# The relationship between income and health: using information on separation allowances to single out causality from correlation

Ann Lecluyse and Tom Van Ourti

Work in progress – August 2005

**Introduction**

Economists agree about the existence of a positive association between health and income. Less clear, however, is the direction of causation. First, it is plausible that income affects health. Second, the reverse is possible. And finally, third factors may influence both health and income. In the underlying paper, we focus on the first mechanism.

Recently, there have been several attempts to deal with the problem of endogeneity. Ettner (1996) examines the effect on different health proxies, such as self-assessed health, daily activity limitations, proxies for alcohol abuse, an 84-point scale of depressive symptoms and others. For the estimation of income effects on health, she draws on two-stage instrumental variable estimation applied to cross-sectional data. Depending on the health outcome, she uses ordered probit, dichotomous probit or two-parts models. Instruments for family income include state unemployment rate, work experience, parental education and spouse characteristics. As the null-hypothesis that income is exogenous is rejected, the IV coefficients are more appropriate than the ordinary estimates, subject to the validity of instruments. In each case, she finds that income remains significant and that its magnitude increases after instrumenting. The latter is rather surprising and the author suggests that this can happen due to measurement error in the income variable.

Meer et al. (2003) also utilize instrumental variables estimation to deal with this, more specifically they draw on the two stage probit model. They work with a five-period interval because only in these waves detailed information about their instrument is available. As dependent variable they use a dichotomous variable derived from self-assessed health. It has a value of 1 for excellent, very good and good values of health and zero otherwise. As income variable they use change in wealth, which is instrumented by the amount of inheritances and gifts (amount larger than US $10,000). Their instrument is strongly correlated with the change in wealth, moreover when it is entered in the health equation, it is not significant. In their results the authors find that the change in wealth becomes insignificant when taking endogeneity into account. These results are robust to the health variable (an alternative was the presence or absence of physical or nervous disability), to the sample used (e.g. balanced or unbalanced) and to some alternative formulations of the model.

An alternative instrument is used by Lindahl (2005). He uses monetary lottery prizes which is a truly exogenous variation in income. A "standardized index of bad health" out of 48 health symptoms is constructed and used as dependent variable, next to this measure, he also uses the health variables separately or in groups and two mortality indicators. The methods to estimate the models are poisson, ordered probit or OLS depending on the nature of the variables. Lindahl only works with a subsample of lottery players because they have other observable characteristics than non-players. When the total sample is used for OLS models, the estimates are much smaller than those of the subsample of players, so extrapolation has to be done with caution. As in Ettner (1996), he finds that the coefficients of IV estimates are higher, but not significanlty, than those of OLS. However, the author doubts that there is measurement error in the income variable as he used information of the tax register. He suggests possible errors in the measurement of control variables. Another result is that there is only a small effect on the coefficients when temporary or permanent income measures are used.

Another possible instrument which, to our knowledge, has previously not been used, is the amount of separation allowances people have to pay when they get divorced. When couples in Belgium get divorced[1], they can be obliged to pay separation allowances and/or maintenance for the ex-partner. These allowances should be used for the housing, living, supervision, raising and education of the children. The ex-partner should be able to proceed the standard of living as within the marriage. According to the law, the

[1] This paragraph is based on two internet sources: www.houvast.be and www.goudi.be.

amount to be paid has to be in proportion to the resources of the parents. Unfortunately, the meaning of "…in proportion to…" is not defined. It follows that in practice the judge decides about the amount of separation allowances (sometimes the ex-partners agree among each other, but this rarely occurs although it is allowed by the law). The earnings of the parents are expressed as percentages of the total earnings, and the same is done for the visiting rights in order to justify the amounts as much as possible. The amount to pay or receive is thus for a large part based on the income difference between the partners. This allowance has to be paid as long as the child has not finished education and at least till it attains the age of majority. As such the amount of separation allowances is expected to be uncorrelated with current health and can be used as an instrument for income.

In the subsample of individuals who have the right to receive or who have to pay allowance, the correlation between the real received separation allowance and SAH (remark: SAH is measured in 5 categories) is 0.021 and the correlation between real paid separation allowance and SAH is -0.0588.

**Data**

In our analysis to estimate the impact of income on health, we use a subsample of the longitudinal Panel Survey of Belgian Households (1994-2002) which only includes people that have the right to receive or have to pay separation allowances, and keep 2,688 observations. This is only a small part of the dataset, which has 55,690 observations over all the periods.

We use different health proxies as dependent variable: self-assessed health (5 categories varying from very bad to very good health), chronic illness (1/0 variable), natural logarithm of the body mass index, physical symptoms (5 categories, 1=never; 5=often) and various mental health variables (each has 5 categories varying from never to often) such as feeling depressed, strange thoughts, not able to concentrate, etc.

Independents are: the logarithm of individual income[2] (in 1988 prices), interaction between age and sex, size of the household, education, region (Flanders, Wallonia and Brussels) and time dummies.

We use the quadratic form of paid (paid and paid²) and received (received and received²) separation allowances, also in 1988 prices, as instrumental variable in order to correct for endogeneity of income.

In Table 3, we give some descriptive statistics of the variables.

**Methods and results**

The properties of the dependent variable determines which kind of regression we will apply. For the self-assessed health, we draw on an interval regression with as external thresholds the Flemish EQ-5D (values: 0, 0.1354, 0.5356, 0.7408, 0.9089 and 1). The natural logarithm of the body mass index is continuous and (approximately) symmetrically distributed, so we use OLS. For the dummy variables we apply a probit regression and finally for the categorical variables we use an ordered probit approach. In each case we allow for clustering on the individual level in the statistical inference.

First, we estimate the conventional regressions without taking into account possible endogeneity of income. The results are presented in column 1 of table1. They show that the coefficient of income is significant at the 10% level in the cases of illness or injuries (ziekte_verwond), the logarithm of the body mass index, easy crying (WA589) and having strange thoughts (WA592). Remarkably, higher individual income is associated with a larger probability on having to reduce daily activities due to sickness or injuries (ziekte_verwond) and it is positively related to body mass index as well. For the other variables, the sign is negative (except in the case of being irritable (WA593) and having problems to sleep (WA583)), however it is not significantly different from zero. In general we can conclude that there is a weak positive relationship between income and health status. However, this may be an overstatement as income can be endogenous. Therefore, we re-estimate the model using IV.

[2] Income data is taken from the ECHP. As the data do not exist for the last wave, we drop this wave and perform the analysis for 8 years only. The full dataset now contains 50,329 observations and the subsample has 2,374 observations.

To check the validity of the instruments, we first check if they are partially correlated with income: in table 2, we see that this is the case by regressing income on the instruments. The coefficients of paid allowance (paid and paid²) are significant, contrary to those of received allowance (received and received²). However, they are jointly significant and as such the first check of partial correlation is satisfied. Second, we estimate the reduced form model (include the instruments in the main regression instead of income). The results can be found in the column with headnumber 3 of table 1. Here, we learn that for emotional or mental problems (emo_geest), the natural logarithm of the body mass index, not able to concentrate (WA587), being irritable (WA593), having need for affirmation (WA594) and physical symptoms (WA596) the coefficients of the instruments are jointly significant while this is not the case for all other variables. Third, we include both the income variable and the instruments in the main regression (columns numbered 4 in table 1). It should be the case that income is significant and the instruments are not significant. This is the case for only two health variables: illness and injuries (ziek_verwond) and easy crying (WA589).

Now we can test the null-hypothesis of exogenous income. Therefore, we use the Rivers-Vuong approach except for the case of the logarithm of body mass index were we applied two stages least squares. The Rivers Vuong method consists of estimating the auxiliary equation (income on all exogenous variables and instruments) and adding the predicted errors of this regression in the main equation (health on income, exogenous variables). The main assumption of this approach is that we have a good instrument, which we have tested in the 3 previous steps. When these predicted errors are significant, it means that we have to deal with endogeneity (Wooldridge, 2002). In our results, shown in the last two columns of table 1, we first see that the coefficients of the income variable are much larger in absolute value in all cases. But, there are 8 health variables for which the error is significantly different from 0: self assessed health (sah), illness or injuries (ziek_verwond), not able to concentrate (WA587), having strange thoughts (WA592), being irretable (WA593), need for affirmation (WA594), feel not well (WA595) and physical symptoms (WA596). Taking into account all the results, we can conclude that either there is no reverse causality (*maybe measurement error*)

As we find no evidence for endogeneity (*maybe we should be more prudent given measurement error etc.*), we now estimate similar models on the total sample (results: see table with “pooled static model”) and include interaction terms between income and marital status to allow for a different income effect according to marital status. We find that the income effect is smaller than in the IV specification using the Rivers-Vuong approach, although it is significant in half of the cases. A second finding is that the interaction terms are jointly significant for all health variables (except for physical symptoms). Third, the differentiated income effect of the several marital status possibilities is in two thirds of the cases larger in absolute values for the interaction between income and being devorced or separated (burstink23).

It has been shown that a modeling strategy that allows for health dynamics is more realistic. Hurd and Kapteyn (2003) exploit differences in the institutional structure between the US and the Netherlands to analyze the dynamics of the income-health gradient. We follow another route, i.e. we use a Mundlak approach to allow for dynamics (see also Contoyannis et al. (2004)). We include the lagged and initial values of the dependent health variable and the time-averages of the other covariates (except of the time dummies). The results show that the impact of income and the interactionterm between income and marital status is lower and less significant than in the case of the static model. Second, approximately the same small effects are found for the mean values of the interaction terms between income and marital status. Third, and remarkably, the sign of the mean values of individual income is in all but five cases the opposite of the sign of transitory income. Further, the absolute values of the mean income variable are in most cases larger.

**Conclusions**

Our IV approach seems to give results that are similar to the existing literature, i.e. IV-income coefficients are larger than OLS coefficients. This is probably due to the fact that our instrument is not a perfect one. Separation allowances are arbitrary set by a judge with the only guideline that it has to be in line with the income *difference* between the partners. As such it is indeed uncorrelated with the health status, however it is also rather uncorrelated with the level of income of the individuals.

We could redo this exercise on different subgroups: young versus old and employer versus employee: see if there is a different impact.

Second: we could try inheritage as an instrument. However, we know that it is not a perfect instrument as sell of a house is also included in the amount.

**References**


Contoyannis P., Jones A.M. and Rice N. (2004), The dynamics of health in the British Household Panel Survey, *Journal of Applied Econometrics*, 19 (4), pp.473-503

Ettner S.L. (1996), New evidence on the relationship between income and health, *Journal of Health Economics*, 15, pp.67-85

Hurd M. and Kapteyn A. (2003), Health, wealth and the role of institutions, *Journal of Human Resources*, 38 (2), pp.386-415

Lindahl M. (2005), Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income, *Journal of Human Resources*, 40 (1), forthcoming.

Meer J., Miller, D.L. and Rosen, H.S. (2003), Exploring the health-wealth nexus, *Journal of Health Economics*, 22, pp.713-730

Wooldridge J.M. (2002), *Econometric analysis of cross section and panel data*, Cambridge: Mass. MIT Press

Table 1

| | 1 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | income | paid | paid$^2$ | received | received$^2$ | P joint sign | income | paid | paid$^2$ | received | received$^2$ | P joint sign | income | Residuals |
| sah | 0.008 | 0.026* | -0.002* | 0.015 | -0.001 | 0.2583 | 0.005 | 0.032+ | -0.002* | 0.01 | -0.00015 | 0.0777 | 0.075+ | **-0.070+** |
| ziek_verwond | **0.020+** | 0.101+ | -0.071* | 0.038 | -0.017 | 0.2148 | **0.019*** | 0.053 | -0.051 | 0.036 | -0.022 | **0.4373** | -0.096 | **0.121+** |
| emo_geest | -0.002 | 0.026 | -0.016 | -0.005 | -0.013 | **0.079** | -0.001 | 0.023 | -0.012 | 0.01 | -0.017 | 0.258 | -0.036 | 0.035 |
| ln(bmi) | **0.018*** | 0.032+ | -0.001 | -0.047+ | 0.011** | **0** | 0.017+ | 0.018 | 0 | -0.048+ | 0.010** | 0 | 0.042 | |
| W*A581 | -0.014 | -0.18 | 0.01 | 0.067 | -0.013 | 0.5573 | 0.005 | -0.199 | 0.012 | 0.024 | -0.014 | 0.2117 | -0.431 | 0.435 |
| W*A582 | -0.059 | -0.023 | 0.003 | -0.147 | 0.016 | 0.795 | -0.058 | -0.002 | 0.001 | -0.114 | 0.015 | 0.953 | -0.081 | 0.023 |
| W*A583 | 0.044 | -0.018 | -0.001 | -0.047 | -0.002 | 0.4775 | 0.059 | -0.107 | 0.004 | -0.066 | -0.005 | 0.1837 | -0.278 | 0.336 |
| W*A584 | -0.076 | -0.151 | 0.01 | -0.007 | -0.01 | 0.2713 | -0.057 | -0.185 | 0.013 | -0.027 | -0.012 | 0.0299 | -0.486 | 0.429 |
| W*A585 | -0.015 | -0.031 | 0.004 | 0.098 | -0.024 | 0.8663 | -0.013 | -0.037 | 0.004 | 0.074 | -0.016 | 0.9176 | -0.064 | 0.051 |
| W*A586 | -0.076 | -0.124 | 0.007 | 0.079 | -0.03 | 0.2848 | -0.064 | -0.108 | 0.006 | -0.004 | -0.014 | 0.2476 | -0.347 | 0.283 |
| W*A587 | -0.073 | -0.097 | 0.002 | 0.158 | -0.038 | **0.0235** | -0.054 | -0.187+ | 0.009 | 0.144 | -0.038 | 0.0673 | -0.491* | **0.437*** |
| W*A1288 | -0.051 | 0.176 | -0.058 | 0.174 | -0.093 | 0.7316 | -0.045 | 0.127 | -0.048 | 0.055 | -0.071 | 0.7287 | -0.225 | 0.181 |
| W*A1289 | -0.02 | 0.081 | -0.006 | 0.237 | -0.076 | 0.7088 | -0.023 | 0.132 | -0.009 | 0.313 | -0.143+ | 0.3406 | 0.108 | -0.134 |
| W*A589 | **-0.177**** | -0.002 | -0.002 | 0.151 | -0.08 | 0.661 | **-0.178**** | 0.103 | -0.009 | 0.15 | -0.082 | **0.7231** | -0.141 | -0.038 |
| W*A590 | -0.012 | -0.242+ | 0.016 | -0.075 | -0.006 | 0.1525 | 0.017 | -0.293+ | 0.019+ | -0.157 | 0.01 | 0.1352 | -0.652 | 0.669 |
| W*A591 | -0.02 | -0.111 | 0.008 | -0.098 | 0.001 | 0.5881 | -0.004 | -0.151 | 0.01 | -0.125 | 0.01 | 0.5501 | -0.361 | 0.357 |
| W*A592 | **-0.089+** | -0.162 | 0.008 | -0.041 | 0.002 | 0.5704 | -0.073 | -0.184 | 0.01 | -0.068 | 0.007 | 0.5259 | -0.481* | **0.409+** |
| W*A593 | 0.065 | -0.091 | -0.001 | 0.138 | -0.035+ | **0** | 0.082 | -0.119 | -0.001 | 0.122 | -0.035+ | 0 | -0.304 | **0.386+** |
| W*A594 | -0.024 | -0.151 | 0.004 | 0.075 | -0.018 | **0.0118** | -0.006 | -0.157 | 0.004 | 0.009 | -0.008 | 0.004 | -0.419+ | **0.413+** |
| W*A595 | -0.02 | -0.251* | 0.014+ | -0.059 | 0.01 | 0.3387 | 0.003 | -0.257+ | 0.014 | -0.057 | 0.009 | 0.4409 | -0.532* | **0.535*** |
| W*A596 | -0.051 | -0.306** | 0.021** | -0.044 | -0.008 | **0.0054** | -0.019 | -0.402** | 0.027** | 0.009 | -0.012 | 0.0025 | -0.818** | **0.800**** |

1: conventional regression; 3: instruments in main regression; 4: instruments and income in main regression; 5: Rivers Vuong

Table 2

| | logr_totalinc_p |
|---|---|
| r_albetbed | 0.424** |
| r_albetbed2 | -0.025** |
| r_alontbed | 0.15 |
| r_alontbed2 | 0.002 |

Table 3

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| wave | 2374 | 7.0430 | 2.0272 | 3 | 10 |
| sah | 2368 | 2.0731 | 0.8062 | 1 | 5 |
| ziek_verwond | 2359 | 0.1072 | 0.3095 | 0 | 1 |
| emo_geest | 2321 | 0.0491 | 0.2162 | 0 | 1 |
| ln(bmi) | 1429 | 3.1679 | 0.1668 | 2.6990 | 4.4653 |
| g_depri | 2367 | 2.4250 | 1.1451 | 1 | 5 |
| g_gntrek | 2365 | 1.7759 | 1.0274 | 1 | 5 |
| g_slaap | 2366 | 2.4260 | 1.2267 | 1 | 5 |
| g_futl | 2366 | 2.9544 | 1.1133 | 1 | 5 |
| g_ntstil | 2365 | 2.3945 | 1.2727 | 1 | 5 |
| g_onders | 2364 | 2.1967 | 1.1157 | 1 | 5 |
| g_ntconc | 2367 | 2.2708 | 1.0348 | 1 | 5 |
| g_zelfm | 2368 | 1.2711 | 0.7006 | 1 | 5 |
| g_dood | 2369 | 1.7944 | 1.0297 | 1 | 5 |
| g_wenen | 2369 | 2.0135 | 1.1266 | 1 | 5 |
| g_pessi | 2366 | 2.1817 | 1.0849 | 1 | 5 |
| g_sombe | 2364 | 2.0550 | 1.0663 | 1 | 5 |
| g_vreem | 2350 | 1.6885 | 0.9505 | 1 | 5 |
| g_prikkel | 2365 | 2.5848 | 1.0504 | 1 | 5 |
| g_beves | 2364 | 2.6997 | 1.1830 | 1 | 5 |
| g_ntgoed | 2369 | 2.3808 | 1.0875 | 1 | 5 |
| g_fysisc | 2359 | 2.1585 | 1.1770 | 1 | 5 |
| log(totalinc_p) | 2022 | 12.9516 | 0.9852 | 6.5041 | 16.9469 |
| leden | 2374 | 2.9587 | 1.4370 | 1 | 9 |
| male3650 | 2374 | 0.2818 | 0.4500 | 0 | 1 |
| male5165 | 2374 | 0.0543 | 0.2267 | 0 | 1 |
| male66200 | 2374 | 0.0185 | 0.1349 | 0 | 1 |
| female035 | 2374 | 0.1681 | 0.3740 | 0 | 1 |
| female3650 | 2374 | 0.3206 | 0.4668 | 0 | 1 |
| female5165 | 2374 | 0.0413 | 0.1990 | 0 | 1 |
| female66200 | 2374 | 0.0088 | 0.0937 | 0 | 1 |
| educ_2 | 2236 | 0.5818 | 0.4934 | 0 | 1 |
| educ_3 | 2236 | 0.2254 | 0.4179 | 0 | 1 |
| educ_4 | 2236 | 0.0962 | 0.2949 | 0 | 1 |
| brussels | 2374 | 0.1099 | 0.3129 | 0 | 1 |
| wallon | 2374 | 0.4452 | 0.4971 | 0 | 1 |
| 1994 | 2374 | 0.0438 | 0.2047 | 0 | 1 |
| 1995 | 2374 | 0.0864 | 0.2809 | 0 | 1 |
| 1996 | 2374 | 0.1276 | 0.3338 | 0 | 1 |
| 1997 | 2374 | 0.1310 | 0.3375 | 0 | 1 |
| 1998 | 2374 | 0.1714 | 0.3770 | 0 | 1 |
| 1999 | 2374 | 0.1554 | 0.3624 | 0 | 1 |
| 2000 | 2374 | 0.1449 | 0.3521 | 0 | 1 |
| (r_albetbed/10000) | 2050 | 0.2709 | 0.7149 | 0 | 17.1898 |
| (r_albetbed/10000)^2 | 2050 | 0.5843 | 7.4126 | 0 | 295.4905 |
| (r_alontbed/10000) | 1852 | 0.3515 | 0.5659 | 0 | 8.0173 |
| (r_alontbed/10000)2 | 1852 | 0.4436 | 2.4947 | 0 | 64.2774 |