NBER WORKING PAPER SERIES

MARKET WORK, WAGES, AND MEN'S HEALTH

Robert Haveman

Mark Stone

Barbara Wolfe

Working Paper No. 3020

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1989

John Bascom Professor of Economics and Director of the Robert M. LaFollette Institute of Public Affairs; Graduate Student of Economics; and Professor of Preventive Medicine and Economics, respectively, University of Wisconsin-Madison. Professors Haveman and Wolfe are Research Affiliates of the Institute for Research on Poverty and acknowledge the partial financial support of the Assistant Secretary of the Department of Health and Human Services for Planning and Evaluation through a grant to the Institute. We thank participants at a workshop of the Institute of Research on Poverty and a session of the Health Economics Research Organization at the Allied Social Science Meetings, 1988, for their helpful comments. This paper is part of NBER's research program in Health Economics and Labor Studies. Any opinions expressed are those of the authors not those of the National Bureau of Economic Research.

NBER Working Paper #3020
June 1989

MARKET WORK, WAGES, AND MEN'S HEALTH

ABSTRACT

In this paper, we investigate the complex interrelationships among worktime, wages and health identified in the Grossman model of the demand for health. We specify a 3-equation simultaneous model designed to capture the time dependent character of these interrelationships, and estimate the model using 8 years of panel data on 882 males aged 22 to 71. The model is estimated using Hansen's generalized methods of moments imposing a weak set of conditions on the error term covariance structure. Using our data, we estimate simpler models with more restrictive assumptions commonly found in the literature, and find substantial differences between these estimates and those from the simultaneous model. For example, the positive relationship between worktime and health found in other studies disappears when the relevant simultaneities are accounted for. Our simultaneous estimates also suggest that worktime spent in environmentally adverse conditions are inversely related to health status, while job related physical exercise retards health deterioration.

Robert Haveman
Mark Stone
Barbara Wolfe
University of Wisconsin
Department of Economics
7432 Social Science
Madison, WI 53706

I. INTRODUCTION

The past two decades have witnessed important and puzzling changes in the reported health status of males. While age adjusted mortality rates have fallen, self-reports of poor health and disability have increased.

Numerous speculations, drawn from various disciplines, have sought to explain the determinants of health status and changes in it. In the economics literature, the limited empirical analysis derives from the theoretical work of Grossman (1972a; 1972b), and focuses on the relationships among wages, worktime, and health. In this model, health is viewed as an endogenously determined capital stock, and changes in it as depending on: 1) the exogenous effects of health depreciation related to age (and intensity of use of health stock, see Muurinen 1982), 2) individual investments in health improvement (e.g., preventive medical care), and 3) individual decisions regarding time allocations (e.g., worktime, leisure, exercise). The benefits of health capital are increased time available for productive activities and direct increases in utility from less pain and suffering. Work to date has analyzed each of these gains independent from its alternative. The investment or productive benefit has received the most attention.

Wages play an important role in the investment version of the Grossman model. Higher wages lead to a proportional increase in the marginal return to health, and at the same time increase the cost of own-time health investments. However, because the production of health involves both time and market work, the marginal return to health investments rises with wages, as does the optimal quantity of health.[1] Simultaneously, better health increases the time available for work and, hence, labor supply.

While worktime is dependent on both wages and health, it may have its own effect on these variables. Worktime beyond the norm may carry with it stress factors that increase the health deterioration factor. Particular conditions at the workplace, such as a hazardous work environment, may also influence health deterioration. Conversely, regular attachment to the work place and the psychological support that it provides are often viewed as retarding health deterioration. Worktime, likewise, embodies experience and on-the-job training, and is associated with higher wages.

In this paper, we attempt to advance understanding of the structure of the complex interrelationships between economic phenomena--worktime and wages--and health. While we build on extant empirical work on this issue, we carry the analysis forward in a number of dimensions. First, we specify a 3-equation simultaneous model, designed to capture the interrelationships among these phenomena. Second, we estimate this model using individual longitudinal micro-data on 882 men observed for 8 consecutive years, yielding over 7000 observations. Third, utilizing the longitudinal nature of the data, we are able to capture the time dependent character of the relationships among the variables. Finally, we estimate the model using Hansen's generalized method of moments technique, hence imposing a weak set of conditions on the error term structure of the system and allowing simultaneous estimation of the three structural equations.

Section 2. of the paper briefly reviews previous attempts to measure the relationships among economic variables and health, emphasizing the shortcomings of these studies. Section 3. presents our model, and describes the estimation technique that we employ. In section 4., our data and the variables used in the estimation are described. The results of our estimation are presented and interpreted in section 5. Because of the complexity of the relationships among the three variables of interest, we indicate the differences in results between our simultaneous model and alternative research studies which fail to capture the full interdependence and impose unrealistic restrictions on the estimation. This is done in section 6. Finally, section 7. concludes.

2. WORK, WAGES AND HEALTH: A BRIEF REVIEW

Following the 1972 study by Grossman, several investigators have attempted to measure the interdependent relationships at issue. Luft

(1972), Grossman and Benham (1974), and Bartel and Taubman (1979) all focussed on the relationship of labor force participation and health. This work suggests that poor health in period t-1 reduces both labor supply and wages in period t, with larger effects found if health in t-1, which is self-reported, is treated as endogenous. These findings indicate the need to distinguish carefully the effect of prior health on work time and wages and the independent effect of worktime, wages, and health in an earlier period on health in a later period. The Grossman-Benham paper, in particular, emphasized the need for a model which views "observed wage rates, hours of work and health [as reflecting] a set of interrelated decisions made by the household."

A few studies have analyzed the determinants of health, focussing on the role of education. Taubman and Rosen (1982) modeled changes in the subjectively-reported health of white men, exploiting the longitudinal nature of the Retirement History Survey. Education was found to be positively related to health, and was robust, even when a variety of other variables expected to be related to health are controlled for. While this result is consistent with the indication in Auster, Leveson and Sarachek (1969) and Kemna (1987) that schooling has an important impact on health, even after controlling for income, and occupation (Kemna), it is likely to be partly attributable to the absence of a full simultaneous equations estimation framework accounting for the interdependence of work, wages, and health. Both Grossman (1976) and Lee (1982) have shown that when health and wages are estimated simultaneously, the independent impact of education on health is substantially reduced.

Lee (1982) and Wagstaff (1986) have pursued the worktime-wages-health interaction most extensively in the context of the theoretical framework of health capital developed by Grossman. The former of these studies estimated the simultaneous effects of health and wages in a structural equations framework, and accounted for the unobservable nature of health capital and the discrete nature of health indicators. Using a sample of older male workers, Lee concluded that wages and health capital are strongly jointly dependent--that wages exert a positive and significant influence on the demand for health, and that good health increases wages. In both cases, the relationships are significant.

Wagstaff emphasizes the different implications of both the investment and consumption variants of Grossman's model. In Wagstaff's estimations, wages are not part of a simultaneous model, but are estimated using an auxiliary equation. Ill-health is found to negatively influence wages. In both the investment and the consumption structural models, health is treated as an unobservable; wages are found to have a positive association with health, while job characteristics (e.g., temperature, mentally and physically demanding work) have negative associations with health, except (surprisingly) for hazardous substances.[2]

These studies, then, have established preliminary estimates of some of the relevant relationships implicit in the Grossman model of health as capital. They have also emphasized the interdependence of decisions regarding health capital and those pertaining to a variety of other economic choices, especially work effort and individual productivity (wages). However, this work is limited in a number of dimensions: in but few cases are the multi-equation systems estimated simultaneously; when simultaneous estimation has been attempted, severe restrictions have had to be placed on the functional form and the related structure of error terms (in the face of a complicated joint covariance matrix related to unobservable components of the process), and in no case has a full simultaneous equation estimation been adopted for the work, wages, health interaction. The few analyses based on longitudinal data have used panels of short time duration, and of a limited age span. Our research reported below extends this previous work. While our estimates are based on a simultaneous model, reflect the imposition of a weak set of conditions on the system's error term structure, and exploit longitudinal data, they continue to be limited by the relatively imprecise and self-reported measure of health and potential problems of selectivity.

3. A SIMULTANEOUS MODEL OF WORK, WAGES AND HEALTH

To capture the interdependent relationship of work, wages and health, we posit the following 3-equation structural model, utilizing time-indexed individual panel data:

$$(1)\quad H_{it} = \beta_0 + \beta_1 W_{it} + \beta_2 PH_{it} + \beta_3 OH_{it} + e_{it}$$

$$W_{it} = \alpha_0 + \alpha_1 H_{it-1} + \alpha_2 w_{it} + \alpha_3 PW_{it} + \alpha_4 OW_{it} + u_{it}$$

$$w_{it} = \pi_0 + \pi_1 H_{it-1} + \pi_2 W_{it-1} + \pi_3 Pw_{it} + v_{it}$$

H = health

W = hours worked

w = wages

PH,PW,Pw = personal characteristics that determine health, worked and wages

OH,OW = occupational characteristics that determine health ours worked

e,u,v = disturbance terms capturing measurement error of the ndent variables and omitted variables

Several aspects of this model preclude application of commonly-used ometric techniques. Since we are interested in estimating the ficients of all three equations, instrumental least squares, or ted information maximum likelihood cannot be used. Assuming that model is fully specified, three-stage least squares (3SLS) or full ormation maximum likelihood (FIML) estimation could be appropriate. ever, 3SLS estimation entails that e,u, and v are homoscedastic and autocorrelated, while joint normality and knowledge of the structure the system disturbance term covariance matrix (which is of dimension) are required for efficient FIML estimates. Moreover, most multaneous equation techniques, such as 3SLS, entail construction of e structural equation parameters from reduced-form estimates.

The generalized method of moments estimation (GMM) developed by ısen (1982) allows direct estimation of the structural equations, le imposing a weak set of conditions on the covariance matrix of tem error terms. GMM estimation rests on the assumption that a set nstruments orthogonal to the disturbance terms of each equation are

available. These orthogonality conditions are used to identify t
structural parameters and obtain consistent estimates under an ar
stationary nonspherical covariance structure. While knowledge of
true covariance structure would yield more efficient estimates, o
complicated multiequation panel data model, the large sample size
the availability of appropriate instruments suggests that the less
restrictive GMM assumptions are preferred.

The econometric techniques used in other empirical analyses of
health-worktime-wages nexus [e.g. OLS, LISREL, and multiple equati
maximum likelihood used by Chirikos and Nestel (1985), Wagstaff (1
and Lee (1982), respectively] are based on the assumption that the
terms have a known--usually homoscedastic--covariance structure.
argue that the assumption of a common disturbance term variance for
individuals may not be appropriate. For example, older individuals
likely to exhibit a higher health variance, and the second moment c
important omitted variables such as productivity or "constitution"
differ between subjects. In these cases, OLS estimates of coeffici
standard errors will be incorrect.

4. DATA AND ESTIMATION

The data that we use are annual observations from 1976-1983 on 8
males with a history of significant labor force attachment from the
Michigan Panel Study on Income Dynamics. The sample includes males
were heads of household from 1974-1983, who were at least 21 years o
age in 1976, and who worked an average of 450 hours per year or more

from 1976 to 1983. The variables used in the study and relevant summary statistics on each are found in Appendix A. The age range of men in the sample is 22 to 71; the range of annual hours worked is 0 to 5496.

The variable used to measure individual health is a self-reported measure based on two survey questions: "Do you have a physical or nervous condition that limits the type of work or the amount of work you can do?" Then if the respondent answers yes, he is asked: "Does it [the condition] limit your work a lot, somewhat, or just a little." As the responses range from 0 (no condition) to 3 (condition limits "a lot"), the health variable reflects the intensity of health limitations. We note that the loss of estimator efficiency resulting from the categorical health measure is at least partially offset by the panel nature of the data set, and by the large sample size. It is assumed that this indicator is an objective measure of health capital, and hence is a determinant of economic productivity. Although self-reported responses have been asserted to reflect ambition or labor market success, the literature suggests that such measures are reliable indicators of health capital [see Kemna (1987); Leigh (1983); Grossman (1972); and Taubman and Rosen (1982)].

The specification of the three equations in our simultaneous system--the health equation, the worktime equation, and the wage equation--is as follows:

$$\text{HEALTH STATUS} = \beta_0 + \beta_1 \text{ WORKTIME} + \beta_2\text{SELF-EMPLOYED} + \beta_3\text{DIVORCED OR SEPARATED} + \beta_4\text{AGE} + \beta_5\text{STRENGTH REQUIRED ON JOB} + \beta_6\text{JOB HAZARDS} + \beta_7\text{EDUCATION} + \varepsilon_1$$

$$\text{WORKTIME} = \alpha_0 + \alpha_1\text{WAGES} + \alpha_2\text{LAGGED HEALTH STATUS} + \alpha_3\text{SELF-EMPLOYED} + \alpha_4\text{NUMBER OF CHILDREN IN FAMILY} + \varepsilon_2$$

$$\text{WAGES} = \pi_0 + \pi_1\text{LAGGED WORKTIME} + \pi_2\text{LAGGED HEALTH} + \pi_3\text{EDUCATION} + \varepsilon_3$$

Worktime is the only endogenous variable among the right hand side variables in the HEALTH STATUS equation.[3] Self-employed status and change to divorced or separated marital status are included to reflect stress on the job and in personal life, respectively. The strength and hazards variables reflect workplace conditions expected to affect health, the first positively and the second negatively. Both are weighted by the months worked on the current job since the effect of job characteristics on health is likely to vary with length of exposure.[4] Age reflects the health depreciation function so emphasized by Grossman, and education is taken as a measure of the efficiency with which individuals combine market goods and time to produce health.

The WORKTIME equation includes both wages and health status (lagged) as right-hand side variables. The health variable is lagged by one period to reflect the dependence of labor supply on prior health. Self-employed status is commonly associated with freely determined work hours, and often hours in excess of the norm. The number of children in the family is a proxy for income needs and, hence, is expected to be positively associated with worktime.

WAGES are taken to depend on prior health and lagged worktime and, consistent with the human capital model, schooling.

Section 2 above discussed some of the large number of variables that were found to be significant determinants of health by previous studies--almost all of which are based on single-equation models. Our approach estimates a comparatively large number of parameters, controlling for the simultaneous interaction of the three endogenous variables, dictating that the impact of only the most important of these variables on health will be identified. Hence, the relatively parsimonious specification of our health equation compared to that of other studies is a consequence of the simultaneous equation framework.

The PSID data set includes several variables which covary with the explanatory variables and, because in many cases they are determined prior to the sample interval, are not likely to be correlated with the model disturbances. Consequently estimation will be enhanced by the set of available measures which meet the criteria for GMM instruments. The education of the head's mother is used as an instrument for all three equations. The health equation estimates are also identified by the wife's age, a dummy variable which equals 1 if the head's parents were poor, the age of the youngest child, dummy variables for residence in the northeast, north central, and south regions, a dummy variable which equals 1 if the head is married, and the squares of wife's age, wife's education, and father's age. Instruments for the worktime equation include the level and square of father's education, the poor parent dummy variable, the age and age squared of the youngest child, the square of the head's mother's education and the county unemployment rate squared. Wife's age, her education, dummy variables for residence in the northeast and south regions, and the county unemployment rate are used as instruments for the wage equation.

5. ESTIMATION RESULTS

We have emphasized the need for simultaneous estimation of the health-worktime-wages process. The first column of Table 1 presents our baseline results for the three-equation model, jointly estimated by the method of moments technique.

Consider first the health status equation. As predicted by the Grossman model, age, a proxy for the natural deterioration process, is associated positively with health limitations, and the relationship is significant. The occurrence of a marital separation or a divorce is also positively and significantly related to health limitations. Education, while associated with better health, is not significant. The variables related to market work and its characteristics are of special interest. The amount of worktime does not appear to be associated with health limitations. Being self-employed, however, is associated with good health and has a t-statistic of 1.49. The nature of the work done does matter. The time spent working on a job designated as having "hazardous" characteristics is significantly associated with health limitations, while the time on a job requiring strength (and proxying for physical exercise while on the job) is associated significantly with good health.

All of the independent variables in the worktime equation are significant and of the predicted signs. Health limitations in prior periods are negatively associated with worktime, as expected. Also as expected, the economic variables of wage and self-employed status are positively associated with worktime, as is the number of children in the family unit.

In the wage equation, both the log of education and the number of hours worked in previous periods are positively associated with wages. The presence of health limitations in prior periods is negatively associated with wages and has a t-statistic of 1.64.

In terms of the health-worktime-wages relationships, then, we conclude that: 1) worktime per se is not a significant contributor to the health deterioration function as defined by Grossman, but 2) continued exposure to job characteristics such as hazards and strength do contribute to observed health status, 3) prior health limitations negatively affect worktime, while wages have a positive effect, 4) prior health limitations have a negative and significant effect on wages, 5) prior worktime is a significant determinant of wages, and 6) while education contributes positively to wages, and hence worktime, its direct contribution to health, though positive, is not significant.

The consequences of not controlling for the joint determination of health, worktime and wages can be seen by comparing the results in column 1 with those in column 2. The estimates in column 2 are for the same three-equation model, but estimated in single equation form, using a "just identified" specification and generalized method of moments estimation procedures.[5] These coefficient estimates are identical to OLS estimates, but the GMM standard errors are adjusted for the nonspherical covariance matrix.

Estimating the model simultaneously (column 1) changes only the sign on the insignificant worktime variable in the health equation of column 2. However, the significance levels and the story that the estimation tells change dramatically.[6] For example, simultaneous estimation changes the nature of the observed worktime-health causality link; assuming independence (non-simultaneity) suggests that worktime is strongly and positively associated with better health, while the simultaneous estimates indicate that worktime does not significantly influence health. Endogenizing hours worked appears to obviate the worktime-health causality link observed by other researchers.[7]

In addition to changes in sign and significance, the magnitude of the coefficients changes substantially between the simultaneous (column 1) and single equation (column 2) estimations. In the full model, all of the independent variables (save hours worked and education) have much larger effects on health status than in the single equation estimates. In the worktime equation, lagged ill health and being self-employed have a substantially larger estimated effect in the full model than in the single equation estimate. Note that the lagged health status coefficient in the worktime equation is larger in absolute value by a factor of three in the simultaneous framework. The single equation estimates appear to substantially bias downward the estimated effects of prior health and education on wages.

Finally, while the column 2 estimates suggest a positive, strong and statistically significant relationship between education and good health, the simultaneous estimates in column 1 indicate a weak and insignificant relationship. This change in the estimated direct influence of education on health is consistent with that estimated by both Grossman and Lee in moving from a single equation reduced form estimate to a simultaneous estimate.[8]

The consequences of relaxing the widely used restrictions on the error term covariance matrix can be seen by comparing the column 2 t-statistics in parenthesis with those from simple individual equation linear (OLS) regression estimates of the model, shown in brackets. The regression standard errors are somewhat different from their GMM counterparts,[9] especially for the wage equation, suggesting that estimating the impact of health on wages using linear regression techniques may lead to unduly strong conclusions regarding the reliability of parameter estimates.

Column 3 of Table 1 shows the results of a simultaneously estimated two-equation model of endogenously determined ill health and worktime. Comparison of these estimates with the three equation results indicates that endogenizing wages does not alter substantially the health equation results. This suggests that while prior health appears to be an important determinant of wages, wages do not appear to be endogenous to the determination of health.[10]

Two-equation ill health--wage results are displayed in column 4 of Table 1. They are presented primarily to permit comparison with Lee's results. Both the health and wage equation estimates are little changed (although they are quite different from the single equation results).

Grossman and Benham used a nonsimultaneous equation framework to estimate the impact of education on health. In their model this linkage operates directly and indirectly via the impact of education on wages and wages on ill health. Because wages were found to impact ill health positively, they found that the negative direct effect of education on ill health was partially offset by a positive indirect education-ill health relationship. The direct and indirect impacts of education on ill health in our model can be examined by decomposing the coefficient of the log-education variable in the reduced form version of the ill health equation:

$$\frac{d(\text{ill health})}{d(\text{log of education})} = \frac{(1 - \alpha_1)\beta_7}{(1 + \beta_1)} + \frac{\alpha_1\beta_1\pi_3}{(1 + \beta_1)} = -.207 + .016 = -.191$$

The first ratio represents the direct effect of education in the health equation, while the second captures the education to wage to worktime impact. Education is positively associated with good health, although the relationship is not statistically significant. Since in our simultaneous equation framework worktime has a small impact on ill health, the overall effect of education on health tends to operate directly rather than indirectly through the schooling-wage relationship. Our result, that the primary impact of education on health is direct, is consistent with that of Grossman and Benham.

Finally, in column 5 of Table 1 we show the full GMM model estimates on the sample that eliminates all nonwhite observations. In the data, whites have better health than nonwhites, work more hours, and have higher wages, suggesting the possibility of quite different interaction patterns. While the general pattern of estimates in column 5 is similar to that in column 1, several exceptions are noteworthy.

First, in the sample of whites only, the coefficient on education in the health equation is about 20 percent of that for the entire sample but is strongly statistically significant. Second, in the whites-only sample, the effect of wages on worktime is somewhat larger than when the model was fit over the entire sample. Third, lagged ill health is not statistically significant in the worktime equation in contrast to the marginal significance in the equation for the entire group. Finally, the significance of both annual hours worked and lagged ill health in the wage equation is substantially reduced in the white-only equation. The overall fit is improved.

6. A COMPARISON OF SIMULTANEOUS AND CONSTRAINED ESTIMATES

Given the significance differences in estimated relationships between simultaneous and constrained specifications revealed in section 5., we reexamine the results of a recent and prominent study of the health-worktime-wages nexus, using the unrestricted covariance matrix-simultaneous equation framework which we prefer.

Wagstaff (1986) estimated two empirical versions of the Grossman model, one for the investment model which more closely resembles the work here and one for the pure consumption version of that framework. In addition to its close resemblance to our own estimates, we use the investment model because it is directed to the role of health in workhours and because (as Wagstaff notes) the high degree of multicollinearity in the consumption estimates "prevent one from obtaining precise estimates of the model's parameters" (p. 226).[11]

The measure of health capital used by Wagstaff was estimated as a latent variable in a LISREL framework, and cross section data were employed. His health capital equation was estimated using a single equation maximum likelihood framework. The results from his estimation are shown in the first column of Table 2.[12] The unobserved health variable employed in the Wagstaff estimate is increasing in good health. His results indicate that years of education is positively related to health, a result consistent with our results and nearly all other estimates of the determinants of health status. Recent education and experience in a health care job early in ones career (a proxy for medical education) are also positively associated with health. Spouse education is similarly positively associated with health status, but the relationship is not significant. Age is negatively related to health, and is the most significant of the Wagstaff variables. The number of family members by age group are not significantly related to health. Being married or cohabiting is positively associated with good health, but is not significant. Living in an urbanized area is inversely related to health status and is significant. Job tenure is positively related to good health although the coefficient is small. Of the work conditions on the job that are significant--physical and mental demands and temperature extremes--all show a negative relationship with health status. Finally, the economic variable--log of wages--is positively associated with good health and is significant.

Our estimation of a health equation which is as similar to that of Wagstaff as our data permit is shown in column 2 of Table 2. The parameter estimates shown there are similar to Wagstaff's with but few exceptions. (Note that while Wagstaff's health variable is increasing in good health, ours is decreasing.) Age, job tenure and marital status have similar relationships in both our and Wagstaff's models, though our marital status variable is significant while his is not. While Wagstaff finds essentially no effect of wage or number of family members on health, we find a negative relationship between number of children and good health and a positive relationship between age of youngest child and good health; both are significant. Our job hazards and temperature variables, like Wagstaff's, suggest an inverse relationship to health status, although we find that jobs that require strength contribute to the absence of health limitations. Finally, like Wagstaff, we find wages to be positively and significantly associated with good health.

We conclude that our linear regression, single equation estimates of the Wagstaff model using our data (which are equivalent to maximum likelihood estimates under a normally distributed disturbance) are generally consistent with those reported by Wagstaff. The question then becomes: "How robust are these estimated relationships to a less restricted set of assumptions?"

To answer this question, we estimated our three equation, simultaneous GMM model, but with Wagstaff's specification for the health status equation. These estimates are shown in column 3 of Table 2.[13] A comparison of columns 2 and 3 indicates major changes from adopting the less restrictive specification. Numerous sign changes occur, and only the age of youngest child variable is significant in the health equation estimated in a simultaneous framework. While education is negatively related to the presence of health limitations and is strongly significant (t-statistic = 11.6) in the linear regression estimate, it falls to insignificance in the simultaneous model. Changes of a similar magnitude are observed for other variables. We conclude that failing to adopt a simultaneous equation framework which admits nonspherical error term covariance structures almost certainly results in biased estimates of the demand for health function, and substantially alters empirical conclusions regarding the health-worktime-wage nexus.

7. CONCLUSION

We have addressed the question of the demand for health in the context of Grossman's theoretic model. As in his framework, worktime and wages are viewed as interdependent with health status. Recognizing this interdependence, we have developed a simultaneous model with an error term covariance structure with few restrictions, designed to capture the relationships involved. This model was estimated using a rich data set involving 8 years of panel data on each of 882 men. Hence, we are able to capture the time dependent nature of the relationship among these variables.

Our baseline model (column 1, Table 1) reveals the complexity of these interrelationships. While some of the relationships found in other studies (e.g., the positive and significant relationship between health limitations and both age and divorce or marital separation, and a negative relationship between education and health limitations) are verified, others (e.g., a positive and significant relationship between worktime and health limitations) are found to be nonexistent. Still other relationships (e.g., between conditions on the job and health status) are revealed for the first time in a simultaneous framework. We judge that the implicit demand for health function shown in Table 1 is the only available estimate that accounts for the interrelationships among health, worktime, and wages that are so emphasized in the Grossman model.

In addition to the estimated relationships themselves, we have demonstrated the necessity of a simultaneous estimation framework in order to obtain reliable estimates of the relationships of interest within the health-worktime-wages nexus. The substantial differences in the estimated coefficients and their significance between our simultaneous GMM model and its restrictive single equation counterpart indicates the need for such estimation. Further verification of this proposition comes from our replication of the results of a recently published health demand function, and our estimation of this same function in a simultaneous model.

The numerous sign and significance level changes associated with the move from a restrictive to a simultaneous GMM estimation indicates the fragility and potential bias incorporated in estimates that neglect the mutual interdependence of health, worktime and wages.

Notes

[1]In the consumption version of the Grossman models or in a mixed consumption-investment model, the impact of wages on health is ambiguous.

[2]Wagstaff notes that "the results [indicating that the coefficients are not reliable estimates] might stem from a failure to estimate a system of structural demand equations for health inputs" (emphasis his, p. 226), and "One might also question the accuracy of the structural parameter estimates on the grounds that the error terms have not been allowed to covary." (p.228)

[3]A model that included wages in the health equation was also estimated and their results are described below. The overall fit of this model is weaker than our preferred estimates.

[4]The rationale for this is the so-called healthy worker effect. It is hypothesized that workers who are ill value health more highly and so avoid hazardous work and that those most affected by hazards leave the job earlier than those least affected. This argues for a non-linear relationship between duration of exposure to work hazards and health. Kemna estimates an increasing negative effect of hazards on health through five years in an occupation and then a decrease, so that the effect at 10 years is the same as the initial level (p.199).

[5]The parameter estimates of each of the just-identified equations set k (the number of parameters per equation) linear combinations of the r (the number of orthogonality conditions) equal to zero. Consider the

simultaneous equation framework where r = r1 + r2 + r3 and K = k1 + k2 + k3. If there are more instruments for each equation than parameter estimates, (ri > ki for i = 1,2,3), then there will be r - k linearly independent combinations of the orthogonality conditions which equal zero only in the limit. These overidentification restrictions imply the sample inner products of instruments and residuals from one equation interact with those of other equations in helping to determine all parameter estimates and standard errors. Under the null hypothesis that the overidentification restrictions are true, the minimized value of the GMM objective function (times the number of observations) is distributed chi-squared with r-k degrees of freedom. The p-value for this statistic is reported in the last row of Table 1.

[6]In the GMM framework, multiple-equation estimates will differ from their single-equation counterparts because of the non-zero sample products of instruments from one equation with disturbances from another. Suppose the disturbance terms of both the health and worktime equations include the omitted variable "constitution" implying that the health equation worktime parameter estimate reported in column 2 of Table 1 is inconsistent due to covariance between the explanatory variable and error term. Controlling for the impact of health on worktime addresses this source of single-equation estimation bias resulting from cross-equation error term covariance.

[7]A possible explanation of the estimated positive relationship between worktime and health in the single equation model may be that healthier persons work more. In a single equation estimate, this

relationship would lead to estimation of a positive worktime-health relationship. The underlying structure would not be that more worktime leads to better health, but simply that those in better health are able to work more than those in poorer health. Including in the estimation the latter relationship--which is at the heart of the investment component of Grossman's model--requires a simultaneous model of worktime and health. In such a model, the observed positive association is obviated.

[8]Lee's model, however, did not include worktime, was restricted to men 45-59, and used cross-sectional data (for 1966).

[9]The just-identified GMM coefficient estimates will be identical to the linear regression estimates, but calculation of standard errors of the former include weighting matrices that account for the nonspherical covariance matrix.

[10]In addition, in unreported estimates, wages were included in the full model health equation. Higher wages were associated with better health, but the relationship was small and insignificant. Moreover, the standard errors of the age and workplace environment variables increased substantially and the overall fit of model was poorer. This result is probably caused by a higher correlation of the instrument-residual inner products from the health and wage equations with the inclusion of wages in the health equation.

[11]Wagstaff estimated both an investment model and a consumption model to compare the estimates under the two specifications. The consumption model is not used for comparison with the model estimated in

this paper for it includes expected lifetime hourly wage, estimated from an auxiliary equation, and initial assets, also estimated from an auxiliary equation, neither of which are available with our data set. Furthermore, comparison of his two sets of results shows that including initial assets and lifetime wage alters the estimated impact of education from a positive association with health to an unexpected negative association. Wagstaff discusses the resulting problems with multicollinearity on p. 223 of his paper--including that the iterations in the computer program did not converge with the inclusion of initial assets and lifetime wage.

[12]Wagstaff's empirical model is based on a simultaneous equation framework which includes demand for medical care equations. Wagstaff notes that his results might be altered by joint estimation of his extra equations. While the underlying Grossman model includes worktime and wages, worktime is not endogenous to the Wagstaff model because household production is homogenous of degree one, and wages are assumed exogenous.

[13]The explanatory and instrumental variables for the workhours and wage equations are identical to those of the model reported in column 1 of Table 1. In addition to the 12 instruments used in the Table 1 three-equation model, the health equation instruments used for these estimates include the square of the air pollution measure, and the cubes of wife's age, father's education, and wife's education.

References

Auster, R., I. Leveson, and D. Sarachek. 1969. "The Production of Health: An Exploratory Study." Journal of Human Resources. 4, 411-36.

Bartel A. and P. Taubman. 1979. "Health and Labor Market Success: The Role of Various Diseases." Review of Economics and Statistics. 56, 2-8.

Chirikos, T. and Nestel, G. 1985. "Further Evidence on the Economic Effects of Poor Health." Review of Economics and Statistics. 67 (February), 61-69.

Crocker, T. Undated. "Procedures Followed in Building Air Quality Data Set." University of Wyoming.

Grossman M. 1972a. The Demand for Health: A Theoretical and Empirical Investigation. New York: Columbia University Press.

________. 1972b. "On the Concept of Health Capital and the Demand for Health." Journal of Political Economy. 80, 223-255.

________. 1976. "The Correlation between Health and Schooling" in Household Production and Consumption. ed. by N. Terleckyj. New York: Columbia University Press. 147-223.

Grossman M. and L. Benham. 1973. "Health, Hours and Wages" in The Economics of Health and Medical Care. ed. by M. Perlman. New York: Wiley.

Hansen, L. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." Econometrica. 50, 1029-1054.

Kemna, Harrie J.M.I. 1987. "Working Conditions and the Relationship between Schooling and Health." Journal of Health Economics. 6, 189-210.

Lee, L.F. 1982. "Health and Wage: A Simultaneous Equation Model with Multiple Discrete Indicators." International Economic Review. 23, 199-221.

Leigh, J. Paul. 1983. "Direct and Indirect Effects of Education on Health." Social Science and Medicine. 17, 227-234.

Luft, H. 1975. "The Impact of Poor Health on Earnings." Review of Economics and Statistics. 57, 43-57.

Muurinen, J. 1982. "Demand for Health: A Generalized Grossman Model." Journal of Health Economics. 1, 5-28.

Taubman, P. and S. Rosen. 1982. "Healthiness, Education and Marital Status" in Economic Aspects of Health. ed. by V. Fuchs. Chicago: University of Chicago Press.

Wagstaff, A. 1986. "The Demand for Health: Some New Empirical Evidence" in Journal of Health Economics. 5, 195-234.

Data Appendix

The data are from the Michigan Panel Study of Income Dynamics. The 882 subjects used in the study meet the following criteria:[a]

1. were heads of households from 1974 to 1983
2. were at least 21 years of age in 1976
3. worked an average of 449 hours or more per year from 1976 to 1983
4. lived in a county in which an EPA air quality reporting station was located

Ill Health	- Head, whether work limited by health (0 if healthy, 1, 2, 3)
Annual work hrs	- Head, annual hours work in thousands
Log of wages	- Head, log of real wages
Divorced or separated	- Head, 1 if divorced or separated
Married	- Head, 1 if married
Age	- Head, age
Months on job	- Head, months at current job
Job - strength	- Head, job DOT strength measure
Job hazards	- Head, job DOT hazard measure
Self-employed	- Head, 1 if work for someone else, 2 if work for both someone else and self, 3 if self-employed
Number children	- Family unit, number of children
Log of education	- Head, log of years of education
Father's education	- Head's father's years of education (0 elementary, ..., 8 graduate school)
Mother's education	- Head's mother's years of education (0 elementary, ..., 8 graduate school)
Parent poor	- Head, 1 if parent's socio-economic status is reported to be poor
Age youngest	- Family unit, age of youngest child
Wife education	- Wife, years education
Wife age	- Wife, age
Northeast	- Family unit, 1 if live in Northeast
South	- Family unit, 1 if live in South
North Central	- Family unit, 1 if live in North Central
Unemployment rate	- County unemployment rate
Average level of sulfur dioxide	- Level of sulfur dioxide in head's county

Variable	Mean	Maximum	Minimum	Standard Deviation
Ill health	0.2126	3.0000	0.0000	0.6763
Annual work hours	2.0523	5.4960	0.0000	0.7898
Log wages	3.5009	4.5152	-4.0000	1.7637
Divorced or separated	0.0333	1.0000	0.0000	0.1794
Age	42.4422	71.0000	22.0000	11.4993
Months on job	78.1899	540.0000	0.0000	89.6078
Job - strength	2.1770	4.0000	0.0000	1.0591
Job hazards	18.2357	100.0000	0.0000	25.4805
Self-employed	1.1451	3.0000	0.0000	0.7288
Number children	1.3454	10.0000	0.0000	1.4107
Log of education	0.3900	1.0000	0.0000	0.4878
Father's education	3.1043	8.0000	0.0000	1.8167
Mother's education	3.1746	8.0000	0.0000	1.5934
Parent poor	2.5261	5.0000	1.0000	1.5409
Age youngest child	4.6443	17.0000	0.0000	5.4630
Wife education	10.7245	17.0000	0.0000	4.6534
Wife age, '83	36.0509	73.0000	0.0000	15.8329
Northeast	0.2058	1.0000	0.0000	0.4043
South	0.3284	1.0000	0.0000	0.4697
North Central	0.2553	1.0000	0.0000	0.4325
County unemployment rate	5.5132	60.0000	1.0000	2.9823
Average level of sulfur dioxide	3.6744	108.0000	0.0000	9.5469
Marital status	0.8766	1.0000	0.0000	0.3292

[a]The PSID Wave 16 for 1976 to 1983 included 2489 males who were heads of household from 1974 to 1983. The subset of 882 subjects used in this study consisted of those individuals that were 21 to 63 years of age in 1976, worked an average of 449 or more hours from 1975 to 1982, and lived in a county monitored by an EPA air quality station. Air quality reporting stations were located in 118 representative counties (Crocker, undated). This procedure was followed in order to include environmental factors in the model. In no specification were the air pollution variables statistically significant, and they often entered with an unexpected sign. Hence, the final specification of our simultaneous model (Table 1) does not include an environmental variable except as an instrument.

Table 1

Estimation Results, Three Equation
Simultaneous Model of Health Status, Worktime, and Wages
(t-statistics in parentheses)

	Simultaneous Weighted GMM		Single Equation Just Identified Weighted GMM and OLS[1]			Two Equation Weighted GMM		Two Equation Weighted GMM		Simultaneous Weighted GMM (Whites)	
Health Status (Ill Health)											
Constant	.132	(.19)	.670	(7.47)	[10.92]	.154	(.21)	.145	(.21)	.127	(.28)
Annual work hours	.111	(.48)	-.185	(-12.50)	[-17.07]	.106	(.45)	.107	(.46)	.172	(.88)
Self-employed	-.364	(-1.49)	-.061	(-5.63)	[-5.28]	-.358	(-1.47)	-.361	(-1.48)	-.070	(-0.45)
Divorced or separated	.933	(2.35)	.149	(1.85)	[2.29]	.940	(2.37)	.941	(2.37)	.978	(2.89)
Age	.017	(3.13)	.009	(11.00)	[12.68]	.017	(3.11)	.017	(3.12)	.019	(3.94)
Strength*months on job	-.002	(-2.86)	-.0003	(-5.56)	[-6.18]	-.002	(-2.85)	-.002	(-2.87)	-.002	(-4.03)
Hazard*months on job	.0001	(2.00)	.00001	(3.74)	[4.08]	.0001	(1.98)	.0001	(1.99)	.00007	(1.08)
Log of education	-.288	(-.62)	-.342	(-4.90)	[-8.39]	-.300	(-.64)	-.296	(-.64)	-.057	(-2.34)
Worktime (Annual Work Hours)											
Constant	.650	(2.11)	1.010	(48.71)	[46.74]	.661	(2.26)			.295	(0.74)
Log of wages	.200	(2.89)	.172	(36.83)	[38.78]	.195	(2.83)			.284	(3.14)
Ill health, lagged	-.362	(-1.66)	-.123	(-10.84)	[-10.73]	-.380	(-1.75)			-.247	(-0.98)
Self-employed	.618	(4.98)	.367	(24.11)	[34.58]	.629	(5.50)			.663	(4.95)
Number of children	.047	(3.93)	.373	(6.74)	[6.94]	.047	(4.00)			.033	(1.75)
Log of Wages											
Constant	1.520	(1.63)	1.760	(9.30)	[15.24]			1.310	(1.49)	.315	(0.16)
Annual work hours, lagged	.633	(3.21)	.802	(17.53)	[32.59]			.763	(4.21)	.508	(1.83)
Ill health, lagged	-.830	(-1.64)	-.234	(-4.75)	[-7.85]			-.731	(-1.40)	-.796	(-1.13)
Log of education	.791	(1.34)	.156	(1.01)	[1.55]			.719	(1.29)	2.080	(1.30)
N	7056		7056			7056		7056		4904	
Minimized value of NxObjective Function	14.8					14.6		12.5		10.4	
Level of significance for chi-squared test statistic of over-identification[2]	0.097					0.041		0.051		0.320	

[1] GMM and OLS t-statistics are in parentheses and brackets respectively. In the just-identified, single-equation case, GMM and OLS estimates are identical.

[2] Under the null hypothesis that the overidentification condition holds, this statistic is distributed chi-squared.

Table 2

A Comparison of Single Equation and Simultaneous Estimates of Health Capital

	Wagstaff Estimates[a]		Linear Regression, Wagstaff Model[b]		Simultaneous, GMM, 3 - Equation Weighted Wagstaff Health Model[b]	
Constant	.11	(.66)	.83	(12.33)	.267	(0.24)
Education in years	.04	(4.56)	-.03	(-11.60)	.003	(0.09)
In school 1976=1	.13	(2.44)	--		--	
Medical education=1	.23	(2.17)	--		--	
Spouse education in years	.01	(1.35)	-.00	(-2.07)	.006	(-0.96)
Spouse medical education=1	.02	(0.14)	--		--	
Male=1	.17	(2.68)	NA		NA	
Age in years	-.03	(-8.97)	.01	(10.08)	.013	(0.96)
Family members < 6	-.05	(-.90)	--		--	
Family members 6-15	-.02	(-.65)	--		--	
Family members > 15	.01	(.98)	--		--	
Age youngest			-.01	(-3.49)	-.022	(-2.01)
Number children			.02	(2.68)	.214	(1.51)
Married or cohabiting=1	.05	(.53)	--		--	
Married=1			-.07	(-2.33)	-.235	(-1.12)
Urbanization index	-.02	(-2.59)	--		--	
Job tenure in years	.01	(3.28)	--		--	
Years on present job			-.00	(-5.88)	.001	(0.49)
Physically demanding job[c]	-.05	(-3.10)	--		--	
Mentally demanding job[c]	-.12	(-7.73)	--		--	
Hazardous substances on job[c]	.02	(1.43)	--		--	
Temperature extremes on job[c]	-.01	(-6.00)	--		--	
Strength			-.07	(-6.59)	-.116	(-0.45)
Job hazards			.00	(2.01)	.015	(0.87)
Temperature			.00	(.75)	-.026	(-0.44)
log of hourly wage	.01	(3.21)	-.08	(16.92)	-.174	(-1.25)
N	2243		7056		7056	
R^2	.26		.14		--	
Minimized value of N* objective function	--		--		38.2	
Level of significance for chi-squared test statistic[d]					0.000	

[a]Dependent variable is an index derived from a LISREL estimation and is increasing in good health.
[b]Dependent variable is Ill Health, and hence is decreasing in good health.
[c]Factor loadings from principal components analysis.
[d]See Footnote 2 of Table 2.