NBER WORKING PAPER SERIES

A LATENT VARIABLE MODEL OF QUALITY DETERMINATION

Paul J. Gertler

Working Paper No. 1750

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 1985

I am indebted to Arthur Goldberger, George Jakubson, Salih Meftci, Robert Porter, and the participants of the NBER conference on Incentives in Government Financing for helpful suggestions. Of course, I am responsible for any remaining errors. The research reported here is part of the NBER's research program in Health Economics and project in Government Budgets. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #1750
November 1985

A Latent Variable Model of Quality Determination

ABSTRACT

Despite substantial interest in the determination of quality, there has been little empirical work in the area. The problem, of course, is the general lack of data on quality. This paper overcomes the data problem by constructing a Multiple Indicator Multiple Cause (MIMIC) model of quality determination. We present a one-factor MIMIC model of quality which derives natural indicators out of the relationship between input demand and output determination. The indicators turn out to be input demands which have been filtered to remove variation due to all factors, except quality and random disturbances. These indicators are measures of input investment in each unit of output or the volume (intensity) of service. The model is identified by defining input demand to be a function of quantity and "total effective output" (quantity times average quality), instead of of quantity and average quality. The model is then applied to the determination of nursing home quality. The model appears to perform quite well, as the results generally conform with economic theory and restrictions implied by the MIMIC structure are accepted in hypothesis tests.

Paul J. Gertler
Department of Economics
State University of New York
Stony Brook, New York 11794

I. INTRODUCTION

Despite substantial interest in the determination of quality, there has been little empirical work in the area.[1] The problem, of course, is the general lack of data on quality. Quality is usually defined as characteristics of goods other than the physical units in which they are sold.[2] For example, automobiles are sold on a per car basis with quality characteristics being durability, miles per gallon of gasoline, comfort of ride, etc. Data on these characteristics can be hard to come by.

This paper overcomes the data problem by constructing a Multiple Indicator Multiple Cause (MIMIC) model of quality determination.[3] The difficulties in implementing the MIMIC model are (i) to find valid indicators of the latent variable (quality), and (ii) to identify the model. We present a one-factor MIMIC model of quality which derives natural indicators out of the relationship between input demand and output determination. The indicators turn out to be input demands which have been filtered to remove variation due to all factors, except quality and random disturbances. These indicators are measures of input investment in each unit of output or the volume (intensity) of service. The model is identified by defining input demand to be a function of quantity and "total effective output" (which is quantity times average quality), instead of of quantity and average quality.

This structure is then used to estimate a model of the determination of nursing home quality. It is estimated with a minimum distance procedure proposed in Chamberlain (1982) for the purpose of estimating panel data models. The model appears to perform quite well, as the results generally conform with economic theory and restrictions implied by the MIMIC structure are accepted in hypothesis tests.

The paper is organized as follows. In section II, we briefly present the MIMIC structure. In section III, we derive the indicators of quality, and identify the model. In section IV, we discuss estimation. In section V, we specify and estimate a model of the determination of nursing home quality. Finally, conclusions are drawn in section VI.

II. THE MIMIC MODEL

The MIMIC model is concerned with estimating an equation of the form

$$(1) \qquad Q_i = \beta_0 + \sum_{j=1}^{J} \beta_j X_{ij} + \epsilon_i,$$

where Q_i is the latent variable, and the X_{ij}'s are causes of Q_i. Although Q_i is not observed, indicators of Q_i are and are related to Q_i via a measurement model of the form

$$(2) \qquad \underline{Y}_i = \underline{\lambda} Q_i + \underline{\eta}_i,$$

where $\underline{Y}_i$ is the vector of indicators, and $\underline{\eta}_i$ is a vector of measurement errors. By substituting (1) into (2) for Q_i, we derive a model which is no longer a function of the latent variable Q_i. That model is a system of equations whose right hand sides are restricted to be proportional to one another. These proportionality restrictions constrain the structure to be a one-factor model of the latent variable, and, in addition to a normalization, identify the parameters in (1) and (2).

III. A MIMIC MODEL OF QUALITY DETERMINATION

The problem is to estimate a reduced form model of quality determination. Let Q_i be an index of the average quality of goods produced by firm i. Then, the reduced form quality equation is

$$(3) \qquad Q_i = \beta_0 + \sum_{j=1}^{J} \beta_j Z_{ij} + \sum_{k=1}^{K} \beta_{J+k} W_{ij} + \epsilon_i, \ (i = 1,\ldots,n),$$

where the Z_{ij}'s are exogenous demand variables, the W_{ij}'s are exogenous supply variables, the β_{ij}'s are unknown parameters, and the ϵ_{ij}'s are independently distributed random disturbances with zero mean. Since quality is not directly observed, the coefficients of (1) cannot be directly estimated. Instead, we exploit the theoretical relationship between input demand and output determination to indirectly identify and estimate these coefficients.

The arguments of a cost function and the corresponding conditional input demand functions are outputs, input prices, and in the short run, capital stock. A firm's output depends on the quantity and quality of goods produced.

Let us define a firm's total "effective" output as quantity times quality.[4] The cost of production, and corresponding conditional input demands may depend upon the quantity-quality mix as well as total "effective" output. Let φ_i be firm i's total "effective" output, and X_i be the physical number of units produced. Then, assuming a quadratic cost function, firm i's conditional input demand functions are

(4) $$Y_{i\ell} = \alpha_{0\ell} + \alpha_{\varphi\ell}\varphi_i + \alpha_{x\ell}X_i + \textstyle\sum_{k=1}^{K}\alpha_{k\ell}W_{ik} + \nu_{i\ell} \quad (\ell = 1,\ldots,L),$$

where the α's are unknown coefficients, and the $\nu_{i\ell}$'s are independently distributed random disturbances with mean zero.

Since a firm's total "effective" output can be measured as quantity times quality, φ_i can be expressed as

(5) $$\varphi_i = X_i Q_i \ .$$

Substitution of (3) into (5) yields the reduced form total "effective" output equation:

(6) $$\varphi_i = \beta_0 X_i + \textstyle\sum_{j=1}^{J}\beta_j Z_{ij} X_i + \sum_{k=1}^{K}\beta_{J+k} W_{ik} X_i + \epsilon_i X_i,$$

and substitution of (6) into (4) gives us the reduced form input demand equations:

(7) $$Y_{i\ell} = \alpha_{0\ell} + \alpha_{x\ell}X_i + \textstyle\sum_{k=1}^{K}\alpha_{k\ell}W_{ik} + \sum_{j=1}^{J}\gamma_{j\ell}Z_{ij}X_i$$

$$+ \textstyle\sum_{k=1}^{K}\gamma_{J+k,\ell}W_{ij}X_i + \eta_{i\ell} \quad (\ell=1,\ldots,L),$$

where

(8) $$\eta_{i\ell} = \alpha_{1\ell}X_i\epsilon_i + \nu_{i\ell},$$

(9) $$\gamma_{j\ell} = \alpha_{\varphi\ell}\beta_j \text{ for all } \ell \text{ and } j.$$

With the imposition of several restrictions on the reduced form input demand functions (7) the coefficients in the reduced form quality equation (3) and in the structural input demand functions (4) can be identified. The first restriction is a normalization to provide quality with a unit of measurement. As of now, it is measured in arbitrary units. If we divide quality by the coefficient on quality in one of the input demand equations, then quality is measured in the same units as that input. This is equivalent to restricting one of the $\alpha_{\varphi\ell}$'s to unity. In addition, there are the cross equation restrictions implied by (9), which require the reduced form input demand functions (7) to be proportional to one another in the exogenous variables that determine total "effective" output in (6).

This model can be interpreted as a MIMIC model. The normalization and proportionality restrictions are similar to the restrictions that arise in one-factor MIMIC models. The normalization bases quality, the latent variable, to the same scale as one of the indicators, and the proportionality constraints restrict the measurement model to be a one-factor model of quality. The indicators are input demands which have been filtered to remove variation due to quantity, input prices, and capital stock. The remaining variation in the filtered input demands is due to quality and random disturbances. The causes are the right hand side variables in (6).

The estimates can be used to predict quality scores for each firm, which are indicies of input quantities normalized to account for differences in quantity, input prices, and capital stock. Therefore, they measure the input investment in each unit of output or the volume (intensity) of service.

IV. ESTIMATION

Instead of maximum likelihood, the method usually used to estimate MIMIC models, the coefficients in the reduced form quality equation (3) and in the structural input demand functions (4) are estimated using a minimum distance procedure described in Chamberlain (1982).[5] This involves estimating the reduced form input demand equations (7) subject to the normalization and proportionality restrictions (9).[6] Chamberlain develops this procedure for panel data models, but it is appropriate for estimating any multivariate model subject to non-linear cross equation restrictions. The advantages of this procedure are: (i) normally distributed disturbances are not required, (ii) general heteroskedaticity is taken into account, (iii) the estimates are fully efficient, and (iv) computational ease. In addition, he derives χ^2 statistics for hypothesis testing.

V. THE DETERMINATION OF NURSING HOME QUALITY OF CARE

A. Nursing Home Quality

Nursing homes provide their patients with a package of goods and services such as medical care, room and board, and social services. Therefore, a nursing home's quality is utility patients derive from consuming its package of goods and services. A home can improve its quality by adjusting the composition of its package so as to be more in accordance with patients' preferences or by increasing the quantity of any commodity. Therefore, if patients prefer nursing services to social activities, the home can improve its quality without raising its operating costs by shifting resources from from social activities to nursing services. The more patients are willing to pay for a particular component and the lower its marginal cost, the greater the equilibrium quantity of that component. If there is an exogenous increase in demand, then the home invests in those components that yield the greatest marginal profit. Therefore, holding input prices constant, increases in service intensity raise patient utility. Since quality is the utility patients derive from consuming a nursing home's package, observed increases in a home's service intensity are tantamount to increases in quality.

B. Data

The model is estimated with data constructed from New York State's 1980 survey of Long Term Care Facilities.[7] The sample consists of 455 nursing homes, of which 288 are proprietary and 167 are "not for profit" institutions. Unless otherwise specified, the variables are daily averages, with the home as the unit of observation. Descriptive statistics are presented in table 1.

The dependent variables in the input demand equations are 100's nursing labor hours, 100's of other labor hours, and a supplies quantity index. The exogenous supply variables are the input prices and capital stock. The input prices are the hourly nursing wage rate, the hourly other labor wage rate, and a supplies price index. Since, the majority of capital owned by a nursing home is the facility itself, capital stock measured as the area of the facility in 100,000's of square feet.

The exogenous demand variables are the per capita income of the people living in the nursing home's market area, the population over age 65 in the nursing home's market area, the proportion of patients in the nursing home whose last residence before entering the nursing home was located in the same county in which the nursing home is located, and an index of health status of the patients in the nursing home. Income is measured in 1000's of dollars, and population in 10,000's of people. The proportion of patients from the same county is a measure of the distance of the nursing home from the family and friends of its patients. Presumably, nursing homes that are located closer to its patients' family and friends are more attractive, ceteris paribus. The health index is really an index of ill-health.

In addition to the exogenous demand and supply variables, we must consider the instruments of government regulation. The two major forms of regulation in the nursing home industry are the Medicaid patient subsidy program and the Certificate of Need (CON) cost containtment program.[8]

Medicaid is an entitlement program established under the Social Security Act to provide the poor with a minimum floor of health services. Through direct subsidies, the Medicaid program makes health care available to individuals who otherwise could not afford it. The Medicaid program reimburses nursing homes for the care of Medicaid eligible patients. Typically, the reimbursements are determined by a "cost plus" method, where the reimbursement per patient is equal to average operating costs plus some return referred to as the Medicaid "plus" factor.

The CON cost containment program requires that, in order to expand an existing nursing home or build a new one, the government must certify that the proposed facility is indeed "needed". Effectively, CON limits the capacity of existing nursing homes and new entry into the market. It was thought that costs could be contained by limiting the available supply of nursing home beds.

Therefore, the policy variables are the Medicaid plus factor, the CON capacity constraint, and the concentration of the home's market area. Since the CON capacity constraint is binding, it is measured as the average daily census of patients in the home, and CON entry policy is captured by a Herfindahl index of the concentration of each home's market.[9] Entry reduces the concentration of a home's market.

C. Results

Due to the CON capacity constraint it was necessary to impose one additional restriction on the model. The CON capacity constraint implies that quantity is exogenous, and therefore, a determinant of quality. Consequently, in order to identify the intercept of the reduced form quality equation, we excluded X_i from one of the input demand functions. This assumption places an implicit restriction on the underlying technology of nursing home production. It requires one of the input demands to depend only on total "effective" output, and not on the quantity-quality mix. Since linear input demand is consistent with a quadratic cost function, this restriction requires one of the second order coefficients on an input price times X_i to be zero.

The model was estimated separately for proprietary and "not for profit" institutions as hypothesis tests rejected pooling.[10] The estimated coefficients and corresponding t-statistics for the reduced form quality equations are presented in table 2, and for the input demand equations in table 3.

The estimates appear to be quite reasonable. There is a substantial amount of precision in the estimates, and hypothesis tests to determine if the normalization, proportionality, and exclusion restrictions are valid were accepted. Specifically, the test statistics are 20.08 for the proprietary sample, 16.68 for the "not for profit" sample, and are both distributed $\chi^2(27)$. The corresponding critical value at the .05 significance level is 40.11.

In addition, the estimates generally conform with economic theory. In the quality equations, the coefficients on input price variables have negative signs when they are significantly different from zero, and the coefficients on capital stock are positive and significant, suggesting that quality is easier to produce the bigger the facility, certerus paribis. Also, the coefficients on the ill-health index, income, and percent of patients form the same county are positive and significantly different from zero, suggesting that increases in demand lead to higher quality. On the other hand, the coefficient on population is negative and significant in the proprietary model.

The policy variables in the quality equations also perform as expected. The coefficients on the Medicaid plus factor are negative and significantly different from zero.[11] In addition, the coefficient on the concentration index in the proprietary sample is negative and significant, suggesting that reductions in competition lower quality. Finally, the coefficients on the total number of patients are negative, suggesting that the marginal cost of quality increases as the number of patients increase.

Finally, the input demand equations also look quite reasonable. The coefficients on quantity and quality are positive and generally significant, and the coefficients on own price are generally negative and significant.

V. CONCLUSIONS

The empirical investigation of the determinants of quality is usually hampered by a lack of data on quality itself. To solve this problem, we have developed a MIMIC model which allows us to indirectly estimate the coefficients of a reduced form quality equation. The approach is to simultaneously estimate a reduced form model of quality and input demand functions. The indicators of quality turn out to be input demands which have been filtered to remove variation due all factors except quality and random disturbances. These indicators are measures of the input investment in each unit of physical output. Identification is achieved by defining input demand to be a function of quantity and "total effective output", instead of of quantity and average quality. The model was successfully applied to a nursing home data set, as the results generally conform with economic theory and restrictions implied by the MIMIC structure are accepted by hypothesis tests.

FOOTNOTES

[1] See Leffler (1982) fore a good discussion of the theoretical quality literature through 1982, and Allen (1984) thereafter. The major source of empirical work has been through the Hedonic pricing literature spurred by Rosen (1974). This Literature requires the assumption of perfectly competitive markets, and in practice is more concerned with how quality characteristics affect the competitive price, than the factors that determine the quality choices.

[2] See Lancaster (1966), and more recently Leffler (1982), for more on the definition of quality.

[3] See Goldberger (1972a), Joreskog and Goldberger (1975), and Robinson (1974) for discussions of MIMIC models, and Aigner et. al. (1984) for a general survey of latent variable models.

[4] This notion of total "effective" output is used in Spady and Friedlander (1978) and in Leffler (1982). It is a quality constant measure of output, and, as Spady and Friedlander point out, is consistant with aggregation theory.

[5] See Goldberger (1972b), Joreskog and Goldberger (1975), and Joreskog and Sorbom (1981) for Maximum-Likelihood approachs.

[6] In many cases (other than regulated industries), quantity may be endogenous. This wouold require the additional specification of a reduced form quantity equation along the lines of (3). In this case, the model would be estimated by instrumental variables. Chamberlain (1982) also considers these cases.

[7] See Gertler (1985b) for a more detailed description of the data.

[8] See Gertler (1985a) and (1985b) for a detailed analysis of the effects of these regulation on the nursing home industry.

[9] See Gertler (1985a) for more on the measursement of the CON capacity constraint.

[10] The test for pooling proprietary and "not for profit" homes in the same model is preformed under the maintained hypothesis that the proportionality, normalization, and exclusion restrictions are valid. The test statistic is 99.74, and is distributed $\chi^2(27)$. The corresponding critical value at the .05 significance level is 40.11.

[11] This point is discussed in detail in Gertler (1985a). Briefly, nursing homes face two types of demand: (i) "private pay" patient demand, which is sensitive to price and quality choices, and (ii) Medicaid demand, which is perfectly elastic at the Medicaid reimbursement rate. Therefore, nursing homes can price discriminate between patients who finance their care privately and patients whose care is financed by Medicaid, but are required to provide the same quality to both types of patients. An increase in the Medicaid "plus" factor raises the marginal profit of a Medicaid patient. Therefore, homes have incentive to substitute Medicaid patients for "private pay" patients. Homes reduce "private pay" demand and operating costs by lowering quality.

REFERENCES

Aigner, D., C. Hsaio, A. Kapteyn, and T. Wansbeek (1984), "Latent Variable Models in Econometrics," in the Handbook of Econometrics Vol. 2, edited by Z. Griliches and M. Intriligator, North-Holland, Amsterdam. pp.1321-1393.

Allen, F. (1984), "Reputation and Product Quality," The Rand Journal of Economics, Vol. 15, pp. 311-327.

Chamberlain, G. (1982), "Multivariate Models for Panel Data," Journal of Econometrics, Vol. 18, pp. 5-46.

Gertler, P. J. (1985a), "Subsidies, Quality, and Regulation in the Nursing Home Industry," Working Paper No. 268, Department of Economics, State University of New York at Stony Brook.

Gertler, P. J. (1985b), "Regulated Price Discrimination and Quality: The Implications of Medicaid Reimbursement Policy for the Nuring Home Industry," Working Paper No. 267, Department of Economics, State University of New York at Stony Brook.

Goldberger, A. S. (1972a), "Structural Equations in the Social Sciences," Econometrica, Vol. 40, pp. 979-1001.

Goldberger, A. S. (1972b), "Maximum-Likelihood Estimation of Regressions Containing Unobservable Independent Variables," International Economic Review, Vol. 13, pp. 1-15.

Joreskog, K. and A. S. Goldberger (1975), "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable," Journal of the American Statistical Association, Vol. 70, pp. 631-639.

Joreskog, K. S. and D. Sorbom (1981), LISERAL V User's Guide, Chicago: National Educational Resources.

Lancaster, K. (1966), "A New Approach to Consumer Theory," Journal of Political Economy, Vol. 74, pp. 132-157.

Leffler, K. D. (1982), "Ambiguous Changes in Product Quality," American Economic Review, Vol. 72, pp. 956-967.

New York State Department of Health, Office of Health Systems Management (1980), "Title XIX Nursing Home Cost Survey."

Robinson, P. M. (1974), "Identification, Estimation, and Large Sample Theory for Regressions Containing Unobservable Variables," International Economic Review, Vol. 15, pp. 680-692.

Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," Journal of Political Economy, Vol.82, pp. 34-55.

TABLE 1
DESCRIPTIVE STATISTICS: MEANS AND STANDARD DEVIATIONS

VARIABLE	ALL HOMES	PROPRIETARY	NOT FOR PROFIT
1. Nursing labor Hours	315.09 (266.27)	313.20 (224.27)	318.36 (327.01)
2. Other Labor Hours	216.18 (176.39)	189.62 (125.88)	261.99 (233.15)
3. Supplies Quantity Index	2.50 (2.04)	2.60 (2.18)	2.32 (1.78)
4. Nursing Hourly Wage	7.82 (2.75)	7.67 (2.84)	8.06 (2.59)
5. Other labor Hourly Wage	8.78 (2.78)	7.97 (2.44)	8.79 (3.23)
6. Supplies Price	1.41 (.10)	1.38 (.03)	1.44 (.15)
7. Capital Stock	6.86 (25.51)	7.04 (31.77)	6.54 (5.84)
8. Per Capita Income	7.15 (1.40)	7.08 (1.44)	7.27 (1.32)
9. Population Over 65	1.07 (.88)	.98 (.87)	1.23 (.87)
10. % Patients Same County	.74 (.26)	.77 (.24)	.70 (.30)
11. Health Status Index	1.43 (.61)	1.51 (.57)	1.31 (.67)
12. Medicaid Plus Factor	15.41 (15.21)	18.35 (13.96)	10.34 (15.96)
13. Total Patients	124.21 (94.04)	124.61 (89.33)	124.01 (101.94)
14. Market Concentration	.12 (.11)	.12 (.12)	.11 (.09)
15. # of Observations	455	288	167

TABLE 2
REDUCED FORM QUALITY EQUATIONS
ESTIMATED COEFFICIENTS AND T-STATISTICS IN PARENTHESES

INDEPENDENT VARIABLE	PROPRIETARY HOMES	"NOT FOR PROFIT" HOMES
1. CONSTANT	.51 (.36)	.21 (.26)
2. TOTAL # OF PATIENTS	-.08 (1.52)	-.14 (1.14)
3. NURSING HOURLY WAGE	-.14 (3.03)	.06 (.81)
4. OTHER LABOR'S HOURLY WAGE	.04 (.86)	-.18 (2.98)
5. SUPPLIES PRICE	.58 (.56)	.69 (1.38)
6. CAPITAL STOCK	.02 (1.70)	.04 (1.88)
7. HEALTH STATUS	1.27 (29.98)	1.62 (23.30)
8. INCOME	.04 (3.00)	.06 (1.96)
9. POPULATION	-.04 (4.01)	-.05 (.66)
10. MEDICAID PLUS FACTOR	-.01 (2.11)	-.01 (2.13)
11. % PATIENTS SAME COUNTY	.19 (2.21)	.29 (2.05)
12. MARKET CONCENTRATION	-1.18 (4.12)	.14 (.21)

TABLE 3
PROPRIETARY INPUT DEMAND EQUATIONS

INDEPENDENT VARIABLE	NURSING HOURS	OTHER LABOR HOURS	SUPPLIES
1. CONSTANT	3.39	-2.92	16.38
	(1.37)	(2.78)	(6.12)
2. TOTAL QUALITY	1.00*	.24	.04
	-	(12.41)	(.75)
3. TOTAL # OF PATIENTS	.00**	.86	2.04
	-	(16.38)	(12.76)
4. NURSING HOURLY WAGE	.02	.00	.14
	(.44)	(.31)	(2.90)
5. OTHER LABOR'S HOURLY WAGE	-.05	-.10	-.11
	(.91)	(3.95)	(1.92)
6. SUPPLIES PRICE	-2.20		-12.01
	(1.22)	(3.45)	(6.14)
7. CAPITAL STOCK	-.03	-.04	-.00
	(1.72)	(9.47)	(.60)

"NOT FOR PROFIT" INPUT DEMAND EQUATIONS

INDEPENDENT VARIABLE	NURSING HOURS	OTHER LABOR HOURS	SUPPLIES
1. CONSTANT	.21	-1.85	6.61
	(.26)	(2.87)	(9.46)
2. TOTAL QUALITY	1.00*	.26	.08
	-	(7.77)	(2.00)
3. TOTAL # OF PATIENTS	.00**	1.21	1.62
	-	(7.61)	(8.89)
4. NURSING HOURLY WAGE	-.17	.07	.01
	(2.75)	(1.92)	(.21)
5. OTHER LABOR'S HOURLY WAGE	.18	-.06	.07
	(2.99)	(1.86)	(1.89)
6. SUPPLIES PRICE	-.83	-.04	-4.91
	(1.08)	(2.76)	(9.08)
7. CAPITAL STOCK	.00	.03	-.01
	(.08)	(1.19)	(.48)

* Coefficient restricted to unity.
** Coefficient restricted to zero.