Journal of Student Financial Aid

Volume 6 | Issue 3 Article 3

12-1-1976

Can Need Analysis Procedures be Simplified Through Stepwise Regression Analysis?

Dwight H. Horch

CAN NEED ANALYSIS PROCEDURES BE SIMPLIFIED THROUGH STEPWISE REGRESSION ANALYSIS?

By Dwight H. Horch

Statement of the Problem

Determination of the amount parents are expected to contribute toward the cost of postsecondary education involves several financial and nonfinancial variables of the family considered by experts in the field to influence ability to pay. It involves an assessment of the family's taxable and nontaxable income, federal taxes (FICA and income tax), state and other taxes, unusual medical and dental expenses, emergency expenses (casualty and theft losses) family size, family assets and debts, and the number of family members in college. These variables are used to determine the parents' effective or available income, from which parents' contribution is calculated.

Over the years, the techniques for measuring parental ability to pay have become progressively refined, as an outgrowth of efforts by the financial aid community to enhance both horizontal and vertical equity. Historically, these improvements have led to more sophisticated computational algorithms, but the number of items used to calculate parents' contribution has not increased significantly since the last 1950s. On the contrary, as a result of the deliberations of the National Task Force on Financial Aid Problems, chaired by Frances Keppels and its development of a common application form, the number of items entering into the computation has been somewhat reduced.

Despite widespread acceptance of current need analysis procedures, recurr-

Dwight Horch is currently Program Director for the Graduate and Professional School Financial Aid Service, which is administered by Educational Testing Service (ETS), Princeton, New Jersey. He is presently serving on the National Task Force Follow-Through Committee on Student Aid Problems.

ing sentiment has been expressed, most recently by the Keppel Task Force, favoring simplification of financial aid forms and computation procedures. The purpose of the present study was to investigate, through stepwise regression analysis, which financial and nonfinancial variables are most effective in predicting parents' contribution.

Other similar studies have been conducted. In 1958, Olsen conducted a study of 811 College Scholarship Service (CSS) applicants to determine whether CSS procedures could be simplified.[1] For this sample, Olsen calculated an abbreviated parental contribution based only on before-tax income and number of dependent children. Regression weights were then developed for the prediction of CSS parents' contribution from the abbreviated contribution and the sum of three allowances: housekeeping, schooling, and other dependents. The best-weighted combination of these two independent variables yielded a .990 correlation with full CSS calculations of contribution. Analysis of discrepancies between predicted contributions and CSS calculations led Olsen to suggest screening out families with net incomé in excess of $11,000. In effect, this resulted in a less skewed distribution of the criterion variable, full calculation of CSS contribution, and in a reduction of the standard error of prediction.

Orwig and Jones (1970) conducted a stepwise regression analysis of a sample of the ACT Family Financial Statement filing population to determine whether need analysis procedures could be simplified.[2] Their study revealed that the asymtotic level of prediction was reached at a multiple correlation of .77 with 10 variables, listed below in the order in which they entered the equation.

1. Parents' Federal Tax
2. Real Estate Equity
3. Parents' Investments
4. Business or Farm Net Worth
5. Parents' Savings
6. Nontaxable Income
7. Home Equity
8. Parents' Trusts
9. Parents' Taxable Income
10. Parents' Other Debts

Orwig and Jones then used these 10 variables, plus selected nonfinancial variables such as family size, to compute an abbreviated ACT contribution, which was then compared with the full contribution. Worthy of note is the fact that their computation of abbreviated contribution did not involve application of regression weights to the independent variable. The authors found a .99 correlation between abbreviated calculations and full calculations, regardless of income group, and mean differences between the two types of calculation ranging from $105 for the lowest income group ($0-$5000) to $403 for the highest income group ($15,000 and over). Orwig and Jones found that two items — parent's nontaxable income and income tax — predicted contribution from *income only,* with a multiple R of .91. They were intrigued with the possibility of using a regression model to predict parents' contribution in their operational program, but suggested inadequacies of such an approach (1) because the regression model would not be computable for families with no nontaxable income and paying no federal tax, (2) because financial data are generally not normally distributed, and (3) because the procedure would be inflexible for individuals with special circumstances, which were accounted for in the full calculation.

Method and Procedure

In the present study, a 5 percent sample, consisting of 45,199 records was systematically selected, after a random start, from the 1973-74 CSS master file, which contained parental information for more than 900,000 students applying for financial aid to be used in the 1974-75 academic year.

An analysis file was then created for the sample, which contained the following information from each student record: parents' net income, other investments, number of children in college, other real estate equity, business or farm net worth, home equity, number of dependent children, housekeeping allowance, bank accounts, medical expenses, state income tax, emergency expenses, debt outstanding, federal income tax, other dependents, indebtedness, and parent's contribution calculated according to CSS formulas in effect for the academic year 1974-75.

Next, student records were grouped according to parents' net (before tax) income, and a stepwise regression analysis was run separately by income group and then for the entire sample. Parents' contribution, computed according to CSS procedures, served as the dependent variable for these analyses, and the remaining variables carried in the sample file were entered as independent variables. These analyses resulted in a predictor battery for each income group and for the entire sample. Separate equations were run for each income group because it was anticipated that a predictor battery computed from the entire sample would overpredict contributions for lower income families, and underpredict for higher income families.

Results and Discussion

An intercorrelation matrix was produced for the entire sample (See Appendix) to gain an overview of the relationships among the independent variables and of the independent variables with the dependent variable, CSS parents' contribution. Table 1, derived from the Appendix, lists the zero order correlations of each independent variable with parents' contribution. Inspection of Table 1 suggests that net income, federal income tax, state income tax, other investments, bank accounts, and home equity correlate most highly with parents' contribution. It should be noted that these correlation coefficients are probably somewhat attentuated due to the likelihood that the distributions of the variables being correlated differ in shape and degree of skewness.

For the stepwise regression analysis run for the entire sample, forward selection program was employed, which selected, at the first step, the most powerful predictor of parents' contribution (that is, the predictor with the highest partial correlation).

TABLE 1
Zero Order Correlations of Independent Variables with CSS Parents' Contribution

Independent Variable	Zero Order Correlation
Number of Dependent Children	—.07
Net Income	.81
Federal Income Tax	.71
State Income Tax	.51
Housekeeping Allowance	.19
Medical Expenses	—.06
Emergency Expenses	.00
Indebtedness	—.02
Other Dependents	—.01
Home Equity	.41
Business or Farm Net Worth	.21
Bank Accounts	.29
Other Investments	.39
Debt Outstanding	.05
Number of Children In College	—.05
Other Real Estate Equity	.28

TABLE 2
Stepwise Prediction of CSS Parents' Contribution
Total Sample
N=45,199

Step	Independent Variable	Contribution to R^2 at this Step	Multiple R at this Step	Standard Error at this Step
1	Net Income	.6480	.81	$1,303.1
2	Other Investment	.0624	.84	$1,181.94
3	Number of Children in College	.0413	.87	$1,094.48
4	Other Real Estate Equity	.0343	.89	$1,016.1
5	Business or Farm Net Worth	.0194	.90	$ 968.9
6	Home Equity	.0165	.91	$ 926.8
7	Number of Dependent Children	.0132	.91	$ 891.9
8	Housekeeping Allowance	.0085	.92	$ 868.7
9	Bank Accounts	.0049	.92	$ 854.9
10	Medical Expenses	.0031	.92	$ 846.3
11	State Income Tax	.0024	.92	$ 839.4
12	Emergency Expenses	.0019	.93	$ 833.8
13	Debt Outstanding	.0007	.93	$ 831.7
14	Federal Income Tax	.0007	.93	$ 829.7
15	Other Dependents	.0001	.93	$ 829.3
16	Indebtedness	.0001	.93	$ 829.3

Successively powerful predictors were then dropped into the equation step. Results of this analysis appear in Table 2. The data in this table indicate that, for the total sample, net income is the most important single predictor of parent's contribution, followed by other investments, number of children in college, other real estate equity, business or farm net worth, home equity, and number of dependent children.

The seven variable predictor battery had a multiple R of 91 with parents' contribution, and the standard error of prediction at the seventh step was $891.90. The remaining independent variables each contribute to less than 1 percent of the criterion variance, and their inclusion in the battery does not raise the multiple R appreciably nor reduce the standard error to any practical extent.

Although the high (.91) multiple R is encouraging, the large standard error of prediction ($892) suggests that an overall predictor battery would not yield very precise predictions of parental contributions for *individuals*. This was not an unexpected finding, and for this reason separate analyses were conducted for each of six subgroups of families. Families were grouped by net (before tax) income as follows:

Group	Net Income
1	$ 0 - 4,999
2	$ 5,000 - 9,999
3	$10,000 - 14,999
4	$15,000 - 19,999
5	$20,000 - 24,999
6	$25,000 - over

The results of these analyses appear in Table 3-8 on the following pages. Not surprisingly, the order of predictors is different for each income group. For lower income families, for example, the two most potent predictors of CSS parents' contribution are other real estate equity and business or farm net worth. This finding is not startling when one considers that the CSS expects very little or no contribution from *income* before $5,000; therefore any calculation of CSS contribution would result only for families in the lowest income group who had substantial assets.

TABLE 3
Stepwise Prediction of CSS Parents' Contribution
for Income Group 1: $0-$4,999
N = 5,595

Step	Independent Variable	Contribution to R^2 at this Step	Multiple R at this Step	Standard Error at this Step
1	Other Real Estate Equity	.6284	.79	$221.6
2	Business or Farm Net Worth	.1015	.85	$189.0
3	Home Equity	.0483	.88	$171.3
4	Net Income	.0205	.89	$163.2
5	Other Investments	.0149	.90	$157.0
6	Bank Accounts	.0102	.91	$157.7
7	Number of Dependent Childern	.0078	.91	$149.3
8	Medical Expenses	.0062	.92	$146.5
9	Federal Income Tax	.0018	.92	$145.7
10	Number of Children in College	.0016	.92	$145.0
11	Emergency Expenses	.0007	.92	$144.7
12	Housekeeping Allowance	.0006	.92	$144.4
13	Debt Outstanding	.0002	.92	$144.3
14	Indebtedness	.0012	.92	$143.8

The same phenomenon was apparent in the next income group ($5,000-$10,000). For this group, other real estate equity is the most powerful single predictor of CSS contribution and is followed by net income.

TABLE 4
Stepwise Prediction of CSS Parents' Contribution
for Income Group 2: $5,000<$10,000
N = 10,866

Step	Independent Variable	Contribution to R^2 at this Step	Multiple R at this Step	Standard Error at this Step
1	Other Real Estate Equity	.4845	.69	$537.7
2	Net Income	.1380	.78	$460.1
3	Number of Dependent Children	.0876	.84	$403.2
4	Home Equity	.0567	.88	$361.6
5	Business or Farm Net Worth	.0647	.91	$307.3
6	Bank Accounts	.0145	.92	$293.8
7	Medical Expenses	.0129	.93	$281.2
8	Housekeeping Allowance	.0122	.93	$268.7
9	Other Investments	.0083	.94	$260.0
10	Emergency Expenses	.0044	.94	$255.0
11	Debt Outstanding	.0025	.94	$252.5
12	Number of Children in College	.0020	.94	$250.2
13.	State Income Tax	.0011	.94	$249.0
14	Other Dependents	.0010	.94	$247.8
15	Indebtedness	.0006	.94	$247.2
16	Federal Income Tax	.0002	.94	$247.0

TABLE 5
Stepwise Prediction of CSS Parents' Contribution
for Income Group 3: $10,000<$15,000
N = 13,407

Step	Independent Variable	Contribution to R^2 at this Step	Multiple R at this Step	Standard Error at this Step
1	Home Equity	.2265	.48	$735.3
2	Number of Dependent Children	.1554	.62	$657.3
3	Business or Farm Net Worth	.1460	.73	$574.5
4	Net Income	.1083	.50	$504.3
5	Other Investments	.0606	.83	$460.4
6	Other Real Estate Equity	.0348	.86	$433.2
7	Housekeeping Allowance	.0301	.87	$408.2
8	Medical Expenses	.0202	.88	$390.5
9	Bank Accounts	.0186	.89	$373.6
10	Number of Children in College	.0120	.90	$362.2
11	Federal Income Tax	.0084	.91	$354.0
12	Emergency Expenses	.0083	.91	$345.7
13	Debt Outstanding	.0047	.91	$340.9
14	State Income Tax	.0023	.91	$338.5
15	Other Dependents	.0013	.92	$337.2
16	Indebtedness	.0002	.92	$337.0

TABLE 6
Stepwise Prediction of CSS Parents' Contribution
for Income Group 4: \$15,000<\$20,000
N = 9,207

Step	Independent Variable	Contribution to R^2 at this Step	Multiple R at this Step	Standard Error at this Step
1	Number of Dependent Children	.1938	.44	\$1,039.6
2	Home Equity	.1374	.57	\$ 946.9
3	Business or Farm Net Worth	.1280	.68	\$ 851.5
4	Net Income	.0822	.74	\$ 784.2
5	Number of Children in College	.0773	.79	\$ 715.0
6	Other Investments	.0693	.83	\$ 646.8
7	Other Real Estate Equity	.0443	.86	\$ 599.2
8	Housekeeping Allowance	.0326	.87	\$ 561.6
9	Bank Accounts	.0229	.89	\$ 533.6
10	Federal Income Tax	.0186	.90	\$ 509.7
11	Medical Expenses	.0141	.91	\$ 490.8
12	Emergency Expenses	.0074	.91	\$ 480.6
13	State Income Tax	.0056	.91	\$ 472.8
14	Debt Outstanding	.0029	.91	\$ 468.6
15	Other Dependents	.0010	.92	\$ 467.2
16	Indebtedness	.0000	.92	\$ 467.2

TABLE 7
Stepwise Prediction of CSS Parents' Contribution
for Income Group 5: \$20,000<\$25,000
N = 3,957

Step	Independent Variable	Contribution to R^2 at this Step	Multiple R at this Step	Standard Error at this Step
1	Number of Children in College	.3259	.57	\$1,522.5
2	Other Investments	.1741	.71	\$1,311.5
3	Home Equity	.0680	.75	\$1,219.2
4	Other Real Estate Equity	.0579	.79	\$1,134.8
5	Net Income	.0516	.82	\$1,053.8
6	Bank Accounts	.0410	.85	\$ 984.7
7	Business or Farm Net Worth	.0367	.86	\$ 918.4
8	Number of Dependent Children	.0325	.89	\$ 855.4
9	Federal Income Tax	.0218	.90	\$ 810.6
10	Housekeeping Allowance	.0132	.91	\$ 782.0
11	Medical Expenses	.0122	.91	\$ 754.7
12	State Income Tax	.0063	.92	\$ 740.3
13	Emergency Expenses	.0062	.92	\$ 725.8
14	Debt Outstanding	.0023	.92	\$ 720.4
15	Other Dependents	.0008	.92	\$ 718.5
16	Indebtedness	.1741	.92	\$ 718.5

TABLE 8
Stepwise Prediction of CSS Parents' Contribution
for Income Group 6: over $25,000
N = 2,167

Step	Independent Variable	Contribution to R^2 at this Step	Multiple R at this Step	Standard Error at this Step
1	Net Income	.4429	.67	$3,835.1
2	Number of Children in College	.2074	.81	3,039.0
3	Other Investments	.1189	.88	2,469.1
4	Business or Farm Net Worth	.0450	.90	2,215.9
5	Federal Income Tax	.0274	.92	2,046.5
6	Other Real Estate Equity	.0097	.92	1,983.6
7	Bank Accounts	.0082	.93	1,928.7
8	Number of Dependent Children	.0070	.93	1,880.4
9	Home Equity	.0061	.93	1,837.3
10	Emergency Expenses	.0046	.94	1,804.4
11	State Income Tax	.0036	.94	1,778.5
12	Housekeeping Allowance	.0025	.94	1,759.9
13	Debt Outstanding	.0024	.94	1,742.5
14	Medical Expenses	.0016	.94	1,730.6
15	Indebtedness	.0003	.94	1,728.5
16	Other Dependents	.0000	.94	1,728.5

Looking across income groups, however, it is curious that only in the highest income group is net income the best single predictor.

Income Group		Best Single Predictor
$ 0	$ 5,000	Other Real Estate Equity
$ 5,000	$10,000	Other Real Estate Equity
$10,000	$15,000	Home Equity
$15,000	$20,000	Number of Dependent Children
$20,000	$25,000	Number of Children in College
over	$25,000	Net Income

This finding underscores the importance of including asset items, in addition to income, in batteries for predicting parental contribution.

As noted earlier, the purpose of the present study was to identify a subset of variables that might be used, with regression weights developed for specific subgroups, to predict parental contributions.

However, if the predictor battery for each subgroup contains a unique set of predictors, then it is conceivable that it would be necessary to collect all of the current financial and nonfinancial items, and a streamlining of input items would not be accomplished. As a consequence, the presence of the top seven* independent variables for the total sample in income subgroup predictor batteries was examined. The results of this analysis appear in Table 9.

*Chosen because the symptotic level was reached at the seventh interation.

TABLE 9
Presence of Top Seven Overall Predictors
in Income Subgroup Predictor Batteries

Independent Variable	Income Group					
	1 (Lowest)	2	3	4	5	6 (Highest)
Net Income	X	X	X	X	X	X
Other Investments	X	O	X	X	X	X
Number of Children in College	O	O	O	X	X	O
Other Real Estate Equity	X	X	X	X	X	X
Business or Farm Net Worth	X	X	X	X	X	X
Home Equity	X	X	X	X	X	O
Number of Dependent Children	X	X	X	X	O	X

X denotes presence of variable in predictor battery
O denotes absence of variable in predictor battery

The data in Table 9 suggest considerable commonality of predictor items across income subgroups. Differential weights developed for these seven items for each subgroup would very likely yield subgroup equations with multiple Rs in the low 90s.

One troubling finding of the stepwise regression analyses, however, is the level of standard errors of prediction, which are summarized in Table 10.

These data suggest that prediction of parents' contribution is considerably more accurate for families with income under $10,000 than for those with income in excess of $10,000. But even for the below $10,000 income families, predicted contributions that vary by ± $150 to ± $300 from the criterion may result in inequities for *individuals*, even though the procedure would be equitable for the group as a whole.

TABLE 10
Standard Errors of Prediction for
Subgroup Predictor Batteries

Income Subgroup	SE
$ 0 - $ 4,999	$ 149.3
$ 5,000 - $ 9,999	$ 281.2
$10,000 - $14,999	$ 408.2
$15,000 - $19,999	$ 599.2
$20,000 - $24,999	$ 918.2
$25,000 - over	$1,928.7

The magnitude of the standard errors arising from linear regression may occur, in part, because the equations the CSS uses to calculate parents' contribution are curvilinear; that is, the need analysis relationships between income, assets and contribution are nonlinear. Moreover, the large standard errors of prediction for higher income families are probably also due, in part, to the propensity of higher income families to have highly skewed distributions of assets.

For these reasons, one is unlikely to find a reduced set of variables that excludes family assets but accurately predicts parents' contribution for higher income families. This finding is probably not too startling, especially in view of the results of Olsen's study. Olsen found that her prediction system was improved when higher income (over $11,000) families were eliminated.

Perhaps even more curious is the fact that no expense items were found among the top five best predictors for any of the income groups.

In conclusion, it does not appear as though stepwise regression analysis holds much promise as a technique for simplifying need analysis procedures for individual and applicants, if one accepts the prepositions that (1) a simplified method of determining parental contribution should yield results for individuals that are not greatly dissimilar from those arising from current procedures, and (2) need analysis procedures should allow credit for a family's unusual or extraordinary expenses.

REFERENCES

1. Margaret Olsen, *Notes on a study of Refined CSS Computing Procedures.* Educational Testing Service, 1958. (Unpublished Study)
2. M. D. Orwig and Paul K. Jones, *Can Need Analysis be Simplified?* ACT Research Report No. 33 American College Testing, Iowa City, Iowa, March 1970.

APPENDIX

Intercorrelation Matrix[1] for PCS Variables
Total Sample

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Number of Dependent Children																
2. Net Income	15															
3. Federal Income Tax	-03	85														
4. State Income Tax	04	65	68													
5. Housekeeping Allowance	00	38	31	21												
6. Medical Expenses	02	-03	-05	-04	-04											
7. Emergency Expenses	01	06	04	04	00	06										
8. Indebtedness	00	00	00	00	00	01	02									
9. Other Dependents	-03	02	00	00	00	02	02	00								
10. Home Equity	02	35	29	22	09	04	03	-03	00							
11. Other Real Estate Equity	00	08	06	06	00	02	00	00	00	09						
12. Business or Farm Net Worth	03	04	02	07	05	02	-01	00	-01	-08	12					
13. Bank Acounts	09	18	17	13	-02	03	04	-01	00	20	06	03				
14. Other Investments	00	18	21	13	-03	03	03	00	00	14	14	10	12			
15. Debt Outstanding	03	09	06	05	03	05	04	63	00	06	03	-01	-02	08		
16. Number of Children in College	40	18	10	10	06	05	03	00	00	09	02	02	01	04	03	
17. Parents' Contribution	-07	81	71	51	19	-06	00	-02	-01	41	28	21	29	39	05	-05

[1] Decimal points excluded