	Dependent Variables				
	Orig Net Farm Income	Edit Net Farm Income	Orig Gross Farm Income	Edit Scaled Hours	Orig Total Scaled Hours
HR69*C/E	.9509 (.81)	-.2472 (-.30)	.9892 (.45)	.4225 (.20)	.7743 (2.27)[c]
$HR69^2$*C/E	-.00019 (-.70)	.00008 (.42)	-.00057 (-1.14)	-.00043 (-.90)	-.0002 (-2.5)[c]
C/E	29.44 (.02)	686.0 (.73)	-332.4 (-.13)	-83.83 (-.03)	-1000.0 (-2.55)[c]
D71*C/E	-623.1 (-.77)	-487.9 (-.86)	-796.0 (-.52)	-501.6 (-.34)	33.47 (.14)
D72*C/E	-731.7 (-.89)	-262.5 (-.45)	-988.7 (-.64)	-718.5 (-.48)	218.2 (.92
AGE55	-119.9 (-1.05)	-162.8 (-2.00)[c]	-408.6 (-1.89)[a]	-457.1 (-2.19)[c]	-13.41 (-.40)
RACE*C/E	-721.6 (-.98)	-409.0 (-.79)	1696.0 (1.22)	1312.0 (.98)	411.4 (1.93)[a]
R^2	.17	.28	.44	.50	.61

Table D.3. Continued

	Dependent Variables				
	Edit Total Scaled Hours	Orig Adj Scaled Hours	Edit Adj Scaled Hours	Orig Scaled Crop Hours	Edit Scaled Crop Hours
Constant	1741.0	1247.0	1321.0	1474.0	1474.0
HR69	.5858 (1.24)	.4033 (2.01)[c]	.3227 (1.86)[a]	.9912 (4.39)[b]	.8796 (4.12)[b]
HR69^2	.00007 91.4)	.0000 (.00)	.00003 (.75)	-.00003 (-.6)	.00002 (.4)
AGE	-23.04 (-2.64)[b]	-14.32 (-1.88)[a]	-15.98 (-2.46)[c]	-19.63 (-2.15)[c]	-21.74 (-2.56)[c]
EDUC	33.64 (1.67)[a]	21.32 (1.21)	13.38 (.89)	30.69 (1.46)	31.51 (1.61)
DEBT RATIO	.8457 (.44)	-2.109 (-1.32)	-.4623 (-.33)	-3.176 (-1.70)[a]	.4185 (.23)
NET FARM EQUITY	-.0042 (-.59)	-.0071 (-1.48)	.00093 (.18)	-.0078 (-1.37)	-.0139 (-1.99)[c]
Δ OFF-FARM HRS	-.2501 (-2.45)[c]	-.0704 (-.88)	-.1329 (-1.75)[a]	-.1759 (-1.84)[a]	-.2176 (-2.19)[c]
D71	-23.83 (-.15)	-75.20 (-.53)	46.41 (.38)	-98.58 (-.58)	-36.17 (-.23)
D72	-48.83 (-.29)	-141.5 (-.98)	-61.41 (-.50)	-204.1 (-1.19)	-54.51 (-.34)

Table D.3. Continued

	Dependent Variables				
	Edit Total Scaled Hours	Orig Adj Scaled Hours	Edit Adj Scaled Hours	Orig Scaled Crop Hours	Edit Scaled Crop Hours
# Males 13-15	322.1 (2.62)[b]	249.2 (2.33)[c]	303.2 (3.31)[b]	219.7 (1.72)[a]	283.0 (2.37)[c]
# Females 13-15	292.9 (2.24)	129.1 (1.13)	190.9 (1.96)[c]	136.2 (1.00)	248.5 (1.95)[a]
# Males 16-20	143.0 (1.63)[a]	184.7 (2.42)[c]	128.1 (1.96)[c]	165.3 (1.81)[a]	124.4 (1.46)
# Females 16-20	121.0 (1.26)	151.9 (1.82)[a]	78.02 (1.09)	112.1 (1.12)	90.53 (.97)
# Males 21-60	-71.80 (-.40)	2.673 (.02)	3.685 (.03)	32.13 (.17)	28.23 (.16)
# Females 21-60	-48.05 (-.39)	4.198 (.04)	8.185 (.09)	-41.58 (-.33)	-45.21 (-.38)
RACE	-459.1 (-2.85)[b]	-308.2 (-2.24)[c]	-513.8 (-3.45)[b]	-515.6 (-3.18)[b]	-519.4 (-2.72)[b]
AGE55	76.55 (2.35)[c]	33.26 (1.20)	53.72 (2.21)[c]	49.41 (1.49)	71.64 (2.26)[c]
HR69*C/E	1.030 (3.28)[b]	.4687 (1.69)[a]	.6066 (2.59)[b]	.2422 (.80)	.4454 (1.59)
$HR69^2$*C/E	-.0003 (-4.29)[b]	-.00009 (-1.5)	-.00014 (-2.8)[b]	-.0001 (-1.43)	-.00018 (-2.57)[c]

Table D.3. Continued

	Dependent Variables				
	Edit Total Scaled Hours	Orig Adj Scaled Hours	Edit Adj Scaled Hours	Orig Scaled Crop Hours	Edit Scaled Crop Hours
C/E	-976.1 (-2.7)[b]	-527.7 (-1.65)[a]	-532.2 (-1.98)[c]	-457.9 (-1.36)	-470.3 (-1.52)
D71*C/E	-32.09 (-.15)	102.0 (.53)	-42.43 (-.26)	-16.05 (-.07)	-19.84 (-.09)
D72*C/E	25.41 (.11)	170.7 (.88)	66.26 (.40)	193.9 (.84)	34.0 (.16)
AGE55*C/E	-31.65 (-1.01)	-22.22 (-.82)	-44.43 (-1.90)[a]	-17.13 (-.53)	-28.88 (-.95)
RACE*C/E	282.6 (1.41)	63.01 (.36)	180.2 (1.20)	329.6 (1.62)	246.6 (1.29)
R^2	.66	.48	.60	.62	.68

Table D.4. Iowa coefficient estimates for a selected model explaining work disincentives for selected dependent variables contrasting the original and edited data bases

	Dependent Variables				
	Orig Net Farm Income	Edit Net Farm Income	Orig Gross Farm Income	Edit Scaled Hours	Orig Total Scaled Hours
Constant	13070.0	8936.0	10130.0	9299.0	751.4
HR69	-1.459 (-.68)	1.357 (.79)	13.67 (3.63)[a]	15.65 (4.61)[a]	.9793 (4.47)[a]
$HR69^2$	.00052 (1.16)	-.00023 (-.68)	-.0016 (-2.05)[b]	-.00208 (-3.10)[a]	-.00005 (-1.00)
AGE	-109.5 (-2.58)[a]	-90.00 (-2.41)[b]	-446.1 (-5.99)[a]	-393.1 (-5.33)[a]	-8.191 (-1.89)[c]
EDUC	-264.0 (-1.53)	-247.7 (-1.66)[c]	316.5 (1.04)	-132.0 (-.45)	-12.44 (-.70)
DEBT RATIO	-34.76 (-2.82)[a]	-17.76 (-1.52)	99.89 (4.61)[a]	107.9 (4.69)[s]	-1.806 (-1.44)
NET FARM EQUITY	-.0475 (-2.02)[b]	.0165 (.92)	.0653 (1.58)	.2063 (5.81)[a]	.0031 (1.29)
Δ OFF-FARM HRS	-.1215 (-.13)	-1.090 (-1.16)	-.7654 (-.45)	-2.300 (-1.24)	-.0745 (-.75)

[a]Significant at the 1 percent level.

[b]Significant at the 5 percent level.

[c]Significant at the 10 percent level.

TABLE D.4. Continued

	Dependent Variables				
	Orig Net Farm Income	Edit Net Farm Income	Orig Gross Farm Income	Edit Scaled Hours	Orig Total Scaled Hours
D71	268.3 (.30	465.6 (.58)	426.3 (.27)	1607.0 (1.02)	46.27 (.50)
D72	4277.0 (4.70)[a]	3617.0 (4.54)[a]	6916.0 (4.33)[a]	7103.0 (4.52)[a]	-119.4 (-1.29)
AGE55	182.0 (1.11)	37.40 (.26)	783.7 (2.72)[a]	516.0 (1.83)[c]	6.221 (.37)
HR69*C/E	5.540 (2.11)[b]	2.911 (1.31)	-9.904 (-2.15)[b]	-9.433 (-2.14)[b]	.1161 (.43)
$HR69^2$*C/E	-.0014 (-2.64)[a]	-.00049 (-1.11)	.0014 (1.51)	.00154 (1.77)[c]	-.00006 (-1.2)
C/E	-4840.0 (-1.58)	-3224.0 (-1.22)	11680.0 (2.18)[b]	11758.0 (2.25)[b]	-62.12 (-.20)
D71*C/E	-41.01 (-.03)	263.5 (.23)	944.0 (.41)	-82.15 (-.04)	-88.29 (-.66)
D72*C/E	-1919.0 (-1.46)	-649.6 (-.56)	-1357.0 (-.59)	-1024.0 (-.45)	-66.26 (-.50)
AGE55*C/E	-233.7 (-1.19)	-269.2 (-1.55)	-586.3 (-1.70)[c]	-762.8 (-2.23)[b]	-27.59 (-1.38)
R^2	.21	.21	.50	.57	.69

Table D.4. Continued

	Dependent Variables				
	Edit Total Scaled Hours	Orig Scaled Crop Hours	Edit Scaled Crop Hours	Orig Scaled Livestock Hours	Edit Scaled Livestock Hours
Constant	887.9	216.2	275.9	395.8	691.4
HR69	1.176 (6.93)[a]	1.031 (4.58)[a]	1.048 (4.59)[a]	.6895 (4.09)[a]	.9009 (6.26)[a]
HR69^2	-.00006 (-2.0)[b]	-.00004 (-.4)	-.00005 (-.5)	.00002 (.33)	-.00002 (-.4)
AGE	-9.206 (-2.50)[b]	-5.201 (-2.79)[a]	-5.308 (-2.89)[a]	-2.354 (-.58)	-3.597 (-1.01)
EDUC	-39.32 (-2.68)[a]	4.731 (.64)	.8030 (.11)	-.1381 (-.01)	-28.80 (-2.06)[b]
DEBT RATIO	.6933 (4.65)[a]	.2456 (.43)	.4134 (.71)	-2.535 (-2.14)[b]	-.0324 (-.03)
NET FARM EQUITY	-.00134 (-.76)	.00053 (.52)	.00005 (.06)	.00081 (.36)	-.0019 (-1.13)
Δ OFF-FARM HRS	-.2877 (-3.11)[a]	-.0407 (-.92)	-.0856 (-1.80)[c]	-.1183 (-1.26)	-.2544 (-2.85)[a]
D71	15.02 (.19)	21.10 (.53)	12.46 (.32)	25.87 (.30)	3.156 (.04)
D72	-122.8 (-1.56)	-18.43 (-.47)	-25.45 (-.65)	-99.79 (-1.14)	-96.57 (-1.28)

Table D.4. Continued

	Dependent Variables				
	Edit Total Scaled Hours	Orig Scaled Crop Hours	Edit Scaled Crop Hours	Orig Scaled Livestock Hours	Edit Scaled Livestock Hours
AGE55	5.690 (.40)	8.384 (1.15)	4.849 (.70)	-1.420 (-.09)	-3.370 (-.26)
HR69*C/E	-.3481 (-1.58)	.1059 (.39)	.1011 (.38)	.0034 (.02)	-.1909 (-1.07)
$HR69^2$*C/E	.00005 (1.25)	-.00008 (-.73)	-.00007 (-.64)	-.00005 (-.71)	.00006 (1.00)
C/E	349.3 (1.34)	25.35 (.15)	10.49 (.06)	-102.8 (-.75)	-59.55 (-.50)
D71*C/E	-72.52 (-.64)	-17.05 (-.30)	-10.81 (-.19)	-74.38 (-.59)	-63.76 (-.58)
D72*C/E	26.27 (.23)	-33.35 (-.59)	-30.28 (-.53)	-36.33 (-.29)	54.01 (.50)
AGE55*C/E	-29.95 (-1.75)[c]	-21.55 (-2.55)[b]	-19.58 (-2.38)[b]	-7.028 (-.39)	-5.739 (-.36)
R^2	.80	.79	.79	.61	.75