Is there a wage premium for dangerous jobs?

Franco Insalaco[1]

Economic theory asserts that jobs with unpleasant characteristics offer higher wages as compensation. This suggests that jobs that put workers at a greater risk of being physically hurt should offer higher wages once other characteristics have been controlled for. This hypothesis is tested using data from Wave 1 from the British Household Panel Survey. The results run counter to economic theory.

Software used: STATA 5.0, SPSS 8.0, Excel 97, Word 97

Word count: 4, 991 (excluding cover page, references and appendix)

[1] The British Household Panel Survey data used in this paper was made available through the Data Archive. The data were originally collected by the ERSC Research Centre in Micro-Social Change at the University of Essex. Material from the Labour Force Survey is Crown Copyright; has been made available by the Office for National Statistics through The Data Archive and has been used by permission. Niether the ONS nor The Data Archive bear any responsibility for the analysis or interpretation of the data reported here.

Is there a wage premium for dangerous jobs?

Introduction

Background

1. This paper examines whether the level of physical danger in the job affects wages using data from the British Household Panel Survey. Economic theory suggests that, other things being equal, a higher level of danger will result in higher wages. This study uses accidents at work as a proxy for measuring danger. Studies that have measured danger in this way have come to conflicting conclusions. A US study found that there was a wage premium for danger while a British study found that there was not.

Theory

2. The theory of compensating differentials asserts that, other things being equal, jobs with unpleasant characteristics will offer higher wages. Failure to do so will cause recruitment difficulties in these relatively unpleasant jobs. This implies that jobs that put workers at greater risk of physical injury will offer a wage premium to compensate for the danger. Therefore a variable that measures the danger in a job should have a positive and statistically significant coefficient if it is entered as a regressor in a wages equation.

Previous research

3. Studies that have attempted to measure the premium for risk have typically used the number of *fatal* work-related accidents as a measure of risk. The author is aware of only two studies that have used non-fatal accidents as a risk measure. Both of these papers included both fatal and non-fatal risk proxies as regressors in the same wage equation. Garen (1988) uses US data and finds that both variables have positive and significant effects. Arabsheibani & Marin (1996) use British data taken from the General Household Survey (GHS) and find that while fatal risk has a positive and significant effect, non-fatal risk tends to have a negative and insignificant effect on wages.

Objective of current study

4. The current study examines whether there is a wage premium for *non-fatal* risk in the British labour market. The methodology used closely follows Arabsheibani & Marin (1996) but uses different sources for both the estimation sample and the measurement of non-fatal risk.

Data

Sample

5. The sample used to estimate the wages equation comes from Wave 1 (1991) of the British Household Panel Survey (BHPS). The main reason why the BHPS is chosen is because, unlike the Labour Force Survey, it has a Goldthorpe-Hope Index (GH) score for each respondent. This is important, as GH is one of the regressors used in Arabsheibani & Marin (1996).

6. The measure of risk in each occupation is calculated from information given in the Labour Force Survey (LFS). In every winter quarter since 1993 respondents in the LFS are asked whether they have had a work-related accident in the past 12 months. As there are 150,000 respondents in each LFS and five years of data are available (Winter 1993 to

Winter 1997) a maximum of 750,000 respondents would have been in a position to give information on work-related accidents.[2]

7. Arabsheibani & Marin (1996) obtained information on risk from a smaller pool of respondents. The information came from three years of the GHS (1987-1989). There are approximately 20,000 respondents in each GHS therefore only a maximum of 60,000 respondents would have been in a position to give information on work accidents. The larger pool of information in the current study should provide more precise estimates of the level of jobrisk than was available in the earlier study.

8. As is common in this type of study, only male employees are present in the estimation sample. There are two reasons for this:

- Women are excluded because of difficulties in ascertaining the occupation of women involved in fatal job accidents.[3] Although this problem is not relevant to the current study, women are excluded in order to be comparable to Arabsheibani & Marin (1996).
- The self-employed are excluded, as it is reasonable to assume that they have a different attitude to risk compared with employees. We would expect the self-employed to be less risk-averse as they have chosen to incur greater risk through owning their own business.

Earnings (LOGREAL)

9. Wages is measured as the real gross wage per hour. Both wages and hours refer to quantities that are normally done. Gross wages are used rather than net wages as this is the variable that employees negotiate with their bosses over. The wages have been put into real wages using September 1991 as the base. This was done using the retail price index.[4]

10. When fitting wage equations, the log of wages is entered as the dependent variable. This is because taking the log tends to fit the data better than using the unadjusted figures.

Measuring the risk of having a work-related accident

11. Using non-fatal accidents to proxy for jobrisk has advantages and disadvantages compared with using fatal accidents as the proxy. Its major advantage is that it is easier to collect information on. Information is collected every year in the Labour Force Survey. In contrast information on fatal accidents are only published once a decade by the OPCS (now the ONS).

12. The major disadvantage of using a proxy based on non-fatal accidents is that they are a less precise measure than fatal accidents. The information only relates to the number of accidents. It does not indicate how serious each accident is. In contrast each fatal accident is broadly similar (i.e. each accident ends in death for the unfortunate worker).

13. The formula to measure job risk in occupation j is:

$$JOBRISK_j = (1000/5)*\Sigma_i ACC_{ij}/\Sigma_i EMP_{ij} \qquad i=1993,\ldots, 1997$$

[2] Not everyone would have been asked the question as not all respondents were in work.

[3] See Marin & Psacharopoulos (1982)

[4] Although the current fashion is to examine the retail price index excluding mortgage payments when measuring inflation, this is to exclude the effect of monetary policy on the figures. For a worker changes in the level of their mortgage payments does have an effect on their standard of living and so should be included when calculating real wages.

where ACC_{ij}= Number of work-related accidents in the past 12 months that occurred to male employees in occupation j who had been in their current job for at least 12 months

EMP_{ij}= Number of male employees who worked in occupation j and had been in their current job for at least 12 months

14. The formula was used for the following reasons: firstly respondents were asked about accidents that had occurred in the past 12 months. If a person who had an accident has been in their job for less than a year it is possible that the accident occurred in a different occupation to the one they are currently in. In order to eliminate the risk of this occurring, only information on people who have been in their current job for at least a year are included in the calculation of the measure.

15. The figure is divided by 5 in order to get an annual figure. Jobrisk is a flow rather than a stock variable. The amount of risk in a job is proportional to the amount of time spent there. Therefore the variable is rescaled to refer to one year. As the variable is calculated from five years of data, this means that the formula has to be divided by 5.

16. As the proportion of people in each occupation who have suffered an accident is very small, implying that the coefficient of this variable will also be very small, the formula is multiplied by 1,000 in order to get reasonably large coefficients (e.g. getting a coefficient of 0.2 rather than 0.0002).

17. The final issue to consider concerning the construction of the jobrisk variable is the choice of occupational classification to use. Occupations are measured using the 1990 Standard Occupational Classification (SOC). This classification is available at three levels of detail. The three-digit level is the most detailed and defines 374 occupations. The two-digit level groups the occupations into 74 categories. The less detailed classification is the one-digit level that only has nine categories.

18. The choice as to which classification level to use is important. If the classification is too broad then there are very few categories and a lot of heterogeneous jobs are grouped together. This can result in the jobrisk variable having an insignificant coefficient when in fact there is a significant effect.

19. If the classification used is too detailed then there will be very few people in each occupation. This can also lead to misleading estimates as sampling variability can have a large influence.

Years of schooling (S)

20. The human capital model[5] on which wage equation regressions are based requires the inclusion of a variable that records the years of schooling the respondent has undergone. Schooling is used as a proxy for the respondent's productivity. The greater the years of schooling undertaken the greater the productivity of the respondent.

21. Specifying the effect of education can be misleading, as it does not take account of what the schooling consists of. For example an individual who has undergone 17 years of schooling and gained a Master's degree is assumed to be less productive than an

---

[5] See Berndt (1991) Chapter 5 for more details.

individual who has undergone 18 years of schooling but whose highest qualifications are A-levels. Despite this disadvantage, years of schooling will be used in the wages regression as this is the conventional way of taking education into account.

22. S (the 'years of schooling' variable) is not given in the dataset. It has to be constructed from three variables: ASCEND (age respondent left school), AFEEND (age respondent finished further education) and AQFEDHI (respondent's highest qualification). Subtracting five from the age at which the respondent finished education derives S. The age at which the respondent finished their education is usually taken to be AFEEND or (if AFEEND is missing), ASCEND. However a check is made with the highest qualification that the respondent has gained. This is to screen out those cases where the respondent has completed a period in work before returning to education. For example if an individual's highest qualification is A-level but the age at which they finished further education is 28 it would be a mistake to impute that the person had 23 years of schooling. Rather the S score for this individual should reflect the qualification that they have gained. Failure to do could lead to a misleading impression of the individual's productivity.

Experience (EXP and EXPSQ)

23. The human capital model asserts that the respondent's productivity is also affected by work experience. It is the convention to measure years of experience as

$$EXP = AGE - S - 5$$

24. This assumes that the respondent has never spent any time out of work. As we are restricting our attention to males this is a reasonable assumption to make. Anecdotal evidence by labour market researchers suggest that almost all males who are currently in employment have never been out of work or been out for a small period of time. This assumption would have been more difficult to maintain if women were included in our sample as women spend significant amounts of time out of the labour market.

25. Even allowing for this, EXP suffers from a measurement problem in that it does not distinguish between formal training and on-the-job training. One would expect formal training to have a greater effect on the respondent's productivity. However data limitations make it difficult to construct a variable that takes account of the formal training that the respondent has undertaken through their working life.

26. Work experience has two effects on productivity (and wages):

- The accumulation of experience increases the individual's productivity.
- As the individual gets older, their productivity diminishes (for example because it takes longer for them to perform the same task).

27. These two effects are allowed for by including a squared experience term (EXPSQ). In regressions the squared term is expected to have a negative coefficient. This means that at low levels, an increase in experience has a positive effect on wages but at high levels the marginal effect of experience is negative.

Union membership (UNION)

28. This variable indicates whether the respondent is a member of their workplace union. It is a (0,1) variable that takes the value 1 if the respondent is a union member. We would

expect this variable to have a positive effect on wages as unions increase the bargaining power of workers in securing higher wages.

The Goldthorpe-Hope Index (GH)

29. This variable is a cardinal scale that measures the 'desirability' of an occupation. The higher the score the more desirable the occupation.[6] Economic theory would suggest that this variable should have a negative effect on wages. However the two studies involving Alan Marin (Marin & Psacharopoulos (1982) and Arabsheibani & Marin (1996)) have found that GH has a positive effect on wages. As a result the variable has been interpreted as a proxy for abilities that segment the labour force.

The statistical model

30. When determining the effect of a variable on wages it is conventional to use a reduced form equation with both supply-side and demand-side variables included as regressors. This is because researchers have found it difficult to estimate a structural model due to the problems in finding identifying restrictions and choosing an appropriate utility function.

31. The fact that a reduced form equation is estimated suggests that a large number of regressors could be used. However researchers typically estimate 'stripped down' models where the independent variables consist of solely S, EXP, EXPSQ and the variables that are of direct interest to the researcher. The rationale for this is that variables from cross-section data are likely to be orthogonal. Therefore omitting significant variables will not bias the results. This is the approach used in Arabsheibani & Marin (1996).

32. The equation that will be estimated will be the same specification used in Arabsheibani & Marin (1996):

$$LOGREAL_i = \alpha_i + \beta_1 * S_i + \beta_2 * EXP_i + \beta_3 * EXPSQ_i + \beta_4 * UNION_i + \beta_5 * GH_i + \beta_6 * JOBRISK_i \quad (1)$$

33. The only differences between the two specifications are that firstly, Arabsheibani & Marin (1996) used monthly wages as the dependent variable and made no allowances on hours worked. This was because the GHS did not have an appropriate hours worked variable; there was no information on the number of overtime hours worked. Secondly the measure of risk is slightly different between the two specifications. Further information on the differences in the risk measure is given later when the estimates of the two equations are compared.

Results

The specification of the dependent variable

34. Some initial regressions suggested that there were outliers in the wages data. In order to get rid of these outliers all individuals with a recorded wage rate of less than £1 and greater than £35 per hour were removed from the sample. This eliminates approximately 25 observations from the data leaving almost 2,500 in the sample.

35. The first step in the construction of the model is choosing the appropriate dependent variable. The literature suggests that wages equations should be estimated using the log

[6] See Goldthorpe and Hope (1974) for more details on how the index was constructed.

of wages as the dependent variable. It is asserted that this provides a better fit. Table 1 verifies whether this is true.

36. In Table 1 the R^2, Akaike Information Criterion (AIC) and the Schwartz Bayesian Information Criterion (SBC) are reported for two sets of regressions derived from (1). The first set of regressions use REALWAGE as the dependent variable. REALWAGE is the real hourly wage rate in levels. The second set of equations uses LOGREAL (the log of REALWAGE) as the dependent variable. Each set consists of three regressions: one for each measure of risk. JOBRISK (3 digit) refers to occupational risk recorded at the three-digit level of occupations. JOBRISK (2 digit) and JOBRISK (1 digit) refer to risk at the two-digit and one-digit levels of occupation respectively.

37. Table 1 indicates that using LOGREAL as the dependent variable substantially improves the fit of the model. The LOGREAL equations have AIC and SBC values of approximately –3,900 and R^2s of around 0.45. In contrast the REALWAGE equations have substantially lower AIC and SBC values of approximately –6,100. The corresponding R^2s are also substantially lower having values of approximately 0.39. Hence the data suggests that the log of the wage rate is the appropriate dependent variable for the model.

Table 1: How the choice of dependent variable affects the fit of the wage equation

	Model A	Model B	Model C	Model D	Model E	Model F
Realwage				√	√	√
Logreal	√	√	√			
Jobrisk(3digit)	√			√		
Jobrisk(2digit)		√			√	
Jobrisk(1 digit)			√			√
R^2	0.4545	0.4549	0.4531	0.3919	0.3933	0.3903
AIC	-3865.06	-3881.58	-3883.31	-6105.97	-6130.88	-6133.47
SBC	-3869.9	-3886.43	-3888.16	-6110.82	-6135.73	-6138.33

*= Each regression includes constant, S, EXP, EXPSQ, TU and GH as regressors. √ indicates whether variable is included in regression.

Comparison of the wage equation estimates

38. Table 2 compares the estimates of the following wage equations:

- Model 1: The estimates from Arabsheibani & Marin (1996). As stated earlier, this study was carried out on the 1985 GHS. It should be noted that the measure of risk used is slightly different to those of the current study.
    - Firstly, the occupational classification used was the three-digit measure of the 1980 classification (the current study uses the 1-digit, 2-digit and 3-digit measures of the1990 classification).
    - Secondly, the data on accidents in each occupation comes from the 1987, 1988 and 1989 GHS (these were the only times that a question on work-related accidents was asked in the GHS); the data for the current study comes from the Winter LFSs of 1993, 1994, 1995, 1996 and 1997.
    - Finally, Arabsheibani & Marin (1996) measured risk in a way that took into account the age structure of employees in each occupation. Specifically "*for each occupation group the measure is the observed number in that group minus the number expected given the age structure of workers in the group, divided by the number of employees*

in the group." The age structure of employees is not taken into account in the current study.

Although these changes will affect the size of the coefficient they should affect neither the sign nor the statistical significance (at the 5% level) of the risk variable.

- Model 2: The equation from the current study that measures risk using the three-digit level classification of occupations.

- Model 3: The equation from the current study that measures risk using the two-digit level classification of occupations.

- Model 4: The equation from the current study that measures risk using the one-digit level classification of occupations.

Table 2: Estimates of the wage equation

	Model 1	Model 2	Model 3	Model 4
	Coefficients (t-ratios in brackets)			
S	0.0415 (11.823)	0.0408 (10.561)	0.0406 (10.539)	0.0412 (10.649)
Exp	0.0372 (18.643)	0.0456 (21.571)	0.0457 (21.721)	0.0451 (21.516)
Expsq	-0.0006 (-15.516)	-0.0008 (-18.175)	-0.0008 (-18.315)	-0.0008 (-18.115)
GH	0.0106 (19.906)	0.0117 (17.921)	0.0111 (15.929)	0.0113 (15.044)
TU	0.1159 (4.325)	0.1467 (8.777)	0.1456 (8.773)	0.1440 (8.637)
Jobrisk (3 digit)		-0.0034 (-3.001)		
Jobrisk (2 digit)			-0.0055 (-3.6)	
Jobrisk (1 digit)				-0.0046 (-2.237)
Jobrisk (3 digit; 1980)	-0.7E-06 (-0.366)			
Constant	3.6748 (75.014)	0.3802 (6.715)	0.4353 (7.027)	0.4094 (5.755)
	Misspecification tests (7) (P-values)			
Normality of residuals		.00001	.00001	.00001
Heteroskedasticity		0	0	0
Functional Form		0	0	0

[7] The normality of the residuals is tested using the Shapiro-Francis test for normality. Heteroskedasticity is tested by estimating an auxiliary regression of the squared residuals on all the regressors in the equation and testing whether all of the coefficients are zero (Cook-Weisburg test); and functional form is tested using Ramsey's RESET test where the square, cube and 4th power of the fitted values are included as regressors in the auxiliary equation. Further details of these tests are given in STATA (1997) under the entries for 'swilk' and 'fit'.

Table 2: Estimates of the wage equation

	Model 1	Model 2	Model 3	Model 4
		Measures of fit		
R^2	.313	0.4545	0.4549	0.4531
AIC		-3865.06	-3881.58	-3883.31
SBC		-3869.9	-3886.43	-3888.16
No. of observations	3601	2424	2434	2434

39. Looking at Table 2 we can see that all four models have very similar coefficients and t-ratios for the variables s, exp and expsq. In addition the sign of these coefficients correspond with economic theory: more years of schooling increases an employee's wage rate while the effect of an additional year of experience in the labour market will depend on the individual's accumulated experience. If the accumulated experience is high then additional experience can reduce their wage rate. The opposite is true for someone with relatively little experience.

40. The coefficients and t-ratios of GH (the Goldthorpe-Hope variable) are also similar in the four models. However the sign of the coefficient is the opposite of what one would expect. The higher the value of GH the more desirable is the occupation, therefore one would expect, other things being equal, that being in a more desirable occupation would be associated with a fall in the wage rate. However GH has a positive coefficient. A possible explanation for this could be sample selection bias. If the desirability of an occupation is taken into account when an individual chooses their job then there is a non-random sample of people in each occupation. This can lead to biased coefficients when estimating using OLS. Unfortunately trying to estimate a wages equation that takes into account sample selection bias is beyond the scope of this study.

41. The coefficient of TU (whether the individual is a member of a trade union) is also similar between the four models although the three models of the current study have t-ratios for this variable that are almost double the corresponding t-ratio from Model 1. As one would expect trade union membership is associated with higher wage rates.

42. Turning our attention to the risk variable we see that all four models record a very small negative coefficient. All 3 wages equations estimated on the BHPS (Models 2-4) report the coefficient of risk to be significantly different from zero. There are three possible reasons for this counter-intuitive result:

- *Sample selection bias*: It may be that risk-loving individuals are attracted into dangerous jobs. As they are risk-loving they are willing to pay a premium in order to be employed in a dangerous job. This explanation is highly unlikely to be correct. This is because studies that have measured risk by the incidence of *fatal* accidents have found that dangerous jobs do offer a positive wage premium[8]. While these studies have found that sample selection bias does occur, the nature of the bias is that OLS estimates coefficients are biased downwards but are still positive and significant.
- *Measurement error*: It could be the case that the way risk is measured is too blunt. The JOBRISK variable only records the proportion of people in an occupation who have had an accident. It does not record the severity of the accident. It may be the case that if risk was recorded in a way that took into account the severity of accidents (for example the

[8] For example Garen (1988), Sandy & Elliot (1996) and Arabsheibani & Marin (1996).

proportion of working days lost in an occupation through work-related accidents) then it would record a positive and significant coefficient. However this does not explain the tendency for the jobrisk variable to be negative (both studies) and significant (this study).

- *Lack of choice for employees*: It may be the case that occupations that have a high proportion of non-fatal accidents tend to be 'poor' jobs that have bad working conditions and pay low wages. The nature of the demand and supply for labour in these jobs may be such that workers do not receive a premium for working in a dangerous environment. However this argument is undermined by the fact that when risk is measured by the incidence of fatal accidents, there is a positive wage premium.

43. Arabsheibani & Marin (1996) did not report any fit or misspecification statistics for their model with the exception of R^2. Therefore our comparison is limited to models 2-4. The fit of the three models is almost identical: although model 2 (which measures risk using the 3-digit classification of occupations) has a slightly better fit in terms of both the AIC and SBC. All three models suffer from non-normal errors, heteroskedasticity and a misspecified functional form.

44. An examination of the residuals of the models reveals that the non-normality is not caused by outliers but is due to thick tails. For example the residuals of model 2 has a kurtosis score of 4.8 compared with the value of 3 that a normally distributed variable would have. This means that little can be done to make the residuals normal. However non-normal errors do not have serious implications for our model as, because of the Central Limit Theorem, all series have an asymptotically normal distribution.

45. The problems of heteroskedasticity and functional form are more serious. A potential reason why the model failed both tests is that its functional form is misspecified which is also causing the model to have heteroskedastic errors. Graphing the residuals of the models against the regressors does not reveal a clear picture as to what is causing the misspecification. This may be partly due to the fact that 2,500 observations are used in the model.

46. A more analytical method of finding out the cause of the misspecification is to include all cross-product and squares of the regressors as additional variables in the regression. Table 3 shows the coefficients of selected variables from this regression together with the P-values for the Heteroskedasticity and RESET tests. All the other variables in the regression were either insignificant or had coefficients that were extremely close to zero.

Table 3: Selected results from the equations with cross-products and squares

	Model 2	Model 3	Model 4
	Coefficients (t-ratios in brackets)		
S	.1199 (3.045)	.1241 (3.016)	.1245 (2.801)
EXP	.1585 (9.754)	.1543 (8.841)	.1589 (8.038)
TU	.4566 (3.712)	.3614 (2.749)	.3696 (2.370)
JOBRISK (3 digit)	-.0039 (-.464)		
JOBRISK (2 digit)		-.0040 (-.283)	

Table 3: Selected results from the equations with cross-products and squares			
	Model 2	Model 3	Model 4
JOBRISK (1 digit)			.01647 (.611)
Misspecification tests (P-values)			
Heteroscedasticity	0	0	0
Functional Form	.0180	.0145	.001

47. As Table 3 shows the expanded regressions have the effect of making the risk variables insignificant (although interestingly the risk coefficient for model 4 becomes positive). In addition the effect of trade union membership almost trebles to an implausibly high value. Finally the expanded regressions still fail the heteroscedasticity and functional form tests. In summary the expanded regressions still fail the misspecification tests and have the side effect of greatly altering the values of some of the original regressors. While it is possible to expand the regressions still further by including the variables that make up the cube of the fitted values this would have the disadvantage that many of the variables would become difficult to interpret.

48. Leaving the functional form problem unresolved, the next method that can be used to improve the model is to correct the regressions for heteroscedasticity. This is done using Iterated GLS. This technique works by performing a regression, calculating case weights based on the absolute residuals and regressing again using those weights. Iterations stop when the change in weights falls below a certain value[9]. The results are given in Table 4.

Table 4: 'Robust regression' estimates of the wages equation (10)			
	Model 2	Model 3	Model 4
Coefficients (t-ratios in brackets)			
S	0.0415 (11.532)	0.0412 (11.485)	0.0420 (11.673)
EXP	0.0438 (22.249)	0.0440 (22.435)	0.0434 (22.218)
EXPSQ	-0.0008 (-18.966)	-0.0008 (-19.152)	-0.0008 (-18.921)
GH	0.0121 (19.925)	0.0117 (17.991)	0.0121 (17.219)
TU	0.1375 (8.824)	0.1374 (8.881)	0.1350 (8.703)
JOBRISK (3 digit)	-0.0023 (-2.15)		
JOBRISK (2 digit)		-0.0038 (-2.624)	
JOBRISK (1 digit)			-0.0021 (-1.104)
Constant	0.3740 (7.088)	0.4143 (7.174)	0.3705 (5.594)

[9] The technique is described in more detail under the 'rreg' entry in STATA (1997).
[10] Please note that no fit measures are produced for the robust regression as the residual sum of squares are not calculated for this model as they are considered to be inappropriate.

49. Table 4 shows that correcting for heteroscedasticity has little effect on the estimates. The coefficients and t-ratios are very similar to the corresponding OLS estimates in Table 2. The one exception is the jobrisk variable where, for each model, there is a noticeable fall in the t-ratio. However in models 2 and 3, the jobrisk variable is still significant. The jobrisk variable in model 4 becomes insignificant. Given that model 4 uses a very broad measure of occupational risk (occupations are grouped into only nine categories), this suggests that using the 1-digit classification of occupations is inadequate as each defined category of occupations includes heterogeneous jobs.

Conclusions

50. Studies have shown that there is a statistically significant premium for working in dangerous jobs if the level of danger is measured by the incidence of fatal accidents at work. This study has investigated whether the same story holds when the level of danger in a job is measured by the incidence of non-fatal accidents.

51. In line with Arabsheibani & Marin (1996) the results suggest that danger has a negative effect on wages. This result is counter-intuitive and is probably due to the way that the incidence of non-fatal accidents is measured in an imprecise way with no allowance made for the severity of accidents. While it is likely that the estimates may have been affected by sample selection bias this does not explain the negative coefficient of jobrisk as measuring danger by the incidence of fatal accidents produces a positive coefficient even when sample selection bias is ignored.

52. The study has also raised questions about the way wages equations are estimated. Despite using a standard specification drawn from the human capital model, the resulting model suffers from non-normal errors, heteroscedasticity and a misspecified functional form. This suggests that more work should be done in future to develop a model that does not suffer from these problems.

References

Arabsheibani, G.R. and A. Marin (1996) Stability of estimates for the compensation of danger, mimeo

Berndt ,E.R. (1991) The Practice of Econometrics, Addison Wesley

Garen, J. (1988) Compensating wage differentials and the endogeneity of job riskiness, Review of Economics and Statistics 70

Goldthorpe, J.H. and K. Hope (1974) The social grading of occupations: A new approach and scale, Oxford: Clarendon Press

Marin, A and G. Psacharopoulos (1982) The reward for risk in the labour market: Evidence from the United Kingdom and a reconciliation with other studies, Journal of Political Economy 90(4)

Sandy, R and R.F. Elliot (1996) Unions and risk: their impact on the compensation for fatal risk, Economica 63

STATA Corporation (1997) STATA Release 5 Reference manual, STATA Press

APPENDIX: List of variables in STATA data file

AHID household identification number

APNO person number

ADOIM date of interview: month

ASEX sex

ASCEND school leaving age

AFEEND further education leaving age

AJBSOC occupation (soc): current main job (3-digit classification)

AJBSEMP employee or self-employed

AJBHRS no. of hours normally worked per week

ATUIN1 member of workplace union (called TU in paper)

AAGE age at date of interview

AQFEDHI highest educational qualification

APAYGU usual gross pay per month

AJBHGS hope - goldthorpe scale: present job (called GH in paper)

JOB100 Variable used to help create AJBMAJM

JOB10 Variable used to help create AJBMINM

AJBMAJM Occupation in main job (Single digit SOC)

AJBMINM Occupation in main job (Two digit SOC)

NOMWAGE Nominal hourly wage

REALWAGE Real hourly wage (at Sep 91 prices)

RISKMAIN JOBRISK at 3-digit level of classification (called JOBRISK (3 digit) in paper)

RISKMINM JOBRISK at 2-digit level of classification (called JOBRISK (2 digit) in paper)

RISKMAJM JOBRISK at 1-digit level of classification (called JOBRISK (1 digit) in paper)

SCHLE School leaving age

S years of schooling

EXP years of work experience

EXPSQ years of work experience squared

LOGREAL Log of real hourly wage

Below are the higher power and cross-product variables used in the expanded regression

ssq=s*s

sexp=s*exp

sexpsq=s*expsq

stu=s* atuin1

sminm=s* riskminm

smajm=s* riskmajm

smain=s* riskmain

sgh=s* ajbhgs

exp3=exp*expsq

exptu=exp* atuin1

expminm=exp* riskminm

expmajm=exp* riskmajm

expmain=exp* riskmain

expgh=exp* ajbhgs

exp4=expsq* expsq

expsqtu=expsq* atuin1

exp2minm=expsq* riskminm

exp2majm=expsq* riskmajm

exp2main=expsq* riskmain

expsqgh=expsq* ajbhgs

```
tuminm= atuin1* riskminm

tumajm= atuin1* riskmajm

tumain= atuin1* riskmain

tugh= atuin1*  ajbhgs

minmsq=  riskminm* riskminm

majmsq=  riskmajm* riskmajm

mainsq=  riskmain* riskmain

minmgh=  riskminm*  ajbhgs

majmgh=  riskmajm*  ajbhgs

maingh=  riskmain*  ajbhgs

ghsq=   ajbhgs*  ajbhgs
```

SN: 4189 - SPSS SYNTAX FILES

The following six SPSS syntax files used to create derived variables for the STATA data file were supplied by the depositor. The data were originally created in SPSS format (not supplied to the Data Archive). The syntax files were run on said SPSS data file, and the resulting file was converted to STATA format.

A version of the STATA file in SPSS format was created by the Data Archive and is available to users.

1. accident.sps

This file was run on data from Winter quarters of the 1993-97 Labour Force Surveys (LFS). It was used to collect the information on the incidence of non-fatal accidents by occupation.

```
USE ALL.
COMPUTE filter_$=(inecaca >= 1 & inecaca <= 4 & sex = 1 & empmon >= 12
&
  uresmc < 20).
VARIABLE LABEL filter_$ 'inecaca >= 1 & inecaca <= 4 & sex = 1 & empmon
>= 12'+
 ' & uresmc < 20 (FILTER)'.
VALUE LABELS filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
CROSSTABS
  /TABLES=socmain  BY accdnt
  /FORMAT= AVALUE TABLES
  /CELLS= COUNT .
CROSSTABS
  /TABLES=socmajm  BY accdnt
  /FORMAT= AVALUE TABLES
  /CELLS= COUNT .
CROSSTABS
  /TABLES=socminm  BY accdnt
  /FORMAT= AVALUE TABLES
  /CELLS= COUNT .
FREQUENCIES
  VARIABLES=socmain socmajm socminm
  /ORDER  ANALYSIS .
FILTER OFF.
USE ALL.
EXECUTE .
USE ALL.
COMPUTE filter_$=(inecaca >= 1 & inecaca <= 4 & sex = 1 & uresmc < 20).
VARIABLE LABEL filter_$ 'inecaca >= 1 & inecaca <= 4 & sex = 1 & uresmc
< 20'+
 ' (FILTER)'.
VALUE LABELS filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
```

```
EXECUTE .
FREQUENCIES
  VARIABLES=socmajm wchjb secjob
  /ORDER  ANALYSIS .
CROSSTABS
  /TABLES=socmajm  BY secjob
  /FORMAT= AVALUE TABLES
  /CELLS= COUNT .
```

2. riskmain.sps

This file was used to allocate the 3-digit jobrisk values (calculated from the LFS data) to the BHPS data.

```
IF (ajbsoc = 100) riskmain = 2.30 .
IF (ajbsoc = 101) riskmain = 1.74 .
IF (ajbsoc = 102) riskmain = 2.87 .
IF (ajbsoc = 103) riskmain = 3.11 .
IF (ajbsoc = 110) riskmain = 5.97 .
IF (ajbsoc = 111) riskmain = 7.86 .
IF (ajbsoc = 112) riskmain = 3.03 .
IF (ajbsoc = 113) riskmain = 5.17 .
IF (ajbsoc = 120) riskmain = 2.16 .
IF (ajbsoc = 121) riskmain = 3.10 .
IF (ajbsoc = 122) riskmain = 3.02 .
IF (ajbsoc = 123) riskmain = 2.96 .
IF (ajbsoc = 124) riskmain = 4.36 .
IF (ajbsoc = 125) riskmain = 5.00 .
IF (ajbsoc = 126) riskmain = 2.32 .
IF (ajbsoc = 127) riskmain = 1.23 .
IF (ajbsoc = 130) riskmain = 1.53 .
IF (ajbsoc = 131) riskmain = 2.21 .
IF (ajbsoc = 132) riskmain = 2.10 .
IF (ajbsoc = 139) riskmain = 2.40 .
IF (ajbsoc = 140) riskmain = 6.54 .
IF (ajbsoc = 141) riskmain = 5.02 .
IF (ajbsoc = 142) riskmain = 8.76 .
IF (ajbsoc = 150) riskmain = 7.09 .
IF (ajbsoc = 151) riskmain = 18.18 .
IF (ajbsoc = 152) riskmain = 7.23 .
IF (ajbsoc = 153) riskmain = 25.00 .
IF (ajbsoc = 154) riskmain = 0.00 .
IF (ajbsoc = 155) riskmain = 0.00 .
IF (ajbsoc = 160) riskmain = 12.16 .
IF (ajbsoc = 169) riskmain = 9.72 .
IF (ajbsoc = 170) riskmain = 4.34 .
IF (ajbsoc = 171) riskmain = 7.59 .
IF (ajbsoc = 172) riskmain = 2.35 .
IF (ajbsoc = 173) riskmain = 8.14 .
IF (ajbsoc = 174) riskmain = 7.50 .
IF (ajbsoc = 175) riskmain = 10.30 .
IF (ajbsoc = 176) riskmain = 9.40 .
IF (ajbsoc = 177) riskmain = 5.13 .
IF (ajbsoc = 178) riskmain = 6.42 .
```

```
IF (ajbsoc = 179) riskmain = 5.70 .
IF (ajbsoc = 190) riskmain = 2.72 .
IF (ajbsoc = 191) riskmain = 2.86 .
IF (ajbsoc = 199) riskmain = 3.89 .
IF (ajbsoc = 200) riskmain = 5.07 .
IF (ajbsoc = 201) riskmain = 3.19 .
IF (ajbsoc = 202) riskmain = 2.74 .
IF (ajbsoc = 209) riskmain = 2.72 .
IF (ajbsoc = 210) riskmain = 6.21 .
IF (ajbsoc = 211) riskmain = 8.93 .
IF (ajbsoc = 212) riskmain = 6.82 .
IF (ajbsoc = 213) riskmain = 6.33 .
IF (ajbsoc = 214) riskmain = 3.66 .
IF (ajbsoc = 215) riskmain = 8.57 .
IF (ajbsoc = 216) riskmain = 3.00 .
IF (ajbsoc = 217) riskmain = 8.66 .
IF (ajbsoc = 218) riskmain = 5.05 .
IF (ajbsoc = 219) riskmain = 8.25 .
IF (ajbsoc = 220) riskmain = 3.14 .
IF (ajbsoc = 221) riskmain = 2.22 .
IF (ajbsoc = 222) riskmain = 0.00 .
IF (ajbsoc = 223) riskmain = 2.00 .
IF (ajbsoc = 224) riskmain = 27.59 .
IF (ajbsoc = 230) riskmain = 1.78 .
IF (ajbsoc = 231) riskmain = 2.78 .
IF (ajbsoc = 232) riskmain = 0.00 .
IF (ajbsoc = 233) riskmain = 5.55 .
IF (ajbsoc = 234) riskmain = 4.16 .
IF (ajbsoc = 235) riskmain = 15.00 .
IF (ajbsoc = 239) riskmain = 4.74 .
IF (ajbsoc = 240) riskmain = 0.00 .
IF (ajbsoc = 241) riskmain = 4.82 .
IF (ajbsoc = 242) riskmain = 0.63 .
IF (ajbsoc = 250) riskmain = 1.33 .
IF (ajbsoc = 251) riskmain = 0.82 .
IF (ajbsoc = 252) riskmain = 1.35 .
IF (ajbsoc = 253) riskmain = 3.19 .
IF (ajbsoc = 260) riskmain = 1.79 .
IF (ajbsoc = 261) riskmain = 0.00 .
IF (ajbsoc = 262) riskmain = 5.25 .
IF (ajbsoc = 270) riskmain = 0.00 .
IF (ajbsoc = 271) riskmain = 10.53 .
IF (ajbsoc = 290) riskmain = 0.00 .
IF (ajbsoc = 291) riskmain = 11.76 .
IF (ajbsoc = 292) riskmain = 4.49 .
IF (ajbsoc = 293) riskmain = 16.18 .
IF (ajbsoc = 300) riskmain = 5.49 .
IF (ajbsoc = 301) riskmain = 12.88 .
IF (ajbsoc = 302) riskmain = 9.49 .
IF (ajbsoc = 303) riskmain = 3.03 .
IF (ajbsoc = 304) riskmain = 11.01 .
IF (ajbsoc = 309) riskmain = 14.18 .
IF (ajbsoc = 310) riskmain = 1.83 .
IF (ajbsoc = 311) riskmain = 17.54 .
IF (ajbsoc = 312) riskmain = 3.40 .
IF (ajbsoc = 313) riskmain = 7.27 .
IF (ajbsoc = 320) riskmain = 2.98 .
```

```
IF (ajbsoc = 330) riskmain = 0.00 .
IF (ajbsoc = 331) riskmain = 5.66 .
IF (ajbsoc = 332) riskmain = 12.81 .
IF (ajbsoc = 340) riskmain = 10.47 .
IF (ajbsoc = 342) riskmain = 13.33 .
IF (ajbsoc = 343) riskmain = 6.45 .
IF (ajbsoc = 344) riskmain = 4.76 .
IF (ajbsoc = 345) riskmain = 0.00 .
IF (ajbsoc = 346) riskmain = 19.72 .
IF (ajbsoc = 347) riskmain = 8.33 .
IF (ajbsoc = 348) riskmain = 11.11 .
IF (ajbsoc = 349) riskmain = 25.00 .
IF (ajbsoc = 350) riskmain = 1.18 .
IF (ajbsoc = 360) riskmain = 6.04 .
IF (ajbsoc = 361) riskmain = 2.15 .
IF (ajbsoc = 362) riskmain = 2.37 .
IF (ajbsoc = 363) riskmain = 1.68 .
IF (ajbsoc = 364) riskmain = 1.71 .
IF (ajbsoc = 370) riskmain = 6.52 .
IF (ajbsoc = 371) riskmain = 7.05 .
IF (ajbsoc = 380) riskmain = 2.43 .
IF (ajbsoc = 381) riskmain = 4.98 .
IF (ajbsoc = 382) riskmain = 5.13 .
IF (ajbsoc = 383) riskmain = 22.22 .
IF (ajbsoc = 384) riskmain = 7.28 .
IF (ajbsoc = 385) riskmain = 6.19 .
IF (ajbsoc = 386) riskmain = 7.84 .
IF (ajbsoc = 387) riskmain = 29.03 .
IF (ajbsoc = 390) riskmain = 0.00 .
IF (ajbsoc = 391) riskmain = 5.42 .
IF (ajbsoc = 392) riskmain = 10.91 .
IF (ajbsoc = 393) riskmain = 9.45 .
IF (ajbsoc = 394) riskmain = 3.33 .
IF (ajbsoc = 395) riskmain = 16.67 .
IF (ajbsoc = 396) riskmain = 1.14 .
IF (ajbsoc = 399) riskmain = 3.64 .
IF (ajbsoc = 400) riskmain = 3.23 .
IF (ajbsoc = 401) riskmain = 5.02 .
IF (ajbsoc = 410) riskmain = 1.94 .
IF (ajbsoc = 411) riskmain = 5.13 .
IF (ajbsoc = 412) riskmain = 16.35 .
IF (ajbsoc = 420) riskmain = 6.52 .
IF (ajbsoc = 421) riskmain = 4.35 .
IF (ajbsoc = 430) riskmain = 4.81 .
IF (ajbsoc = 440) riskmain = 13.51 .
IF (ajbsoc = 441) riskmain = 16.35 .
IF (ajbsoc = 450) riskmain = 66.67 .
IF (ajbsoc = 451) riskmain = 0.00 .
IF (ajbsoc = 452) riskmain = 0.00 .
IF (ajbsoc = 459) riskmain = 0.00 .
IF (ajbsoc = 460) riskmain = 7.41 .
IF (ajbsoc = 461) riskmain = 0.00 .
IF (ajbsoc = 462) riskmain = 0.00 .
IF (ajbsoc = 463) riskmain = 7.09 .
IF (ajbsoc = 490) riskmain = 5.00 .
IF (ajbsoc = 491) riskmain = 0.00 .
IF (ajbsoc = 500) riskmain = 16.69 .
```

```
IF (ajbsoc = 501) riskmain = 13.99 .
IF (ajbsoc = 502) riskmain = 16.83 .
IF (ajbsoc = 503) riskmain = 30.67 .
IF (ajbsoc = 504) riskmain = 13.29 .
IF (ajbsoc = 505) riskmain = 11.70 .
IF (ajbsoc = 506) riskmain = 15.90 .
IF (ajbsoc = 507) riskmain = 8.57 .
IF (ajbsoc = 509) riskmain = 16.50 .
IF (ajbsoc = 510) riskmain = 15.38 .
IF (ajbsoc = 511) riskmain = 20.41 .
IF (ajbsoc = 512) riskmain = 10.91 .
IF (ajbsoc = 513) riskmain = 11.90 .
IF (ajbsoc = 514) riskmain = 30.00 .
IF (ajbsoc = 515) riskmain = 20.96 .
IF (ajbsoc = 516) riskmain = 19.64 .
IF (ajbsoc = 517) riskmain = 7.22 .
IF (ajbsoc = 518) riskmain = 0.00 .
IF (ajbsoc = 519) riskmain = 14.43 .
IF (ajbsoc = 520) riskmain = 6.98 .
IF (ajbsoc = 521) riskmain = 15.52 .
IF (ajbsoc = 522) riskmain = 6.15 .
IF (ajbsoc = 523) riskmain = 12.25 .
IF (ajbsoc = 524) riskmain = 21.13 .
IF (ajbsoc = 525) riskmain = 10.38 .
IF (ajbsoc = 526) riskmain = 4.57 .
IF (ajbsoc = 529) riskmain = 13.37 .
IF (ajbsoc = 530) riskmain = 36.54 .
IF (ajbsoc = 531) riskmain = 11.97 .
IF (ajbsoc = 532) riskmain = 18.68 .
IF (ajbsoc = 533) riskmain = 26.40 .
IF (ajbsoc = 534) riskmain = 26.90 .
IF (ajbsoc = 535) riskmain = 23.67 .
IF (ajbsoc = 536) riskmain = 12.12 .
IF (ajbsoc = 537) riskmain = 24.49 .
IF (ajbsoc = 540) riskmain = 20.90 .
IF (ajbsoc = 541) riskmain = 23.33 .
IF (ajbsoc = 542) riskmain = 22.95 .
IF (ajbsoc = 543) riskmain = 15.28 .
IF (ajbsoc = 544) riskmain = 31.58 .
IF (ajbsoc = 550) riskmain = 3.85 .
IF (ajbsoc = 551) riskmain = 0.00 .
IF (ajbsoc = 552) riskmain = 14.81 .
IF (ajbsoc = 553) riskmain = 3.81 .
IF (ajbsoc = 554) riskmain = 7.73 .
IF (ajbsoc = 555) riskmain = 5.03 .
IF (ajbsoc = 556) riskmain = 6.90 .
IF (ajbsoc = 557) riskmain = 16.13 .
IF (ajbsoc = 559) riskmain = 20.51 .
IF (ajbsoc = 560) riskmain = 1.28 .
IF (ajbsoc = 561) riskmain = 12.61 .
IF (ajbsoc = 562) riskmain = 10.53 .
IF (ajbsoc = 563) riskmain = 17.95 .
IF (ajbsoc = 569) riskmain = 15.73 .
IF (ajbsoc = 570) riskmain = 17.60 .
IF (ajbsoc = 571) riskmain = 15.44 .
IF (ajbsoc = 572) riskmain = 20.69 .
IF (ajbsoc = 573) riskmain = 11.11 .
```

```
IF (ajbsoc = 579) riskmain = 12.96 .
IF (ajbsoc = 580) riskmain = 11.82 .
IF (ajbsoc = 581) riskmain = 26.87 .
IF (ajbsoc = 582) riskmain = 11.32 .
IF (ajbsoc = 590) riskmain = 20.29 .
IF (ajbsoc = 591) riskmain = 0.00 .
IF (ajbsoc = 592) riskmain = 0.00 .
IF (ajbsoc = 593) riskmain = 0.00 .
IF (ajbsoc = 594) riskmain = 14.55 .
IF (ajbsoc = 595) riskmain = 13.25 .
IF (ajbsoc = 596) riskmain = 8.96 .
IF (ajbsoc = 597) riskmain = 41.67 .
IF (ajbsoc = 598) riskmain = 18.18 .
IF (ajbsoc = 599) riskmain = 19.90 .
IF (ajbsoc = 600) riskmain = 19.81 .
IF (ajbsoc = 601) riskmain = 9.45 .
IF (ajbsoc = 610) riskmain = 34.82 .
IF (ajbsoc = 611) riskmain = 24.30 .
IF (ajbsoc = 612) riskmain = 26.39 .
IF (ajbsoc = 613) riskmain = 2.11 .
IF (ajbsoc = 614) riskmain = 8.89 .
IF (ajbsoc = 615) riskmain = 11.25 .
IF (ajbsoc = 619) riskmain = 10.26 .
IF (ajbsoc = 620) riskmain = 18.79 .
IF (ajbsoc = 621) riskmain = 9.61 .
IF (ajbsoc = 622) riskmain = 6.36 .
IF (ajbsoc = 630) riskmain = 12.63 .
IF (ajbsoc = 631) riskmain = 12.00 .
IF (ajbsoc = 640) riskmain = 30.77 .
IF (ajbsoc = 641) riskmain = 14.49 .
IF (ajbsoc = 642) riskmain = 25.84 .
IF (ajbsoc = 644) riskmain = 12.14 .
IF (ajbsoc = 650) riskmain = 0.00 .
IF (ajbsoc = 651) riskmain = 0.00 .
IF (ajbsoc = 652) riskmain = 38.71 .
IF (ajbsoc = 659) riskmain = 10.53 .
IF (ajbsoc = 660) riskmain = 4.00 .
IF (ajbsoc = 661) riskmain = 100.00 .
IF (ajbsoc = 670) riskmain = 0.00 .
IF (ajbsoc = 671) riskmain = 0.00 .
IF (ajbsoc = 672) riskmain = 16.44 .
IF (ajbsoc = 673) riskmain = 17.58 .
IF (ajbsoc = 690) riskmain = 17.39 .
IF (ajbsoc = 691) riskmain = 2.25 .
IF (ajbsoc = 699) riskmain = 10.53 .
IF (ajbsoc = 700) riskmain = 0.00 .
IF (ajbsoc = 701) riskmain = 1.94 .
IF (ajbsoc = 702) riskmain = 8.00 .
IF (ajbsoc = 703) riskmain = 4.35 .
IF (ajbsoc = 710) riskmain = 4.96 .
IF (ajbsoc = 719) riskmain = 6.27 .
IF (ajbsoc = 720) riskmain = 8.03 .
IF (ajbsoc = 721) riskmain = 4.57 .
IF (ajbsoc = 722) riskmain = 0.00 .
IF (ajbsoc = 730) riskmain = 8.51 .
IF (ajbsoc = 731) riskmain = 12.34 .
IF (ajbsoc = 732) riskmain = 6.63 .
```

```
IF (ajbsoc = 733) riskmain = 9.30 .
IF (ajbsoc = 790) riskmain = 24.24 .
IF (ajbsoc = 791) riskmain = 6.90 .
IF (ajbsoc = 792) riskmain = 0.00 .
IF (ajbsoc = 800) riskmain = 15.38 .
IF (ajbsoc = 801) riskmain = 3.85 .
IF (ajbsoc = 802) riskmain = 13.33 .
IF (ajbsoc = 809) riskmain = 20.15 .
IF (ajbsoc = 810) riskmain = 26.09 .
IF (ajbsoc = 811) riskmain = 19.35 .
IF (ajbsoc = 812) riskmain = 29.27 .
IF (ajbsoc = 813) riskmain = 0.00 .
IF (ajbsoc = 814) riskmain = 21.43 .
IF (ajbsoc = 820) riskmain = 17.20 .
IF (ajbsoc = 821) riskmain = 30.41 .
IF (ajbsoc = 822) riskmain = 20.41 .
IF (ajbsoc = 823) riskmain = 19.23 .
IF (ajbsoc = 824) riskmain = 14.58 .
IF (ajbsoc = 825) riskmain = 19.39 .
IF (ajbsoc = 826) riskmain = 5.71 .
IF (ajbsoc = 829) riskmain = 20.65 .
IF (ajbsoc = 830) riskmain = 25.29 .
IF (ajbsoc = 831) riskmain = 19.35 .
IF (ajbsoc = 832) riskmain = 11.76 .
IF (ajbsoc = 833) riskmain = 21.62 .
IF (ajbsoc = 834) riskmain = 22.99 .
IF (ajbsoc = 839) riskmain = 23.70 .
IF (ajbsoc = 840) riskmain = 10.50 .
IF (ajbsoc = 841) riskmain = 17.11 .
IF (ajbsoc = 842) riskmain = 24.32 .
IF (ajbsoc = 843) riskmain = 39.29 .
IF (ajbsoc = 844) riskmain = 29.63 .
IF (ajbsoc = 850) riskmain = 9.38 .
IF (ajbsoc = 851) riskmain = 21.67 .
IF (ajbsoc = 859) riskmain = 21.43 .
IF (ajbsoc = 860) riskmain = 8.16 .
IF (ajbsoc = 861) riskmain = 16.98 .
IF (ajbsoc = 862) riskmain = 15.16 .
IF (ajbsoc = 863) riskmain = 9.64 .
IF (ajbsoc = 864) riskmain = 0.00 .
IF (ajbsoc = 869) riskmain = 12.40 .
IF (ajbsoc = 870) riskmain = 11.43 .
IF (ajbsoc = 871) riskmain = 15.58 .
IF (ajbsoc = 872) riskmain = 20.38 .
IF (ajbsoc = 873) riskmain = 11.30 .
IF (ajbsoc = 874) riskmain = 8.36 .
IF (ajbsoc = 875) riskmain = 50.00 .
IF (ajbsoc = 880) riskmain = 19.57 .
IF (ajbsoc = 881) riskmain = 22.92 .
IF (ajbsoc = 882) riskmain = 21.40 .
IF (ajbsoc = 883) riskmain = 5.71 .
IF (ajbsoc = 884) riskmain = 12.50 .
IF (ajbsoc = 885) riskmain = 12.12 .
IF (ajbsoc = 886) riskmain = 17.82 .
IF (ajbsoc = 887) riskmain = 15.14 .
IF (ajbsoc = 889) riskmain = 22.89 .
IF (ajbsoc = 890) riskmain = 16.67 .
```

```
IF (ajbsoc = 891) riskmain = 15.78 .
IF (ajbsoc = 892) riskmain = 11.88 .
IF (ajbsoc = 893) riskmain = 10.29 .
IF (ajbsoc = 894) riskmain = 22.22 .
IF (ajbsoc = 895) riskmain = 26.87 .
IF (ajbsoc = 896) riskmain = 12.47 .
IF (ajbsoc = 897) riskmain = 22.82 .
IF (ajbsoc = 898) riskmain = 29.91 .
IF (ajbsoc = 899) riskmain = 19.61 .
IF (ajbsoc = 900) riskmain = 14.40 .
IF (ajbsoc = 901) riskmain = 18.07 .
IF (ajbsoc = 902) riskmain = 16.78 .
IF (ajbsoc = 903) riskmain = 20.45 .
IF (ajbsoc = 904) riskmain = 25.90 .
IF (ajbsoc = 910) riskmain = 8.00 .
IF (ajbsoc = 911) riskmain = 40.00 .
IF (ajbsoc = 912) riskmain = 25.00 .
IF (ajbsoc = 913) riskmain = 23.53 .
IF (ajbsoc = 919) riskmain = 14.29 .
IF (ajbsoc = 920) riskmain = 18.75 .
IF (ajbsoc = 921) riskmain = 14.71 .
IF (ajbsoc = 922) riskmain = 25.00 .
IF (ajbsoc = 923) riskmain = 19.21 .
IF (ajbsoc = 924) riskmain = 18.75 .
IF (ajbsoc = 925) riskmain = 15.74 .
IF (ajbsoc = 930) riskmain = 13.64 .
IF (ajbsoc = 931) riskmain = 24.00 .
IF (ajbsoc = 932) riskmain = 22.22 .
IF (ajbsoc = 933) riskmain = 31.25 .
IF (ajbsoc = 934) riskmain = 11.11 .
IF (ajbsoc = 940) riskmain = 22.59 .
IF (ajbsoc = 941) riskmain = 7.61 .
IF (ajbsoc = 950) riskmain = 24.32 .
IF (ajbsoc = 951) riskmain = 10.53 .
IF (ajbsoc = 952) riskmain = 16.00 .
IF (ajbsoc = 953) riskmain = 17.48 .
IF (ajbsoc = 954) riskmain = 12.90 .
IF (ajbsoc = 955) riskmain = 9.38 .
IF (ajbsoc = 956) riskmain = 14.71 .
IF (ajbsoc = 957) riskmain = 7.59 .
IF (ajbsoc = 958) riskmain = 8.39 .
IF (ajbsoc = 959) riskmain = 17.39 .
IF (ajbsoc = 990) riskmain = 14.31 .
IF (ajbsoc = 999) riskmain = 4.65 .
EXECUTE .
```

3. riskmajm.sps

This file was used to allocate the 1-digit jobrisk variable (calculated from the LFS) to the BHPS data.

```
IF (ajbmajm = 1) riskmajm = 5.30 .
IF (ajbmajm = 2) riskmajm = 4.48 .
IF (ajbmajm = 3) riskmajm = 6.26 .
IF (ajbmajm = 4) riskmajm = 8.44 .
IF (ajbmajm = 5) riskmajm = 16.48 .
IF (ajbmajm = 6) riskmajm = 19.09 .
IF (ajbmajm = 7) riskmajm = 6.55 .
IF (ajbmajm = 8) riskmajm = 16.77 .
IF (ajbmajm = 9) riskmajm = 16.85 .
EXECUTE .
```

4. riskminm.sps

This file was used to allocate the 2-digit jobrisk values (calculated from the LFS) to the BHPS sample.

```
IF (ajbminm = 10) riskminm = 2.32 .
IF (ajbminm = 11) riskminm = 6.39 .
IF (ajbminm = 12) riskminm = 2.93 .
IF (ajbminm = 13) riskminm = 2.25 .
IF (ajbminm = 14) riskminm = 6.99 .
IF (ajbminm = 15) riskminm = 10.39 .
IF (ajbminm = 16) riskminm = 11.98 .
IF (ajbminm = 17) riskminm = 6.50 .
IF (ajbminm = 19) riskminm = 3.71 .
IF (ajbminm = 20) riskminm = 3.60 .
IF (ajbminm = 21) riskminm = 6.02 .
IF (ajbminm = 22) riskminm = 3.80 .
IF (ajbminm = 23) riskminm = 4.36 .
IF (ajbminm = 24) riskminm = 1.06 .
IF (ajbminm = 25) riskminm = 1.60 .
IF (ajbminm = 26) riskminm = 3.51 .
IF (ajbminm = 27) riskminm = 4.04 .
IF (ajbminm = 29) riskminm = 9.49 .
IF (ajbminm = 30) riskminm = 10.33 .
IF (ajbminm = 31) riskminm = 3.56 .
IF (ajbminm = 32) riskminm = 2.98 .
IF (ajbminm = 33) riskminm = 8.74 .
IF (ajbminm = 34) riskminm = 10.89 .
IF (ajbminm = 35) riskminm = 1.18 .
IF (ajbminm = 36) riskminm = 2.74 .
IF (ajbminm = 37) riskminm = 6.94 .
IF (ajbminm = 38) riskminm = 7.24 .
IF (ajbminm = 39) riskminm = 5.84 .
IF (ajbminm = 40) riskminm = 3.88 .
IF (ajbminm = 41) riskminm = 3.80 .
IF (ajbminm = 42) riskminm = 6.41 .
IF (ajbminm = 43) riskminm = 4.81 .
```

```
IF (ajbminm = 44) riskminm = 16.20 .
IF (ajbminm = 45) riskminm = 4.35 .
IF (ajbminm = 46) riskminm = 5.29 .
IF (ajbminm = 49) riskminm = 4.99 .
IF (ajbminm = 50) riskminm = 13.82 .
IF (ajbminm = 51) riskminm = 18.15 .
IF (ajbminm = 52) riskminm = 13.48 .
IF (ajbminm = 53) riskminm = 21.87 .
IF (ajbminm = 54) riskminm = 21.30 .
IF (ajbminm = 55) riskminm = 8.06 .
IF (ajbminm = 56) riskminm = 12.49 .
IF (ajbminm = 57) riskminm = 17.19 .
IF (ajbminm = 58) riskminm = 20.61 .
IF (ajbminm = 59) riskminm = 14.84 .
IF (ajbminm = 60) riskminm = 18.48 .
IF (ajbminm = 61) riskminm = 23.58 .
IF (ajbminm = 62) riskminm = 14.43 .
IF (ajbminm = 63) riskminm = 12.31 .
IF (ajbminm = 64) riskminm = 19.85 .
IF (ajbminm = 65) riskminm = 21.05 .
IF (ajbminm = 66) riskminm = 5.26 .
IF (ajbminm = 67) riskminm = 16.12 .
IF (ajbminm = 69) riskminm = 9.98 .
IF (ajbminm = 70) riskminm = 2.53 .
IF (ajbminm = 71) riskminm = 5.34 .
IF (ajbminm = 72) riskminm = 7.59 .
IF (ajbminm = 73) riskminm = 10.18 .
IF (ajbminm = 79) riskminm = 6.80 .
IF (ajbminm = 80) riskminm = 18.29 .
IF (ajbminm = 81) riskminm = 21.82 .
IF (ajbminm = 82) riskminm = 19.05 .
IF (ajbminm = 83) riskminm = 22.59 .
IF (ajbminm = 84) riskminm = 14.34 .
IF (ajbminm = 85) riskminm = 17.56 .
IF (ajbminm = 86) riskminm = 11.73 .
IF (ajbminm = 87) riskminm = 16.65 .
IF (ajbminm = 88) riskminm = 16.30 .
IF (ajbminm = 89) riskminm = 17.49 .
IF (ajbminm = 90) riskminm = 16.74 .
IF (ajbminm = 91) riskminm = 22.17 .
IF (ajbminm = 92) riskminm = 17.17 .
IF (ajbminm = 93) riskminm = 24.00 .
IF (ajbminm = 94) riskminm = 18.23 .
IF (ajbminm = 95) riskminm = 12.75 .
IF (ajbminm = 99) riskminm = 13.69 .
EXECUTE .
```

5. schooling years.sps

This file shows the commands used to create the years of schooling variable used in the study.

```
RECODE
  afeend aqfedhi ascend  (Lowest thru 0=SYSMIS)  .
IF (aqfedhi = 1 & afeend  <=  27) schle = afeend .
IF (aqfedhi = 1 & afeend > 27) schle = 25 .
IF (aqfedhi = 1  & MISSING(afeend)) schle = 25 .
IF (aqfedhi = 2 & afeend  <=  24) schle = afeend .
IF (aqfedhi = 2 & afeend > 24) schle = 21 .
IF (aqfedhi = 2  & MISSING(afeend)) schle = 21 .
IF (aqfedhi = 3 & afeend  <=  22) schle = afeend .
IF (aqfedhi = 3 & afeend > 22) schle = 22 .
IF (aqfedhi = 3  & MISSING(afeend)) schle = 22 .
IF (aqfedhi = 4 & afeend  <=  22) schle = afeend .
IF (aqfedhi = 4 & afeend > 22) schle = 20 .
IF (aqfedhi = 4  & MISSING(afeend)) schle = 20 .
IF (aqfedhi = 5 & afeend  <=  21) schle = afeend .
IF (aqfedhi = 5 & afeend > 21) schle = 21 .
IF (aqfedhi = 5  & MISSING(afeend)) schle = 21 .
IF (aqfedhi = 6 & afeend  <=  20) schle = afeend .
IF (aqfedhi = 6 & afeend > 20) schle = 18 .
IF (aqfedhi = 6  & MISSING(afeend)) schle = ascend .
IF (aqfedhi = 7 & afeend  <=  17) schle = afeend .
IF (aqfedhi = 7 & afeend > 17) schle = 17 .
IF (aqfedhi = 7  & MISSING(afeend)) schle = ascend .
IF (aqfedhi = 8 & afeend  <=  19) schle = afeend .
IF (aqfedhi = 8 & afeend > 19) schle = 19 .
IF (aqfedhi = 8  & MISSING(afeend)) schle = ascend + 1 .
IF (aqfedhi = 9 & afeend  <=  17) schle = afeend .
IF (aqfedhi = 9 & afeend > 17) schle = 17 .
IF (aqfedhi = 9  & MISSING(afeend)) schle = ascend .
IF (aqfedhi = 10 & afeend  <=  20) schle = afeend .
IF (aqfedhi = 10 & afeend > 20) schle = 20 .
IF (aqfedhi = 10  & MISSING(afeend)) schle = 18 .
IF (aqfedhi = 11 & afeend  <=  18) schle = afeend .
IF (aqfedhi = 11 & afeend > 18) schle = ascend + 2 .
IF (aqfedhi = 11  & MISSING(afeend)) schle = ascend .
IF (aqfedhi = 12 & afeend  <=  18) schle = afeend .
IF (aqfedhi = 12 & afeend > 18) schle = 17 .
IF (aqfedhi = 12  & MISSING(afeend)) schle = ascend .
VARIABLE LABELS schle 'School leaving age' .
COMPUTE s = schle - 5 .
VARIABLE LABELS s 'years of schooling' .
COMPUTE exp = aage - schle .
VARIABLE LABELS exp 'years of work experience' .
COMPUTE expsq = exp*exp .
VARIABLE LABELS expsq 'years of work experience squared' .
EXECUTE .
```

The above commands led to nine cases where experience is negative and a further 45 where experience is equal to zero. All these cases concerned very young people (mean age of 17.63).

The respondents with negative experience were all 18 or 19 years old. Of these eight had 'Other higher QF' as their highest qualification, while the ninth had 'teaching qualification' as their highest qualification. For the nine with negative experience, their experience and experience squared will be recoded to zero.

```
RECODE
  exp  (Lowest thru -1=0)  .
COMPUTE expsq = exp*exp .
VARIABLE LABELS expsq 'years of work experience squared' .
EXECUTE .
```

6. var_define.sps

This file shows the commands used to create the major derived variables in the study.

```
RECODE
  apaygu ajbhrs  (Lowest thru 0=SYSMIS)  .
COMPUTE job100 = ajbsoc/100 .
COMPUTE job10 = ajbsoc/10 .
IF (ajbsoc > 0) ajbmajm = TRUNC(job100) .
VARIABLE LABELS ajbmajm 'Single digit SOC' .
IF (ajbsoc > 0) ajbminm = TRUNC(job10) .
VARIABLE LABELS ajbminm Two digit SOC' .
RECODE
  atuin1  (1=1)  (2=0)  (Lowest thru 0=0) (MISSING=0) .
COMPUTE nomwage = apaygu / (4*ajbhrs) .
VARIABLE LABELS nomwage 'Nominal hourly wage' .
IF (adoim = 9) realwage = nomwage*134.6/134.6 .
IF (adoim = 10) realwage = nomwage*134.6/135.1 .
IF (adoim = 11) realwage = nomwage*134.6/135.6 .
IF (adoim = 12) realwage = nomwage*134.6/135.7 .
VARIABLE LABELS realwage 'Real hourly wage (at Sep 91 prices)' .
COMPUTE logreal = LN(realwage) .
VARIABLE LABELS logreal 'Log of real hourly wage' .
RECODE
  ajbhgs  (Lowest thru 0=SYSMIS)  .
EXECUTE .
```

Data Archive, October 2000