5-2001

Scale Economies in Public Education: Evidence from School Level Data

Ryan C. Bosworth
Utah State University

SCALE ECONOMIES IN PUBLIC EDUCATION:

EVIDENCE FROM SCHOOL LEVEL DATA

and

FARM HOUSEHOLD WEALTH: MEASUREMENT

STRUCTURE, AND DETERMINANTS

by

Ryan C. Bosworth

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Economics

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

2001

ACKNOWLEGEMENTS

I would like to express sincerest thanks to Professor Tyler J Bowles. Without his continual guidance, this thesis would not have been possible. I also thank him for the personal interest he has taken in my education, and for the excellent advice.

Thanks are also due to Professor Chris Fawson (I love Economics!), Professor W Cris Lewis, Professor David Aadland, Professor Dwight Israelsen, Professor Kenneth Lyon, Professor Basudeb Biswas and Department Head Keith Criddle. Each of these wonderful educators has helped me greatly

Thanks to Frank Caliendo, Jeremy Kidd, Kristian Mortensen, and Michelle Crook for making Economics fun.

Thanks to my parents for instilling an understanding of the value of education.

Thanks to my wife for understanding, support, and encouragement.

PAPER 1

SCALE ECONOMIES IN PUBLIC EDUCATION:

EVIDENCE FROM SCHOOL LEVEL DATA

Abstract

This paper uses a school level, panal data set to investigate the existence and extent of possible scale economies in the production of public education in Wyoming. We find that scale economies exist, and that an equitable funding mechanism must account for these scale effects.

I. Introduction

The structure of school finance regimes in the United States has been a subject of much political and legal debate over the past three decades. Court rulings have required many states to restructure school financing methods in order to pursue some concept of equality.[1] Achieving equality of spending is, of course, a simple matter. Developing a funding mechanism that provides for equality of educational opportunity, however, is difficult since such a system, by definition, must allow for cost differences across schools and districts.

In the context of education, the cost of achieving a given output (e.g., a high school graduate who can gain admittance into a state university) may differ across schools and districts for various reasons: The average level of household income may vary across schools and districts; teachers and administrators may be more skilled in one district than another; or school size may affect average cost. The focus of this paper is on this latter source of possible cost differences; i.e., does school size affect the average cost of producing education? If it does, a school financing scheme must account for these economies or diseconomies of scale if the objective is equal educational opportunities.

The contribution of this paper to the literature on school finance is not in the question asked, per se, but rather in the data and techniques used to address the question. A data set specific to the state of Wyoming that contains expenditure details *at the school level* is used to analyze economies of scale.[2]

The *Journal of Education Finance* devoted a special issue (Winter 1997) to the collection of school-level finance data. Busch and Odden (1997) noted that while policymakers increasingly are ". . . concerned about resource allocation within districts," the ". . . grim reality . . . is that only small amounts of detailed, school-level data are available." Several of the papers in this special study noted the necessity of school- or site-specific data in addressing questions of equity and efficiency (see Berne, Stiefel, and Moser; Monk; Farland; Goertz; and Cohen):

> Regardless of whether we focus on efficiency or effectiveness, we should be trying to measure the relationship between inputs or resources and outputs or outcomes. And more than that, we should measure whether the relationship is such that more could be achieved with the same resources or not. These are not easy concepts to measure. Thus far, efforts to measure efficiency or effectiveness too often have focused exclusively on the input side, or when they have measured input/output relationships, ***the district and not the school*** has been the unit of analysis. (Berne, Stiefel, and Moser, p. 247, emphasis added)

The present study uses a data set that contains rather detailed expenditure data[3] by school for 16 Wyoming school districts. These data are used to address the question of equity: If small schools receive the same funding per student as large schools, will the students in the small schools receive an equal education? If economies of scale are present, the answer is no.

Since an understanding of cost and efficiency is important to understanding issues of equity and economies of scale, the following section is devoted to these concepts. The third section presents an empirical model designed to estimate economies of scale at the school level. The fourth section discusses and applies estimation techniques for the model and presents the results. The fifth section discusses and tests for group or district effects. The final section contains a summary and the conclusions.

II. Problems of Cost and Efficiency

Discussions of cost and efficiency in education are plagued by the fundamental problems of measurement and definition. First, relatively little is known about what economists call the production function of education.[4] More specifically, little is known about how class size specifically affects education quality. According to Hoxby (1999, p. 2-3), ". . . it would be accurate to describe class size policies as highly controversial among researchers, who disagree about whether reducing class size actually improves student achievement" (also see Hoxby (2000), and Betts (1995)). Unless researchers understand and can formally specify a production function, rigorous discussions of cost and efficiency are difficult. Secondly, economists mean something very specific with the term cost. Cost is what must be given up to produce a well-defined unit of output. Two problems are encountered in discussing and analyzing costs in public education: (1) output is difficult to define and measure, and (2) given the lack of competitive markets, expenditures are observed rather than costs.[5] Fortunately, the issues of economies of scale and equity can be discussed in education financing without getting excessively bogged down in definitions of cost and efficiency. The following hypothetical example will help clarify this point.

Consider two school districts, A and B, both of whom have a number of elementary schools. Assume Figures 1a and 1b represent observed average expenditure (AE) per student (vertical axis) for each of the different elementary schools in the districts. District B may be spending more per student than district A because the funding formula provides more funds to B compared to A and administrators can afford to have smaller class sizes in district B.[6] Indeed, the choice of class size is critical in determining how far the average expenditure curve is located from the origin. Not knowing how class size affects quality, the difficulty in measuring education output, and the difficulty in controlling for

Figure 1a. (District A)

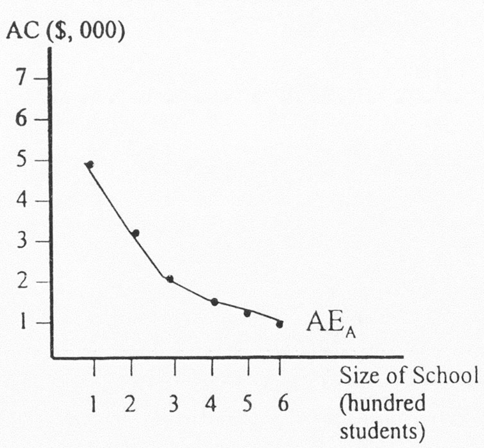
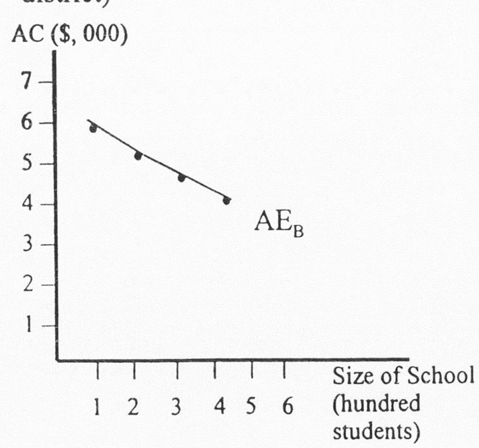


Figure 1b. (District B—smaller district)



environmental factors that affect the educational process, make it very difficult to say that district B is less efficient than A, that district A is efficient in an absolute sense, or that the average expenditure curve for district A reflects the cost of education while district B's average expenditure curve does not. Indeed, the average cost curve of providing education may be closer to the origin than the average expenditure curve for district A.

Although we may not be able to infer from the average expenditure data in Figure 1a that district A is efficient or that this expenditure curve reflects the cost of education, we may reasonably infer that economies of scale exist given one apparently reasonable assumption: *district administrators attempt to distribute resources within their districts to achieve equitable outcomes.* That is, administrators in district A would want students in the elementary school of 100 to be receiving approximately the same education as those in the school of 600. If this is true and if the district is spending $5,000 per student in the smaller school compared to $1,000 in the larger school, then the reasonable conclusion is that it costs more per student in the small school to provide this education; i.e., economies of scale exist. The focus of this study is to analyze school level expenditure data and test for economies of scale. Implicitly, the assumption is made that district administrators do indeed attempt to distribute resources within their districts to achieve equitable outcomes.

III. Empirical Model

Since the instant focus is on the data set and econometric techniques rather than the theoretical derivation of an appropriate cost function specification, the interested reader is referred to in Downes and Pogue (1994) and Chakraborty et al. (2000) for the theoretical justification for a log-linear cost function with per student cost as the dependent variable and output, input prices, and school or district attributes as explanatory variables. We start with the model specified by Chakraborty et al. (2000) and justify implemented changes.

Chakraborty et al. (2000) posit the following cost function:[7]

$$\ln C_{it} = \alpha + \alpha_1 \ln Q_{it} + \alpha_2 \ln P_{it} + \alpha_3 \ln S_{it} + \varepsilon_{it}$$

where C_{it} = cost per student in district i at time t; Q_{it} = measure of output; P_{it} = measure of input prices, and S_{it} = a vector of variables that measure those attributes of the school district that influence cost. Chakraborty et al. use the proportion of students graduating in each district as their measure of output (Q), the 20-year average teacher salary as the input price (P), and the number of schools and number of students in the district as the elements of S.

Although Chakraborty et al use the proportion of students graduating in each district as the measure of output, they noted that "most studies of educational production relationships measure output by standardized achievement test scores." The current study uses test scores as the output measure.[8]

As a measure of input prices, the current study uses the average teacher salary in the district for each type of school. elementary, middle, and high school. The current study uses school size and income level of school patrons as the two variables in the vector S_{it}, the attributes of *the school* that influence cost. Thus, the general specification of our model is

(1) $\ln Cost_{it} = \alpha_0 + \alpha_1 Scores_{it} + \alpha_2 Salary_{it} + \alpha_3 Income_{it} + \alpha_4 \ln Size_{it} + \varepsilon_{it}$

where $Cost_{it}$ = operating expenditures per student in school i for period t; $Scores_{it}$ = test score for school i for period t; $Salary_{it}$ = average teacher salary for school i for period t; $Income_{it}$ = patron income in school i for period t; $Size_{it}$ = average daily membership for school i for period t; and ε_{it} = error term.

Other researchers have used similar cost functions in evaluating economies of scale. Early studies include Riew (1966) and Cohn (1968). More recent studies that have used various cost functions to estimate economies of scale include Riew (1986), Monk (1990), and Lewis and Chakraborty (1996, 2000). The present contribution is the use of school level data and econometric techniques.

IV. Estimation Procedures and Results

A typical panel data set includes observations in two dimensions: across time and across individuals. The current data set includes observations across time and across two different cross-sections: districts and schools. This is a rich but rather complex type of data set and requires econometric techniques slightly more sophisticated than simple linear regression. As a test of robustness, the parameters of equation (1) are estimated using three separate techniques or models: (1) least-squares dummy variable model (LSDV), (2) fixed effect model, and (3) pooled model.

A. Least-Square Dummy Model Results

The least-squares dummy variable (LSDV) model provides a framework for estimating the parameters of a model using a three-dimensional panel data set. The LSDV model assumes that differences across groups (e.g., schools and districts) can be measured or accounted for by different constant terms. The LSDV version of equation (1) is as follows:

$$(2) \quad \ln Cost_{it} = \alpha_0 + \alpha_1 Scores_{it} + \alpha_2 Salary_{it} + \alpha_3 Income_{it} + \alpha_4 lnSize_{it} + \sum_{i=1}^{19} \beta_i d_i + \varepsilon_{it}$$

where d_1 through d_3 are dummy variables for type of school (i.e., elementary, middle, or high school, and d_4 through d_{19} are dummy variables to account for the 16 different school districts represented in the sample. In terms of Figures 1a and 1b above, the inclusion of the dummy variables allow the cost functions for different districts to be located different distances from the origin. In essence, the inclusion of dummy variables allows for the identification of the effect of school size on cost per student.

Equation (2), as written, cannot be estimated due to perfect multicollinearity.[9] To avoid this problem, it is common to drop one dummy variable from each set of dummies. Thus, the model estimated is as follows:

(3) $$\ln Cost_{it} = \alpha_0 + \alpha_1 Scores_{it} + \alpha_2 Salary_{it} + \alpha_3 Income_{it} + \alpha_4 \ln Size_{it} + \sum_{i=1}^{17} \beta_i d_i + \varepsilon_{it}.$$

Results are reported in Table 1.

TABLE 1

LSDV MODEL REGRESSION RESULTS

Dependent Variable: Cost per student

$R^2 = 0.73$

Independent Variables	Coefficient	*t*-Statistic	White *t*-Statistic
Constant	8.7687	19.1148	19.9860
Test score	0.0015	2.4809	2.2374
Salary	0.0001	2.4210	2.5301
Income	-0.0023	-1.3991	-1.6327
School size	-0.2061	-8.2879	-6.7776

The results clearly indicate the existence of economies of scale in the production of public education in Wyoming. The coefficient of interest, log of school size, indicates that a 1 percent change in school size is associated with approximately a 0.2 percent change in cost per student.

Next equation (3) is tested for heteroscedasticity. Equation (3) is a classical regression model and the standard tests of heteroscedasticity apply. The Breusch-Pogen test (see Greene (1997), Ch. 12) was chosen and yields a test statistic of 15.78, which strongly suggests heteroscedasticity.

If the disturbance term is heteroscedastic, the ordinary least squares (OLS) estimators are still unbiased and consistent, but not efficient. Furthermore, the OLS standard errors of the parameter estimates are biased. Given the evidence of heteroscedasticity, two solutions are pursued: (1) without making any assumptions about the nature of the heteroscedasticity, calculate the OLS estimators but use White's procedure to obtain the unbiased estimates of the standard errors and calculate the corrected *t*-statistics; and (2) make a plausible assumption about the nature of the heteroscedasticity and estimate the model using generalized least squares (GLS). The *t*-statistics calculated using the White standard errors are reported in Table 1. Estimates based on GLS are reported in Table 2. The GLS model was implemented based on the assumption the variance of the disturbance term is proportional to the log of school size.

TABLE 2

LSDV MODEL CORRECTED FOR HETEROSCEDASTICITY

Dependent Variable: Cost per student

$R^2 = 0.939$

Independent Variables	Coefficient	*t*-Statistic
Constant	8.8172	18.7935
Test scores	0.0044	2.8713
Salary	0.0001	2.2601
Income	-0.0024	-1.4151
School size	-0.2095	-8.2535

After correcting for heteroscedasticity and estimating the parameters of the LSDV model, the school size coefficient, again, has the anticipated sign and is significant. Again, these school level data suggest economies of scale are present in providing public education in Wyoming.

B. Fixed Effects Model

As an alternative to the LSDV model discussed above, a fixed effects version of equation (3) was estimated. The fixed effects model allows for a separate intercept term for each school in the sample. Hence, the fixed effects model is an extension of the LSDV model. Rather than have a dummy variable for each district and school type, however, in the fixed effects model there is a dummy variable for each school.

The fixed effects model in the current context is a slightly different approach to estimating economies of scale. By allowing for a different constant term for each school, we are estimating the effects of changes in school size *over time* on average cost. Essentially, the estimate of α_4 in equation (3) provided by the fixed effect model is a weighted average of the effect of changes in school size over the four-year sample period for the 63 different schools.

The results from the fixed effects model are reported in Table 3. The two variables, test scores and income, were highly collinear with the constant term for each of the 63 schools and, hence, were dropped from equation (3) when estimating the fixed effects model. Although the economies of scale parameters is much larger than was estimated using the LSDV model, it still has the anticipated sign and is statistically significant.

C. Pooled Model

The LSDV model includes 17 dummy variables to account for the different school districts and types of schools (e.g., elementary vs. middle school).

TABLE 3

FIXED EFFECTS MODEL (63 INDIVIDUALS (SCHOOLS), 4 YEARS)

Dependent Variable: Cost per student

$R^2 = 0.330$

Independent Variables	Coefficient	*t*-Statistic
Salary	0.0001	3.139
School size	-0.7883	-11.969

The fixed effects model essentially has 63 dummy variables to account for the 63 different schools contained in the sample. There is a corresponding loss in degrees of freedom in both models.[10] The pooled model is an alternative model with fewer parameters to estimate and, hence, designed to conserve degrees of freedom.

In terms of equation (3), the pooled model contains two dummy variables to distinguish school type and a variable called allocation percentage in addition to the variables, scores, salary, income, size, and constant terms. The allocation percentage variable is the percentage of each district's general fund expenditures allocated to individual schools. For example, the District Superintendent's salary is a district expenditure and would not be allocated to an individual school. Of course, the allocation percentage variable is the same for all schools within the same district and, hence, is similar to a district dummy variable and will capture district level effects on cost per student. However, unlike the district dummy variable in the LSDV model, the allocation percentage varies across time. Table 4 reports the parameter estimate from the pooled model. As was the case with the LSDV model, there was evidence of heteroscedasticity in estimating the pooled model. Hence, White *t*-statistics are reported in Table 4, and Table 5 reports parameter estimates from a pooled model corrected for heteroscedasticity.

Table 6 provides a summary of the estimates of the economies of scale parameter from the different models.

V. Significance of Groups Effects

Next, statistical tests are applied within the context of the LSDV model to determine if there are group effects; i.e., do districts have different cost functions? If the 16 dummy variables representing the 16 school districts in the sample are dropped from the equation (3), the restriction is imposed that $\beta_i = \beta_0$ for $i = 4, 5, 6, \ldots, 19$. The hypothesis that all these parameters are all equal is tested with the following *F* test:

$$F = \frac{(SSE^* - SSE)/J}{SSE/(n-k)}$$

TABLE 4

POOLED MODEL RESULTS

Dependent Variable: Cost per student

$R^2 = 0.51$

Independent Variables	Coefficient	*t*-Statistic
Constant	10.4873	34.6371
Test scores	0.0026	1.4701
Salary	-0.0001	-2.3420
Income	-0.0007	-0.5205
School size	-0.2303	-8.5035

TABLE 5

POOLED MODEL CORRECTED FOR HETEROSCEDASTICITY

Dependent Variable: Cost per student

$R^2 = 0.89$

Independent Variables	Coefficient	*t*-Statistic
Constant	10.6667	30.5135
Test scores	0.0029	1.8079
Salary	-0.0001	-2.5130
Income	-0.0010	-0.5515
School size	-0.2316	-8.6594

TABLE 6

SUMMARY OF ESTIMATES OF ECONOMIES OF SCALE PARAMETER

Model	Coefficient on School Size	t-Statistic	Reference
LSDV	-0.2061	-6.7776	Table 1
LSDV corrected for heteroscedasticity	-0.2095	-8.2535	Table 2
Fixed effects	-0.7883	-11.9690	Table 3
Pooled model	-0.2303	-8.5035	Table 4
Pooled model corrected for heteroscedasticity	-0.2316	-8.6594	Table 5

where SSE^* = the sum of squared errors from the restricted regression (i.e., d_4 through d_{19} dropped from equation (3)); SSE = the sum of squared errors from the complete LSDV model; J = number of restrictions (i.e., J = 16); n = number of observations; k = number of parameters estimates in the complete LSDV model.

Applying the above F-test yields a test statistic of 5.8422. The critical value of the 1 percent level is approximately 2.00. Thus, there is a strong indication of district effects. In terms of Figures 1a and 1b above, districts have cost functions located different distances from the origin.

VI. Summary and Conclusions

Reference to Table 6 indicates that the economies of scale parameter is rather independent of the econometric model used to estimate the average cost equation. Based on this analysis, there is strong evidence of economies of scale in Wyoming public education. Furthermore, there is evidence of district effects; i.e., the location of the average cost function in the output/average cost plane depends on the district. This implies that cost studies need to control for these district effects when comparing schools from different districts.

VII. Appendix

Sixteen Wyoming's school districts provided expenditure and enrollment data by school for four years: the academic years 1994-95, 1995-96, 1996-97, and

1997-98. Expenditures are coded by object (e.g., salaries vs. materials) and by function (e.g., instruction vs. instructional support). School size ranges from 3 to over 1,500, with approximately 80 schools represented.[11]

Standardized test results are available for every public school in Wyoming. Students in grades 4, 8, and 11 are tested in three categories: reading, writing, and mathematics. National percentile scores for each school and for each category are provided. For a measure of school output, this study uses the average of the national percentile scores.

As a measure of the income level of the students families, this study uses the percentage of the studentbody not eligible for federal free and reduced lunch programs. As a measure of input prices, the average teacher salary in each district for each type of school (e.g., elementary, middle, high school) is used.

References

Betts, Julian. "Is There a Link Between School Inputs and Earnings? Fresh Serating of an Old Lliterature." In Gary Burtless, ed., *Does Money Matter? The Link Between Schools, Student Achievement, and Adult Success*, Washington, D.C.: The Brookings Institute, 1995.

Busch, Carolyn, and Allan Odden. "Introduction to the Special Issues: Improving Educational Policy and Results with School-Level Data–A Synthesis of Multiple Perspectives." *Journal of Education Finance* 22(Winter 1997):225-245.

Berne, Robert, Leanna Stiefel, and Michele Moser. "The Coming Age of School-Level Finance Data." *Journal of Education Finance* 22(Winter 1997):246-254.

Campbell County School District et al. v. State of Wyoming et al. 907 P 2d 1238 (1995) Wyo.

Chakraborty, Kalyan, Basudeb Biswas, and W. Cris Lewis. "Economies of Scale in Public Education: An Econometric Analysis." *Contemporary Economic Policy* 18(2, April 2000):238-247.

Cohen, Matthew C. "Issues in School-Level Analysis of Education Expenditure Data." *Journal of Education Finance* 22(Winter 1997):255-279.

Cohn, Elchanan. "Economies of Scale in Iowa High School Operations." *Journal* *of Human Resources* (Fall 1968):422-434.

Downes, T.A., and T. F. Pogue. "Adjusting School Aid Formulas for the Higher Cost of Educating Disadvantaged Students." *National Tax Journal* (March 1994):422-434.

Farland, Gary. "Collection of Fiscal and Staffing Data at the School-Site Level." *Journal of Education Finance* 22(Winter 1997):280-290.

Goertz, Margaret E. "The Challenges of Collecting School-Based Data." *Journal of Education Finance* 22(Winter 1997):291-302.

Greene, William H. *Econometric Analysis*. 3rd edition. Upper Saddle River, NJ: Prentice Hall, 1997.

Hanushek, Eric A. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24(September 1986):1141-1177.

Hoxby, Caroline M. "Improving Student Achievement: Examining the Impact of Teacher Quality and Class Size." Testimony prepared for the U.S. House Subcommittee on Postsecondary Education, Training, and Life-Long Learning, Thursday, April 29, 1999.

Hoxby, Caroline M. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics* 15(4, November 2000):1239-1285.

Lewis, W. Cris, and Kalyan Chakraborty. "Scale Economies in Public Education." *Regional Analysis and Policy* 26(1, 1996):23-35.

Monk, David H. "Educational Costs and Small Rural School." *Journal of Education Finance* 16(Fall 1990):213-225.

Monk, David H. "The Organization and Reorganization of Rural Schools." In *The Rural Education in America*, A. De Young (ed.), New York: Garland Press, 1990.

Monk, David H. "Challenges Surrounding the Collection and Use of Data for the Study of Finance and Productivity." *Journal of Education Finance* 22(Winter 1997): 303-316.

Riew, John. "Economies of Scale in High School Operations." *Review of Economics and Statistics* (August 1966):280-287.

Riew, John. "Scale Economies, Capacity Utilization, and School Costs: A Comparative Analysis of Secondary and Elementary Schools." *Journal of Education Finance* (Spring 1986):433-446.

Sarravo v., Priest. 487 P 2d 1241 (1971) Cal.

Endnotes

1. For an early ruling requiring less inequality in spending, see Sarravo v. Priest, 5 Cal. 3d 584, 487 P. 2d 1241 (1971). For a recent ruling requiring the state to develop a funding scheme to achieve "equal educational opportunity," see Campbell County School District et al. v. State of Wyoming et al. 907 P. 2d 1238 (1995) Wyo.

2. There is a rather large school finance literature dealing with economies of scale (see, for example, Chakraborty, Biswas, and Lewis (2000)). The empirical analysis in these studies, however, generally depends upon expenditure data at the district level and, hence, misses the important issue of economies of scale at the school level.

3. Expenditures are delineated by both functions (e.g., instruction vs. instructional support) and object (e.g., salaries vs. materials and supplies).

4. A production function specifies the technical relationships among inputs and output.

5. See Hanushak (1986) for a discussion of the lack of a well-defined education production function. See Monk (1990) for a concise discussion of costs vs. expenditures.

6. It is not uncommon for smaller districts to receive more funds per student than larger districts. In our hypothetical example, district B is smaller than

A.

7. Chakraborty et al. explain that their specification of the education cost function relies upon that specified by Downes and Pogue (1994).

8. For additional detail concerning the data set, see the appendix.

9. The problem arises from the obvious fact that $\sum_{i=1}^{3} d_i = \sum_{i=4}^{19} d_i = 1$.

10. The number of degrees of freedom generally equals the number of observations in the sample less the number of parameters estimated. Conserving degrees of freedom is preferable since the standard errors (i.e., precision) of the parameter estimates are usually smaller, the larger is the number of degrees of freedom.

11. Technically, a data set with observations across units (e.g., schools) and across time is referred to as a *panel* data set. Such data sets are a rich source of information and are popular in economic research.

PAPER 2

FARM HOUSEHOLD WEALTH: MEASUREMENT, STRUCTURE, AND DETERMINANTS

Abstract

This paper uses the 1992 Survey of Consumer Finances public release data set, as well as the 1992 (Wave 1) Health and Retirement Survey data set to document the differences in farm household and nonfarm household wealth. We also attempt to provide possible explanations for the differences. This paper has important policy implications considering the extent of income transfer programs to farm households and the structure of the U.S. tax code.

Introduction

This paper documents in detail the differences in farm household and nonfarm household wealth structure and attempts to shed light on potential explanations for the differences. Particular attention is paid to households at or near retirement.

A better understanding of farm household wealth should be useful in the farm policy debate. Knowledge concerning the level and structure of farm household wealth is relevant to the decision of whether to transfer wealth and income from nonfarm households to farm households. Furthermore, knowledge of the determinants of farm household wealth should help guide policy designed to affect farmers' saving decisions. Whether an average farmer's wealth at retirement is largely determined by income and/or circumstances beyond his control as opposed to the decision of how much to save, is an important policy question. If retirement wealth is relatively unaffected by saving decision, policies designed to promote wealth accumulation through promoting saving (e.g., estate tax repeal) will be ineffective (see Venti and Wise for more on this point).

Data

Household wealth is not a simple parameter to measure. There is a long list of asset types and the valuation of specific assets is problematic (e.g., business interests, real estate, defined benefit pension plans, social security benefits, etc.).[1] Fortunately for researchers interested in wealth issues, two household surveys have made a concentrated effort to gather detailed household wealth information: *The Survey of Consumers Finances* (*SCF*) and the *Health and Retirement Survey* (*HRS*).[2] A brief description of the general characteristics of each data set is provided below along with a description of how each survey identifies farm households.[3]

Survey of Consumer Finances

Table 1 presents a summary of the characteristics of these two data sets. Table 2 provides the major wealth variables collected by each survey The *SCF* is a cross-sectional data set conducted every three years by the Federal Reserve. The focus is on household wealth with all ages of the household head included. The *SCF* allows researchers to identify farm households through the following questions:

a. Where does respondent live? (possibilities include farm or ranch)

b. Do you operate a farming or ranching business on this property?

c. What is the value of farmland and buildings?

d. Do you work for someone or are you self-employed?

e. What kind of business do you work in?

We use question b above to identify farm households in the *SCF* survey

A strength of the *SCF* is the detailed questions on financial assets, nonfinancial assets, and liabilities. A weakness, perhaps, is that all ages of household heads are included. Given the sample size and the lack of a focus on a certain age group, it is difficult to come to strong conclusions about the structure of wealth of given groups while controlling for age. Finally, the *SCF* oversamples high net worth families and thus provides meaningful estimates of population parameters.[4]

Health and Retirement Study

The *HRS* is sponsored by the Michigan Center on the Demography of Aging. It is similar to the *SCF* in terms of the detailed wealth information it collects. Similar to the *SCF*, it allows researchers to identify farm families with the following questions:

a. Does respondent live on a farm or ranch?

b. Do you own this farm/ranch; do you own part of it; do you rent it?

c. What kind of business or industry do you work in—that is, what do you do or make at the place where you work?

d. Do you work for someone else, are you self-employed, or what?

We use question b above to identify farm households in the *HRS* data set. For purposes of comparing farm household wealth with other groups, an advantage of the *HRS* is its focus on individuals at or near retirement. At the time of wave 1 of the survey (1992), the average age of the respondent was 56. Furthermore, the sample is much larger than the *SCF* and the *HRS* is a panel data set. For questions concerning the structure of wealth of comparable groups at or near retirement and how these individuals are or will fare during retirement, the *HRS* survey probably provides more focused information than the *SCF*.

Descriptive Statistics

Before presenting wealth data, Figures 1a and 1b show total household income from the *SCF* and *HRS* data sets, respectively. Mean farm household income from both surveys was approximately $52,000. Mean nonfarm household income was approximately $38,000 in the *SCF* data set and $50,000 in the *HRS*. The average age of farm and nonfarm households is 52.7 and 48.4 in the *SCF* survey, respectively, and 56.6 and 56.1 in the *HRS* data set, respectively. Comparing the results from the *HRS* to the *SCF*, were on average the respondents are younger, shows an interesting result: Nonfarm household income is higher in the *HRS* survey while farm household income is lower in the *HRS* relative to the *SCF*.

For comparison, the *Structural and Financial Characteristics of U.S. Farms. 2001 Family Farm Report* (hereafter *Family Farm Report*) estimates mean income of $59,700 in 1998 for farm households and $51,900 for all U.S. households. (Data from the *SCF* and *HRS* are for 1992.) Hence, the mean income estimates from the *SCF* and *HRS* appear consistent with other sources.

Farm household income and age are higher than comparable variables for nonfarm households but it is our hypothesis that these differences do not explain the wealth dispersion between farm and nonfarm households. Before this hypothesis is formally tested, Figures 2a through 5b present detailed information on the structure of farm and nonfarm household wealth.

The *SCF* and *HRS* surveys, respectively, yield net worth estimates of farm households of approximately $650,000 and $435,000. The *SCF* and *HRS* surveys result in mean net worth estimates for nonfarm households of approximately $180,000 and $260,000, respectively, in 1992. (The *Family Farm Report* estimates net worth of family farm households at approximately $500,000 and reports that the mean net worth of all U.S. households was $282,500 in 1998.) Again, the results appear consistent with other sources. The remarkable result is the large wealth difference between farm and nonfarm households.

Figures 3a and 3b report financial asset wealth for the two different types of households (see Table 2 for the definitions of financial wealth in the two data sets). These two tables show that the difference in farm and nonfarm financial wealth are not as great as the differences in total net worth. Indeed, the *HRS* implies that nonfarm families have slightly more financial wealth than nonfarm families.

Figures 4a and 4b report estimates of retirement account balances. Both data sets imply that nonfarm households have larger retirement account balances. Finally, Figures 5a and 5b illustrate the wealth distribution.

Determinants of Wealth Dispersion

Farm households have significantly higher levels of wealth than nonfarm households. Differences in income levels, investment choice, and age do not appear on the surface to explain the dispersion between farm household and nonfarm household wealth. It is our hypothesis that the explanation for higher farm household net worth is farm household saving behavior.

To test the hypothesis that farm households choose to save more than nonfarm households we follow Venti and Wise and "attribute to saving choice the dispersion that remains after accounting for . . .circumstances that limit or enhance resources." The following specification is used to control for factors, other than saving choice, that determine wealth:[5]

(1) $$\text{Net worth} = \alpha + \beta_1 \cdot \text{Age} + \beta_2 \cdot \text{Amount of inheritances} + \beta_3 \cdot \text{Income} + \varepsilon .$$

As a preliminary procedure, equation (1) is estimated using (a) the entire sample, (b) farm households, and (c) nonfarm households and then the Chow test is applied.[6] Table 3 reports the results of the three estimates. The F-statistic (i.e., Chow test statistic) is 14.34. Therefore, we reject the hypothesis that the coefficients of equation (1) are equivalent across the two subsamples. The conclusion that farm household wealth is not affected by income, age, and inheritances equivalently to how nonfarm household wealth is affected by these variables implies that saving behavior, the variable left out of equation (1), also is fundamentally different across the two equations.

For additional insight into possible differences in saving behavior, the coefficients from equation (1), estimated using the total sample, were used to calculate predicted farm household wealth. Based on these estimated population parameters and farm household characteristics, farm household net worth is predicted at $255,300.[7] The fact that observed farm household net worth is $433,699 implies income, inheritances, and age do not explain the dispersion in wealth between farm households and nonfarm households. Our preliminary conclusion is that it is saving behavior that explains this dispersion.

Conclusions

Farm households have higher net worth than nonfarm households. Differences in income, inheritances, and age do not appear to explain the difference. Based on our preliminary analysis, we attribute higher farm household wealth to the saving behavior of farm households.

References

Hoppe, R.A. (editor). "Structural and Financial Characteristics of U.S. Farms: 2001 Family Farm Report." Economic Research Service, U.S. Department of Agriculture, Agriculture Information Bulletin No. 768, Washington, D.C., May 2001.

Juster, F.T., J.P. Smith, and F. Stafford. "The Measurement and Structure of Household Wealth." *Labour Economics* 6 (1999):253-275.

Smith, J.P. "Racial and Ethnic Differences in Wealth." *Journal of Human Resources* 30(1995): S158-S183.

Venti, S.F., and D.A. Wise. "Choice, Chance, and Wealth Dispersion." National Bureau of Economic Research, Working Paper 7521, February 2000.

Endnotes

1. See Juster, Smith, and Stafford (1999) for a discussion of the methodological issues that arise in measuring household wealth.

2. Two other national surveys contain wealth models but with less detail that the *SCF* and *HRS* surveys: *The Panel Study of Income Dynamics* (*PSID*) and the *Survey of Income and Program Participation* (*SIPP*). Furthermore, both the *PSID* and *SIPP* contain very few observations from the top of the wealth distribution and hence fail to produce reliable estimates of the the wealth distribution (Juster, Smith, and Stafford 1999).

3. Smith (1995) provides a more detailed discussion of the *HRS* data set; Juster, Smith, and Stafford (1999) discuss the *SCF* survey.

4. Given that the U.S. wealth distribution is extremely positively skewed and the paucity of observations of high-wealth households, oversampling of high-wealth households is necessary. Without such oversampling, the sample "may routinely miss virtually everyone of the top end of he wealth distribution" (Juster).

5. Venti and Wise use a similar specification to test for saving behavior.

6. See Greene, pp. 349-353, for a discussion of the Chow test.

7. The mean net worth for the entire sample (i.e., farm and nonfarm households) was $241,919 and $234,450, respectively.

Table 1. A Summary of the Characteristics of Major Data Sets that Contain Information on the Structure of Wealth of U.S. Households

Characteristic	*SCF*	*HRS*
1. Sponsor	Federal Reserve	Michigan Center on the Demographics of Aging
2. Unit of observation	Household	Household
3. Cohorts covered	All	Individuals at or close to retirement
4. Oversample	Wealthy	African Americans, Hispanics, residents of Florida
5. Sample size	4,500 households	7,600 households
6. Type of data set	Cross-section	Panel
7. Identify age?	Yes	Yes
8. Identify farm operators?	Yes	Yes
9. Identify self-employed?	Yes	Yes
10. Nature and value of financial assets	Detailed	Detailed
11. Nature and value of nonfinancial assets	Detailed	Detailed
12. Questions on IRA/KEOGH account balances?	Yes	Yes
13. Questions on defined benefit pensions?	Yes	Yes
14. Questions on Social Security benefits?	Yes	Yes

Table 2. Asset Categories of the *HRS* and *SCF* Data Sets

HRS[1]	*SCF*[1]
1. Housing equity	1. Liquid assets
2. Vehicles	2. CDs
3. CDs and government bonds	3. Total mutual funds
4. Checking, savings, & money market accounts	4. Stocks
5. Stocks, mutual funds, & investment trusts	5. Bonds
6. Bonds & bond funds	6. Retirement assets
7. Business equity	7. Savings bonds
8. IRAs and KEOGHs	8. Cash value of life insurance
9. Other assets	9. Other managed assets
	10. Other financial assets
	11. Vehicles
	12. Houses
	13. Other residential real estate
	14. Net equity in nonresidential real estate
	15. Business interests
	16. Other nonfinancial assets
Financial assets: 3 + 4 + 5 + 6 + 8	Financial assets: Sum of 1-10
Retirement accounts: 8	Retirement accounts: 6

[1]In general, the *HRS* asks for asset values net of associated debt. The *SCF* asks for gross values and contains another section that gathers detailed debt information.

Table 3. Test for Structural Differences in Wealth Equation: Farm Households versus Nonfarm Households

Coefficient*	Total Sample	Farm Households	Nonfarm Households
Constant	-507,872 (-8.46)*	-22,670 (-0.06)	-523,860 (-8.68)
Age	8,766 (8.28)	4,144 (0.62)	8,906 (8.36)
Inheritances	1.318 (15.62)	1.062 (1.23)	1.314 (15.76)
Income	4.763 (48.46)	3.968 (4.34)	4.781 (49.10)

*t-statistics are in parentheses.

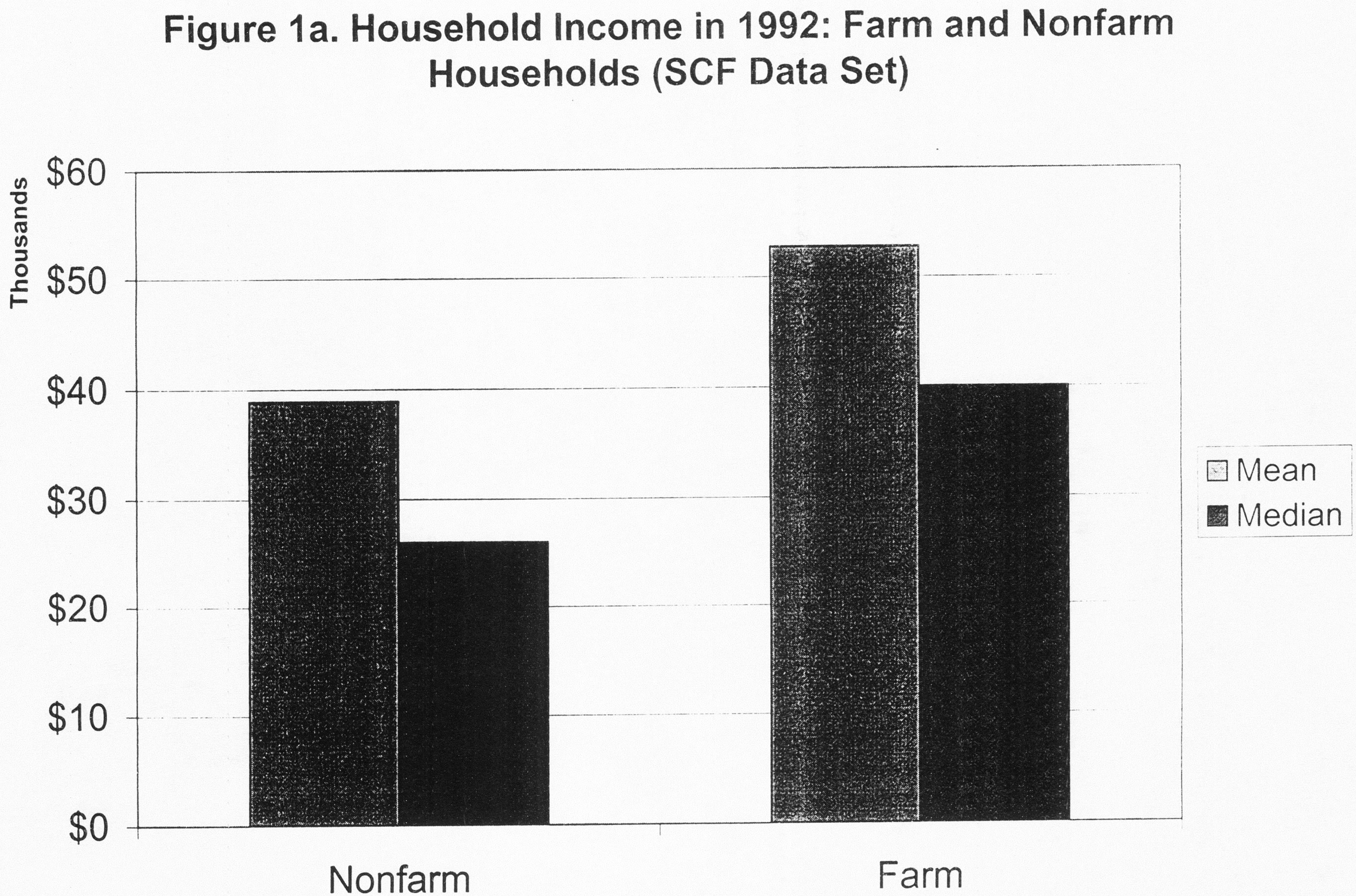


Figure 1a. Household Income in 1992: Farm and Nonfarm Households (SCF Data Set)
Thousands
$60
$50
$40
$30
$20
$10
$0
Nonfarm
Farm
Mean
Median

Figure 1b. Household Income in 1992: Farm and Nonfarm Households (HRS Data Set)

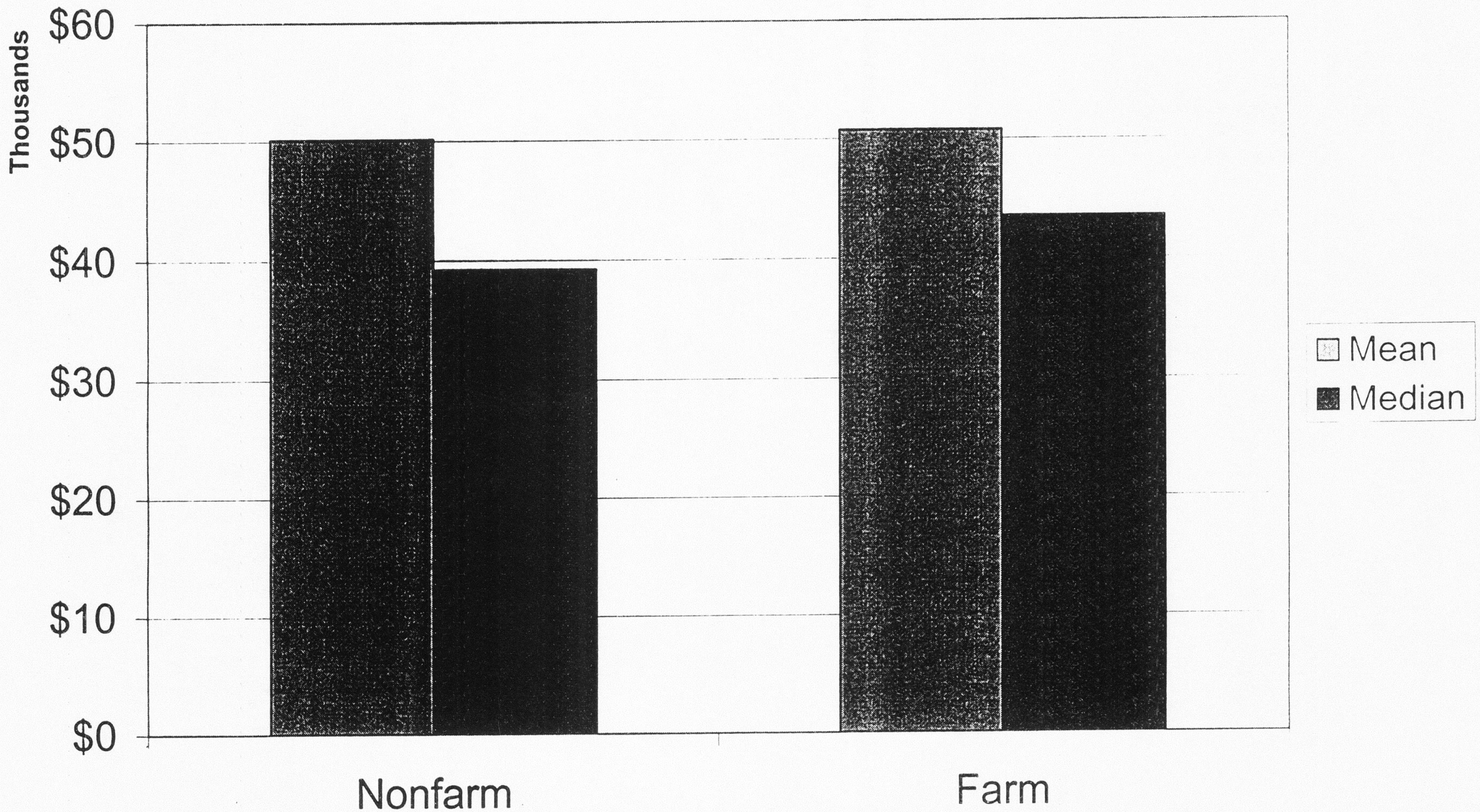


Figure 2a. Net Worth: Farm and Nonfarm Households
SCF Data Set

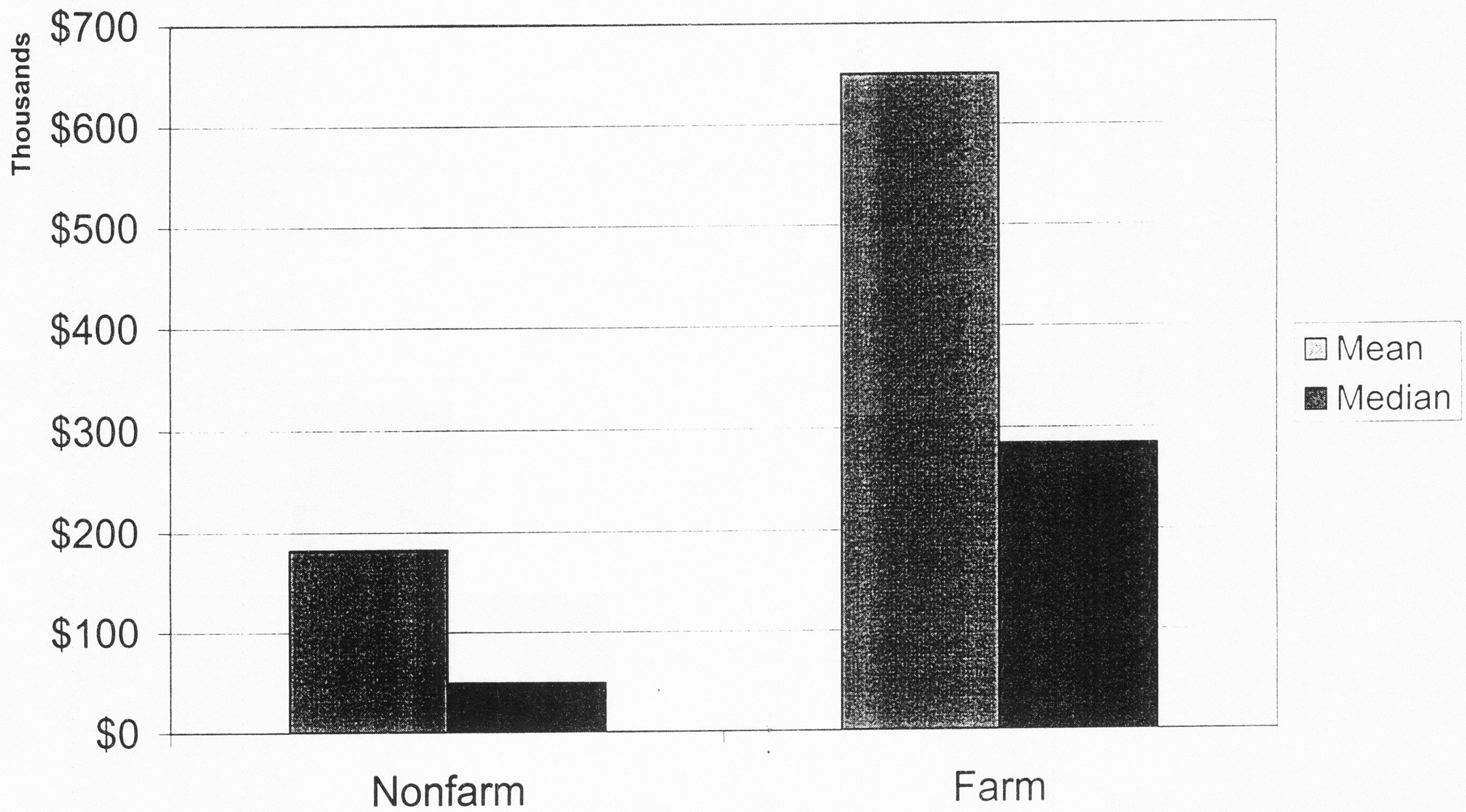


Figure 2b. Net Worth: Farm and Nonfarm Households
HRS Data Set

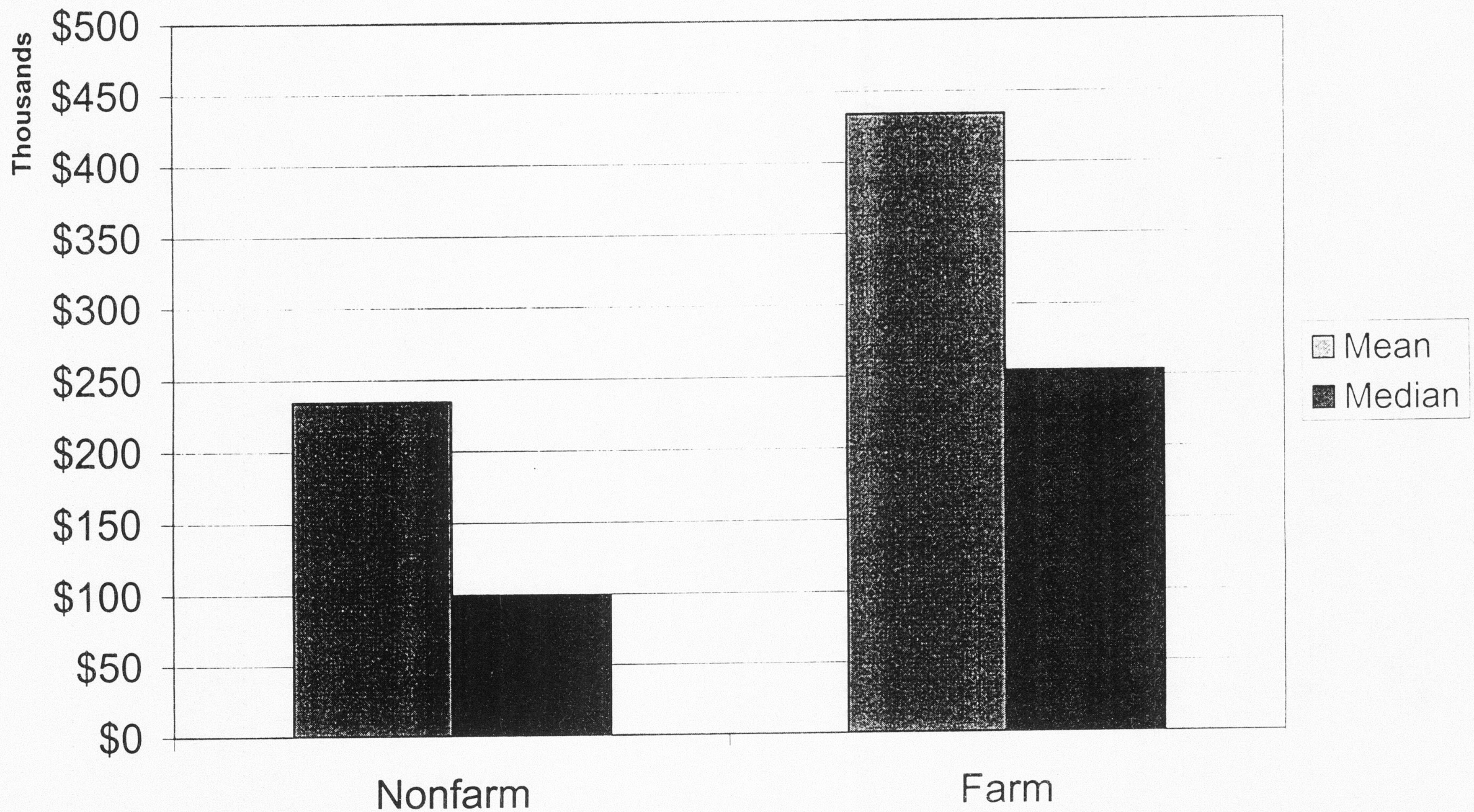


Figure 3a. Value of Financial Assets: Farm and Nonfarm Households (SCF Data Set)

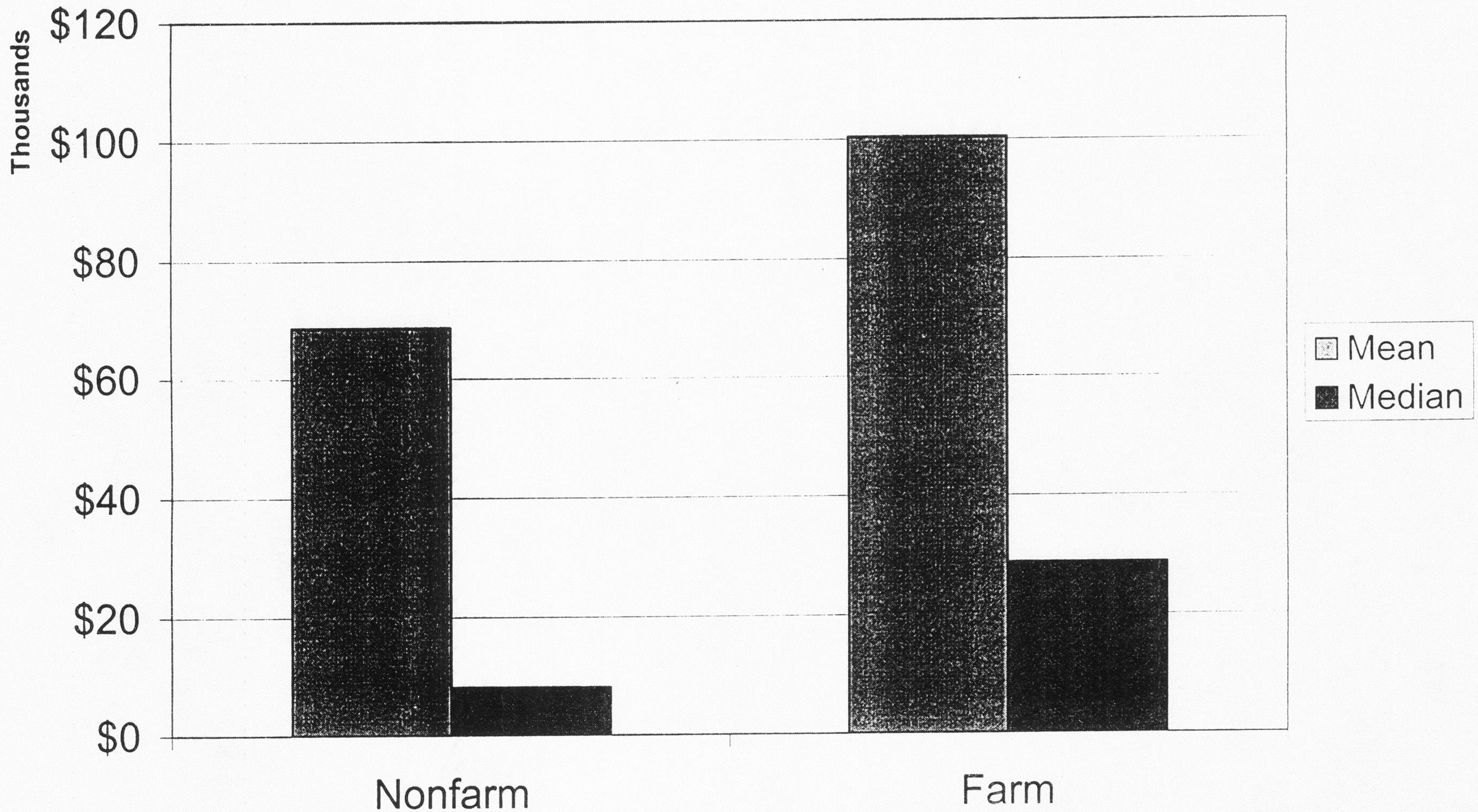


Figure 3b. Value of Financial Assets: Farm and Nonfarm Households (HRS Data Set)

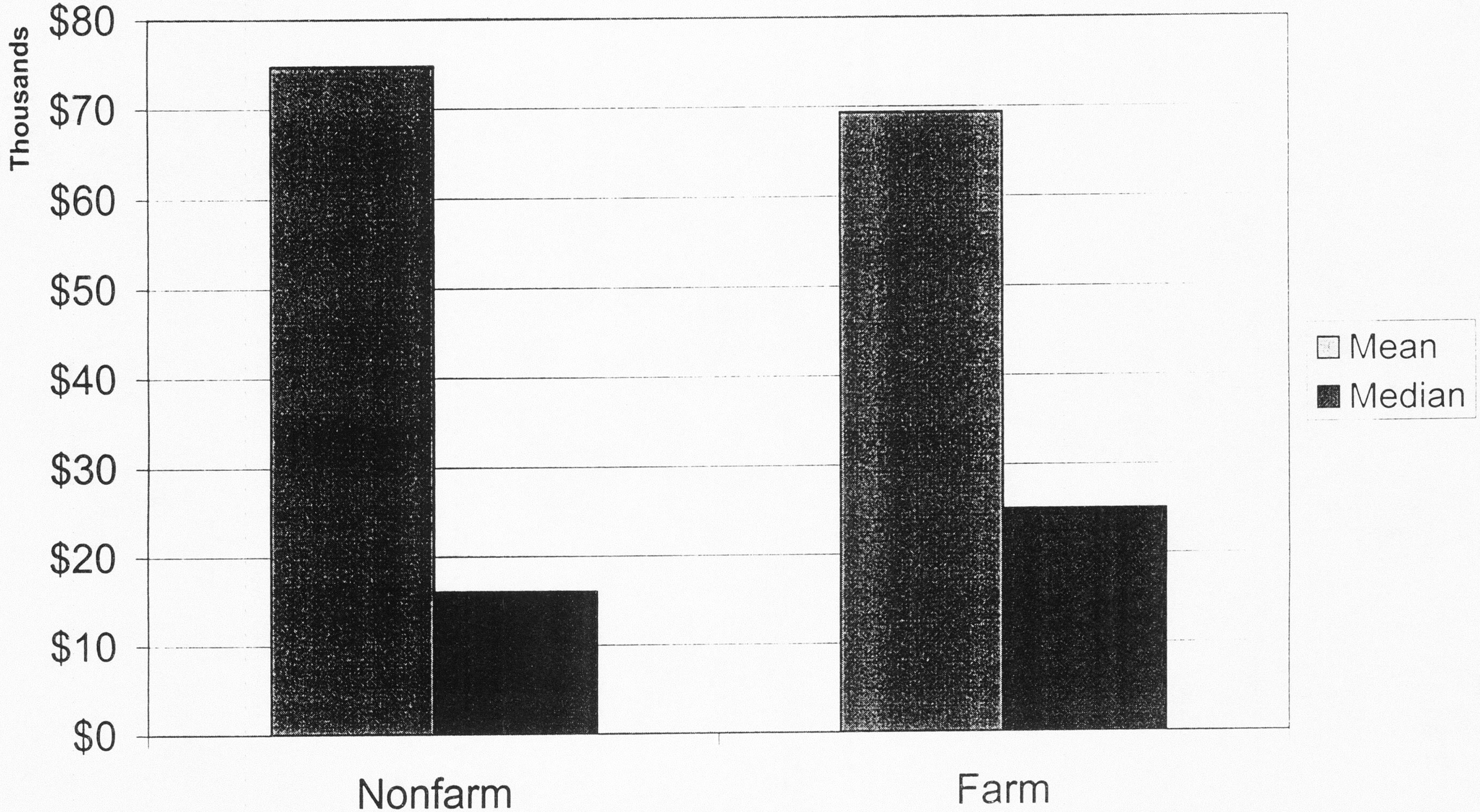


Figure 4a. Value of Retirement Accounts: Farm and Nonfarm Households (SCF Data Set)

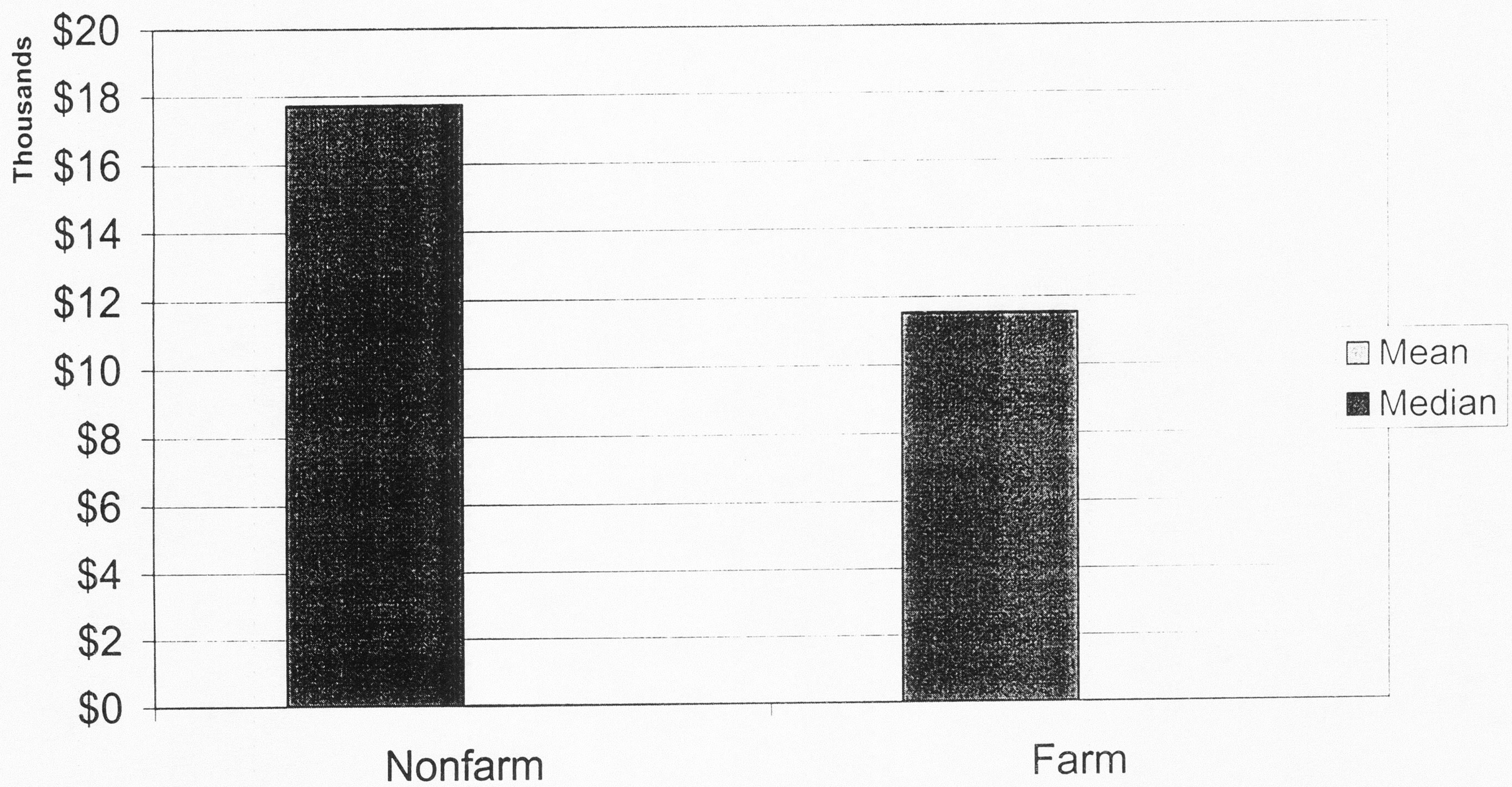


Figure 4b. Value of Retirement Accounts: Farm and Nonfarm Households (HRS Data Set)

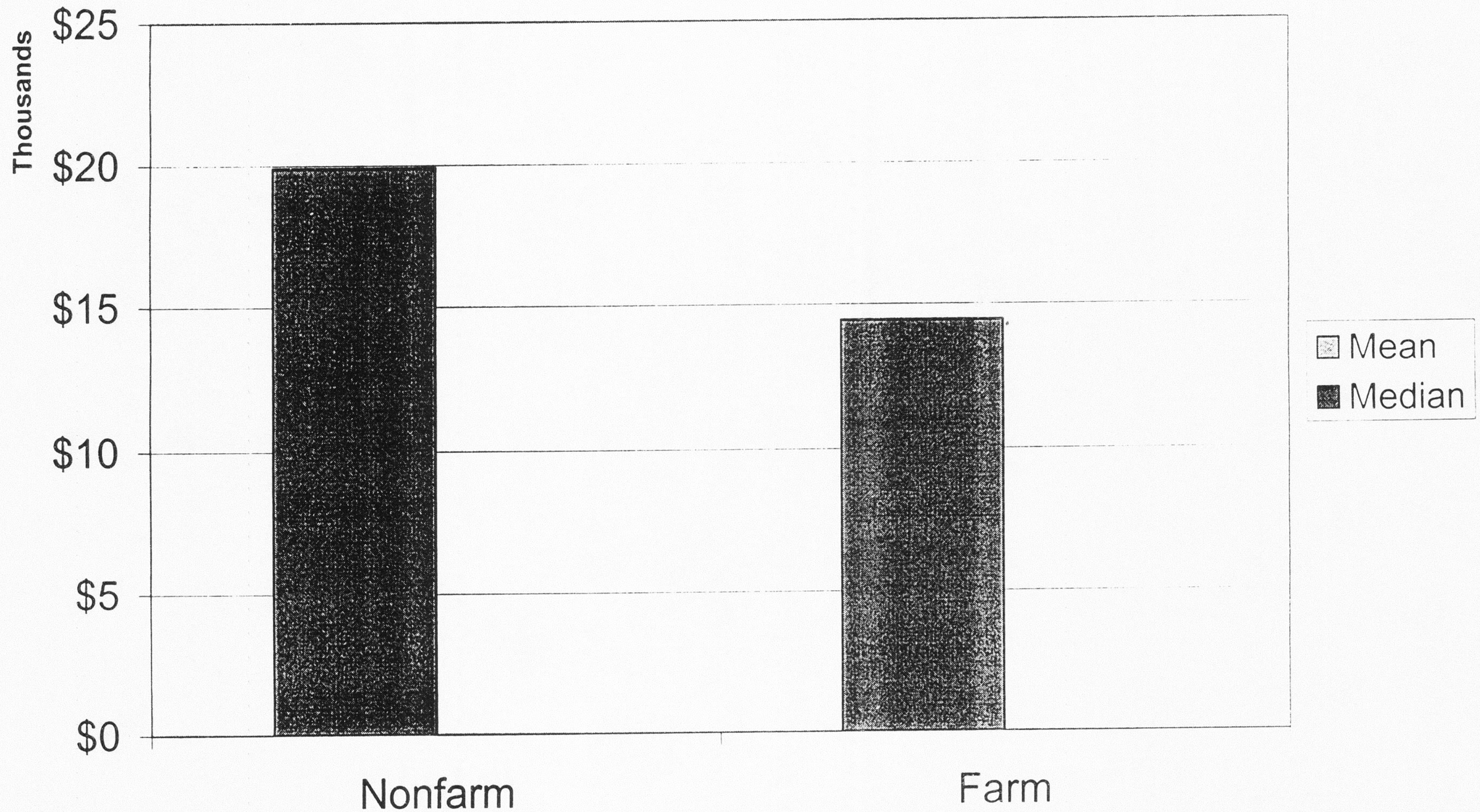


Figure 5a. Distribution of Household Mean Wealth
SCF Data Set

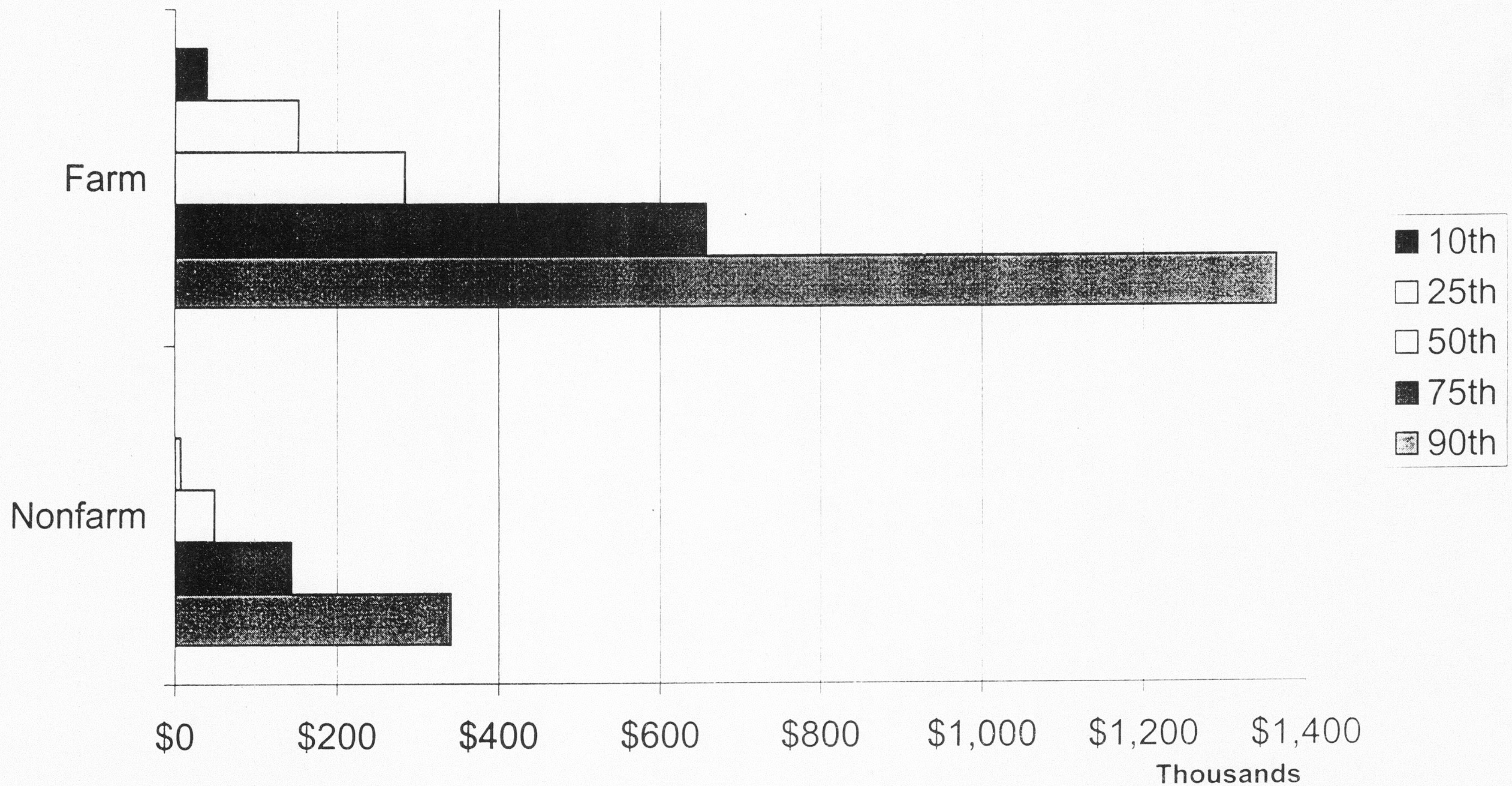


Figure 5b. Distribution of Household Mean Wealth
HRS Data Set