

Gender differences in the effect of health on wages in Britain

Lynn Gambin
May 2004

ABSTRACT. The impact of income and earnings on health has been well-examined in the health economics literature while the impact of health on wages has been less studied. Even rarer in previous work is the possible difference between the influences of health on wages for men versus women. In this paper, I attempt to fill this apparent gap in the literature. I augment the well-established earnings function to include a number of health indicators and estimate equations for men and women using eleven waves of the British Household Panel Survey. I consider a range of estimation procedures, including pooled ordinary least squares, random and fixed effects, and Hausman-Taylor instrumental variables approaches. The impact of health is found to differ slightly by sex and is more strongly related to women's wages more than men's.

Lynn Gambin
Department of Economics & Related Studies
University of York
York, UK
YO10 5NG
+ 44 8707 585622
Email: lmg110@york.ac.uk

1. INTRODUCTION

Differences in the health of men and women have been long noticed in developed countries. Statistics indicate that overall mortality rates in the UK have been higher for men than for women over the 20th century and that women continue to exhibit lower mortality. Additionally, morbidity appears higher for women than for men in developed countries. The types of illness and main causes of death also differ by gender – for men the main causes of death include cardiovascular diseases, cancers, and accidents, while for women they are breast cancer and cancers of the genitor-urinary system. Women also are heavier users of professional health care services than men, even after controlling for reproduction-related services.¹

There are also significant gender-related inequalities in income and earnings. Men, on average, earn more than women. There are also important differences regarding labour market participation patterns. The difference between male and female earnings and income has been examined quite extensively in the labour economics literature. Oaxaca (1973) explores a persistent earnings gap between male and female full-time urban workers using decomposition analysis. He discovers significant differences in the wage structure for males and females with respect to common characteristics such as race, marital status, and experience. With such differences in the wages and health of the sexes it is reasonable to hypothesize that the impact of health on wages could be different for men than for women.

It seems somewhat of an oversight that to date relatively little has been done to examine differences between men and women regarding the influence of health on earnings. The

¹ See Hayes and Prior (2003) for further information on this.

use of an economics framework for this task is even less apparent. More work has been done on gender inequalities in health and on gender inequalities in earnings. However, a combination of these two issues has not been a focus of many, if any, studies. Studying these two inequalities together is an important extension to the current base of knowledge relating socioeconomic factors and health. In this paper, I intend to examine gender inequalities in both health and wages using the economics framework. As there are significant differences between the sexes regarding both earnings and health it is important to investigate whether there is a difference by gender in the relationship between these two factors. The focus here is the influence of an individual's health on wages and related differences between men and women. In later papers other specifications will be used to further delve into the relationship between income, health and gender.

Many studies have addressed the impact of wages and income on health at both the individual and aggregate levels. The idea that income positively affects health is a common finding of studies and continues to be examined. The explanations for such a relationship are quite plausible. People with higher wages can afford lifestyles features (nutrition, exercise, etc) which can contribute to higher levels of health. Higher income also presents the ability for individuals to afford better quality and more health care to improve their health in times of sickness. Wages in particular, may have an effect on a person's health through many avenues. For example, lower wages may result in a person working additional hours to achieve some desired level of income and working excessive hours may adversely impact on health. Lower wages may also be prevalent in certain

types of work which may present additional hazards to health.² Obviously, the link between health and wages is not an uncomplicated one.

The impacts of health on wages are potentially as plausible as those presented for the reversed direction and require thorough study to really understand the complete relationship between health and wages³. Health may influence wages in a number of ways. There may be discrimination against unhealthy or disabled individuals in the labour market. Such discrimination may be exhibited in the offer of lower wages or of less employment opportunities to those with apparent health problems. Employers may believe that some health states would prevent an employee from being fully productive. Additionally, the employer may associate particular health states with other (unobservable) characteristics of the individual that would result in low productivity. Health also impacts individual labour income in that bouts of sickness may produce absenteeism, lowering opportunities for training and perhaps shortening tenure.

In addition to there being a small number of studies done on the impact of health on wages, most of the analyses that have been undertaken are limited in the conclusions they draw due to technical shortcomings such as the use of cross-sectional data and associated problems of endogeneity. The analysis presented here aims to overcome some of these limitations by using panel data and addressing endogeneity through an instrumental variables estimator presented by Hausman and Taylor (1981). I use panel and instrumental variable estimators on a sample from the British Household Panel Survey (BHPS) Waves 1 – 11 and estimate an earnings function to examine the influence of

² Compensating wage differentials would argue against this. In highly dangerous occupations such as fire-fighting, piloting, etc. wages would be higher to compensate for risk-taking by the worker. However, in factories in some industries, wages could be less than compensatory for the risk of losing limbs, or exposure to harmful chemicals, loud noises, etc.

³ Mushkin 1962

health on wages. Previously, health has been largely mistreated as a variable in such a context. In some models health has been treated as strictly exogenous⁴ or has been omitted from the analysis. As I have stated, it is as likely that wages influence health as it is that health influences wages. Both wages and health may be correlated with other factors affecting the relationship between the two. Thus, in this analysis I assume health to be endogenous in the model of wages. Previous analyses using the BHPS panel have estimated wage models but have omitted health as an explanatory variable.⁵ Again, our analysis overcomes this omission.

The paper proceeds as follows. In section 2, some of the previous analyses of the impact of health on wages or earnings are discussed. Section 3 outlines the empirical model, the earnings function, and the estimation procedures I intend to use. In section 4, there is a description of the data and variables used in the analysis. The results of the model estimation are presented in section 5. Finally, section 6 provides a conclusion and offers some potential extensions.

2. PREVIOUS WORK

Most of the work done to determine the effects of health on wages or earnings has been based, theoretically, on the concept of health being an investment in human capital as pointed out by Grossman (1972). In this framework, health can be viewed as a consumption or investment good. People desire good health for the utility gained directly from being in a good health state (consumption element) but also for the utility gained

⁴ Hausman and Taylor (1981) treat health as strictly exogenous in their analysis.

⁵ Previous analyses by Baltagi and Khanti-Atom (1990) and Kim and Polachek (1993) fail to include any indicators of health as covariates in their estimated earnings functions. Disney and Gosling (1998), Harkness (1999), and Hildreth (1999) utilize the BHPS in their analyses but also omit health as an explanatory variable; they are instead interested in the affects of public sector employment, gender, and unionsation on wages, respectively.

from the impact of health on market and non-market activities (investment element). In the Grossman model, health is treated as a component of human capital in which people will invest in order to increase the number of hours one can work as well as one's level of productivity. It is this treatment of health that motivates studying the link between health and earnings. A higher health stock is expected, according to the Grossman model, to increase earnings through enabling greater hours of work and may also influence earnings through greater productivity of the worker and thus, higher wages.

Luft (1975) presents one of the few papers that compare the influence of health on earnings by sex. The relationship between various components of earnings, including wages, and indicators of illness and disability are examined through fairly simple mean analysis and regression analysis using a cross-section of adults from the US. Luft finds that the impact of poor health on earnings is greater for women than for men but that this differs slightly by race. The paper does not focus greatly on the gender differences but goes on to look at the aggregate impact of illness and disability on income.

Using data on male US-born veteran twins, Bartel and Taubman (1979) find that the presence of various diseases decreases labour supply and wage rates. Berkowitz *et al* (1983) study the impact of health on wages, labour supply and annual earnings of white adult males. They use generalized least squares estimation and employ eight dichotomous physical impairment variables such as blindness, deafness, mobility problems, and mental difficulties, as indicators of health capital. Six of the eight indicators are shown to have significant negative effects on wages in their analysis. Thus, the data used by Berkowitz *et al* supports the idea of expanding the Grossman human capital theory to include not only the labour supply effects of health but also the wage effects. The results

of the paper are encouraging but the analysis is limited in that it uses cross-sectional data and is limited to white adult males. Lee (1982) also bases the analysis on Grossman and uses a sample of adult males from the US. Lee examines the joint dependence of wages and health using a simultaneous equations model with a three-stage estimation procedure. The findings are that wages and health are strongly jointly dependent. While wages have a strong positive affect on the demand for health, good health is also found to increase market productivity and thus wages.

Like these previous studies, Haveman *et al* (1994) analyse the health of only male adults. The authors develop a simultaneous model of work, wages and health with three structural equations. The data is taken from the US Panel Study of Income Dynamics and covers the years from 1976 to 1983. They use a categorical health variable. Their findings are consistent with previous studies. They find that lagged ill-health reduces wages and find this effect to be larger once they have accounted for endogeneity.

One of the more recent studies examining the influence of health on income is by Adams *et al* (2003). It uses data on elderly individuals from the US and examines plausible causal pathways between health and income. Socioeconomic status in their analysis is represented by wealth and so the study is not as similar to the analysis presented in our paper as I would like. However, Adams *et al.*, using panel data, do find that they cannot reject the hypothesis of a direct causal link from health to socioeconomic status. The results then can substantiate the logic of undertaking an analysis such as ours as well as provide some guidance in terms of technical analysis. Additionally, Adams *et al* propose various ways of testing for causality and can prove useful in directing such

investigations. Unfortunately, specific results of their study do not translate to the general population due to the relatively narrow scope of their sample.

The most recent relevant and similar study is from Contoyannis and Rice (2001). They use the first six waves of the BHPS and estimate the earnings function using various estimators. Contoyannis and Rice split their sample by sex and while gender inequalities are not their main focus, they do discuss some of the differences between men and women and recognize the importance of partitioning the sample by sex. They use single equation fixed effects and random effects instrumental variable estimators. Included in the estimations are variables indicating as marital status, work experience, age, occupational class, unionisation, and education. Their results show a significant impact of psychological well-being on the hourly wage for men and a significant impact of self-assessed health on women's wages.

3. EMPIRICAL MODEL

The first step in disentangling the relationship between earnings and health and the potential variations in the relationship across the sexes is to examine this association using the earnings function. I estimate the natural logarithm of hourly wage as a function of health (indicated by more than one variable) and other variables to control for age, education, and so on. Through this approach I can isolate the association between earnings and health and examine whether this relationship differs by gender.

I estimate the earnings function separately for men and women. Both full-time and part-time employees are included in the estimation but a separate set of estimations were performed using only full-time employees as well. Four models are estimated using each

subsample: (1) pooled ordinary least squares model (OLS); (2) fixed-effects, or within, model (FE); (3) random effects model (RE); and (4) Hausman-Taylor instrumental variables model (HT).

I describe the earnings function and each model in more detail here. Firstly, the earnings function is well-accepted part in labour economics work. It has developed with human capital theory and over the years there have been a number of amendments to the original specification. Mincer's model of earnings (1974) has been used as the framework for estimating the role of returns to schooling and work experience. It has also been used to consider the difference in these returns between men and women. The standard form of the Mincer earnings model relates earnings to schooling and experience as follows:

$$\ln[w(s, x)] = \alpha_0 + \beta_s s + \beta_0 x + \beta_1 x^2 + e \quad (1)$$

where w is hourly wage, s is level of schooling, x is work experience and e is a mean zero residual with $E(e | s, x) = 0$. The coefficients on s and x represent the returns to schooling and work experience, respectively. In equation (1) it is assumed that the constant term and slope coefficients are identical for all individuals. However, a more general model would not make such a restriction and would allow returns to schooling and experience, as well as the constant term, to differ across individuals as follows:

$$\ln[w(s_i, x_i)] = \alpha_{0i} + \beta_{si} s_i + \beta_{0i} x_i + \beta_{1i} x_i^2 + e_i \quad (2)$$

Equation (2) represents the random coefficient model of the earnings function.

Three important implications of Mincer's model are (i) log-earnings experience profiles are parallel across schooling levels; (ii) log-earnings age profiles diverge with age across schooling levels; and (iii) the variance of earnings over the lifetime has a U-shaped pattern. These implications of Mincer's model have been largely supported by the literature.⁶

For the analysis here, it is necessary to further augment the standard Mincer specification (1), as it has been in much work subsequent to the original. Most relevant here is to include a relationship between health and earnings in the specification. Additionally, I must allow the model to account for the panel nature of the data to be used and for other important variables to be included. Doing so produces the following:

$$W_{it} = X_{it}\beta + Z_i\gamma + a_i + \epsilon_{it}, \quad i = 1, \dots, N, \quad t = t_i, \dots, T_i \quad (3)$$

where i represents each individual and t each time period for which there is data on the individual. Rather than the time periods ranging from 1 to T , because I am using an unbalanced panel, it is necessary to indicate that individuals do not necessarily appear in all time periods and the total number of periods for which there is information on a respondent is individual-specific. W_{it} is the natural logarithm of hourly wages, X_{it} is a vector of time-varying, individual specific regressors including age, work experience, health and others. Z_i is a vector of time-invariant individual-specific regressors which includes ethnicity. a_i and ϵ_{it} are disturbance terms. a_i is time-invariant, individual-specific and is assumed to be $iid N(0, s_a^2)$ whereas ϵ_{it} is time-varying and assumed $iid N(0, s_\epsilon^2)$. I assume that the time-varying disturbances, ϵ_{it} , are not correlated with the explanatory

⁶ Heckman, Lochner and Todd (2003) examine the evidence obtained through use of Mincer's earnings function and find it largely supportive of the model, even with modifications for variables not considered in the original specification.

variables or with the other disturbances, a_i . The time-invariant, individual effects, a_i , may be correlated with some of the explanatory variables. It is necessary to separate them into exogenous or endogenous, for instrumental variables estimation. I can separate X and Z into $X = (X_1, X_2)$ and $Z = (Z_1, Z_2)$. I assume, *a priori*, that X_1 and Z_1 are exogenous while X_2 and Z_2 are assumed to be endogenous.

In all estimations there are some assumptions, other than model-specific assumptions, that should be pointed out. The first model estimation undertaken here is pooled OLS. In this procedure, the panel structure of the data is ignored and all observations are pooled together. OLS is unbiased and consistent as long as the errors (v_{it}) are not correlated with the observable explanatory variables (x_{it}), that is, OLS will achieve a consistent estimates if $E(x_{it}' v_{it}) = 0$ where $v_{it} = a_i + \epsilon_{it}$. This is however, a restrictive assumption. Inference using OLS requires estimation of the robust covariance matrix and robust test statistics. Pooled OLS will be inefficient as there is within-individual correlation of the errors. I am not maximizing the benefit of the information inherent in the panel by using OLS and so I next move to estimation procedures that do exploit the panel nature of the data.

Panel estimation allows us to control for individual heterogeneity and improve efficiency. Similar to the OLS estimator, but taking advantage of the panel data is the second model used – the random effects model (RE). Like OLS, RE puts the individual effect, a_i , in the error term. It imposes stricter assumptions than does pooled OLS – strict exogeneity in addition to orthogonality between the individual effects and the observable explanatory variables. A stronger conditional mean independence is necessary to substantiate inference. The strict exogeneity assumption can be stated as $E(\epsilon_{it}/x_i, a_i) = 0$ and the orthogonality assumption as $E(a_i|x_i) = E(a_i) = 0$. For consistency, I also need the usual

rank condition for GLS: $\text{rank } E(\mathbf{X}_i' \mathbf{O}^{-1} \mathbf{X}_i) = K$ where \mathbf{O} is the unconditional variance matrix of \mathbf{v}_i , $\mathbf{O} = E(\mathbf{v}_i \mathbf{v}_i')$. Efficiency of RE requires that $E(\mathbf{v}_i \mathbf{v}_i' | \mathbf{x}_i, a_i) = s_u^2 \mathbf{I}_T$ (conditional variances are constant and conditional covariances are zero), and $E(a_i^2 | \mathbf{x}_i) = s_a^2$ (same as $\text{Var}(a_i | \mathbf{x}_i) = \text{Var}(a_i)$ – homoskedasticity assumption on the unobserved effect, a_i). When all previous conditions are satisfied, RE is asymptotically equivalent to generalised least squares (GLS).

The random effects model may be reasonable to use in this context as the sample is drawn from a large population. It also saves a lot of degrees of freedom. However, there is a major drawback of the RE model. The RE model assumes that the random error associated with each cross-sectional unit is uncorrelated with any of the other regressors. This assumption is likely to be unrealistic. Failure to meet this assumption results in biased estimates.

The fixed effects, or within, estimator is the third method employed. The FE estimator is unbiased and consistent as N and/or $T \rightarrow \infty$ even if a_i is correlated with the regressors. The FE estimator is however, likely to be inefficient. The first assumption of the FE model is that of strict exogeneity of the explanatory variables conditional on a_i : $E(\mathbf{v}_{it} | \mathbf{x}_i, a_i) = 0$. Unlike in the case of RE, here I do not assume orthogonality. In FE, $E(a_i | \mathbf{x}_i)$ is allowed to be any function of \mathbf{x}_i . FE is more robust than RE however, in FE I cannot include time-constant factors in \mathbf{x}_{it} . Under this first assumption of strict exogeneity FE results in unbiased estimates. The second assumption of FE is the standard rank condition on the matrix of time-demeaned explanatory variables:

$\text{rank} \left(\sum E(\ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it}) \right) = \text{rank} [E(\ddot{\mathbf{X}}_{it}' \ddot{\mathbf{X}}_{it})] = K$. This assumption illustrates why time-constant variables are not permitted in FE analysis as if an element of \mathbf{x}_{it} that does not

vary over time for any i , then $\check{\mathbf{X}}_{it}$ would contain a column of zeros for all i . To ensure efficiency of FE requires a third assumption: $E(\epsilon_{it} | \mathbf{x}_i, a_i) = 0$ which implies that the idiosyncratic errors have a constant variance across all periods and are serially uncorrelated.

Finally, I estimate the relationship between health and earnings using the Hausman-Taylor instrumental variables estimator. This model fits panel-data random effects models in which some of the covariates are correlated with the unobserved individual-level random effect as follows:

$$W_{it} = \beta_1 X_{1it} + \beta_2 X_{2it} + \gamma_1 Z_{1i} + \gamma_2 Z_{2i} + a_i + \epsilon_{it}, \quad (4)$$

$$i = 1, \dots, N, \quad t = t_i, \dots, T_i$$

It assumes that a subset of the explanatory variables are correlated with the individual-level random effects, a_i , but that none of the explanatory variables are correlated with the idiosyncratic error, ϵ_{it} . This model can help to overcome the limitations of the RE and FE estimates. RE will produce inconsistent estimates if there is correlation between the individual effect and FE does not allow time-constant variables.

Unfortunately, the HT estimator requires strong exogeneity assumptions but it is more precise than the previous estimators. For the HT model, I assume strict exogeneity conditional on a_i . Another assumption is that $E(\mathbf{z}_i' a_i) = 0$, the individual unobserved effect is not correlated with the observed, time-constant explanatory variables.

Additionally, in the HT model I assume that $E(\mathbf{z}_i' a_i) = \mathbf{0}$ and $E(\mathbf{x}_{it}' a_i) = \mathbf{0}$ for all t .

Together with $E(\mathbf{z}_i' a_i) = 0$, these assumptions provide orthogonality conditions. These assumptions also allow for the definition of a set of instruments for estimation. The HT

estimator is consistent. It is also more efficient than the FE estimator if the model is overidentified and the explanatory variables are correctly classified as exogenous and endogenous. If some of the assumed exogenous variables are indeed correlated with a_i , then the estimator is inconsistent. A Hausman test between the FE and HT estimators can be used to evaluate this.

4. DATA AND VARIABLES

The data used in this paper were taken from waves 1 to 11 of the British Household Panel Survey (BHPS). The BHPS contains a vast array of information pertaining to households and individuals. The dataset contains rich information on socio-demographic, health and occupational characteristics of the respondents. It is a longitudinal survey of private households in England, Wales and Scotland. The survey was designed to be nationally representative of the population of Great Britain. It is an annual survey of all adult (16+) members of the selected households. The first wave of the BHPS was conducted in 1991. Respondents in the initial wave were interviewed in successive waves as were new members of the households. In cases where individuals moved from their original household to a new household (not initially covered by the survey) the individual as well as all other adult members of the new household were interviewed in subsequent waves. In following all individuals from the first wave and expanding the scope of respondents as individuals move, the aim is to keep the sample as representative as possible. Information was collected at both the household and individual levels. Such items as total household income, number of children in the household and so on would be collected at the household level. Individual-level items include individual income, occupational status, health status, mobility and many more.

The working sample used in this analysis comprises respondents from all eleven waves of the survey and thus covers the period from 1991 to 2001, inclusively. Included in the sample, however, are only those respondents who conform to the particular criteria that make them useful for the examination undertaken in this paper. Most importantly, only individuals giving valid responses for all items being used in the estimation were included. Additionally, selection of the sample also depended on employment status. Self-employed individuals were excluded from the sample. Those still in formal education were also eliminated. I include all individuals who satisfy the specified conditions in *any* wave of the BHPS and not only those who do so in all survey waves. The result of this is that the analysis that follows is based on an unbalanced panel of individuals. While a balanced sample would certainly help to simplify analysis, one positive aspect of utilizing an unbalanced panel is that the sample size is considerably larger and more information is retained. The unbalanced nature of the panel used has not posed any significant problems in this paper, however, the fact that an unbalanced panel is being used should be kept in mind during estimation and interpretation of results.

As the focus in this paper is on the affects of health on earnings (productivity) of men and women and not on labour supply issues, the sample is restricted to respondents indicating that they are employed at the time of the survey. Both part-time and full-time workers are included and estimation is performed using both part-time and full-time employees as well as using only full-time employees. By including part-time employees in the estimation, I can examine whether particular health profiles are associated with reductions in the amount of hours worked by individuals. The working sample resulting from omitting respondents in the above manner contains 20,387 males and 20,653 females who are either full-time or part-time employees, and 19,739 males and 13,555

females who are full-time employees only. Using these samples I estimate the earnings function, developed largely by Mincer (1958, 1974), which is common to the labour economics literature.

4.1. DEPENDENT VARIABLE

To investigate the influence of health on labour productivity and how it may differ by gender, I use hourly wages of an individual as the dependent variable. Inconveniently, there is no pre-constructed variable for hourly wage in the BHPS and so I have constructed one using the information that is available in the survey. The hourly wage (*wageall*) has been calculated using available data on a person's usual gross monthly pay, and usual hours worked per month including both regular and overtime hours.⁷ The average hourly wage was calculated as a weighted average of the hourly wage a person receives in his/her current main job and, where applicable, in his/her second paid job.⁸ The weights used are equal to the proportion of total hours worked per month attributable to each job. The dependent variable used in our model estimation is the natural logarithm of this constructed hourly wage and is presented in the model specification in section 3 as *lnwage*.

The average hourly wage for the full sample of males is £8.28 while for females it is £6.42. Amongst only full-time employees, the average wage is £8.32 for men and £6.77 for women. As expected, men, on average, have higher hourly wages than women. This may be due to a number of factors such as the selection of more men than women into

⁷ Variable descriptions are seen in Table 1 and the mean values of variables of interest are given in Table 2.

⁸ For all subsamples, the mean of the constructed hourly wage was greater than the mean of the hourly wage in the main job. Including the second job in constructing the average hourly wage allows us to use the wage variable as a proxy for an individual's 'maximum productivity' as those individuals receive lower levels of pay in their primary job may be able to compensate by obtaining perhaps higher paid secondary employment. Those with higher paid primary employment would be less likely to seek a second job.

higher paying occupations whether due to official or unofficial constraints on entry. This illustrates the fact that other factors such as occupational class, education, age and so on need to be included in our analysis so I can separate the impacts of health on wages more clearly. For both men and women, the (constructed) average wage rate is higher in the sample of only full-time employees than it is in the sample of all workers. 34.4% of all females in the sample were classified as part-time employees while for men the figure is only 3.2%. Women may exhibit lower rates of full-time employment than men because of the traditional role of women as the main caregiver for children or family members. These figures also highlight the fact that although this analysis is focused on earnings and tries to diminish the issue of labour supply it is not entirely ignorable. I have excluded unemployed and self-employed persons from our sample however labour supply decisions still remain a factor in the analysis as once a decision to participate in the labour market has been made individuals can then choose to work more or less hours in accordance with health status and health shocks.

4.2. TIME-VARYING VARIABLES

It is important to distinguish between the explanatory variables which are time-varying and those that are time-invariant. As the affects of health on wages is the main interest here, I start by describing the time-varying health variables used in the estimation. I hypothesize that that these health variables are endogenous. Three indicators are used in the analysis. The first indicator of health is self-assessed health. From the BHPS I use two items to obtain the self-assessed health status of individuals. For all waves except wave 9, I use the item *hlstat* while for the ninth wave I use item *hlsfl*. The reason for using a different item for self-assessed health in wave nine is that for that wave *hlstat* was not included. The two items differ in the phrasing of the question of self-assessed health

and also in the categorical answers provided in the questionnaire.⁹ Both items provide five different categories which respondents can use to indicate their own assessment of their overall health.

Self-assessed health is a very subjective indicator of an individual's health, however, many studies have used the self-assessed health indicator as their main health variable in various analyses relating health and income (Luft 1975, Haveman *et al* 1994). It has been shown to be a reliable indicator as they are highly correlated with medically determined health status¹⁰. Using such a subjective measure seems intuitively acceptable as one's perception of health may be as important as one's clinically assessed health in affecting wages due to the psychological implications that one's own opinion of well-being may reflect non-clinical problems. Also, a person's perception of their own health may be more important than actual health in determining how much effort one puts into working. However, limitations of subjective measures of health should be kept in mind when interpreting results.

In preliminary estimations (not reported in this paper), all five categories of self-assessed health were included. From the self-assessed health items I created two dummy variables. The first self-assessed health dummy variable, *sahex*, equals one where self-assessed health is indicated as excellent and zero otherwise; the second variable, *sahgd*, is equal to one if self-assessed health is indicated as very good or good and zero otherwise. In our sample, some 29.6% of males and 24.7% of females indicate excellent health while 47.3% of males and 49.3% of females indicate good/very good health. According to the

⁹ In wave 9, the categories for self-assessed health are: excellent, very good, good, fair and poor. In all other waves the categories are: excellent, good, fair, poor and very poor. We have grouped good from all waves except wave 9 and very good from wave 9 together in the variable *sahgd*.

¹⁰ See Currie & Madrian 1999 for more thorough review of relevant studies and for references regarding validity of health measures.

investment theory of human capital, better health is reasonably expected to improve a worker's productivity and consequently wages. I expect the coefficients on both of these variables to be positive with that of *sahex* being greater than the *sahgd* coefficient.

The second health variable is derived from the General Health Questionnaire (GHQ) included in the BHPS. The variable *hlghq1* corresponds to the survey item *hlghq1*.

Originally, this item was developed to be a screening instrument for psychiatric illness but in this context it is used as another measure of subjective well-being. *hlghq1* is a composite index number that ranges from 0 to 36 and is decreasing in psychological well-being. In our sample the average value of *hlghq1* is 11.4 for women and 10.0 for men which agrees with the general finding of higher morbidity in women than men but the difference is small here¹¹. I expect the coefficient on this variable to be negative as poorer psychological well-being could reasonably be associated with lower earnings.

Psychological well-being can influence a worker's productivity as it may influence their effort and perception of their roles. Additionally, employers with knowledge of workers' psychological problems may offer lower wages to those with lower psychological well-being in anticipation of lower productivity and increased absenteeism.

The third health indicator (fourth variable) is *hlprb1* which indicates whether or not an individual has indicated the presence of any health problems that limit activity or work. This variable is equal to one if a person indicates that they have no health problems and equals zero if they have one or more such problems. In the sample of all workers, 55.5% of men indicate that they have no health problems while only 45.9% of women have no problems.

¹¹ See Griffin *et al* (2002).

There are a number of other time-varying variables included in the analysis. The impact of unions on earnings is an important factor to consider in estimating a wage function as significant influences of unionsation on wages have been found in a number of studies.¹² I use two variables to account for union influence on earnings.¹³ The first variable, *covmem*, indicates whether there is a union present in a person's workplace and he/she is a member of that union. The second variable, *covnon*, indicates that a union is present in the workplace but that the individual is not a member. The presence of a union in a workplace is associated with higher wages for employees, whether or not they are members of that union. I expect the coefficients on both *covmem* and *covnon* to be positive. I also expect that being a member of a workplace union would be associated with higher wages than not being a union member. In our sample, just over 30% of men and women belonged to a union in their workplace while just 15.3% and about 19.8% of women were covered by a union in the workplace but were not members of the union.

Age and experience are also included in the estimation. I include the levels as well as the squares of these variables (*age*, *agesq*, *exp*, *expsq*).¹⁴ These are typical factors to include when estimating a wage function. Age should capture the effects of general experience as well as job-specific experience on earnings. The average age of both men and women in our full sample is about 38 years, while in the sample with full-time employees only the average age of men is 37.8 years and that of women is 37 years. Experience, as I use it here, is not given directly in the BHPS data. I calculate experience as the number of years an individual has been working at his/her current (main) job. On average, in our sample,

¹² Jakubson (1991) uses a panel of men from the US and finds that there is a positive increase in wages of 5 – 8% associated with unionsation. Freeman (1984) confirms that despite bias problems in some estimations, unionsation has a definite impact on wages.

¹³ This treatment of unionsation indicators is used by Hildreth (1999).

¹⁴ *age* and *exp* indicate the levels of age and job-specific experience in years, respectively. *agesq* (*expsq*) is equal to the square of the person's age (experience) divided by ten.

men have more years of experience than do women. In the full sample the average number of years of experience for men is 5.6 years and for women it is 5.2 years. In the sample of full-time employees the average for men is 5.8 years while women have almost one year less of experience with an average of 4.9 years.

A variable indicating the sector in which a person is employed (*jobpriv*) is included. It is important to specify whether an employee is in the private sector or the public sector as this has been shown to influence wages. Differences in the influence of sector on wages have been observed across the genders as well by Disney and Gosling (1998) and Dustmann and van Soest (1997). They also find that this variable may be endogenous and so I assume *jobpriv* to be endogenous here. 80.2% of the males in our sample and just 61.3% of females work in the private sector. A variable indicating the size of a person's workplace (indicated by the number of employees), *JBSIZE*, is also included. *JBSIZE* is a vector of dummy variables corresponding to a number of different ranges of numbers of employees (these can be seen in Table 1). Green *et al* (1996) predict a positive relationship between employer size and wages. They also observe differences in this relationship depending on union status and gender.

A vector of binary variables, *OCCLASS*, representing occupational classes (*prof*, *manag*, *skllnm*, *skllm*) is also included in the estimation. These variables are also assumed to be endogenous as individuals may select into employment as such classes of worker on the basis of unobservable characteristics that are also linked to wages. In our sample, proportionally more men than women work in professional occupations and in skilled manual and unskilled occupations. Roughly the same proportion of women as men work in managerial occupations but more women work in skilled non-manual occupations.

A dummy variable, *jobpt*, is included to distinguish between full- and part-time workers. This variable is equal to one if a person works less than thirty hours per week and zero if he works thirty or more hours. The expected sign of the coefficient on *jobpt* is not entirely straightforward. A positive sign could arise if it happens that those who work part-time do so because the wage in that job is sufficiently high for them to reach a desired level of income. A negative coefficient may reflect that many part-time jobs require low skill levels and attract workers with less opportunity to find higher paying employment.

Also included is a vector of indicators of marital status (*divsep*, *widow*, *nvrmar*) called MARSTAT. The theory of household production tells us that people do not make labour decisions in isolation but instead there is interaction between members of the same household. Changes in marital status may also impact on individual earnings. Married or cohabiting individuals may have more freedom over their labour market activity as their partner's wages may be sufficient to sustain the household. Similarly, a newly divorced or separated individual may be compelled to find employment or increase working hours to compensate for the loss of his/her partner's income. A greater proportion of women than men in our sample are divorced or separated, while a greater proportion of men than women have never been married. Proportionally more women than men are widowed. In the full-time employed sample, about 75% of the sample are married or living as a couple.

Another variable included that relates to the make-up of a household or family is *kids04* which measures the number of children between 0 and 4 years of age. Harkness (1996)

found a significant effect of the presence of young children in a household on earnings. In the Harkness paper there was a positive coefficient for men and a negative one for women. Over 80% of men and women in the sample have no children between the ages of 0 to 4 years.

A vector of regional variables (REGION) is also included to account for differences in costs and wages across geographic areas of the United Kingdom. The final time-varying variable included is a vector of time dummies (WAVE). Including these should control for aggregate productivity effects and inflation.

4.3. TIME-INVARIANT VARIABLES

One of the time-invariant and exogenous variables included in this analysis is ethnicity (*race*). This variable is a dummy variable taking the value of 1 if an individual is white and 0 otherwise. About 97% of individuals in our sample are white. Also time-invariant in our sample is the highest level of educational attainment. This item is assumed to be endogenous as it normally is in such an analysis. In our estimations, I use a binary variable (*deg*) indicating whether a person has a degree (*deg* = 1) or not (*deg* = 0). This follows the procedure of Contoyannis and Rice (2001).¹⁵ 85% of men in our sample and 87.9% of women in our sample have obtained lower educational qualifications than a degree.

Rather than including a variable for gender in the estimation, I have decided to split the overall sample by sex. I expect the coefficients on the variables to differ between men

¹⁵ Including all categories was explored but we decided for the same reasons as Contoyannis and Rice (2001) to use degree as a binary indicator of educational attainment.

and women and since this is one of the main interests I have then separating the sample is a logical procedure. Previously, Contoyannis and Rice (2001), Hildreth (1999), Harkness (1996), and Kim and Polachek (1993) also split their samples and find that the coefficients differ by sex.

5. RESULTS AND DISCUSSION

The results that are discussed in this section are based on the estimation of the following reduced form equation:

$$\begin{aligned} \ln wage = & a_0 + a_1 age + a_2 agesq + a_3 exp + a_4 expsq + a_5 jobpriv + \\ & a_6 white + a_7 degree + a_8 kids04 + a_9 jobpt + \\ & a_{10} covmem + a_{11} covnon + ? OCCLASS + \\ & \mathbf{d} \text{ MARSTAT} + ? \text{ JBSIZE} + ? \text{ REGION} + ? \text{ WAVE} + \\ & \mu_1 sahex + \mu_2 sahgd + \mu_3 hlghq1 + \mu_4 hlprb1 \end{aligned} \quad (5)$$

*bold typeface indicates a vector of coefficients.

Both equations were estimated separately for men and women. I also estimated the equation for full-time employees only (with *jobpt* removed).

Tables 3 and 4 show the estimates obtained for equation (5) using the four estimation procedures outlined in section 2. The standard errors are shown in parentheses. Also, the estimates of the coefficients for size of employer, regions and waves are omitted from the tables as those estimates are not of great interest in this paper and the estimated values and signs of these coefficients are what I expected prior to estimation.

As stated in section 2, estimation of the earnings function was carried out using a number of estimation procedures: ordinary least squares, random effects, fixed effects and the Hausman-Taylor model. In all cases, that is, for the full sample and for the fully employed sample, as well as for males and females, the results of the OLS and RE

estimation are quite similar. The results obtained using FE and HT are also much alike. This is an expected observation.

5.1. MALES

I look more closely at the results for males first. In the case of the full sample of males (20,387 observations), both age and experience coefficients indicate a convex relationships with individual earnings as the coefficient on the original variable is positive while the coefficient on its square is less than zero. This indicates that wages increase with age and experience up to a certain point after which there is a negative relationship. The coefficient on *jobpt* is positive. The coefficients on the variables indicating occupational class show a clear gradient of increasing hourly wage from lower classes (unskilled (the reference group), skilled manual) to the higher occupational classes (management, professionals). The results also show that *lnwage* is higher for those males working in private sector jobs than for those not in the private sector. The coefficients on the union variables (*covnon*, *covmem*) are greater than zero with the effect of being a covered member greater than being a covered non-member. White males have higher hourly wages than do men of other races. Those with a higher degree earn more than those with lower levels of educational attainment. In all models, the estimated coefficients on *divsep* and *nvrmar* indicate that the natural logarithm of wages is less for men indicating that they are not currently married or widowed. The lowest wages are attributed to those who have never been married. An unexpected result is that the coefficient for *widow* is positive in all four models indicating that on average, widowed men earn more than married men. The coefficient on *kids04* is positive for all models.

Most important, for our purposes, are the estimates for the coefficients on the four health variables (*sahex*, *sahgd*, *hlghq1*, *hlprb1*). In the OLS and RE models, the estimated coefficients for *sahex* and *sahgd* are positive while the coefficient on *hlghq1* is estimated as less than zero in all models. This is the expected result for *sahex* and *sahgd* as one would expect men with excellent or good health to have higher wages than those with lower levels of health. Additionally, as expected, the coefficient for *sahex* is greater than that for *sahgd* (where positive). The result for *hlghq1* is also as I would expect. Higher values of *hlghq1* indicate worse health and so one would expect a negative coefficient on this variable. The health problems variable, *hlprb1*, like the two SAH variables, is positive in the OLS and RE estimates but negative (albeit statistically insignificant) in the FE and HT models. In the OLS model, where the coefficient on *hlprb1* is positive and significant, the result is as expected as it indicates that the absence of health problems is associated with higher logarithm of hourly wage.

In the sample including only full-time employed males (19,739 observations), the results were similar to those of both full-time and part-time employed men. The relationship between the natural logarithm of wages and age, the square of age, experience and the square of experience is what one expected – a convex relationship between wages and age and between wages and experience. Again, the expected gradient is apparent across occupational classes as all coefficients for the classes are positive and they increase from *skllnm* to *prof*. Private sector employees are also shown to have higher wages than non-private employees. As in the first set of results, the union variables appear to have the expected influence of increasing earnings. In all four models the coefficient on *covmem* is greater than that for *covnon* thus reinforcing the idea that the presence of a union in workplace increases wages overall and that being a member of that union increases

wages above those of non-members. White males have higher earnings than non-whites on average. Males with higher degrees also have higher wages than those with less education. The results corresponding to the marital status variables are quite the same as those in the first sample of men – divorced or separated men and single, never married men have lower wages than presently married men. Once again, the result of widowed men earning more than currently married men, *ceteris paribus*, is present. The coefficient on *kids04* is positive as it was in the first set of results.

In this sample, the estimated coefficients for the health variables are not estimated as similarly across all models as they were using the first sample. Here, the coefficient on *hlghq1* is negative and corresponds to our expectations in all four models. Again, for the OLS and RE models the coefficients on *sahex*, *sahgd* and *hlprb1* are positive but, in the FE and HT models these coefficients become negative (but are statistically insignificant). As expected, the result of the coefficient corresponding to excellent health being greater than that for good health holds in all models.

5.2. FEMALES

For females, the results are somewhat different than the men's results. In the full sample of females (20,653 observations) the results indicate that there is a convex relationship between wages and age, as well as between wages and experience. The coefficient on *jobpt* is negative only for the OLS estimation and was greater than zero in the other three models. As expected, and as was the case for males, the gradient across occupational classes was apparent for this sample of full-time and part-time employed women. Contrary to the results for men, for women in this sample the coefficient on *jobpriv* is negative indicating that women employed in the private sector earn less than those in the

public sector, *ceteris paribus*. The presence of a union in a workplace also increases wages for female workers and even more so when the woman is a member of the union. White women also earn more, on average, than women of other races. Higher education is also associated with higher earnings for women as the coefficient for *deg* is greater than zero. In the HT model, the coefficient on *deg* is the greatest at 1.0241. The estimated coefficients for the marital status variables are somewhat variable across the different models. The coefficient on *widow* is negative for the OLS and RE models but is positive for the FE and HT models. The coefficient on *divsep* is positive for all but the OLS model. The coefficient for *nvrmar* is negative in all models. The result of divorced or separated women earning more than married women may be due to the fact that for such women working at better paying jobs or working multiple jobs may be necessary as they lack the income of a partner that would be present if married or cohabiting. The presence of children under the age of four years is associated with higher *lnwage* in all models except the FE model. Finally, the association between the health variables and the logarithm of wages corresponds to our expectations. Having good or excellent health is associated with higher earnings for women in this sample (except in the FE model where the coefficients on *sahex* and *sahgd* are less than zero) while higher GHQ scores, indicating worse health, are associated with lower earnings. The coefficient for *hlprbl* is positive and statistically in all models.

In the case of full-time employed women only (13,555 observations) the convex relationships between earnings and age and earnings and experience continue to hold. The gradient across occupational classes is also apparent again however in the FE and HT models the management class has a slightly higher coefficient than the professional occupational class. Again, the coefficient on *jobpriv* is negative for women. Workplace

unions and membership in these unions is also associated with higher earnings for women in this second sample. White women in this sample also have higher earnings than non-white females. The coefficient on *deg* is positive. There is a negative association between each of never married women and widowed women with the natural logarithm of wages in all models. Being divorced or separated is on average, associated with lower earnings in all models except the FE model where the coefficient on *divsep* is positive. As in all previous results, the coefficient for *kids04* is positive. The signs on the health variables for this final estimation are the same as those for the full sample of women but the magnitudes are somewhat different.

Age and experience have higher coefficients for males than females in our sample. This may reflect that men can move up the pay scale more easily than women can. It may also be indicative of women having less continuity in working as they may have to stop work at various points in their careers to have children and to care for family members. The unionsation variables are estimated to have greater impact on the earnings of women than men. Working in the private sector results in a greater gain in earnings for men than for women as indicated by the positive coefficient on *jobpriv* for males and the negative coefficients for females. Women, it seems, earn more on average in the public sector than do their female counterparts in the private sector. Men, on the other hand, seem to be able to take advantage of enhanced opportunities in the private sector and earn more there on average than their male counterparts in the public sector. The occupational class has various impacts on men and women. The coefficients on professional, managerial, skilled non-manual and skilled-manual occupations are greater for women than men. The results regarding marital status are quite mixed. Widowed men appear to earn more than married

men while for women the reverse is true.¹⁶ All coefficients for divorced or separated men are negative but for women only the OLS estimate is negative and the other models indicate that women's wages increase given divorce or separation, *ceteris paribus*. This may be due to the necessity to obtain supplementary income that changing from married or cohabiting to living without a partner presents. For both sexes, the coefficient on being single is less than zero. Having young children in the household affects both the wages of men and women significantly but the difference between the sexes regarding this is not great. The difference in wages between white men and non-white men is larger than the difference between white women and non-white women. Finally, higher education seems to present a greater gain in earnings to women than to men as the coefficients on DEG are greater for females.

In all of the estimations, I obtain the expected signs on the estimated coefficients of the health variables (except in those cases previously noted). The coefficients on *sahex* and *sahgd* are positive while the coefficients on *hlghql* are negative. Additionally, the expected gradient of the coefficient on *sahex* being greater than that on *sahgd* is observed. The estimated coefficients on the self-assessed health variables are greater for women than men on the whole. In the full sample, the OLS coefficient on *sahex* is 0.0133 greater for females than males; in the RE model the difference is 0.0117. For *sahgd*, the coefficient is 0.0047 greater for women than men in OLS and 0.0102 greater in RE. The estimated coefficients for *hlghql* are negative in all cases. The estimate is bigger in magnitude for females than males in OLS but is greater for males in all other models. The differences between the coefficients on *hlghql* are very small between the sexes and the statistical significance of the estimates is very low in some instances. *hlprb1* is

¹⁶ Note that in the FE and HT models on the full sample of females and in the FE model for the full-time sample there were positive coefficients on *widow*.

statistically significant for males in only the OLS model, while it is significant and positive for women in all models. In the case of OLS, the coefficient is 0.0094 higher for men than women.

These estimates show that health impacts the hourly wages of the individuals in our sample. Of particular interest is the difference between the impact on males and females. I observe that self-assessed health seems to impact wages more for women than for men. Holding all other variables constant, an indication of excellent health ($sahex = 1$) for males would increase the hourly wage from £1.012 to £1.042 per hour while for women the change would be from £1.024 to £1.056¹⁷. For good health ($sahgd = 1$), ceteris paribus, there would be an increase of hourly wages of £1.021 for men in the OLS estimates and from £1.012 to £1.025 for women. There is little difference between the sexes according to the impact of Likert Scores ($hlghq1$) on hourly earnings. An indication of no health problems ($hlprb1 = 1$) for men would result in an increase of £1.030 per hour for men and an increase ranging from £1.015 to £1.020 for women.

6. CONCLUSION

The impact of health on wages has not received sufficient attention in the economics literature. While more studies focus on the influence of wages and income on health it is equally important to examine the link from health to earnings. The analysis undertaken in this paper extends the work exploring this relationship using panel data and various estimates of the earnings function. In addition, this analysis also focuses attention on the difference between men and women regarding the influence of health on wages. Using all eleven waves of the BHPS, I find that there are differences in the relationship between

¹⁷ The estimates of the marginal effects shown are calculated using only statistically significant estimates

self-assessed health, presence of health problems and psychological well-being scores with hourly wages. The wages of both men and women are affected by health but from the coefficients estimated here appear to be slightly greater affect for women.

Contoyannis and Rice (2001) find that self-assessed health impacts the wages of women while GHQ Likert Scores affect men's wages. Here I find that both health indicators affect the wages of both sexes. However, the additional health variable used in this analysis, the absence of health problems, has a slightly higher impact on wages for men than women.

To improve and extend the findings of this paper a number of steps can be undertaken.

The inclusion of additional health indicators could increase the validity of our results. As well, dynamic specifications and other models such as those proposed by Amemiya and MaCurdy (1986) and Breusch, Mizon and Schmidt (1989) would also be useful. In order to undertake such models a balanced panel would be more appropriate than the unbalanced panel used here. Additionally, simultaneity bias may be present in the current analysis and addressing this would be an important improvement. Another improvement would be to extend the sample used to both unemployed and self-employed individuals. As an extension to the work done here, decomposition analysis, as presented by Oaxaca (1973, 1994) or Macpherson and Hirsch (1995), would be useful in addressing how much of the gender gap in wages is attributable to individual health and other factors.

Table 1: Variable labels and definitions

Label	Definition
<i>wageall</i>	average hourly wage
<i>sahex</i>	self-assessed health (1 = excellent, 0 = otherwise)
<i>sahgd</i>	self-assessed health (1 = good/very good, 0 = otherwise)
<i>hlghq1</i>	General Health Questionnaire Likert Scale score
<i>hlprb1</i>	health problems (1 = no health problems, 0 = 1 or more)
<i>covmem</i>	union status: 1 = covered union member, 0 = otherwise
<i>covnon</i>	union status: 1 = covered non-member, 0 = otherwise
<i>age</i>	age in years
<i>exp</i>	number of years in current job
<i>jobpriv</i>	sector of employment: 1 = private, 0 = public
JBSIZE	size of employer: number of employees
<i>jb39</i>	1 = 3 – 9 employees, 0 = otherwise
<i>jb1024</i>	1 = 10 – 24 employees, 0 = otherwise
<i>jb2549</i>	1 = 25 – 49 employees, 0 = otherwise
<i>jb5099</i>	1 = 50 – 99 employees, 0 = otherwise
<i>jb100199</i>	1 = 100 – 199 employees, 0 = otherwise
<i>jb200499</i>	1 = 200 – 499 employees, 0 = otherwise
<i>jb500999</i>	1 = 500 – 999 employees, 0 = otherwise
<i>jb1000p</i>	1 = 1000 or more employees, 0 otherwise
<i>prof</i>	occupational class: 1 = professional
<i>manag</i>	occupational class: 1 = managerial
<i>skllnm</i>	occupational class: 1 = skilled non-manual
<i>skllm</i>	occupational class: 1 = skilled manual
<i>divsep</i>	marital status: 1 = divorced or separated
<i>widow</i>	marital status: 1 = widowed
<i>nvrmar</i>	marital status: 1 = single, never married
<i>kids04</i>	number of children in household age 0 – 4 years
<i>white</i>	ethnicity: 1 = white, 0 = otherwise
<i>deg</i>	highest educational attainment: 1 = degree or higher, 0 = otherwise
REGION	regional indicators
<i>london</i>	1 = London area, 0 = otherwise
<i>southe</i>	1 = southeast England, 0 = otherwise
<i>southw</i>	1 = southwest England, 0 = otherwise
<i>eanglia</i>	1 = East Anglia, 0 = otherwise
<i>midlands</i>	1 = Midlands, 0 = otherwise
<i>northwest</i>	1 = northwest England, 0 = otherwise
<i>northeast</i>	1 = northeast England, 0 = otherwise
<i>wales</i>	1 = Wales, 0 = otherwise
<i>scot</i>	1 = Scotland, 0 = otherwise
<i>jobpt</i>	part-time employment: 1 = part-time employee, 0 = full-time employee

Table 2: Variable means

	Males		Females	
	full sample	full-time	full sample	full-time
	20,387 obs	19,739 obs	20,653 obs	13,555 obs
<i>wageall</i>	8.2844	8.3204	6.4187	6.7735
<i>hlprb1</i>	0.5550	0.5602	0.4594	0.4712
<i>sahex</i>	0.2964	0.2976	0.2468	0.2578
<i>sahgd</i>	0.4735	0.4733	0.4933	0.4890
<i>hlghq1</i>	10.0386	10.0473	11.4217	11.3561
<i>age</i>	38.1781	37.8331	38.4843	36.9530
<i>exp</i>	5.7180	5.7646	5.2082	4.9526
<i>divsep</i>	0.0461	0.0455	0.0892	0.0971
<i>widow</i>	0.0060	0.0050	0.0210	0.0175
<i>nvrmar</i>	0.1973	0.1961	0.1479	0.1963
<i>kids04</i>	0.1841	0.1872	0.1302	0.0823
<i>deg</i>	0.1464	0.1480	0.1211	0.1506
<i>white</i>	0.9724	0.9727	0.9727	0.9678
<i>jobpriv</i>	0.7985	0.8018	0.6170	0.6132
<i>covmem</i>	0.3228	0.3299	0.3129	0.3594
<i>covnon</i>	0.1533	0.1509	0.1978	0.1880
<i>jb39</i>	0.1302	0.1262	0.1732	0.1323
<i>jb1024</i>	0.1394	0.1369	0.1810	0.1612
<i>jb2549</i>	0.1244	0.1255	0.1441	0.1451
<i>jb5099</i>	0.1270	0.1283	0.1054	0.1193
<i>jb100199</i>	0.1151	0.1160	0.0982	0.1125
<i>jb200499</i>	0.1516	0.1542	0.1069	0.1220
<i>jb500999</i>	0.0789	0.0803	0.0567	0.0678
<i>jb1000p</i>	0.1042	0.1061	0.0986	0.1170
<i>prof</i>	0.0645	0.0755	0.0254	0.0317
<i>manag</i>	0.3068	0.3122	0.3101	0.3834
<i>skllnm</i>	0.1321	0.1291	0.3846	0.3605
<i>skllm</i>	0.3094	0.3135	0.0873	0.0874
<i>london</i>	0.0788	0.0790	0.0874	0.0974
<i>southe</i>	0.1700	0.1696	0.1707	0.1716
<i>southw</i>	0.0805	0.0802	0.0726	0.0639
<i>eanglia</i>	0.0363	0.0358	0.0314	0.0291
<i>midlands</i>	0.1560	0.1575	0.1524	0.1501
<i>northwest</i>	0.0919	0.0926	0.0890	0.0909
<i>northeast</i>	0.1369	0.1363	0.1358	0.1283
<i>wales</i>	0.0989	0.0987	0.0955	0.0963
<i>scot</i>	0.1425	0.1420	0.1558	0.1636
<i>jobpt</i>	0.0318	--	0.3437	--

Table 3: Regression results using *lnwage* as dependent variable in full-time and part-time sample.

	MALE				FEMALE			
	OLS	RE	FE	HT	OLS	RE	FE	HT
<i>hlprb1</i>	0.0291 (5.24)	0.0049 (0.95)	-0.0056 (-0.99)	-0.0055 (-1.04)	0.0197 (3.55)	0.0176 (3.22)	0.0144 (2.34)	0.0152 (2.60)
<i>sahex</i>	0.0410 (5.18)	0.0121 (1.76)	-0.0048 (-0.66)	-0.0011 (-2.16)	0.0543 (6.69)	0.0238 (3.28)	0.0005 (0.06)	0.0022 (0.29)
<i>sahgd</i>	0.0203 (2.98)	0.0018 (0.33)	-0.0079 (-1.36)	-0.0054 (-0.79)	0.0250 (3.75)	0.0120 (2.11)	0.0044 (0.74)	0.0047 (0.82)
<i>hlghq1</i>	-0.0008 (-1.29)	-0.0011 (-2.20)	-0.0010 (-1.91)	-0.0082 (-1.52)	-0.0018 (-3.36)	-0.0005 (-1.11)	-0.0002 (-0.31)	-0.0001 (-0.30)
<i>age</i>	0.0578 (33.78)	0.0753 (38.33)	0.0659 (7.91)	0.0895 (35.69)	0.0359 (18.90)	0.0420 (20.89)	0.0389 (4.49)	0.0507 (19.01)
<i>agesq</i>	-0.0656 (-31.47)	-0.0850 (-36.08)	-0.0988 (-27.79)	-0.0998 (-33.53)	-0.0435 (-18.13)	-0.0505 (-20.62)	-0.0546 (-14.11)	-0.0588 (-18.48)
<i>exp</i>	0.0071 (5.97)	0.0054 (5.17)	0.0043 (3.62)	0.0050 (4.78)	0.0091 (6.04)	0.0049 (4.00)	0.0014 (0.99)	0.0036 (2.90)
<i>expsq</i>	-0.0133 (-3.15)	-0.0092 (-2.49)	-0.0079 (-1.82)	-0.0091 (-2.39)	-0.0241 (-3.58)	-0.0109 (-2.14)	-0.0032 (-.54)	-0.0069 (-1.31)
<i>widow</i>	0.0565 (1.39)	0.1163 (2.90)	0.1881 (3.62)	0.1703 (3.87)	-0.0120 (-.65)	-0.0042 (-.18)	0.0164 (0.52)	0.0087 (0.33)
<i>divsep</i>	-0.0602 (-4.60)	-0.0463 (-3.48)	-0.0326 (-2.08)	-0.0397 (-2.89)	-0.0097 (-1.07)	0.0048 (0.44)	0.0289 (2.09)	0.0176 (1.49)
<i>nvrmar</i>	-0.1202 (-15.02)	-0.0797 (-8.60)	-0.0372 (-3.16)	-0.0490 (-4.82)	-0.0378 (-4.72)	-0.0476 (-4.76)	-0.0725 (-5.39)	-0.0611 (-5.34)
<i>kids04</i>	0.0260 (4.44)	0.0131 (2.50)	0.0108 (1.93)	0.0119 (2.35)	0.0695 (8.97)	0.0229 (3.46)	-0.0092 (-1.27)	0.0062 (0.93)
<i>covnon</i>	0.0196 (2.48)	0.0354 (4.73)	0.0340 (4.14)	0.0331 (4.47)	0.0460 (6.02)	0.0393 (5.04)	0.0371 (4.19)	0.0377 (4.69)
<i>covmem</i>	0.1094 (15.47)	0.1045 (13.49)	0.0928 (10.09)	0.0947 (11.70)	0.1236 (16.74)	0.1216 (14.73)	0.0944 (9.43)	0.1019 (11.30)
<i>jobpt</i>	0.0194 (0.84)	0.0994 (6.71)	0.1600 (9.47)	0.1366 (9.13)	-0.0574 (-8.85)	0.0165 (2.57)	0.0748 (9.97)	0.0478 (7.10)
<i>jobpriv</i>	0.0539 (7.19)	0.0239 (2.55)	-0.0022 (-.18)	0.0032 (0.28)	-0.0608 (-9.03)	-0.0644 (-8.17)	-0.0330 (-3.23)	-0.0379 (-3.98)
<i>prof</i>	0.5397 (44.23)	0.2750 (21.70)	0.0858 (5.84)	0.0907 (6.64)	0.6060 (30.24)	0.3558 (18.72)	0.1016 (4.56)	0.1308 (6.21)
<i>manag</i>	0.4670 (55.09)	0.2545 (28.27)	0.0886 (8.13)	0.0938 (9.28)	0.4688 (55.40)	0.3019 (32.88)	0.1112 (9.58)	0.1317 (12.03)
<i>skllnm</i>	0.2312 (25.01)	0.1245 (12.67)	0.0161 (1.39)	0.0195 (1.81)	0.2217 (31.94)	0.1426 (17.09)	0.0227 (2.12)	0.0373 (3.68)
<i>skllm</i>	0.1178 (15.95)	0.0658 (8.78)	0.0278 (3.31)	0.0302 (3.86)	0.0739 (7.38)	0.0571 (5.53)	0.0315 (2.70)	0.0369 (3.33)
<i>white</i>	0.0650 (3.76)	0.0587 (2.17)	<i>dropped</i>	0.2727 (5.03)	0.0549 (3.43)	0.0416 (1.53)	<i>dropped</i>	0.1867 (3.84)
<i>deg</i>	0.1661 (18.65)	0.2562 (18.49)	<i>dropped</i>	0.8895 (7.95)	0.2158 (22.31)	0.2903 (20.87)	<i>dropped</i>	1.0241 (11.42)

t-statistics/z-statistics in parentheses. Significance at 5% level in bold.
robust OLS estimation

Table 4: Regression results using *lnwage* as dependent variable in full-time only sample.

	MALE				FEMALE			
	OLS	RE	FE	HT	OLS	RE	FE	HT
<i>hlprb1</i>	0.0302 (5.45)	0.0058 (1.12)	-0.0040 (-.72)	-0.0042 (-.82)	0.0187 (2.90)	0.0163 (2.81)	0.0155 (2.44)	0.0154 (2.61)
<i>sahex</i>	0.0398 (5.05)	0.0099 (1.46)	-0.0054 (-.75)	-0.0058 (-.88)	0.0479 (5.07)	0.0219 (2.87)	0.0072 (0.89)	0.0075 (1.00)
<i>sahgd</i>	0.0197 (2.87)	0.0007 (0.12)	-0.0078 (-1.37)	-0.0079 (-1.50)	0.0269 (3.38)	0.0147 (2.48)	0.0101 (1.64)	0.0101 (1.76)
<i>hlghq1</i>	-0.0003 (-.47)	-0.0009 (-1.87)	-0.0010 (-1.88)	-0.0010 (-2.08)	-0.0016 (-2.56)	-0.0008 (-1.75)	-0.0007 (-1.42)	-0.0007 (-1.40)
<i>age</i>	0.0584 (33.49)	0.0756 (36.88)	0.0639 (7.80)	0.0906 (35.26)	0.0544 (24.99)	0.0573 (23.80)	0.0458 (5.17)	0.0644 (22.49)
<i>agesq</i>	-0.0662 (-31.00)	-0.0848 (-33.86)	-0.1014 (-27.93)	-0.1004 (-32.49)	-0.0682 (-24.31)	-0.0699 (-22.73)	-0.0698 (-16.56)	-0.0768 (-21.44)
<i>exp</i>	0.0074 (6.59)	0.0057 (5.51)	0.0040 (3.45)	0.0047 (4.59)	0.0080 (4.55)	0.0038 (2.93)	0.0004 (0.26)	0.0020 (1.55)
<i>expsq</i>	-0.0155 (-4.05)	-0.0128 (-3.43)	-0.0093 (-2.15)	-0.0108 (-2.85)	-0.0204 (-2.59)	-0.0105 (-1.99)	-0.0038 (-.66)	-0.0065 (-1.27)
<i>widow</i>	0.0695 (1.79)	0.0674 (1.55)	0.0567 (1.03)	0.0699 (1.49)	-0.0356 (-1.54)	-0.0338 (-1.24)	0.0013 (0.04)	-0.0154 (-.55)
<i>divsep</i>	-0.0623 (-4.76)	-0.0469 (-3.58)	-0.0338 (-2.21)	-0.0395 (-2.95)	-0.0142 (-1.35)	-0.0048 (-.41)	0.0062 (0.44)	-0.0016 (-.14)
<i>nvrmar</i>	-0.1215 (-15.18)	-0.0786 (-8.60)	-0.0324 (-2.81)	-0.0446 (-4.49)	-0.0314 (-3.84)	-0.0486 (-4.97)	-0.0556 (-4.55)	-0.0478 (-4.61)
<i>kids04</i>	0.0292 (4.98)	0.0164 (3.21)	0.0143 (2.63)	0.0152 (3.10)	0.0452 (4.36)	0.0170 (2.01)	0.0077 (-0.85)	0.0129 (1.59)
<i>covnon</i>	0.0193 (2.46)	0.0345 (4.64)	0.0326 (4.00)	0.0320 (4.40)	0.0434 (4.58)	0.0299 (3.47)	0.0261 (2.73)	0.0324 (3.80)
<i>covmem</i>	0.1068 (15.12)	0.0993 (13.00)	0.0853 (9.41)	0.0876 (11.05)	0.1150 (13.09)	0.0950 (10.58)	0.0642 (6.10)	0.0824 (8.81)
<i>jobpriv</i>	0.0547 (7.31)	0.0230 (2.46)	0.0005 (.04)	0.0027 (0.23)	-0.0500 (-5.97)	-0.0686 (-7.47)	-0.0229 (-1.94)	-0.0346 (-3.23)
<i>prof</i>	0.5248 (44.64)	0.2521 (20.14)	0.0660 (4.55)	0.0703 (5.29)	0.6035 (28.39)	0.2807 (14.88)	0.0379 (1.75)	0.0624 (3.11)
<i>manag</i>	0.4566 (54.71)	0.2396 (26.81)	0.0755 (6.99)	0.0802 (8.10)	0.4762 (46.49)	0.2499 (23.36)	0.0583 (4.39)	0.0770 (6.26)
<i>skllnm</i>	0.2281 (24.63)	0.1182 (12.01)	0.0111 (0.96)	0.0146 (1.37)	0.2676 (28.66)	0.1440 (13.81)	0.0050 (0.38)	0.0197 (1.62)
<i>skllm</i>	0.1122 (15.16)	0.0570 (7.68)	0.0189 (2.28)	0.0209 (2.75)	0.1057 (8.45)	0.0691 (5.90)	0.0356 (2.76)	0.0413 (3.43)
<i>white</i>	0.0646 (3.78)	0.0589 (2.16)	<i>dropped</i>	0.2536 (4.38)	0.0809 (4.76)	0.0683 (2.25)	<i>dropped</i>	0.1431 (2.93)
<i>deg</i>	0.1676 (19.06)	0.2604 (18.93)	<i>dropped</i>	0.8561 (7.45)	0.2018 (20.50)	0.2862 (18.90)	<i>dropped</i>	0.5362 (6.01)

t-statistics/z-statistics in parentheses. Significance at 5% level in bold.
robust OLS estimation

REFERENCES

- Adams P, Hurd M, McFadden D, Merrill A, Ribeiro T (2003) Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics*, 112: 3-56
- Amemiya T, MaCurdy T (1986) Instrumental variable estimation of an error-components model. *Econometrica*, 54: 869-880
- Baltagi B, Khanti-Atom S (1990) On efficient estimation with panel data: An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics*, 5: 401-406.
- Bartel A, Taubman P (1979) Health and labour market success: The role of various diseases. *The Review of Economics and Statistics*, 61: 1-8
- Berkowitz M, Fenn P, Labrinos J (1983) The optimal stock of health with endogenous wages. *Journal of Health Economics*, 2: 139-147
- Breusch T, Mizon GE, Schmidt P (1989) Efficient estimation using panel data. *Econometrica*, 57: 695-700
- Currie J, Madrian BC (1999) Health, health insurance and the labor market in *Handbook of Labor Economics*, Volume 3, Ashenfelter and Card, eds. Elsevier Science B.V.
- Contoyannis P, Rice N (2001) The impact of health on wages: Evidence from the British Household Panel Survey. *Empirical Economics*, 26: 599-622.
- Disney R, Gosling A (1998) Does it pay to work in the public sector? *Fiscal Studies*, 19: 347-374
- Dustmann C, van Soest A (1997) Wage structures in the private and public sectors in West Germany. *Fiscal Studies*, 18: 225-247
- Ettner S. (1996) New evidence on the relationship between income and health. *Journal of Health Economics*, 15: 67-85
- Freeman R (1984) Longitudinal analyses of the effects of trade unions. *Journal of Labor Economics*, 2: 1-26
- Green F, Machin S, Manning A (1996) The employer size-wage effect: Can dynamic monopsony provide an explanation? *Oxford Economic Papers, New Series*, 48: 433-455
- Griffin J, Fuhrer R, Stansfeld S, Marmot M (2002) The importance of low control at work and home on depression and anxiety: do these effects vary by gender and social class? *Social Science and Medicine*, 54: 783-798
- Grossman M (1972) On the concept of health capital and the demand for health. *Journal of Political Economy*, 80: 223-255

- Harkness S (1996) The gender earnings gap: Evidence from the UK. *Fiscal Studies*, 17: 1-36
- Hausman J, Taylor W (1981) Panel data and unobservable individual effects. *Econometrica*, 46: 1251-1271
- Haveman R, Wolfe B, Kreider B, Stone M (1994) market work, wages, and men's health. *Journal of Health Economics*, 13: 163-182
- Hayes B, Prior P (2003) *Gender and Health Care in the UK: Exploring the Stereotypes*. Palgrave, Basingstoke
- Heckman J, Lochner L, Todd P (2003) Fifty years of Mincer earnings regressions. NBER Working Paper no. 9732
- Hildreth A (1999) What has happened to the union wage differential in Britain in the 1990s? *Oxford Bulletin of Economics and Statistics*, 61: 5-31
- Jakubson G (1991) Estimation and testing of the union wage effect using panel data. *The Review of Economic Studies*, 58: 971-991
- Kim M, Polachek S (1993) Panel estimates of male-female earnings functions. *Journal of Human Resources*, 29: 406-428
- Lee, L (1982) Health and wages: A simultaneous equation model with multiple discrete indicators. *International Economic Review*, 23: 199-221
- Luft H (1975) The impact of poor health on earnings. *Review of Economics and Statistics*, 57: 43-57
- Macpherson D, Hirsch B (1995) Wages and gender composition: Why do women's jobs pay less? *Journal of Labor Economics*, 13: 426-471
- Mincer J (1958) Investment in human capital and personal income distribution. *Journal of Political Economy*, 66: 281-302
- Mincer J (1974) *Schooling, Experience and Earnings*, New York: NBER Press
- Mushkin S (1962) Health as an investment. *Journal of Political Economy*, 70(5), Part 2: Investment in Human Beings, 129-157
- Oaxaca R (1973) Male-female wage differentials in urban labour markets. *International Economic Review*, 14: 529-536
- Oaxaca R, Ransom M (1994) On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61: 5-21