# Socioeconomic Status and Sickness Absence

# - What do twins tell us about causality?

by William Nilsson*

Department of Economics, Umeå University, Sweden and

Centre for Research on Welfare Economics (CREB), Barcelona, Spain

**Abstract**

The purpose of this study is to empirically investigate causal effects between socioeconomic status and absence from the workplace due to sickness. To be able to conclude that income causally affects health it is important to control for both reverse causality and unobserved heterogeneity. This study uses a Swedish sample of female twins and a semiparametric censored fixed-effects model. Spousal income is correlated in cross-section with the share of total income that comes from benefits due to sickness absence. Results from this twin study indicate that male spousal income, *i.e.* a non-shared environmental influence, does not have a causal effect.

**JEL:** C24, I12, I21

**Keywords:** income, education, health, causality, twins

---

* I wish to thank seminar participants at the Centre for Research on Welfare Economics (CREB), Universitat de Barcelona. I wish also to thank Niklas Rudholm for valuable comments on a draft of the paper. Financial support from Swedish Council for Working Life and Social Research is gratefully acknowledged. All data, except for the identification of the twin sample, come from Statistics Sweden (SCB). The twin information comes from the Swedish Twin Registry (STR). The Swedish Twin Registry is supported by grants from the Swedish Research Council. To contact the author, write to william.nilsson@econ.umu.se.

# 1 Introduction

The relationship between socioeconomic status and health has been empirically investigated in a large number of different populations. A typical result is that a higher socioeconomic status is correlated with better health (Fuchs, 2004). From a policy perspective it is important to know if the relationship is due to causal effects. It is, however, an empirical challenge to find out if low socioeconomic status causally influences health status. Empirical studies that aim to determine whether socioeconomic status causally influences health have to take into account that other, possibly unobserved, variables can influence both socioeconomic status and health. It is also necessary to deal with possible reverse causality, *i.e*. that health causally affects socioeconomic status.

The purpose of this study is to empirically investigate this issue with a method based on a twin sample. Various ideas in the literature have been used in the effort to establish whether the relationship is causal or not. Among them is the one that endeavored to find a suitable instrument that is correlated with income, uncorrelated with the error term, and not in itself an explanatory variable in the equation explaining health status. Lottery winners' circumstances have been examined on the assumption that a lottery win is an exogenous change in the economic situation (Lindahl, 2005). The reunion of Germany has also been used as an exogenous change (Frijters, 2005). Finding a reliable instrument could, however, be very difficult. For this reason, some studies have instead focused on a *change* in health. Buckley *et al*. (2004), for example, studied healthy individuals in the initial period. In that way they tried to minimize the effects that health may have on income. They did, however, point to the potential problem that individuals could have moved in and out of a good health before the initial period in the data, and for this reason have a lower income. These individuals could also have a higher probability of experiencing a poor health again.

Adams *et al*. (2003) studied a change in health status in an attempt to test the absence of a direct causal path. They chose a population of Americans aged 70 and older. Their test relies on the postulation that no hidden factors influence the initial state or the change. This was their reason for choosing an old population, which is such that hidden factors are more likely to have been manifest in the observed covariates. Adams *et al*. (2003) do not control for unobserved heterogeneity. Michaud & van Soest (2004) argue that not doing that could have biased the estimates and the test results, as correlated unobserved heterogeneity could be important. The reason is that early childhood experiences, genetic transmission and other persistent shocks can influence both health and wealth. With a dynamic vector autoregressive panel-data model, Michaud & van Soest (2004) took into account unobserved time invariant heterogeneity. They found no evidence of health being causally influenced by wealth.

The purpose of this study is to investigate whether a low income for a spouse is important as a causal factor in the receipt in Sweden of a large share of income from sickness-related benefits. To deal with unobserved heterogeneity, this study uses information about female twins. The data set includes information about spouses, and the income of spouses is used as an exogenous influence on socioeconomic status. Since the share of income that comes from sickness-related benefits is in many cases zero, the dependent variable is censored. This censoring is dealt with by using the censored regression model with fixed effects developed by Honoré (1992).

This study contributes to the literature in the following ways: First it uses an alternative method for the identification of a possible causal effect between income and health status. Secondly, unobserved heterogeneity is controlled for by using data for twins. The reason is that it is likely that within individual time-series variations in income, for a short time-period, reflect mainly transitory changes, which, in fact, have very little influence on health. Using traditional fixed-effects models can, accordingly, understate the magnitude of the effect in a way similar to measurement error (McKinnish, 2006). Further, the dynamics of health and income are difficult to model correctly, which would be troublesome in a quest for reliable estimates based on the traditional fixed-effects model. It is, for example, difficult to know whether an effect is expected instantly, whether symptoms show only several years later, or whether the health problems are long-lasting. Thirdly, the use in research of twins and information for spouses does not occur in the literature of socioeconomic status and health.

The results indicate that education matters for the share of total income that comes from sickness-related welfare systems. For the female sample, no effect is found from spousal income on the share of total income that comes from the sickness-related welfare system. The next section outlines the theoretical framework. Section 3 presents the data and section 4 the econometric model. The results are detailed in section 5, and the concluding remarks are offered in section 6.

## 2 Theoretical framework

To be able to identify a possible causal effect between income and health, it is important to have an overview of the theoretical motives of an effect.[1] One suggested reason for a causal impact is that low income earners' access to health care might be less or of lower quality (Deaton, 2002). Even in the case of Sweden with its public health care system, we cannot exclude this potential reason as it is possible that even small costs could keep low income earners away from health care services. It is also possible that different groups of individuals are treated differently within the system. The public health care system should, however, have reduced the importance of this argument.

Another reason for a causal impact between income and health is that such impact might be embed in the channel of behaviors such as consumption habits, risk-taking and other activities that produce stress. Smoking, alcohol consumption and lack of physical activity could in themselves explain a worse health status, but they are also related to socioeconomic status (Deaton, 2002).

Even though these explanations could have an effect, they could also be a

[1] Though this theoretical overview discusses income and health, it is important to remember that the empirical investigation concerns sickness absence, which is only an indirect measure of health.

consequence of circumstances like low income, discrimination and social exclusion. Deaton (2002) noted also that a risky behavior may be neither irresponsible nor irrational. The reason for it is that poor people with little human and financial capital have to rely on their health capital in production and consumption. It is possible to wear and tear the physical body in the process of earning income and seeking pleasure in cheap but unhealthy activities.

A recent empirical study of socioeconomic status, lifestyle and health was conducted in the United Kingdom by Contoyannis & Jones (2004). They constructed a multivariate probit model consisting of a structural health equation and reduced-form lifestyle equations. The error terms in the equations were allowed to correlate, and the result was deduced with full information maximum likelihood (FIML) in order to take into account unobserved heterogeneity. They found that "sleeping well, exercising and not smoking in 1984 have dramatic positive effects on the probability of reporting excellent or good self-reported health in 1991" (Contoyannis & Jones, 2004).

Job-related stress can also be an explanation of the relation between income and health. Lack of control over the work situation seems to be an important negative factor for health (Smith, 1999). Sloan *et al.* (2005) identified a possible path along which a low socioeconomic status can influence physical health status. They found that the parasympathetic nervous system, *i.e*. a part of the nervous system that serves, *inter alia*, to slow the heart rate, “may be a mechanism linking the stressful effects of low SES to increased morbidity and mortality.” The reason is that the stress of the experience of low socioeconomic status reduces parasympathetic activity.

It is also possible that money in itself is not the most important facet of this experience. Possibly more important is the awareness that superior social status can accrue from a high-earning job. In that case the disposable income would matter less, and the labor income that the individual achieves would be more important (Fritzell *et al*. 2004). Related to this are theories of the importance for health status of the relative position on the social ladder (Adler *et al*. 1994 and Wilkinson, 1996).

The theoretical reasons mentioned so far that would align income and health in a relationship that is causal involve attitude and activity variables that are capable of having an observable effect within a fairly short period of time. At the same time, they could work cumulatively, in which case the length of the low income period is more significant than the low income itself. There are also theories that emphasize the impact of early childhood or even intrauterine environmental factors (Smith, 1999). These theories underline also the possible importance of even short, critical periods when the body is developing. In this way, environmental effects on intrauterine life can program the occurrence of diseases in adulthood. Undernourished babies do not only show symptoms such as low birth weight but they could manifest disproportional growth of parts of the body also (Godfrey & Barker, 2000). Case *et al*. (2002) studied economic status and health in childhood. They found families' long-term average income to be an important factor in children's health status. Case *et al* (2005) empirically investigated the effects of early childhood health and intrauterine environmental effects on adult health. Effects were found to be present in both contexts. Apart from early childhood and intrauterine environmental influences, the importance for health status of genes has also been suggested (Fuchs, 2004 and Michaud & van Soest, 2004).

In empirical investigation it is, of course, difficult to include control variables for all individual specific information that is correlated with health status in adulthood and socioeconomic status. Not taking unobserved heterogeneity into account can, accordingly, yield inconsistent estimates due to omitted variables. In particular, it is possible that a correlation between income and health is mistakenly concluded as a causal effect, even though it could be due to unobserved heterogeneity. In the literature this explanation for correlation between health and socioeconomic status has been labeled a third variable explanation (Fuchs, 2004).

A simple model, based on Neumark (1999), can illustrate the problem. In that study the focus is on ability bias in the estimation of a return to education. The basic problem of omitted variables is, however, the same. Auld & Sidhu

(2005) do, in fact, also underline the risk for ability bias in health regressions. The role of cognitive ability (intelligence) in health is also studied in Singh-Manoux *et al.* (2005). As a simplification it is here assumed that health, $h$, is linearly related to income ($I$) and fundamental underlying health status, ($FHS$).[2] $FHS$ is assumed to be unobserved and based on, for example, genetics and childhood experiences. Higher values would indicate better fundamental health status.

$$h = \beta I + \lambda FHS + \varepsilon \tag{1}$$

where $p\lim(I \cdot \varepsilon) = p\lim(FHS \cdot \varepsilon) = 0$. $FHS$ includes unobserved heterogeneity, *i.e.* all omitted variables that affect health and may be correlated with income. Equation (1) can be interpreted as a simple model after all other control variables have been partialed out. Further, income is assumed to be correlated to fundamental underlying health status as:

$$I = \gamma FHS + \eta \tag{2}$$

where $p\lim(\eta \cdot \varepsilon) = 0$. If the model were estimated without $FHS$ with OLS, $b_{LS}$ would be a biased estimate of $\beta$ as:

$$p\lim(b_{LS}) = \beta + \lambda\sigma_{FHS,I} / \sigma_I^2 = \beta + \lambda\gamma\sigma_{FHS}^2 / \sigma_I^2. \tag{3}$$

On the assumption that Fundamental Health Status, $FHS$, consists of omitted variables such as health or socioeconomic status in early childhood, we expect $\lambda$ to be positive. If these unobserved variables that would affect health, $h$, positively also affect income positively, *i.e.* $\gamma > 0$, then $b_{LS}$ is biased

[2] In this simple model the assumption that income is linearly related to health is made when illustrating the problem of omitted variables. In empirical applications the possibility of a nonlinear relationship should be kept in mind.

upwards. That is, $b_{LS}$ would be more positive than the true causal effect of income on health.

If $FHS$ is identical for both twins, an unbiased within-twin estimate, $b_{WT}$, can be achieved by differencing equation (1) for two twin siblings and estimating the differenced equation (4) with OLS:

$$\Delta h = \beta \Delta I + \Delta \varepsilon \tag{4}$$

This would be the case where the income for twins is randomly assigned, and information on twins is a good natural experiment. Unfortunately, we cannot exclude the possibility that differences in childhood health between twins in the same family also have affected income in adulthood. Even identical twins are, for example, found to differ in birth weight. Even with the same genes and same family background, we expect health to differ in early childhood. If these differences are important for later differences in the twins' income, this can produce an even larger bias than the one that occurs in the cross-section. The analogue discussion is present in the literature for twin studies dealing with ability bias in estimates of the return to education (Neumark, 1999). In principle, for consistent estimates income would have to be treated as endogenous.

In the Biology literature, information from spouses of twins has been used as identifying information, for spouses bring non-shared environmental influence (Spotts *et al*. 2004). Lykken & Tellegren (1993) studied middle-aged twins in the Minnesota Twin Registry and found that "characteristics of the chooser and the chosen constrain mate-selection only weakly." The final choice of a mate is found to be largely random. This can be exploited empirically. Assuming $\Delta I_S$, *i.e.* that the difference of the income of twins' spouses is orthogonal to both $\Delta FHS$ and $\Delta I$, spouses contribute a non-shared environmental influence that can be used to investigate the causal effect of income. It is, of course, difficult to confirm that $\Delta I_S$ is uncorrelated with $\Delta FHS$, since the latter is unobservable.

It seems reasonable that within twin-pairs differences in childhood health status do not affect differences in spousal adult income. It is, however, possible that the variable is affected indirectly in the process of finding a spouse, and hence, that a zero income variable obtains where there is no spouse. It is, accordingly, important to study the basic properties of the data. It is, for example, possible to study the relation between $\Delta I$ and $\Delta I_S$ and the descriptive statistics for subsamples of individuals with and without a spouse as a guideline for whether the method can be used. In this study, this is done in the section describing data.

Note that to be able to observe an effect from spousal income it is necessary that incomes within a family are, at least to some extent, pooled or shared. In this study only married couples or couples living together who have at least one child in common are included. The degree of pooling of incomes should be higher among these kinds of families than among non-married spouses without common children, especially for the purposes of daily consumption and other health-related behaviors.

It is important to note also that the method neither investigates the causal effect of socioeconomic status in childhood nor other factors that are shared by twins. Further, only a causal effect that works through access to health care, consumption and/or lifestyle can be detected. The method does not investigate the effects of income connected to "own" job-related stress or to individual status. The hypothesis that relative position and income inequality is important is, accordingly, also left aside for future studies. The "relative position" and "income" channels, estimated on "own" earnings, have to be investigated with a suitable instrument, as suggested above.

In this study information on spouses is used as identifying information. The next section presents the available data and details concerning the sickness-related welfare system in Sweden. In particular it will be noted that the dependent variable is censored. Section 4 explains the necessary departure from estimating equation (4) with OLS to achieve consistent estimates.

# 3 Data

The empirical analysis in this study is based on a sample of twins born between 1949 and 1958. The identification of the twins and whether they are monozygotic or dizygotic comes from the Swedish Twin Registry (STR). For the sample of twins, Statistics Sweden (SCB) detected the probable spouses for each of the years 1994 to 1999. A spouse is considered detected if the twin was listed as married or living with a person with whom he/she had a common child. For the twins and the identified spouses, the relevant variables included in the longitudinal database, LOUISE, for the years 1994-1999 are used. The twins were between 41 and 50 years old during the last year of the panel. Only twins who lived until at least 2000 are included in the empirical analysis. Twins who were self-employed in 1998 or 1999 are excluded, as the system for sickness benefits differ for self-employed and employed individuals.

The dependent variable in this study is the percentage of total income that comes from the following welfare systems: sickness benefits, benefits due to work related injuries, benefits due to rehabilitation and to early pension. All of these welfare systems rely on medical confirmation of people’s health conditions. Sickness benefits are granted for sickness that continues beyond 14 days. Sickness benefits for shorter periods of illness are covered by the employer, apart from the first day, for which no benefit is granted.

The amount of benefit is based on a theoretical income, *sjukpenninggrundande inkomst* (SGI), which is calculated based on current or earlier earnings. The lowest possible SGI is 24 percent of a base amount that is set every year by the government. The highest possible SGI is 7.5 times the base amount, which in 1999 was set at 36 400 SEK. Within these limits, SGI is equal to the current or earlier earnings. The amount of benefit was, in 1999, calculated as 80 percent of SGI. Note that it is possible be in receipt of no government benefits, despite taking many days of sick leave during the year, when no single spell of leave is longer than 14 days.

The rules for benefits due to work related injuries and benefits due to rehabilitation are very similar to the rules for sickness benefits. Benefits for early pension are granted only for medical reasons and in cases where a reduced ability to work is permanent. If the reduced ability is not permanent but long-lasting, *i.e.* more than one year, sickness benefit is granted. This kind of sickness benefit follows the same rules as early pension, with the difference that the benefit is allowed for a limited time. “Total income” is, in this study, the sum of labor income, unemployment benefits, study assistance, parental allowance and income due to the above-mentioned sickness-related welfare systems. “Labor income” is compensation from employer/s and includes sickness benefits paid during the first 14 days of a period of absence from the workplace for reasons of sickness.

There are two reasons for basing the dependent variable on these welfare systems. First, these systems of compensation are based on a loss of earnings and they are, accordingly, connected to a loss of production. Secondly, a joint measure of sickness-and-health problems is achieved by using the systems of benefits. More serious health problems will, with the measure, yield a higher fall in production and a larger share of income from these systems. A drawback is, of course, that it is possible to have health problems without receiving benefits. It is therefore important, when drawing conclusions from the results, to keep in mind the nature of the health problems measured.

The purpose of this study is to empirically evaluate the importance of spousal income, for a female sample, of the share of income that comes from sickness-related welfare systems. Education is used as a control and is measured with two different dummy variables, one each for different degrees of completed education, and compared with the compulsory period of school education.

As noted in the theoretical section, using the income of the individual would be problematic even with twin data. Further, since the amount of benefits depends, in a nonlinear way as described above, on current or earlier income, using individual earnings would be difficult. In this study the logarithm of income of the spouse is used as non-shared environmental influence. The transformation to logarithms is used to reduce the importance of outliers. A period of disposable income is used as well, as a measure of disposable income during a single year. The period of income is not intended to measure a permanent income, as this would include expected *future* incomes. When the intention is to estimate causal effects, it does not seem appropriate to use permanent income. A period of low/high income could, however, causally affect the dependent variable. It is possible also that income has a nonlinear effect, and for this reason a squared-income variable is included in some regressions in the empirical part.

When twins do not have an identified spouse, the income variables are coded to zero. This has to be taken into consideration when analyzing the empirical estimates.

[Table 1, about here]

Summary statistics for the samples of monozygotic and dizygotic twins are included in Table 1. The female monozygotic sample consists of 1750 individuals, and accordingly, of 875 twin pairs. About 22.7 percent of the individuals had incomes from sickness-related welfare systems in 1999; 34.4 percent did not have an identified spouse in 1998. The average logarithm of the disposable incomes for the spouses of 1998 was 12.0. “Disposable income” includes, apart from labor earnings, welfare benefits and income from self-employment. The income measure is after-tax income.[3] The consumer price index is used to deflate all income variables across the years to the price level of 2001. About 50 percent of the female sample had in 1998 reached an education level of upper secondary school. Another 34 percent had reached post-secondary school or post-graduate education in 1998.

Table A1 in the Appendix includes across-family correlation coefficients for a set of variables that are averaged for the twin siblings. The across-families correlation coefficient for the share of sickness-related benefits in 1999 and the logarithm of spousal disposable income in 1998 is negative. This is also the case

[3] For details concerning income measures and other variables used, see (in Swedish) *Bakgrundsfakta till arbetsmarknads- och utbildningsstatistiken*, 2002:2. Statistics Sweden.

for the average income for a present spouse between 1994 and 1998.

Table A2 in the Appendix includes correlation coefficients for within-twin differences of the same variables. Fewer coefficients are significantly different from zero in the case of within-twins difference. For the variables measuring spousal income, this suggests that these variables, in fact, are exogenous. Note that neither of the variables measuring spousal income is correlated with the within-twins difference of total income.[4] It is not surprising that in the spouse-identifying methodology these variables are correlated with the within-twins difference of the number of children and the dummy for being married. Remember that the income variables are coded to zero if no spouse is identified, and that spouses are identified only in cases where the couples are married or have children in common. The overall impression from the descriptive statistics and the correlation matrices is that male spouses provide exogenous non-shared environmental influence. The effect from male spousal income on the share of total income that comes from sickness-related benefits is, accordingly, the main interest in this study. The next section describes the econometric model that take into account the censored dependent variable and unobserved heterogeneity.

# 4 Econometric model

As described in the previous section, the dependent variable in this study is the share of total income that comes from sickness-related welfare benefits. A large proportion of individuals have no such benefits during the year. It is, however, possible that they experienced sickness and/or work absence for periods of less than 15 days. The dependent variable is, accordingly, censored, and an estimation of equation (4) with OLS would yield biased estimates.

[4] This was also the case for the variables measured as twin-averages. The correlation coefficient for the twin-average of spousal disposable income in 1998 is, however, different if it is calculated for the subsample of individuals where both twins had an identified spouse in 1998. In that case a positive correlation coefficient of about 0.1 is found. (It is significantly different from zero at 5% significance level.) The correlation coefficient is not found to be significantly different from zero when the same subsample is studied for the within-twins differenced variable.

Honoré (1992) proposes estimators for truncated and censored regression models with fixed effects in panel data for two time periods. The censored estimator is used in this study, with the difference that the fixed effects refer to differences between twin siblings instead of in time. The measure of health is modeled as unobserved latent variables, $H_1^*$ and $H_2^*$ according to:

$$H_t^* = \alpha + X_t \beta + \varepsilon_t \quad \text{for} \quad t = 1, 2, \tag{5}$$

where $t = 1, 2$ indicates twin sibling, 1 respective 2. $X_1$ and $X_2$ are vectors of explanatory variables and $\beta$ is parameters to estimate; $\alpha$ is the fixed effect and corresponds to fundamental health status, $FHS$, in the theoretical section. $\varepsilon_1$ and $\varepsilon_2$ are error terms. In the data $\{(H_{it}, X_{it}):\ t = 1, 2,\ i = 1, ..., n)\}$ is observed where $H_{it} = \max\{0, H_{it}^*\}$ .[5] $n$ refers to number of twin pairs. The difference between the twins is defined in terms of $\Delta X = X_1 - X_2$, and $\Delta H = H_1 - H_2$.

[Figure 1, about here]

Honoré (1992) explains the method graphically.[6] $\varepsilon_1$ and $\varepsilon_2$ in equation (5) are assumed to be independently and identically distributed conditionals of $(X_1, X_2, \alpha)$, which means that the $(H_1^*, H_2^*)$ conditional on $(X_1, X_2, \alpha)$ is symmetrically distributed around the $45^\circ$- line through $(X_1\beta, X_2\beta)$. This is the same as when the $(H_1^*, H_2^*)$ conditional on $(X_1, X_2, \alpha)$ is distributed around the $45^\circ$- line, $LL'$, through $(\Delta X\beta, 0)$. This is true for any $\alpha$ and the distribution of the $(H_1^*, H_2^*)$ conditional on $(X_1, X_2)$ is, accordingly, also distributed around

[5] Note that higher values of $H_t^*$ indicate a worse health status, and that the theoretic model was specified with a "positive" health measure. In either case, a too-large estimate in absolute terms compared with the true causal effect is expected if $FHS$ is omitted.

[6] Figure 1 reproduces the first figure in Honoré (1992) and shows the case where $\Delta X\beta \geq 0$ . Honoré (1992) also shows a second figure for the case where $\Delta X\beta < 0$.

the $45^{\circ}$ - line through $(\Delta X\beta, 0)$.

Honoré defines the orthogonality conditions that must hold under the true parameters for both the actual and the censored dependent variable. First, the probability that $(H_1^*, H_2^*)$ is included in $A(A = A_1 \cup A_2)$ is the same as the probability of $(H_1^*, H_2^*)$ being included in $B(B = B_1 \cup B_2)$. Secondly, the expected distance from $(H_1^*, H_2^*)$ in $A$ to its lower boundary is the same as the expected horizontal distance from $(H_1^*, H_2^*)$ in $B$ to its left boundary. These conditions finally suggest minimization of two objective functions to estimate the parameters, $\beta$. Further details, including necessary assumptions and proofs for consistency, can be found in Honoré (1992). It is worth noting, however, that the assumption that $\varepsilon_1$ and $\varepsilon_2$ in equation (5) are independently and identically distributed is not severely restrictive, since the fixed effect can capture some dependence. Honoré (2002) notes, for example, that the error terms can be "jointly normal, with equal variances but with arbitrary positive correlation." It is not necessary to assume that the disturbances have a particular parametric form. Further, the model does not rely on homoskedasticity across different pairs of twins, *i.e.* across $n$. This semiparametric fixed-effects tobit estimator can easily be used in empirical analysis since Honoré has made a module, Pantob, available for GAUSS. Results for this empirical application can be found in the next section.

# 5 Results

The data are separated in a sample with monozygotic female twins and another sample with dizygotic female twins. The focus is, of course, on the monozygotic sample, as these twins are identical with respect to their genes. While it may be necessary to estimate the model separately in this way, it should be noted that with a fairly high rate of censoring, problems in estimating the fixed-effects censored model could come up. In particular, the covariance matrix of the objective function's second derivative could turn out not of full rank.[7] Honoré's Pantob module for GAUSS provides the option to choose between an absolute value, a quadratic and a polynomial loss function. With the quadratic loss function, the estimation of the covariance matrix of the objective function's second derivative usually works well. The absolute loss function is, on the other hand, more robust to outliers. The polynomial loss function is a combination of the absolute value loss function and the quadratic loss function, where the parameter $\theta$ controls how the different functions are combined. If $\theta$ is set at zero, the quadratic function is used, while if $\theta \to \infty$, the absolute loss function is used. For all estimations in this study, the polynomial loss function is used with $\theta$ set to 3.[8] Numerical derivatives are used with a bandwidth set to 0.2.[9] (See the instructions for Pantob, written by Campbell & Honoré , 1991, for more details concerning the options).

[Table 2, about here]

Results for the female sample of monozygotic twins can be found in the first part of Table 2. When the logarithm of spousal disposable income in 1998 is included, none of the variables are found to have a significant effect. When the model is re-estimated with the logarithm of average spousal income, the dummy variable for post-secondary school and post-graduate education has an estimated

---

[7] This problem also restricts the possibility of including further control variables in this study. Note that the need to include more control variables is less when the twin fixed-effects method is used. Common variables such as age and gender are, of course, not necessary. Regressions are also made when a dummy for being married in 1998 and a variable measuring number of children in 1998 are included as additional control variables. For the monozygotic sample, the coefficients for these variables are never significantly different from zero. These estimates are available on request from the author.

[8] No qualitative difference is found if $\theta$ = 2, $\theta$ = 4, $\theta$ = 10 or $\theta$ = 100. If $\theta$ is set as low as $\theta$ = 0.5 the standard errors are, in general, very high and the coefficients are seldom significantly different from zero. For the dizygotic sample, the covariance matrix of the objective function's second derivative is not of full rank when $\theta$ = 10 or $\theta$ = 100 for the model with the disposable income in 1998 for the spouse. These results are available on request from the author.

[9] Changing the bandwidth does not affect the estimated coefficient, but its standard error. Estimates are also performed with the bandwidth set to 0.1 and 0.3. If the coefficient is significantly different from zero at another significance level when the bandwidth is set to 0.1 or 0.3 the difference is noted in the tables.

coefficient that is significantly different from zero at 1 percent significance level. The magnitude of the coefficient is similar to the first model and indicates a reduced share of total income that comes from sickness-related welfare systems.[10] Note that the coefficients presented refer to possible effects on the unobserved latent variable, $H^*$. No significant effect is found for the income of the spouse. Accordingly, when controlling for unobserved heterogeneity such as genetics and environmental factors shared by twins, no causal effect from spousal income is found.

These models are replicated for the female sample of dizygotic twins. Estimates are included in the lower part of Table 2. Interestingly, the effect of post-secondary school and post-graduate education is estimated to be almost five times the measure for the monozygotic sample.[11] Further, a negative effect is found for spousal income. The effect is significantly different from zero at 5 percent significance level. The differences between the results for the monozygotic sample and the dizygotic sample indicate that the estimates for dizygotic twins are still affected by biases. It seems that the differencing between dizygotic twins has not been able to remove all relevant unobserved heterogeneity. The genetic differences between the twins could, for example, have influenced the degree of education, and a "third variable" explanation could, accordingly, still be present for the more negative estimates.

All of the mentioned results have included spousal income linearly. The models are also estimated for the female sample of monozygotic twins with the addition of the squared income. These estimates are included in the Appendix in Table A3.[12] Upper secondary school is again found to have a significant negative

---

[10]Even though education would not affect health, an effect from higher education on the share of total income due to sickness-related welfare benefits is expected. This is due to the system of benefits that is based on current or earlier incomes. Since the benefits are granted only up to a certain level, education should affect the share through its effect on the current or earlier income. That is, more individuals with higher education should be present above the maximum level of benefits and thus have a lower share.

[11] A Wald test confirms a significant difference of the coefficients between the monozygotic and the dizygotic sample: ( $\chi^2$ (d.f.=1) = 110.57 > 3.84).

[12] Another way to allow for different effects over the income distribution is to multiply the income variable with dummy variables capturing different parts of the distribution. This is also done for this sample, and dummies for quintiles are multiplied with the measure of average

effect. This applies for both models. When the logarithm of the disposable spousal income is included squared, the coefficient of the logarithm of the disposable spousal income is found to be positive, although not significantly different from zero.[13]

When using the new twin-based method of dealing with unobserved heterogeneity, it is relevant to compare the results with results based on estimates in cross-section. Table A4 in the Appendix includes estimates for tobit models where the first twin in each pair is used as a cross-section of individuals. For the female sample of monozygotic twins, all three explanatory variables are estimated to be significantly different from zero with a negative sign. The dummy variables for the different levels of education are found to have substantially larger effects than did the ones noted earlier.[14]

In cross-section even the income of the spouse is found to be negative for both measures of income. Indeed, it seems that the twin method corrects for substantial biases due to unobserved heterogeneity. The estimates for the monozygotic sample can be seen as a lower bound-for-the-true (assumed negative) causal effect. Even though it is possible that twin differences in education continue to be affected by early childhood health differences, the method seems to tighten the possible causal effect. The results also underline the

---

spousal income in the period 1994-1999. The first variable is, accordingly, the logarithm of average spousal income for spouses with income in the first quintile. For all other cases, including when no spouse was present, the variable takes the value zero. Four more variables are constructed in the same way for quintile 2-5. In the regression, none of these variables is significantly different from zero.

[13] When the bandwidth is set to 0.1, the coefficient is significantly different from zero at the 10 percent significance level. Note that a positive coefficient does not necessarily mean a positive effect of income, since the effect from the squared term also should be included. A positive effect could, however, be present in theory if the female partner feels that she can afford a slower recovery from sickness. Note that if this is the case, a Tobit-based model would be too restrictive. The reason is that income is expected to have a negative effect on the probability of having a positive share and a positive effect on the mean share. In a Tobit model the effects, and the included explanatory variables, are assumed to be the same. A zero-inflated model would be an alternative, where the zeros consist of both "true" zeros (*i.e.* healthy individuals) and zeros due to sickness absence of less than 15 consecutive days. An econometrical zero-inflated model, with a continuous dependent variable for positive values and with fixed effects, is, to my knowledge, not yet developed.

[14] Wald tests confirm a significant difference of the coefficients between the fixed effects model and the cross-section model for the female sample of monozygotic twins. ($\chi^2$ (d.f.=1) = 111.02 > 3.84) and $\chi^2$ (d.f.=1) = 257.18 > 3.84). The tests concern the coefficients for the different levels

importance of focusing on monozygotic twins.

It would, of course, be interesting to know if a male sample would produce similar results. The available sample of male monozygotic twins does, however, contain several complications. First, it is only 196 of the 1400 twins (14%) who took sickness related benefits in 1999. Accordingly, the censoring rate is very high. Secondly, it turned out that "who had a spouse identified in 1998" was an important selection: 104 of the 196 with sickness-related benefits (53%) did not have a spouse present in 1998. The average total income of the twins without a spouse present is significantly lower compared with the average total income for twins with a spouse present. (2-group Hotelling's T-squared, F(1,1394) = 61.7012). No such difference is found for the female monozygotic sample.

Finally, within-twin difference of total income is positively correlated with within-spousal differences in income. The correlation coefficients for both measures of income are around 0.12 and significantly different from zero at 1% significance level. If the correlation coefficient is calculated for the subsample where both twin siblings had a spouse identified in 1998, the within-twin difference of total income is found to be *negatively* correlated (-0.1673, significant at 1% significance level) with the within-spouse difference of disposable income in 1998. Also, this subsample included only 58 twin siblings with a difference in the share of sickness-related benefits. Reliable estimates are not expected under these conditions, and the male sample is, accordingly, left aside in this study. For the male sample it would be crucial to model the selection in terms of partnership.

Fritjers *et al*. (2005) used the German reunification to study the causal effect of income on health. The idea is that the fall of the Berlin Wall was unanticipated, and resulted, due to collectively bargained wages, in substantially increased wages in East Germany. Fritjers *et al*. (2005) found a small but significant positive effect of income on health satisfaction. Lindahl (2005) used lottery prizes as an exogenous source of variation in income to estimate the

---

of education in the model with spousal disposable income in 1998.

causal effect of disposable income on health for Swedish data. The measure of health is constructed as a standardized index of bad health based on questions about health symptoms. For the sample of players the coefficient of income, using lottery prizes as instruments, is not significantly different from zero. Its magnitude is, however, similar to its estimate using OLS for the same sample and Lindahl (2005) concludes that OLS "probably gives quite accurate estimates of the causal effect." For the full sample, a 10 percent increase in income is, accordingly, found to improve health with 0.01-0.02 standard deviations.

Zimmerman & Katon (2005) estimated several different models to reveal a possible causal relationship between income and depression in the US. When using a fixed-effects logit and negative binomial regression, the coefficient for the logarithm of income is not significantly different from zero. They conclude that "the often-observed relationship between income and depression is in fact a result of other observed and unobserved variables." The coefficient for financial strain, measured as the logarithm of debts-to-assets ratio, is, in some cases, found to be significantly different from zero. Zimmerman & Katon (2005) also used instruments, such as log of total inheritances and log of time in the current job to estimate an IV probit and negative binomial IV regression. When IV methods are used, the coefficient for financial strain is not found to be significantly different from zero. Zimmerman & Katon (2005) suggest that the instruments could be performing poorly, and that it is possible that a causal effect could still exist.

As mentioned earlier, Case *et al*. (2005) found significant effects of early childhood health on adult health. That result underlines the importance of controlling for unobserved heterogeneity in studies of socioeconomic status and health situation. When variation over time is used to control for unobserved heterogeneity, such as the fixed-effects models used in Zimmerman & Katon (2005) and Michaud & van Soest (2004), the effect of income/wealth on health is not present. It is, however, possible that the within-individual variation in income/wealth over time includes very little signal, and a substantial part of transitory variation is unlikely to influence health**.** When using variation over time to control for unobserved heterogeneity, it is necessary to be careful about the dynamics of health and income, especially if the panel only covers a few years. It is, for example, difficult to know when, and for how long, a possible effect of income on health is present. Even though a low income during the first year in the panel does in fact appear to cause health problems, we do not, in general, know if these problems will be long-lasting. If the health problems in each of the following years reduce the possibility of earning a high income, the identifying variation over time in income would be rather small. It is difficult also to know if the possible health problems occur in the year that follows, or if the symptoms manifest only several years later. Under these circumstances, a fixed-effects model based on variation over time would give a rather blurred picture of the relationship.

This study presents an alternative way to handle unobserved heterogeneity. The method takes advantage of variation between twins. Since monozygotic twins have identical genes and very similar social background, a substantial part of the unobserved heterogeneity is taken into account. For the female sample used, no significant effect is found for spousal income. Post-secondary school and post-graduate education is found to reduce the share of total income that comes from sickness-related welfare systems.

# 6 Concluding remarks

The purpose of this study is to empirically investigate whether socioeconomic status causally influences the share of the total income that comes from sickness related benefits. In studies of health and socioeconomic status it is important to be aware of the possibility that unobserved heterogeneity can influence both health and socioeconomic status. In other words, it is possible that a correlation can be mistaken for a causal effect. It is also important to deal with the possibility of reverse causality, where the health status influences the socioeconomic status. In this study, a new twin-based method is introduced to deal with these two issues. Using monozygotic twins, the method deals with unobserved heterogeneity, and important background factors that can possibly affect both socioeconomic status and health are controlled for. The method provides an important alternative to using time variation to control for unobserved heterogeneity. Spousal income is used as a non-shared environmental influence, and the risk for reverse causality should therefore be rather small. In this application a censored fixed-effects model is estimated, since the dependent variable is censored.

In the empirical analysis, level of education is found to affect the share of income due to sickness-related welfare systems. Estimates for the sample of dizygotic twins indicate, for post-secondary school and post-graduate education, an effect about five times of that found for the monozygotic sample. Since dizygotic twins are not genetically more alike than ordinary siblings, it seems that this estimate is still biased due to unobserved heterogeneity. It is possible that differences in early childhood health status between monozygotic twins have influenced the within twin-pairs difference in education level, even though education usually takes place early in life. In that case, the estimates for the monozygotic sample are still biased, but they provide a lower bound of the assumed negative causal effect. Intrauterine environmental effects on adult health are found in Case *et al*. (2005). Since even monozygotic twins can differ in birth weight, it seems wise to view the estimates as lower bounds.

As mentioned, in this study the income of the spouse is used as a non-shared environmental influence. For the female sample, spousal income was found to have no causal effect on the share of total income that comes from sickness-related welfare systems. This suggests that if it exists, a causal effect that works from income to health (through changes in consumption behavior, risk-taking or access to health care) is not strong enough to be visible in long-term sickness absence. Trivially significant, or non-significant, effects have been found in earlier studies that intend to measure the causal effect of income on health. (See Fritjers *et al*. (2005), Lindahl (2005), Zimmerman & Katon (2005)). Note, however, that this study estimates the possible effect of spousal income. It is possible that “own” earnings causally influence health through, for example, psychological reasons and/or work-related environmental effects. These possible mechanisms are not included in this study and should be investigated with instrument variable techniques.

Note also that the incomes from sickness-related welfare systems refer to long-term absence from work. Different effects can be present for short-term absence from work. It would also be possible to model differently the cases with no sickness-related benefits. A zero-inflated model would be a possible avenue, assuming that the cases with no sickness-related benefits includes both healthy individual and individuals who, while they have been sick, were never sick for more than 14 consecutive days. Another issue concerning the specification is how income is included in the model. Adding squared income can only capture some specific nonlinearity. It is, accordingly, still possible that specification bias exists, and an option would be to include income in the model non-parametrically.

The male sample of monozygotic twins is not used in this study. The reason is that an important selection into partnership was found for the male sample. Descriptive statistics revealed that a large share of the individuals with sickness-related benefits in 1999 did, in fact, not have an identified spouse in 1998. To study only twins who have a spouse does not seem correct, for then one would be examining a non-random subsample. The process of finding a spouse should be modeled to separate a possible effect of having a spouse from a possible effect of having an extra income in the household.

The twins in this study were between 41 and 50 years old 1999. It is, of course, possible that different results are found for different ages. It would, for example, be interesting to study an older population that has not reached the age of retirement: It is expected that more people in such a sample would be found to have sickness-related benefits, and a lower degree of censoring would make estimations easier. It would be interesting also to see an application of the twin method on measures of health status that are not censored.

While the twin method is an attractive path for future research on the causal effects of income on health, several complications are still present. In particular, it would be interesting to include the “within” difference in the twins’ earnings in the model. In that case the causality test would me more general with respect to the possible channels in which income can affect health. The possible endogeneity has, however, to be handled properly. Further, if spouses are used as non-shared environmental influences, the process of finding a spouse should be modeled. It is important also to remember that the value of twin studies depends on the results being informative for the general population and not just for twins.

# 7 References

Adams, P., Hurd, M., McFadden, D., Merrill, A., Ribeiro, T., 2003, Healthy, Wealthy and Wise? Tests for Direct Causal Paths Between Health and Socioceconomic Status. Journal of Econometrics 112, 3-56.

Adler, N. E., Boyce, T., Chesney, M. A., Cohen, S., Folkman, S., Kahn, R. L., Syme, S. L., 1994, Socioeconomic Status and Health: The Challenge of the Gradient. American Psychologist 49(1), 15-24.

Auld, M. C., Sidhu, N., 2005, Schooling, cognitive ability and health. Health Economics 14, 1019-1034.

Bakgrundsfakta till arbetsmarknads- och utbildningsstatistiken, 2002:2. Statistics Sweden.

Buckley, N. J., Denton, F. T., Robb, A. L., Spencer, B. G., 2004, The Transition from Good to Poor Health: an Econometric Study of the Older Population. Journal of Health Economics 23, 1013-1034.

Case, A., Lubotsky, D., Paxson, C., 2002, Economic Status and Health in Childhood: The Origins of the Gradient. American Economic Review 92(5), 1308-1334.

Case, A., Fertig, A., Paxson, C., 2005, The Lasting Impact of Childhood Health and Circumstances. Journal of Health Economics 24, 365-389.

Contoyannis, P., Jones, A. M., 2004, Socio-economic Status, Health and Lifestile. Journal of Health Economics 23, 965-995.

Deaton, A., 2002, Policy Implications of the Gradient of Health and Wealth. Health Affairs 21(2), 13-30.

Frijters, P., Haisken-DeNew, J. P., Shields, M.A., 2005, The causal effect of income on Health: evidence from German reunification. Journal of Health Economics 24, 997-1017.

Fritzell, J., Nermo, M., Lundberg, O., 2004, The impact of income: assessing the relationship between income and health in Sweden. Scandinavian Journal of Public Health 32, 6-16.

Fuchs, V. R., 2004, Reflections on the Socio-economic Correlates of Health. Journal of Health Economics 23, 653-661.

Godfrey, K. M., Barker, D. JP., 2000, Fetal nutrition and adult disease. American Journal of Clinical Nutrition 71(5), 1344-1352.

Honoré , B. E., 1992, Trimmed Lad and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects. Econometrica 60(3), 533-565.

Lindahl, M., 2005, Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as Exogenous Source of Variation in Income. Journal of Human Resources 40(1), 144-168.

Lykken, D. T., Tellegen, A., 1993, Is Human Mating Adventitious or the Result of Lawful Choice? A Twin Study of Mate Selection. Journal of Personality and Social Psychology 65(1), 56-68.

McKinnish, T., 2006, Panel Data Models and Transitory Fluctuations in the Explanatory Variable. In Millimet, D., Smith, J., Vytlacil, E., Advances in Econometrics, Vol. 21. (Forthcoming).

Michaud, P. C., Soest, A. H. O. van, 2004, Health and Wealth of Elderly Couples: Causality Tests Using Dynamic Panel Data Models. CentER Discussion Paper, No. 2004-81.

Neumark, D., 1999, Biases in Twin Estimates of the Return to Schooling. Economics of Education Review 18, 143-148.

Singh-Manoux, A., Ferrie, J. E., Lynch, J. W., Marmot, M., 2005, The role of Cognitive Ability (Intelligence) in Explaining the Association between Socioeconomic Position and Health: Evidence for the Whitehall II Prospective Cohort Study. American Journal of Epidemiology 161(9), 831-839.

Sloan, R. P., Huang, M-H., Sidney, S., Liu, K., Williams, O. D., Seeman, T., 2005, Socioeconomic status and health: is parasympathetic nervous system activity an intervening mechanism? International Journal of Epidemiology 34, 309-315.

Smith, J. P., 1999, Healthy Bodies and Thick Wallets: The Dual Relation Between Health and Economic Status. Journal of Economic Perspectives 13(2), 145-166.

Spotts, E. L., Neiderhiser, J. M., Caniban, J., Reiss, D., Lichtenstein, P., Hansson, K., Cederblad, M., Pedersen, N. L., 2004, Accounting for Depressive Symptoms in Women: a Twin Study of Associations with Interpersonal Relationships. Journal of Affective Disorders 82, 101-111.

Wilkinson, R. G., 1996, Unhealty Societies: The Afflication of Inequality. Routledge, London.

Zimmerman, F. J., Katon, W., 2005, Socioeconomic status, depression disparities, and financial strain: what lies behind the income-depression relationship? Health Economics 14, 1197-1215.

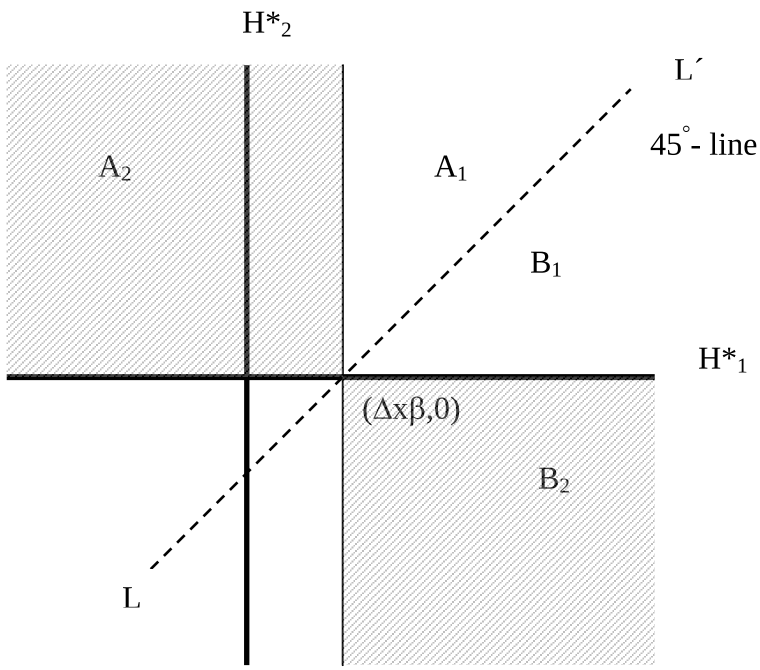


Figure 1. Symmetry conditions used by Pantob. Source: Honoré (1992).

Table 1 Summary Statistics

| Variable | Monozygotic Female Mean | Monozygotic Female Std. Dev. | Dizygotic Female Mean | Dizygotic Female Std. Dev. |
|---|---|---|---|---|
| Share of income due to sickness-related welfare systems | 8.928 | 24.5944 | 8.715 | 24.6997 |
| Share of income due to sickness-related welfare systems given positive share observed | 39.257 | 38.3567 | 39.857 | 39.3803 |
| Compulsary school <= 10 years, 1998 | 0.163 | 0.3693 | 0.175 | 0.3797 |
| Upper secondary school, 1998 | 0.499 | 0.5001 | 0.495 | 0.5001 |
| Post-secondary school and post-graduate education, 1998 | 0.338 | 0.4733 | 0.330 | 0.4703 |
| Married, 1998 | 0.6057 | 0.4888 | 0.5843 | 0.4929 |
| Number of children, 1998 | 1.420 | 1.1057 | 1.4064 | 1.1414 |
| Region, ”Götaland”, 1998 | 0.483 | 0.4999 | 0.4860 | 0.4999 |
| Region, ”Norrland”, 1998 | 0.135 | 0.3417 | 0.1576 | 0.3645 |
| Region, ”Svealand”, 1998 | 0.382 | 0.4859 | 0.3564 | 0.4790 |
| Spouse, ln (disposable income 1998) | 12.044 | 0.8894 | 11.971 | 1.0778 |
| Spouse, ln (average disposable income 1994-1998) | 10.404 | 4.2006 | 10.47 | 4.0635 |
| Number of individuals: | 1750 | | 2360 | |
| Number (percent) of individuals with positive share | 398 (22.7) | | 516 (21.9) | |
| Number (percent) of individuals without spouse 1998 | 585 (33.4) | | 777 (32.9) | |
| Number (percent) of individuals with positive share and without spouse 1998 | 166 (41.7) | | 229 (29.5) | |

*Notes*: summary statistics for spousal disposable income in 1998 are calculated for the subsample of identified spouses in 1998. The average disposable income is calculated as the sum of incomes of *present* spouses during the years 1994-1998 divided by five. The mean in the table refers to a mean for the subsample where a spouse was identified during at least one of the years 1994-1998. The division into three large regions is made on the basis of counties. These large regions bear the following codes of the regional classification NUTS 3, used by the European Community: “Götaland”; SE023, SE041, SE044, SE0A1, SE0A2, SE091-094 “Norrland”; SE063, SE071, SE072, SE081, SE082. “Svealand”; SE010, SE021, SE022, SE024, SE025, SE061, SE062.

Table 2. Female samples, semiparametric censored fixed effects model

| Dependent variable: | Share of total income from sickness and early retirement benefits | | | | | |
|---|---|---|---|---|---|---|
| Monozygotic sample | Female | | | Female | | |
| Variable, measured in 1998 | Beta | std err | t-ratio | beta | std err | t-ratio |
| Upper secondary school, 1998 | -0.6931 | 2.587 | -0.27 | -0.7609 | 1.213 | -0.63 |
| Post-secondary school and post-graduate education, 1998 | -5.070 | 3.829 | -1.32 | -4.900***[a] | 1.336 | -3.67 |
| ln (spouse disposable income, 1998) | 0.0075 | 0.187 | 0.04 | | | |
| ln (spouse average inc.,1994-1998) | | | | -0.0113 | 0.173 | -0.07 |
| | Chi$^2$: 2.7 (p = 43.5 %) | | | Chi$^2$: 14.0 (p = 0.2 %) | | |
| Dizygotic sample | Female | | | Female | | |
| Variable | Beta | std err | t-ratio | Beta | std err | t-ratio |
| Upper secondary school, | -4.073[b] | 3.287 | -1.24 | -1.889 | 2.332 | -0.81 |
| Post-secondary school and post-graduate education | -24.97*** | 2.640 | -9.46 | -23.82*** | 1.917 | -12.43 |
| ln (spouse disposable income) | -1.424**[c] | 0.617 | -2.31 | | | |
| ln (spouse average inc.,1994-1998) | | | | -1.350**[d] | 0.659 | -2.050 |
| | Chi$^2$: 253.2 (p = 0.0 %) | | | Chi$^2$: 486.6 (p = 0.0 %) | | |

*Notes*: Coefficients that are significantly different from zero at 10, 5 and 1 percent levels are marked with *, ** and ***. a) When the bandwidth is set to 0.1 the coefficient is significantly different from zero at the 5 percent level. b) When the bandwidth is set to 0.3 the coefficient is significantly different from zero at the 10 percent level. c) When the bandwidth is set to 0.1 (0.3) the coefficient is significantly different from zero at the 10 (1) percent level. d) When the bandwidth is set to 0.1 or 0.3 the coefficient is significantly different from zero at the 1 percent level.

## Appendix

Table A1. Correlation, variables are measured as average for female monozygotic twin siblings.

| Monozygotic sample, | Share of sickness related benefits | ln(spouse disposable income, 1998) | ln (spouse average income, 1994-1998) |
|---|---|---|---|
| Share of sickness related benefits, 1999 | 1.0000 | | |
| ln (spouse disposable income, 1998) | -0.1397*** | 1.0000 | |
| | 0.0000 | | |
| ln (spouse average income,1994-1998) | -0.1262*** | 0.9481*** | 1.0000 |
| | 0.0002 | 0.0000 | |
| ln (total income, 1998) | -0.3327*** | 0.0338 | 0.0283 |
| | 0.0000 | 0.3203 | 0.4058 |
| Number of children, 1998 | -0.1412*** | 0.4446*** | 0.4679*** |
| | 0.0000 | 0.0000 | 0.0000 |
| Married, 1998 | -0.1269*** | 0.8277*** | 0.8151*** |
| | 0.0002 | 0.0000 | 0.0000 |
| Region, ”Götaland”, 1998 | 0.0148 | 0.0435 | 0.0360 |
| | 0.6610 | 0.1991 | 0.2871 |
| Region, ”Norrland”, 1998 | 0.0144 | 0.0719** | 0.0677** |
| | 0.6713 | 0.0336 | 0.0453 |
| Region, ”Svealand”, 1998 | -0.0256 | -0.0960*** | -0.0853** |
| | 0.4495 | 0.0045 | 0.0116 |
| Post-secondary school and post-graduate education, 1998 | -0.1548*** | 0.0803** | 0.0917*** |
| | 0.0000 | 0.0175 | 0.0066 |
| Upper secondary school, 1998 | 0.0182 | 0.0009 | -0.0159 |
| | 0.5898 | 0.9800 | 0.6396 |
| Compulsory school, 1998 | 0.1780*** | -0.1058*** | -0.0989*** |
| | 0.0000 | 0.0017 | 0.0034 |

*Notes*: The significance level of each correlation coefficient is included on the second row for each variable. Statistical significant correlation at 10, 5 and 1 percent levels are marked with *, ** and ***.

Table A2. Correlation, variables are measured as differences between female monozygotic twins

| Monozygotic sample, | Share of sickness related benefits | ln(spouse disposable income, 1998) | ln (spouse average income, 1994-1998) |
|---|---|---|---|
| Share of sickness related benefits, 1999 | 1.0000 | | |
| ln (spouse disposable income, 1998) | -0.037 | 1.0000 | |
| | 0.2746 | | |
| ln (spouse average income,1994-1998) | -0.0542 | 0.9161*** | 1.0000 |
| | 0.1094 | 0.000 | |
| ln (total income, 1998) | -0.1587*** | -0.0422 | -0.0389 |
| | 0.000 | 0.2146 | 0.2522 |
| Number of children, 1998 | 0.0275 | 0.3465*** | 0.3574*** |
| | 0.4174 | 0.000 | 0.000 |
| Married, 1998 | -0.0502 | 0.7919*** | 0.7653*** |
| | 0.1379 | 0.000 | 0.000 |
| Region, ”Götaland”, 1998 | 0.016 | 0.0246 | 0.0124 |
| | 0.6364 | 0.468 | 0.7131 |
| Region, ”Norrland”, 1998 | -0.0056 | 0.0069 | -0.0102 |
| | 0.8683 | 0.8379 | 0.7622 |
| Region, ”Svealand”, 1998 | -0.011 | -0.0275 | -0.0045 |
| | 0.7459 | 0.4166 | 0.8941 |
| Post-secondary school and post-graduate education, 1998 | -0.0148 | -0.0045 | -0.0128 |
| | 0.6611 | 0.8952 | 0.7055 |
| Upper secondary school, 1998 | -0.0276 | -0.0517 | -0.0406 |
| | 0.4141 | 0.1268 | 0.2307 |
| Compulsory school, 1998 | 0.0596* | 0.0834** | 0.0766** |
| | 0.0781 | 0.0136 | 0.0234 |

*Notes*: The significance level of each correlation coefficient is included on the second row for each variable. Statistical significant correlation at 10, 5 and 1 percent levels are marked with *, ** and ***.

Table A3. Female monozygotic sample, semiparametric censored FE model

| Dependent variable: | Share of total income from sickness and early retirement benefits | | | | | |
|---|---|---|---|---|---|---|
| Monozygotic sample | Female | | | Female | | |
| Variable | beta | std err | t-ratio | Beta | std err | t-ratio |
| Upper secondary school, 1998 | -0.566 | 1.595 | -0.36 | -1.091 | 1.688 | -0.65 |
| Post-secondary school and post-graduate education, 1998 | -4.969*** | 1.158 | -4.29 | -5.099*** | 1.394 | -3.66 |
| ln (spouse disposable income, 1998) | 1.619 | 2.043 | 0.79 | | | |
| ln (spouse disposable income, 1998) – squared | -0.1346 | 0.640 | -0.21 | | | |
| ln (spouse average inc.,1994-1998) | | | | 0.7822[a] | 0.575 | 1.36 |
| ln (spouse average inc.,1994-1998) – squared | | | | -0.0674 | 0.206 | -0.33 |
| | $Chi^2$: 85.5 (p = 0.0 %) | | | $Chi^2$: 32.3 (p = 0.0 %) | | |

*Notes*: Coefficients that are significantly different from zero at 10, 5 and 1 percent levels are marked with *, ** and ***. a) When the bandwidth is set to 0.1 the coefficient is positive and significantly different from zero at the 10 percent level.

Table A4. Female monozygotic sample, tobit cross-section model

| Dependent variable: | Share of total income from sickness and early retirement benefits | | | | | |
|---|---|---|---|---|---|---|
| Monozygotic sample | Female | | | Female | | |
| Variable, measured in 1998. | Beta | std err | t-ratio | Beta | std err | t-ratio |
| Upper secondary school | -23.0499*** | 8.373 | -2.75 | -23.2056*** | 8.365 | -2.77 |
| Post-secondary school and post-graduate education | -41.0228*** | 9.420 | -4.35 | -40.8692*** | 9.410 | -4.34 |
| ln (spouse disp. income, 1998) | -1.5472*** | 0.545 | -2.84 | | | |
| ln (spouse average income, 1994-1998) | | | | -1.6410*** | 0.539 | -3.05 |
| Constant | -14.3113* | 8.330 | -1.72 | -14.013* | 8.244 | -1.70 |
| | LR $chi^2$: 29.64 (p = 0.0%) | | | LR $chi^2$: 30.88 (p = 0.0 %) | | |
| | Pseudo R2 = 0.0100 | | | Pseudo R2 = 0.0104 | | |
| | Log likelihood = -1474.3835 | | | Log likelihood = -1473.7652 | | |

*Notes*: Coefficients that are significantly different from zero at 10, 5 and 1 percent levels are marked with *, ** and ***.