DISCUSSION PAPER SERIES

IZA

IZA DP No. 7651

Evaluating the Role of Individual Specific Heterogeneity in the Relationship Between Subjective Health Assessments and Income

Iván Fernández-Val
Yevgeniya Savchenko
Francis Vella

September 2013

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Evaluating the Role of Individual Specific Heterogeneity in the Relationship Between Subjective Health Assessments and Income

Iván Fernández-Val
Boston University

Yevgeniya Savchenko
Georgetown University

Francis Vella
Georgetown University and IZA

Discussion Paper No. 7651
September 2013

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 7651
September 2013

ABSTRACT

Evaluating the Role of Individual Specific Heterogeneity in the Relationship Between Subjective Health Assessments and Income*

This paper investigates the impact of income on an individual's subjective self-assessment of own health. We employ recently developed methods in the non linear panel data literature to account for the endogeneity of income and the presence of individual heterogeneity. We examine a panel data set of individuals living in Australia and find no statistically significant relationship between income and health responses. Moreover, the evidence suggests that the variation in the individual specific effects, comprising both observed and unobserved time invariant factors, is primarily responsible for the variation across individuals' responses.

JEL Classification: I12, C33, C35

Keywords: subjective health assessments, non linear panel data models, fixed effects

Corresponding author:

Francis Vella
Economics Department
Georgetown University
37th and O Streets, NW
Washington, DC 20057
USA
E-mail: fgv@georgetown.edu

* This paper uses confidentialised unit record file from the Household, Income and Labour Dynamics in Australia (HILDA) survey. The HILDA Project was initiated and is funded by the Commonwealth Department of Family and Community Services (FaCS) and is managed by the Melbourne Institute of Applied Economic and Social Research (MIAESR). The findings and views reported in this paper, however, are those of the authors and should not be attributed to either FaCS or the MIAESR.

1 Introduction

The relationship between an individuals' socio-economic status and quality of health is important from a number of economic perspectives. For example, healthy workers are likely to be more productive, for a variety of reasons, while private and public expenditures on health care are likely to be lower than those for workers in poor health. Thus, for many reasons, it is useful to understand the socio-economic determinants of an individual's health status. While many factors are likely to be important, one potential determinant which has attracted substantial attention in the economics literature is the individual's level of income. The empirical relationship between income and health has been investigated through a variety of econometric techniques and in a number of settings (for example, cross-section studies - Lindahl (2005) for Sweden, Etilé and Milcent (2006) for France; panel studies - Contoyannis et al. (2004), Jones and Wildman (2008) and Carro and Traferri (2012) for the UK, Jones and Schurer (2011) and Frijters et al. (2005) for Germany). However, despite this substantial body of empirical work it appears, as first concluded by Deaton and Paxson (1998), that the relationship is poorly understood. The empirical work subsequent to the Deaton and Paxson paper does not appear to have altered the appropriateness of its conclusion.

One difficulty with evaluating the relationship between health and income is the measurement of health status. Objective health measures are expensive to collect and those available are often non representative of the population of interest due to the non random selection of individuals into medical evaluations (Etilé and Milcent, 2006). As a result, many researchers have used subjective self-assessed health measures. These are frequently considered capable of capturing patterns in objective measures (see, for example, Idler and Benyamini, 1997; Mackenbach et al., 2002; van Doorslaer and Gerdtham, 2003). However, the process of self-assessment is likely to introduce an element of subjectivity into the responses and this may create problems related to individual heterogeneity (see, for example, Etilé and Milcent, 2006; Jones and Wildman, 2008; Jones and Schurer, 2011; Carro and Traferri, 2012). That is, individuals may have different reference points due to social or cultural biases or individual differences. This may result in people with the identical health status, as measured by objective criterion, providing very different self-assessments. Ideally, one would incorporate this individual specific heterogeneity explicitly in the empirical investigation while adopting a fairly unrestricted approach, such as fixed effects procedures. However, until recently this has not been feasible as these subjective measures are generally binary or ordinal and the appropriate econometric procedures associated with explaining such outcomes are not easily adapted to settings with time invariant individual specific heterogeneity. One popular approach in the binary setting is to condition out the individual effects noting that this generally creates difficulties for identifying the marginal effects. An alternative methodology, such as employed in Maurer, Klein and Vella (2011), is to model the unobserved individual specific heterogeneity.

A second difficulty is the direction of causality (see, for example, Smith, 1999). While there is evidence that income affects health (for example, Contoyannis et al., 2004; Frijters et al., 2005; Lindahl, 2005; Jones and Wildman, 2008; Jones and Schurer, 2011; Carro and Traferri, 2012), there is also empirical evidence that an individual's health has an impact on his/her socio-economic status (see, for example, Thomas and Strauss, 1997; Currie and Madrian, 1999; Berhrman and Rosenzweig, 2004; and Black et al., 2007). The issue of endogeneity also arises from the possibility that there are unobservable factors which simultaneously influence both the health response and reported income. Accordingly, in estimating the health/income relationship it is critical that adequate attention is paid to endogeneity. However, endogeneity can be difficult to handle in this context. The ordinal nature of the outcome variable typically means that the health outcome models are not appropriately estimated by least squares methods, and this invalidates the use of instrumental variables procedures. Moreover, the subjective nature of the outcome variable requires that one should account for the role of time invariant individual heterogeneity which is potentially correlated with both health outcomes and the income measure.

In this paper we follow Carro and Traferri (2012) and employ recent advances in the non linear fixed effects panel data literature to account for individual specific heterogeneity in subjective health responses. While that paper focuses on dynamics in health assessment, we focus on both time invariant and time varying endogeneity. We examine Australian data noting that while several papers estimate an impact of health status on poverty, wages, and labor force participation using Australian data (see, for example, Buddelmeyer and Cai, 2009; Cai, 2007; Cai et al., 2008), we are not aware of any work which studies the impact of income on self-assessed health in Australia. The following section provides a brief review of the literature. Section 3 presents the empirical model and estimates a model which includes an explicit role for time invariant individual heterogeneity. The section also examines the objects of primary interest and comments on the role of unobserved heterogeneity in explaining the observed distribution of health responses. Finally it extends the model to allow for time varying endogeneity. Section 4 provides some discussion and Section 5 concludes.

2 Literature Review

Table 1 summarizes the results for a selection of studies which investigate the impact of income on different measures of health status. They incorporate a range of data sets and a variety of estimation and identification strategies. The notable conclusions are the following. Lindahl (2005) identified the impact of income on health in a cross-sectional setting in Sweden using lottery winnings as an exogenous source of variation. That paper reported that a 10 percent income increase improved a constructed health index by 4-5 percent of a standard deviation and decreased the probability of dying within 5 or 10 years after the interview by 2-3 percentage points. Etilé and Milcent (2006) studied the impact of income on self-assessed health in France and showed that there was a considerable individual heterogeneity in health reporting depending on the individual's level of income. The authors rejected the hypothesis of homogenous correlations of income with different self-assessed status in favor of heterogenous correlations. Several panel studies have explicitly addressed issues related to individual specific heterogeneity. Meer et al. (2003) used the Panel Study of Income Dynamics (PSID) to study the effect of income on the probability of being healthy in the United States. The authors transformed self-reported health on a scale from 1 to 5 measure of health into the dichotomous variable "healthy" which was equal to one if an individual reported health status as excellent, very good or good. They found a small positive effect of wealth on self-assessed health although this effect became statistically insignificant once the endogeneity of wealth was accounted for. In their preferred specification an USD 250,000 increase in household wealth change lead to 2.2 percentage points increase in probability of reporting excellent or good health from baseline probability of 81 percent. However, this study did not address the problems arising from the presence of individual heterogeneity inherent in health self-assessment. Contoyannis et al. (2004) used eight waves of British Household Panel Survey (BHPS) 1991-1998 to estimate the impact of household income on self-reported health using a dynamic ordered probit model with random individual specific effects. They found that unobserved heterogeneity accounts for about 30 percent of the self-reported health variation. Moreover, they found a positive and statistically significant effect of income on self-reported health. Carro and Traferri (2012) employed 16 waves of the BHPS 1991-2006 to estimate a dynamic ordered probit with individual fixed effects and fixed effects in the cutoff points. They derived a modified MLE bias-corrected estimator with state dependency. They found a small but statistically significant effect of income on self-assessed health. Jones and Wildman (2008) used eleven waves of BHPS 1991-2001 to estimate the impact of household income and relative household income on dichotomized measures of self-assessed health. They controlled for individual heterogeneity by alternatively using random effects, Mundlak (1978) type correlated random effects and Hausman and Taylor (1981) style instruments. The authors found that there was a strong positive relationship between self-assessed health and income which was robust to different specifications, but the relationship between health and relative deprivation (relative income) depended on the specification. Frijters et al. (2005) and Jones and Schurer (2011) using German Socio-Economic Panel (GSOEP) explicitly controlled for individual heterogeneity by applying conditional fixed effects logit (see Rasch 1960 and Chamberlain 1980). They documented a very small impact of household income on self-assessed health. For example, Frijters et al. (2005) found that a one log point increase in income lead to only a 0.083 improvement in health satisfaction of East German men and 0.067 improvement in health satisfaction of West German men. They found a statistically significant effect of income on health. Jones and Schurer (2011) looked at interaction of income and age categories and found that income was not statistically significant for most age categories when controlling for individual-specific effects, but income was associated with higher self-satisfaction at the lower end of the health distribution.

Ideally one would capture the presence of time invariant individual heterogeneity with a fixed effects procedure. However, subjective assessments frequently result in ordinal outcomes and estimating such models with fixed effects typically introduces an incidental parameter problem. For example, consider the following model:

$$H^*_{it} = \alpha_i + \beta X_{it} + u_{it} \quad (1)$$

$$H_{it} = f(H^*_{it}) \quad (2)$$

where H^*_{it} is a latent subjective measure of individual i's health at time t with observed ordinal counterpart H_{it} noting that the mapping is determined by the censoring function $f(.)$; X denotes a vector of explanatory variables; β is a vector of unknown parameters; α_i is an individual specific fixed effect; and u_{it} is an error term. In estimating this model one would ideally include individual specific intercepts which capture the individual specific heterogeneity and also allow the α to have some correlation with X. However, this requires that one address the incidental parameter problem (see, for example, Neymann and Scott, 1948). While this incidental parameter has been overcome via a variety of bias correction methods (for a survey see Arellano and Hahn, 2007) such an approach, with the exception of Carro and Traferri (2012), is not adopted in the studies outlined above. Many previous studies adopt a random effects treatment of α which does not capture neither the subjectivity of the responses nor the possibility that they are correlated with the elements of X. Some studies employ "correlated random effects" but this requires a parameterization of the individual effects. Other studies account for endogeneity via instrumental variables methods but this requires ignoring the ordinal nature of the outcome.

One strand of the literature dichotomizes the ordinal variable and estimates the model by the conditional fixed effects procedure of Rasch (1960) and Chamberlain (1980). As an ordinal variable with more than two outcomes can be dichotomized in multiple ways, the authors propose various ways of combining the multiple sets of estimates associated with the various dichotomizations. One shortcoming with this methodology is that it does not provide estimates of the individual effects as they are conditioned out. This reduces the capacity to explore issues related to the estimation of marginal effects and the role of the fixed effects themselves.

An important recent paper which employs fixed effects but retains the ordinal structure is Carro and Traferri (2012). That paper also introduces the added complication of incorporating dynamics via a lagged dependent variable while allowing for fixed effects in the separation points. Carro and Traferri (2012) paper is an important contribution in this literature as it is the first to account for the ordinal nature of the health measure while incorporating individual specific effects which are potentially correlated with the other explanatory variables. Our paper contributes to the literature but the focus differs from that of Carro and Traferri. While their paper focuses on allowing for fixed effects when evaluating the role of state dependence, our paper focuses on how the presence of fixed effects affects the health income relationship. We also incorporate a role for time varying endogeneity in addition to the fixed effects. Thus, our study provides evidence on a different dimension to the work of Carro and Traferri (2012).

3 Empirical Model and Results

3.1 Data Description

We analyze data from the 2001-2008 waves of the Household, Income and Labour Dynamics in Australia (HILDA) Survey. HILDA is a nationally representative household-based panel study comprising 7,682 households (19,914 individuals) in the inception year. The survey began in 2001 and is conducted annually. It contains detailed demographic characteristics and information on family structure, employment history, education, income, health, wellbeing, attitudes and values. To construct a balanced panel, we restrict the sample to individuals aged between 17 and 65 years in all periods who are in the labor force with non-missing age, education and health variables.[1] This reduces our sample to N=2,503 individuals.

The dependent variable of primary interest is the response to the question: "In general, would you say your health is?" and the responses take the values from 1 to 5 with 1 being poor, 2–fair, 3–good, 4–very good, 5–excellent.[2] Table 2 reports the distribution of these responses across the 8 time periods. The table indicates that very few report that they are in very poor health and relatively few report that they are in fair health or excellent health. The vast majority of the sample indicate that they are in good or very good health.

Around 88 percent of the sample reported some change in health status over the eight year

[1] The age requirement imposes the individuals must be aged between 17 and 57 years in the first period.

[2] In the questionnaire the scale is in reverse order 5 being poor, 4–fair, 3–good, 2–very good, 1–excellent. We reversed the order to have "the more the better" interpretation of health status.

period. These changes are likely to reflect, in part, changes in the values of the conditioning variables that determine health status. However, approximately 12 percent provide the same response over the 8 year period and this may reflect a substantial time invariant individual component. This may also suggest an element of subjectivity in the individual's response. That is, the individual reports the same response irrespective of his/her health. The lack of objectivity creates a bias in the absence of appropriate attempts to account for it, while the time invariant objective evaluation will also lead to a bias when this evaluation is influenced by the individual's other explanatory variables.

Below we employ the jackknife bias correction method proposed by Hahn and Newey (2004). The procedure estimates the parameters of the model over the full panel of 8 years and then re-estimates the model 8 additional times where each time a different time period is excluded. Identification of the model parameters requires specific variation in the data, and excluding the observations which do not display this variation reduces the sample size to 2,411 observations. Moreover, as small cell sizes can be problematic in discrete choice estimation, we aggregate the lower two categories. This produces an outcome with four values. This aggregation resulted in additional 55 observations which showed inadequate variation, and excluding them this produces a balanced panel of 2,356 observations which are distributed across outcomes as reported in Table 3. One might suspect that the reduction in the sample size may result in some form of selection bias. This will not be an issue provided that the form of the selection is time invariant as this will be captured by the fixed effects procedure.

Tables 4-6 present summary statistics for the pooled sample and male and female subsamples. Men comprise 56 percent and women 44 percent of the final sample. The average age in the sample is 42.41 years old and the average level of education is 13.78 years. Ninety nine percent of the sample report being employed. An employed individual has the average experience level of 25.92 years and has 8.58 years of tenure. Thirty two percent of employed people belong to a union. Tables 5-6 indicate that the summary statistics for men and women are generally similar. The only individual characteristic which is significantly different across gender is marital status with a higher proportion of men reporting that they are married. While the employment rate seems high, the official unemployment rate in this period was around 4 percent. This corresponds to the unemployment rate in the unbalanced panel for these data. The unemployment rate decreases when we examine the balanced panel.

A critical aspect of our study is our measure of income. We employ real household weekly income. This reflects the income available to the household and should capture general economic welfare and is the measure typically employed in the empirical studies discussed in section 2. Real household weekly income was constructed as a sum of wages, pensions and other allowances, and income from business, dividends, interest or rent of all members of household. The average real household income in the sample is equal to AUD 1,501 noting there is substantial variation across households. Table 7 reports the path of real household weekly income of our sample over the sample period. The real household income increased from AUD 1,315 in 2001 to AUD 1,599 in 2008. Mean household income for the female subsample is around AUD 104 lower than that of male subsample. This seems reasonable since we observe that wage income of women is lower than that of men, and women in our sample are less likely to be married. Thus, if the wage represents a substantial share of the household income, a female single household could be expected to have lower household income than a male single household. The mean real hourly wage in 2005 prices was AUD 25.34 (see Table 8). On average, men earned higher hourly wages than women, AUD 26.84 vs. AUD 23.52 respectively, but women displayed more variation in wages (see Tables 5-6).

3.2 Estimating the Determinants of Health Status

In estimating the relationship between health and income we begin with previously used modeling strategies. Write the model of primary interest as:

$$H_{it}^{*} = \beta X_{it} + \theta I_{it} + u_{it} \quad (3)$$

$$H_{it} = f(H_{it}^{*}) \quad (4)$$

where H_{it}^{*} is a latent measure of the individual's health with observed ordinal counterpart H_{it} determined by the censoring function $f(.); I_{it}$ is a measure of income; X denotes a vectors of explanatory variables including education, marital status, age, number of children, gender, employment and location; β, θ and δ are unknown parameters; and u_{it} is a normally distributed iid error term. One can estimate the model by pooled ordered probit by assuming there are no individual specific effects and the income variable, in addition to the other explanatory variables, is exogenous. The associated log likelihood function is:

$$L = \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{J} D_{itj} \left\{ \ln \left(\Phi(\mu_j + \beta X_{it} + \theta I_{it}) - \Phi(\mu_{j-1} + \beta X_{it} + \theta I_{it}) \right) \right\}$$

where Φ denotes the cdf of the standard normal distribution; the $\mu' s$ are separation points to be estimated; and the D_{itj} are indicator functions denoting that individual i responds outcome j in time period t. Identification of the parameters does not require an individual change categories within the sample period as there is no distinction between cross and within individual variation. The standard errors are adjusted to account for having repeated observations on the same individual.

The estimates from this model are shown in column 1 of Table 9 and a number of features are worth noting. First, a number of variables are statistically significant. While it is not possible to make general statements on the basis of the estimated signs of the coefficients (see Crawford et al., 1998), we can conclude that the probability of reporting the highest (lowest) health status increases (decreases) with education level, being married and being female. Not unexpectedly, the probability of reporting the highest (lowest) outcome decreases (increases) with age. Second, there is a statistically significant relationship between real household income and health status. Moreover, the evidence suggests that the highest health outcome is associated with increasing income.

While ordered probit provides consistent estimates under the assumptions outlined above, most empirical studies decompose the error term to reflect the panel nature of the data. Ideally, one would allow for individual specific effects but the incorporation of such a feature in this literature has been limited. Accordingly, we employ the approach which has generally been used and decomposed the error as:

$$u_{it} = \alpha_i + \varepsilon_{it} \tag{5}$$

where the α_i is assumed to be an individual specific time invariant error term and the ε_{it} is an idiosyncratic term. By assuming that these components are each normally distributed and mutually independent, with variances 1 and σ_α^2 respectively, we can estimate the model by random effects ordered probit. The log likelihood function has the form:

$$L = \int_{-\infty}^{\infty} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{J} D_{itj} \left\{ \ln \left(\Phi(\mu_j + \beta X_{it} + \theta I_{it} + \alpha_i) - \Phi(\mu_{j-1} + \beta X_{it} + \theta I_{it} + \alpha_i) \right) f(\alpha) \partial \alpha \right\}$$

where $f(\alpha)$ is the density of α which is integrated out in estimation. While the use of random effects in this type of model is popular, it is again worth highlighting that the only benefit over the pooled ordered probit estimates is an efficiency gain.[3] The inclusion of the random effects does not allow for individual heterogeneity in responses and does not allow for correlation

[3]The random effects ordered probit estimator does not require that individuals change categories in the sample period for identification purposes.

between the explanatory variables and the α_i. The estimates are reported in column 2 of Table 9. The results are generally similar to those in column 1 with respect to statistical significance and sign. The magnitudes of the coefficients are not directly comparable across specifications due to the respective normalizations imposed. The increase in efficiency does not appear to occur and this may reflect that the assumptions of the model are not satisfied.

A difficult challenge here is accounting for individual specific heterogeneity in the interpretation of the questions and the corresponding responses. Also, some of the explanatory variables may be correlated with the individual specific time invariant error term. One way to account for these concerns is to estimate the model by a fixed effects procedure. The existing papers in this literature which have employed fixed effects procedures have used a variety of strategies which involve dichotomizing the outcome variable and then employing the Chamberlain conditional logit estimator. However this approach, in the absence of additional assumptions, is uninformative with respect to the marginal effects which are clearly the objects of primary interest. Rather than by-pass the incidental parameter problem by dichotomizing the outcome of interest, we follow the example of Carro and Trafferi (2012) and estimate the model, using the original characterization of the dependent variable, by fixed effects ordered probit and bias adjust the estimates.

We estimate the following model which combines a number of the equations above:

$$H_{it} = f(\beta X_{it} + \theta I_{it} + \alpha_i + \varepsilon_{it})$$

noting that we retain the assumption that ε_{it} is normally distributed but relax any distributional assumptions for the $\alpha_i's$. We also allow the α_i to be correlated with the explanatory variables. This is an important departure from the procedures generating the results in columns (1) and (2) of table 9. The log likelihood function is now:

$$L = \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{J} D_{itj} \left\{ \ln(\Phi(\mu_j + \beta X_{it} + \theta I_{it} + \alpha_i) - \Phi(\mu_{j-1} + \beta X_{it} + \theta I_{it} + \alpha_i)) \right\}$$

where the $\alpha's$ are now parameters to be estimated. As the estimator acknowledges that there are repeated observations on the same individual, and estimates a parameter to capture this form of heterogeneity, it is unable to estimate the parameters of explanatory variables which are time invariant. Moreover, it requires within individual variation in responses to contribute to the estimation of the slope parameters. However, observations which are always in the same category, provided they are not in either the lowest or highest, contribute to the estimation of the separation points.

Column 3 of Table 9 reports the unadjusted fixed effects ordered probit estimates for the slope parameters. Column 4 reports the corresponding Hahn and Newey bias corrected estimates. The bias correction takes the form:

$$\pi_a = T\pi_u - (T-1) \sum_{t=1}^{T} \frac{\pi_t}{T}$$

where $\pi = [\beta : \theta]$, where π_u are the estimates in column (3) and the π_t are the estimates from estimating the model while omitting the t^{th} panel. The covariance matrix is evaluated at these bias corrected coefficients and is calculated using the Newton-Raphson method as described in Greene (2001). We employ the Hahn and Newey jackknife procedure rather than the analytical correction due to ease of implementation and on the basis of the simulation evidence that suggests that it works well. Due to the stringent conditions it imposes on the time series properties of the explanatory variables, we also employed the more flexible split sample procedure of Dhaene and Jochmans (2010)For the models estimated in this section of the paper the two methods provided almost identical results.[4]

[4]The split sample procedure requires breaking the sample in two and this can be problematic when T is

Before contrasting the adjusted fixed effects estimates with the pooled and random effects estimates, it is valuable to compare the biased and bias corrected estimates. In general, the estimates are quite different and this is consistent with the simulation evidence in this literature (see, for example, Hahn and Newey, 2004; Fernández-Val, 2009). The coefficient of primary interest, namely that on the income variable, increases from -.0003 to -.0002 although each estimate is within sampling error of the other. This description is also applicable to the other slope coefficients noting that in this specification the only statistically significant variable, at the conventional levels of testing, is the "number of children".

Focus on the comparison of the biased adjusted fixed effects estimates with those in column (2) noting that the statistically significant estimate of ρ in the random effects estimates indicates that these estimates are preferred over the pooled ordered probit estimates. However, before doing so, it is valuable to examine which of the fixed effects and the random effects estimates should be preferred. Given the assumptions associated with the respective models, it is possible to apply the Hausman (1978) testing framework. That is, under the null hypotheses of the exogeneity of income and the normality of the individual components the random effect estimator is consistent and efficient. However, if either of these assumptions is violated the estimator is inconsistent. In contrast, the biased adjusted fixed estimator is not efficient under the null hypothesis but consistent under the alternative. Thus we construct a test of the form:

$$H = \left(\pi_a^{FE} - \pi^{RE}\right)' V\left(\pi_a^{FE} - \pi^{RE}\right)^{-1} \left(\pi_a^{FE} - \pi^{RE}\right)$$

where $V\left(\pi_a^{FE} - \pi^{RE}\right)$ is the difference in the estimated covariance matrices. The test value of 60.62 indicates that the random effects specification is rejected in favor of the fixed effects

not large. While this was not an issue in the models where we adjusted only for time invariant endogeneity it provided an unreasonable estimate in the next section of the paper due to one of the estimates of the a slope parameter in one of the subsamples. For this reason we employ the Hahn and Newey adjustment throughout the paper.

model.

Consider the implications of preferring the fixed effects model over the random effects specification. First, there is a difference in the magnitude of the effect of income on self reported health. Second, the magnitude and statistical significance of the individual effects, which are not reported here, indicates a remarkable variation in the individual effects. While we acknowledge that the fixed effects estimates are based only on 8 observations per individual we further explore the magnitude of these effects in the following section.

Consider now the substantive implications of our results. The rejection of the random effects and pooled specifications in favor of the fixed effects specifications has a number of important implications. First, it appears that the statistical significance of many of the variables in the first two columns of Table 9 which are time invariant simply reflects that the model should have included fixed effects. While the coefficients on the time invariant variables are unidentified in the fixed effects procedure, the Hausman test suggests that the fixed effects characterization is more appropriate. While the test may also reflect that some of the explanatory variables are also contaminated with time invariant endogeneity, it is also likely that the individual specific fixed effect is more effective in capturing the role of these time invariant explanatory variables. That is, the exclusion of the fixed effects may, as they do here, result in the coefficients on these time invariant explanatory variables displaying statistical significance when they are simply proxying the fixed effects. The same appears to be true for variables which show little time variation. While the fixed effects specification is identifying these effects from the deviation from individual mean behavior, the other specifications employed here are also incorporating cross individual variation. The evidence here suggests that the variation being exploited in the random effect and the pooled models which are leading to the statistical significance is the variation across individuals.

The biased corrected estimates of the individual fixed effects are generally statistically significant and the point estimates display a great deal of variation. While we delay a discussion of the implications of these individual effects for the associated effects on conditional probabilities, a number of points are worth raising here. First, the lack of movement across the health outcomes, reflected in the earlier tables, indicates that a substantial proportion of the variation across individuals' reported health status, at any specific time, is likely to be captured by the individual effect. In fact, the results here suggest that with the exception of a weak "marital status" effect the only variable which influences the outcome is the number of children. Second, it is clear that in the absence of the fixed effects, this substantial variation in the fixed effects will be somewhat captured by the included explanatory variables. This would result, as it does here, with a conclusion which attributes too large of an effect to a variable such as income. Finally, what do the "large" individual effects reflect? In the case of subjective health assessments it is possible that they are capturing a number of effects. For example, they may simply reflect individual specific heterogeneity in the meaning of the questions and the responses.

3.3 Marginal Effects

While the difference in estimates across specifications is enlightening, it is more important to focus on the marginal effects associated with the various models. The objects of primary interest are the conditional probabilities and how they change in response to changes in the conditioning variables. These are reported in Table 10. The marginal effects indicate how the probability of reporting certain health outcome changes with a change in income.

Before we report the marginal effects, we comment on how each has been calculated and the assumptions underlying these calculations. First, define the probability of reporting a certain health status m at time t for each individual i as $g_m(x_{it}\beta + \alpha_i) = P(y_{it} = m|x_{it}, \alpha_i)$. Then we can calculate a marginal effect with respect to x^k for each person at each time period as:

$$\mu_{mkit} = \frac{\partial g_m(x_{it}\beta + \alpha_i)}{\partial x_{it}^k} = g_m'(x_{it}\beta + \alpha_i)\beta_k.$$

To find an overall marginal effect with respect to x^k, we take an average of all marginal effects over individuals and time:

$$\mu_{mk} = \frac{1}{NT}\sum_{i,t}\mu_{mkit}.$$

Table 10 reports the average marginal effects for the various outcomes for changes in the level of real household income. The marginal effects for the ordered probit model are straightforward to estimate as are those for the fixed effects estimates.[5] However, the marginal effects for the random effects ordered probit model require some treatment of the α_i. Given the nature of the random effects estimator, we integrate out the α. More explicitly, letting $\sigma_\alpha^2 = \sigma^2$, we have $\alpha_i|X_{it} \sim N(0, \sigma^2)$. Then:

$$\begin{aligned} P(y_{it}|x_{it}) = &\int \big[\Phi\big(\frac{c_m - x_{it}\beta}{\sqrt{1+\sigma^2}}\big) - \Phi\big(\frac{c_{m-1} - x_{it}\beta}{\sqrt{1+\sigma^2}}\big)\big] f(\alpha) d\alpha = \\ &= \Phi\big(\frac{c_m - x_{it}\beta}{\sqrt{1+\sigma^2}}\big) - \Phi\big(\frac{c_{m-1} - x_{it}\beta}{\sqrt{1+\sigma^2}}\big). \end{aligned}$$

This gives an individual marginal effect of:

$$\overline{\mu}_{mkit} = \frac{\partial P(y_{it} = m|x_{it})}{\partial x_{it}^k} = \frac{\beta_k \big[\Phi'\big(\frac{c_m - x_{it}\beta}{\sqrt{1+\sigma^2}}\big) - \Phi'\big(\frac{c_{m-1} - x_{it}\beta}{\sqrt{1+\sigma^2}}\big)\big]}{\sqrt{1+\sigma^2}}.$$

Several features of this table of marginal effects are worth noting. First, the magnitude of the average marginal effects is very small using all three approaches. For example, an increase of real household income by AUD 100 is associated with only 0.05 percent decline in

[5]The marginal effects for the fixed effect model were biased adjusted using the Hahn and Newey correction using the following formula $\mu_{mk}^a = T\mu_{mk}^u - (T-1)\sum_{t=1}^{T}\frac{\mu_{mk}^t}{T}$ where μ_{mk}^u are the marginal effects calculated using the uncorrected estimates and μ_{mk}^t are the marginal effects calculated from the uncorrected estimates from estimating the model while omitting t^{th} panel.

the probability of reporting health status as poor or fair.[6] Second, while the marginal effects in the pooled and the random effects estimates are small, they are large, in absolute value, in comparison to those associated with the fixed effects estimates. This further highlights the importance of accounting for unobserved individual time invariant heterogeneity.

While the effects are very small, we acknowledge that they may simply reflect the small change which is being imposed in the income level. We now examine how the probability of self-reporting a certain health status changes as we move individuals along the income distribution. More explicitly, we assign an individual his/her own characteristics but assign them different levels of income. That is we compute $\widehat{\beta}X_{it}+\widehat{\alpha}_i$, where the $\widehat{\alpha}_i's$ denote the fixed effects estimates, and then add $\widehat{\theta}I^p$ where I^p is the value of income at the p^{th} percentile of the income distribution. We then calculate the individual probabilities of reporting a certain health status based on this index and take the average over the individual probabilities. As Table 11 shows, the probability of reporting any type of health status is almost invariant to the individual's location in the income distribution. We highlight that we do not conclude from this evidence that income does not affect health. We do conclude, however, that income does not appear to have anything beyond a negligible effect on subjective health assessments.

Given the model specification and the lack of statistical significance of the remaining explanatory variables, it is valuable to explore what generates the variation in the observed distribution of health responses. When some of the variation in the responses is due the subjective interpretation of the responder, it is likely that this will be captured in the estimates of the individual fixed effects. While acknowledging that the estimate of the fixed effects also capture other time invariant influences, it is nevertheless interesting to see the impact of changing the value of the individual effect on the estimated probabilities of the various health outcomes. This is done in Table 12. This table provides the predicted prob-

[6] While this appears a small change, it does represent about 8 percent of weekly income in 2001 and around 7 percent in 2008.

abilities by calculating the index $\widehat{\beta}X_{it} + \widehat{\theta}I_{it}$, where the $\widehat{\beta}$ and the $\widehat{\theta}'s$ are the fixed effects slope estimates, for each individual and then estimate the probability of self-reporting each health status when the index is supplemented by values from the empirical distribution of the individual fixed effects $\widehat{\alpha}^P$. We then average over individual probabilities. The results are striking. For example, the probability of self-reporting the health status as a poor or fair (y=1) evaluated at the average value of the explanatory variable index and the 5th percentile of the individual fixed effects is 42 percent. Moving to the 10th percentile of the individual fixed effects reduces the probability of reporting a poor or a fair health status to 25 percent. Similar changes can be observed for the other probabilities. For example, the probability of reporting an excellent health goes from almost zero percent at the 5th percentile to 0.2 percent at 25th percentile and to 14.5 at 75th percentile. A comparison of various other entries in the table provides a similar conclusion.

The clear conclusion from an examination of these tables of predicted probabilities is that the major determinant of the distribution of responses of health is the distribution of the individual specific fixed effects. One component of the estimated individual fixed effect will capture the socio-economic time invariant factors such as gender. However, there is nothing in the pooled or random effects estimates which suggests this effect is dominating the individual fixed effect. Another reasonable conclusion is that even in the absence of subjective assessments, the health distribution is determined by factors which are simply not captured in the data. However, note that this result would not necessarily be restricted to this study as the variables we are including as explanatory variables are typical of what are employed.

3.4 Time Varying Endogeneity of Income

We now employ two procedures which incorporate both time invariant and time varying endogeneity of income. We do so to investigate whether our inability to detect an effect from income is due to the endogeneity of income through time varying heterogeneity. We employ a random effects treatment proposed by Vella and Verbeek (1999) and the more robust fixed effects procedure of Fernández-Val and Vella (2011). For both estimators we supplement the health outcomes equation with a representation of the income process. The model has the form:

$$\begin{aligned} H_{it} &= f(\beta X_{it} + \theta I_{it} + u_{it}) \\ I_{it} &= \delta Z_{it} + v_{it} \end{aligned} \tag{6}$$

where Z is a vector of explanatory variables which contains X and at least one additional explanatory variables to ensure identification; δ are unknown parameters; and the v_{it} are unobserved error terms. We impose the following decomposition on the error terms:

$$\begin{aligned} u_{it} &= \alpha_i + \varepsilon_{it} \\ v_{it} &= \theta_i + \eta_{it} \end{aligned} \tag{7}$$

where we assume that the α_i and θ_i are individual specific time invariant terms and the ε_{it} and η_{it} are idiosyncratic terms. We allow for a correlation across equations in the time invariant components and we assume the idiosyncratic components are normally distributed with potentially non zero correlation $\rho_{\varepsilon\eta}$. Thus, the endogeneity of income now operates through correlated time invariant components and correlated time varying components.

First, we assume the α_i and θ_i are normally distributed random error terms. We then follow Vella and Verbeek (1999) and estimate the income equations by random effects MLE to get the parameters δ, σ_θ and σ_η. The following correction terms are then included as additional explanatory variables in the health status equation:

$$E[u_{it}|v_{it}, X_{it}] = \tau_1 v_{it} + \tau_2 \overline{v}_i$$

where $\overline{v}_i = T^{-1} \sum_{t=1}^{T} v_{it}$ and τ_1 and τ_2 are unknown parameters. This equation implies that the conditional expectation of u_{it} is a linear function of conditional expectation of v_{it}. The conditional expectation of v_{it} is:

$$E[v_{it}|Z_{it}] = \int (\theta_i + E[\eta_{it}|Z_{it}, \theta_i]) f(\theta_i|Z_{it}) d\theta_i$$

where $f(\theta_i|Z_{it})$ is the conditional density of θ_i and $E[\eta_{it}|Z_{it}, \theta_i])$ is a the cross-sectional generalized residual from equation 6. In this particular setting the corrections are straightforward to compute as they are functions of least squares type residuals. The main equation is then estimated by random effects ordered probit.

This approach suffers from the criticism that the individual effects are independent of the other explanatory variables. Accordingly, to remain in this fixed effects setting, we follow Fernández-Val and Vella (2011). We relax the distributional assumptions for the individual effects in the two equations and estimate the first step by fixed effects least squares. We then compute the control function:

$$\widehat{v}_{it} = I_{it} - \widehat{\delta} Z_{it} - \widehat{\theta}_i$$

and estimate the health status equation by fixed effects ordered probit.

The results from these estimation strategies are reported in Table 14. We use the same income measure as above, but to ensure we had appropriate variables to include in the Z_{it} vector, our use of tenure, experience and union status variables reduced the sample of 2,138 due to missing values for these variables. The first step for the random effects and fixed effects procedures are reported in Table 13 along with the pooled OLS estimates.

The first step results, shown in Table 13, indicate that many of the proposed determinants of income are statistically significant in the income equation when it is estimated by pooled OLS or random effects. However, when the model is estimated by fixed effects the only statistically significant effects are those operating through the experience profile and union membership. Education has a statistically significant effect but it is very small in magnitude and has the unexpected sign. Once again there is a very different story told by the two different treatments of the individual effects.

The results for the health equation are provided in Table 14. We provide the pooled ordered probit results, with and without correction terms, for comparison although we do not focus on them. The results for the random effects estimation of the health equation indicate there is no statistically significant role for income. The estimates of the coefficients on the two control functions are statistically insignificant revealing no sign of endogeneity. The bias corrected fixed effects estimates, provided in the final column of that table, produce similar conclusions. That is, there is no effect from income on health and there is no evidence of time varying endogeneity above that associated with that of the time invariant form.

4 Discussion

From our empirical evidence we now draw a number of conclusions and also consider their implications. First, our results indicate that the more reasonable characterization of the individual specific effects in this particular is that of fixed, rather than random, effects. Given the subjective nature of the outcome variable, this is not a surprising result. What is more surprising, however, is that the estimates impact of income on subjective health assessments is very small, in terms of magnitude, but also statistically insignificant. In contrast the estimated individual specific components, reflecting time invariant individual heterogeneity, as captured by our estimates of the individual fixed effects, are not only important statistically but also important from a substantive viewpoint. That is, whereas changing the income level appears to have very little impact on the probability distribution of responses, changing the value of the individual fixed effect, by an amount which is consistent with the data, has drastic changes on the probability distribution of the responses.

It is important to consider the implication of these results. First, consider the estimated negligible income effect of income on health. This result is not entirely surprising given the conclusion of Deaton and Paxson (1998) we discuss above and the literature review in Section 2. It appears that for developed countries the effect of income on health assessments is small. This is not unreasonable. A more "controversial" finding is that it appears that none of the other explanatory variables are statistically significantly and everything is explained by the "fixed effect". Note that the fixed effect captures not only the role of time invariant individual specific unobserved heterogeneity but also time invariant individual specific factors. Thus, in the fixed effects models considered here the role of the explanatory variables is identified by the variation from the individual mean. However, it would appear that for the sample considered here, it is possible that there is insufficient movement from the mean to capture these effects. Alternatively, these variables do not have statistically significant relationship with the dependent variable. It is also the case that the mean effect of the variables, such as income, is captured in the individual effect. Accordingly, we explore what proportion of the variation in the fixed effects across individuals can be explained by variation in the means of the explanatory variables. This is explored in Table 15

In Table 15 we report the results from regressing the 2,356 estimates of the fixed effects on the individual time invariant variables (column 1) and the individual time invariant variables and the individual means for the time varying variables (column 2). The results are essentially the same, so we will focus on the second column. Note that while the values of the fixed effects at the various percentiles are shown in Table12, the mean is 1.61 and the standard deviation is 1.17. The minimum value is -1.41 and the maximum value is 4.53. A number of interesting features are apparent from the table. Recalling that table 12 clearly suggests that a higher fixed effect value is associated with a “better” health distribution, we see that males generally have a lower fixed effect as do older people and people with a large number of children. In contrast, higher fixed effects are associated with being married and having more education. Note that all the effects of all these variables are small. There are also statistically significant region effects with the most notable being associated with those individuals living in the areas of Queensland located outside Brisbane. This effect is likely to capture that individuals may be relocating in these areas as they age. Perhaps the most interesting effect is the relationship between the fixed effect and income. As income enters the model directly, it is difficult to give the coefficient a direct interpretation, but it is clear that a large increase in income would result in a large increase in the fixed effect with the implications for the health distribution implied in Table 12.

To explore how income might affect the responses through its impact on the fixed effect we use our results from Table 12 to explore what increase is required in income to move the individual effect from its value at the 5th percentile to the 25th percentile noting that Table 12 indicated that this produced a drastic change in the distribution of response. Our results indicate that the required change is AUD 13,175. Moreover, going from the 25th percentile to the 50th percentile required AUD 12,425 and from the 50th to the 75th required AUD 10,788. These results that the interpretation of the effect of income is somewhat subtle. That is, the effect identified from the ”within individual” variation is essentially zero. However, perhaps this could be interpreted as transitory income. In contrast the changes in ”permanent income” captured by the component of the individual effect appears to suggest that there is some role for income in determining subjective assessments of health. However, it should be noted that the level of income which is required to generate a large change in responses is very large.

An important feature of Table 15 is the remarkably low R squared associated with the two specifications. Even with the expanded set of regressors in column 2 the R squared is 5.6 percent. This indicates that while there is a statistically significant relationship between the fixed effects and a number of the regressors, the regressors collectively explain very little of the variation in the fixed effects. Thus, a high proportion of the fixed effects comprise individual specific features which cannot be explained by the data. This may simply reflect the subjective nature of the dependent variable which the fixed effects themselves are trying to explain. However, irrespective of the interpretation of the fixed effects, it clearly highlights the role of unobserved heterogeneity in this context.

The empirical evidence from accounting for both time varying and time invariant heterogeneity is less significant on its own but in partnership with the earlier evidence presents a compelling picture. That is, the relationship between income and self-assessed health is very tenuous. Moreover, the relationship does not appear to be strengthened when the potential endogeneity of income is accounted for.

5 Conclusion

This paper investigates the impact of income on an individual's subjective self-assessment of their own health employing recently developed methods in the non linear panel data literature to account for the endogeneity of income and the individual heterogeneity which is potentially relevant to any outcome which is based on self-assessment. The results of the paper are striking in that there appears to be no statistical relationship between income and health responses, the implied impact of income on health is remarkably small. We acknowledge that since our preferred model is estimated by fixed effects procedure our effects from income are estimated from within individual variation. As this may reflect transitory income it is perhaps not surprising that the effects are negligible. This is supported by the evidence that the permanent income levels, operating through income's relationship with the fixed effects, appear to have an impact, albeit small, on the subjective health responses.

References

[1] Arrellano,M., and J.Hahn. (2007). "Understanding Bias in Nonlinear Panel Models: Some Recent Developments."In: R. Blundell, W. Newey, and T. Persson (eds.): Advances in Economics and Econometrics, Ninth World Congress, Volume III, Cambridge University Press, 381-409.

[2] Behrman, J. R., and M. R. Rosenzweig. (2004). "Returns to birthweight." *Review of Economics and Statistics*, 86(2): 586–601.

[3] Black, S. E., Devereux, P. J., and K. Salvanes. (2007). "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes. " *The Quarterly Journal of Economics*, 122(1): 409–439

[4] Buddelmeyer, H., and L. Cai. (2009). "Interrelated dynamics of health and poverty in Australia."IZA Discussion Paper No. 4602

[5] Cai, L. (2009). "Effects of health on wages of Australian men." *Economic Record*, 85(270): 290–306.

[6] Cai, L., Mavromaras, K.G., and U.Oguzoglu. (2008). "The Effects of Health and Health Shocks on Hours Worked. "IZA Discussion Paper No. 3496.

[7] Carro, J.M., and A. Traferri. (2012). "State Dependence and Heterogeneity in Health Using a Bias-Corrected Fixed Effects Estimator." *Journal of Applied Econometrics*, forthcoming.

[8] Chamberlain., G. (1980). "Analysis of Covariance with Qualitative Data." *Review of Economic Studies*, 47(2): 225–238.

[9] Contoyannis, P., Jones, A. M., and N. Rice. (2004). "The dynamics of health in the British Household Panel Survey." *Journal of Applied Econometrics*, 19(4): 473–503.

[10] Contoyannis, P., and N. Rice. (2001). "The impact of health on wages: Evidence from the British Household Panel Survey" *Empirical Economics*, 26(4): 599–622

[11] Crawford, D. L., Pollak, R. A., and F. Vella. (1998). "Simple inference in multinomial and ordered logit." *Econometric Reviews*, 17(3): 289–299.

[12] Currie, J., and B.C. Madrian. (1999). Health, health insurance and the labor market. Handbook of labor economics, 3, 3309–3416.

[13] Deaton, A. (2002). "Inequality, Race, and Health. "http://www.nber.org/reporter/spring03/health.html

[14] Deaton, A. S., and C.H. Paxson. (1998). "Aging and inequality in income and health." *The American Economic Review*, 88(2): 248–253.

[15] Dhaene, Geert, and Koen Jochmans. (2010) "Split-panel jackknife estimation of fixed-effect models." Unpublished manuscript.

[16] Etilé, F., and C. Milcent. (2006). "Income-related reporting heterogeneity in self-assessed health: evidence from France. " *Health Economics*, 15(9): 965–981.

[17] Fernández-Val, I. (2009). "Fixed effects estimation of structural parameters and marginal effects in panel probit models. " *Journal of Econometrics*, 150(1): 71–85.

[18] Fernández-Val, I., and F. Vella. (2011). "Bias corrections for two-step fixed effects panel data estimators. " *Journal of Econometrics*, 163(2): 144–162.

[19] Frijters, P., Haisken-DeNew, J. P., and M. A. Shields. (2005). "The causal effect of income on health: Evidence from German reunification." *Journal of Health Economics*, 24(5): 997–1017

[20] Greene, W.H. (2001). "Fixed and Random Effects in Nonlinear Models."NYU Working Paper No. EC-01-01.

[21] Hahn, J., and W. Newey. (2004). "Jackknife and analytical bias reduction for nonlinear panel models."*Econometrica*, 72(4): 1295–1319.

[22] Hausman, J. (1978). "Specification tests in econometrics. "*Econometrica*, 46(6): 1377–1398.

[23] Hausman, J., and W. Taylor. (1981). "Panel data and unobservable individual effects."*Econometrica*, 49(6): 1377—1398.

[24] Idler, E. L., and Y. Benyamini (1997). "Self-rated health and mortality: a review of twenty-seven community studies. "*Journal of Health and Social Behavior*, 38(1): 21–37.

[25] Jones, A. M., and J. Wildman. (2008). "Health, income and relative deprivation: Evidence from the BHPS. "*Journal of Health Economics*, 27(2): 308–324.

[26] Jones, A. M., and S. Schurer. (2011). "How does heterogeneity shape the socioeconomic gradient in health satisfaction? "*Journal of Applied Econometrics*, 26(4): 549–579.

[27] 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): 144–168.

[28] Mackenbach, J. P., Simon, J. G., Looman, C. W., and I. M.Joung. (2002). "Self-assessed health and mortality: could psychosocial factors explain the association?" *International Journal of Epidemiology*, 31(6): 1162–1168.

[29] Maurer, J., Klein, R., and Vella, F. (2011). "Subjective health assessments and active labor market participation of older men: evidence from a semiparametric binary choice model with nonadditive correlated individual-specific effects." *Review of Economics and Statistics*, 93(3): 764–774.

[30] Meer, J., Miller, D. L., and H.S. Rosen. (2003). "Exploring the health–wealth nexus. "*Journal of Health Economics*, 22(5): 713–730.

[31] Mundlak, Y. (1978). "On the pooling of time series and cross-section data."*Econometrica* 46: 69–85.

[32] Neyman, J., and E.L. Scott. (1948). "Consistent Estimates Based on Partially Consistent Observations. "*Econometrica* 16(1): 1–32.

[33] Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests. *Copenhagen: Danmarks Paedagogiske Institut.*

[34] Smith,J. (1999). "Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status" *The Journal of Economic Perspectives*, 13(2): 145–166

[35] Thomas,D. and J. Strauss. (1997). "Health and wages: evidence on men and women in urban Brazil" *Journal of Econometrics*, 77(1): 159–185.

[36] Van Doorslaer, E., and U. G.Gerdtham. (2003). "Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data."*Social Science and Medicine*, 57(9): 1621–1630.

[37] Vella, F., and M.Verbeek. (1999). "Two-Step Estimation of Panel Data Models with Censored Endogenous Regressors and Selection Bias. "*Journal of Econometrics*, 90(2): 239–263.

Table 1: **Summary of the literature**

Author	Country	Health measure	Years	Method used	Results
Carro and Traferri (2012)	UK	Self-assessed health (3 categories)	Panel 1991-2006	Panel approach: • dynamic - health measure in current period depends on health in the previous period; • ordered probit; • fixed individual effects; • fixed effects in cutoff points; • bias correction in the first order conditions.	• There is a substantial positive state dependence in self-assessed health even after controlling for unobserved heterogeneity; • Positive and significant effect of income on health; • Small magnitude of income marginal effects: the average marginal effects evaluated at different age and gender groups were never greater than 1.3 percentage points
Contoyannis et al.(2004)	UK	Self-assessed health (5 categories)	Panel 1991-1998	Panel approach: • dynamic - health measure in current period depends on health in the previous period; • ordered probit; • random individual effects; • allowed for the correlation between observed regressors and individual effects by reparametrizing individual effect as a function of means of observed regressors	• There is a substantial positive state dependence in self-assessed health and unobserved permanent heterogeneity; • Positive and often significant effect of income on health • Permanent income (mean income) has a much greater impact on self-assessed health than transitory income (current income)
Etile and Milcent (2006)	France	Self-assessed health (6 categories); synthetic clinical measure of health	Cross-section 2001	Cross-section approach: • employed ordered probit model for self-assessed health; • used clinical measure of health to decompose the effect of income on self-assessed health into clinical health and heterogeneity bias.	There is substantial income-related reporting heterogeneity in self-assessed measure of health: • income has a significant effect on subjective health through clinical health for individuals in the bottom of the income distribution who report poor health • income has a significant negative reporting effect on subjective health for the richest individuals reporting good health • individuals between medium labels of health (fair and good) are the most affected by the reporting heterogeneity
Frijters et al.(2005)	Germany	Self-assessed health (10 categories)	Panel 1984-2002	Panel approach: • Conditional fixed effects logit (Chamberlain 1980) • Control for income endogeneity by using unification of Germany as natural experiment	• Income has a significant positive effect on health, but the magnitude of the effect is very small
Jones and Schurer (2011)	Germany	Self-assessed health (5 categories)	Panel 1984-2005	Panel approach: • Conditional fixed effects logit (Chamberlain 1980) • Robustness checks: pooled ordered logit, random-effects logit • Income is interacted with several age groups	• Find that imposing homogeneous relationship between income and health satisfaction is too restrictive • Controlling for individual specific characteristics decreases impact of income on health self-assessment
Jones and Wildman (2008)	UK	Dichotomized measure of self-assessed health; Measure of psychological well-being (score)	Panel 1991-2001	Panel approach: • Include income and relative income in the regressors • Random effects • Robustness checks: correlate random effects with observed regressors (Mundlak 1978 and Hausman and Taylor 1981) • Robinson semi-parametric estimation of income effect	• Positive significant effect of income on health, but small in magnitude; • Negative effect of relative deprivation on health, but the significance depends on specification
Lindahl (2005)	Sweden	Standardized Index of Bad Health based on 48 self-reported questions about health symptoms; Mortality within five and ten years	Panel 1968, 1974, and 1981	Cross-section approach: • OLS of health in 1981 on average income (including lottery winnings) between 1967 and 1981 and other covariates measured in 1968. • IV (2SLS) of the same specification as above where the instrument is average lottery winnings between 1969 and 1981	10 percent in family income • improves health by 4-5 percent of a standard deviation • decreases the probability of dying by 2-3 percentage points
Meer et al.(2003)	USA	Dichotomized measure of self-assessed health	Panel 1984, 1989, 1994, 1999	Panel approach: • Probit of health on change in household wealth, initial wealth, past health and other controls • Change in wealth is treated as endogenous, instrument – inheritance/ large gifts received within last five years • Does not control for individual heterogeneity	• Find that the effect of change in wealth on health is small in the magnitude • Impact of change in wealth on health in the short run becomes insignificant if endogeneity of wealth is addressed • Do not rule out long term impact of wealth on health

Table 2: **Distribution of health status by year**

year/health status		1	2	3	4	5
2001	N	11	199	918	1,047	328
	%	0.44	7.95	36.68	42	13.1
2002	N	17	192	893	1,075	326
	%	0.68	7.67	35.68	43	13.02
2003	N	15	200	930	1,058	300
	%	0.6	7.99	37.16	42.27	11.99
2004	N	15	215	924	1,046	303
	%	0.6	8.59	36.92	41.79	12.11
2005	N	17	216	910	1,047	313
	%	0.68	8.63	36.36	41.83	12.5
2006	N	19	204	907	1,055	318
	%	0.76	8.15	36.24	42.15	12.7
2007	N	15	194	892	1,061	341
	%	0.6	7.75	35.64	42.39	13.62
2008	N	11	216	916	1,005	355
	%	0.44	8.63	36.6	40.15	14.18

Table 3: **Distribution of health status on the 1 to 4 scale by year: subsample of individuals whose health status changed in categories 1 and 4 in the whole sample and each 7-year subsample**

year/health status		1	2	3	4
2001	N	156	917	1,045	238
	%	6.62	38.92	44.35	10.1
2002	N	158	889	1,068	241
	%	6.71	37.73	45.33	10.23
2003	N	165	925	1,056	210
	%	7	39.26	44.82	8.91
2004	N	178	921	1,043	214
	%	7.56	39.09	44.27	9.08
2005	N	181	907	1,044	224
	%	7.68	38.5	44.31	9.51
2006	N	171	903	1,048	234
	%	7.26	38.33	44.48	9.93
2007	N	160	885	1,057	254
	%	6.79	37.56	44.86	10.78
2008	N	177	910	998	271
	%	7.51	38.62	42.36	11.5

Table 4: **Summary statistics, pooled data: whole sample**

Variable	Obs	Mean	Std. Dev.	Min	Max
Weekly HH income, AUD in 2005 prices	18848	1501	1380	-13006	43354
Hourly individual wages, AUD in 2005 prices	15204	25.34	17.85	0.31	904.16
Male	18848	0.56	0.50	0	1
Married	18848	0.65	0.48	0	1
Age, years	18848	42.41	10.10	17	65
Education, years	18848	13.78	2.54	0	18
Number of kids	18848	0.99	1.17	0	8
Employed	18848	0.99	0.12	0	1
Experience, years	18552	25.92	10.27	0.08	51
Tenure, years	18579	8.58	8.52	0.02	50
Union	18848	0.33	0.47	0	1

Table 5: **Summary statistics, pooled data: men**

Variable	N	Mean	Std. Dev.	Min	Max
Weekly HH income, AUD in 2005 prices	10512	1547	1232	-13006	23101
Hourly individual wages, AUD in 2005 prices	8357	26.84	18.27	0.31	904.16
Married	10512	0.70	0.46	0	1
Age, years	10512	42.39	9.95	17	65
Education, years	10512	13.85	2.42	0	18
Number of kids	10512	1.07	1.22	0	8
Employed	10512	0.98	0.12	0	1
Experience, years	10400	25.77	10.19	0.5	50
Tenure, years	10350	9.06	8.95	0.02	49
Union	10512	0.32	0.47	0.00	1

Table 6: **Summary statistics, pooled data: women**

Variable	N	Mean	Std. Dev.	Min	Max
Weekly HH income, AUD in 2005 prices	8336	1443	1546	-1739	43354
Hourly individual wages, AUD in 2005 prices	6847	23.52	17.15	0.53	816.40
Married	8336	0.59	0.49	0	1
Age, years	8336	42.44	10.29	17	65
Education, years	8336	13.68	2.69	0	18
Number of kids	8336	0.88	1.09	0	5
Employed	8336	0.99	0.11	0	1
Experience, years	8152	26.10	10.37	0.08	51
Tenure, years	8229	7.97	7.91	0.02	50
Union	8336	0.33	0.47	0.00	1

Table 7: **Average weekly household income, AUD in 2005 prices**

year	women	men	all
2001	1263	1356	1315
2002	1407	1514	1467
2003	1400	1467	1438
2004	1417	1502	1465
2005	1502	1587	1550
2006	1512	1640	1584
2007	1531	1640	1592
2008	1510	1671	1599

Table 8: **Average individual hourly wages, AUD in 2005 prices**

year	women	men	all
2001	21.75	24.88	23.50
2002	22.23	25.16	23.83
2003	22.58	25.46	24.17
2004	22.90	25.68	24.43
2005	23.82	27.12	25.65
2006	24.08	27.68	26.03
2007	25.64	28.26	27.07
2008	25.07	30.49	28.07

Table 9: **Estimation results: income exogenous**

Dep Var: Health (1-4)	Pooled oprobit (1)	RE oprobit (2)	FE oprobit no correction (3)	FE oprobit Hahn and Newey (4)
real hh income, hndrd AUD	0.003	0.002	-0.0002	-0.0003
	(0.002)	(0.001)	(0.001)	(0.001)
education	0.032	0.048	-0.008	-0.008
	(0.007)	(0.009)	(0.019)	(0.019)
married dummy	0.103	0.087	0.085	0.076
	(0.037)	(0.039)	(0.050)	(0.049)
age	-0.010	-0.010	0.002	0.001
	(0.002)	(0.002)	(0.004)	(0.004)
number of children	-0.009	0.031	0.045	0.040
	(0.014)	(0.015)	(0.020)	(0.019)
male dummy	-0.147	-0.233		
	(0.034)	(0.053)		
employed	0.037	-0.088	-0.101	-0.107
	(0.093)	(0.086)	(0.089)	(0.087)
c1	1.376		2.287	2.062
	(0.157)		(0.025)	(0.023)
c2	2.789		4.659	4.188
	(0.158)		(0.035)	(0.031)
C1		-2.101		
		(0.192)		
C2		-0.015		
		(0.191)		
C3		2.132		
		(0.192)		
rho		0.559		
		(0.009)		
Region	Yes	Yes	No	No
Number of observations	2,356	2,356	2,356	2,356

Table 10: **Average marginal effects for real household income (hundred AUD)**

ME	Pooled Oprobit	RE Oprobit	FE Oprobit
$\bar{\mu_1}$	-0.0005	-0.0002	0.00003
	(0.0003)	(0.0001)	(0.0001)
$\bar{\mu_2}$	-0.0009	-0.0003	0.00004
	(0.0004)	(0.0002)	(0.0002)
$\bar{\mu_3}$	0.0008	0.0003	-0.00003
	(0.0005)	(0.0002)	(0.0001)
$\bar{\mu_4}$	0.0006	0.0002	-0.00003
	(0.0004)	(0.0001)	(0.0001)

Table 11: **Conditional probability of self-reported health status evaluated at different income percentiles**

Percentiles/ probabilities	$mean(P(y=1 \mid (x_{it}^{-inc}\beta + alpha_i + inc_{jth}\beta^{inc}))$	$mean(P(y=2 \mid (x_{it}^{-inc}\beta + alpha_i + inc_{jth}\beta^{inc}))$	$mean(P(y=3 \mid (x_{it}^{-inc}\beta + alpha_i + inc_{jth}\beta^{inc}))$	$mean(P(y=4 \mid (x_{it}^{-inc}\beta + alpha_i + inc_{jth}\beta^{inc}))$
5th	0.082	0.375	0.431	0.112
10th	0.082	0.375	0.431	0.112
15th	0.082	0.375	0.431	0.112
20th	0.082	0.375	0.431	0.112
25th	0.082	0.375	0.431	0.112
30th	0.082	0.375	0.431	0.112
35th	0.082	0.375	0.431	0.112
40th	0.082	0.375	0.431	0.112
45th	0.082	0.376	0.431	0.112
50th	0.082	0.376	0.431	0.112
55th	0.082	0.376	0.431	0.112
60th	0.082	0.376	0.431	0.112
65th	0.082	0.376	0.431	0.112
70th	0.082	0.376	0.431	0.112
75th	0.082	0.376	0.431	0.112
80th	0.082	0.376	0.431	0.111
85th	0.082	0.376	0.430	0.111
90th	0.082	0.376	0.430	0.111
95th	0.082	0.376	0.430	0.111

Table 12: **Conditional probability of self-reported health status evaluated at different fixed effects percentiles**

Percentiles/ probabilities	α^{Pth}	$mean(P(y=1\|$ $(x_{it}\beta+\alpha^{jth}))$	$mean(P(y=2\|$ $(x_{it}\beta+\alpha^{jth}))$	$mean(P(y=3\|$ $(x_{it}\beta+\alpha^{jth}))$	$mean(P(y=4\|$ $(x_{it}\beta+\alpha^{jth}))$
5th	0.271	0.424	0.545	0.031	0.000
10th	0.763	0.247	0.668	0.085	0.000
15th	1.025	0.173	0.694	0.133	0.001
20th	1.158	0.141	0.695	0.163	0.001
25th	1.350	0.102	0.682	0.214	0.002
30th	1.569	0.069	0.647	0.281	0.004
35th	1.714	0.051	0.613	0.330	0.005
40th	1.944	0.031	0.546	0.412	0.010
45th	2.130	0.020	0.483	0.480	0.017
50th	2.344	0.012	0.407	0.553	0.028
55th	2.525	0.007	0.343	0.608	0.041
60th	2.693	0.005	0.286	0.651	0.058
65th	2.893	0.002	0.224	0.688	0.085
70th	3.050	0.002	0.181	0.705	0.113
75th	3.207	0.001	0.143	0.711	0.145
80th	3.419	0.000	0.101	0.700	0.199
85th	3.601	0.000	0.072	0.674	0.254
90th	3.951	0.000	0.035	0.588	0.377
95th	4.372	0.000	0.013	0.445	0.542

Table 13: **Estimation results: endogenous income, first stage**

Dep var: HH real income, hndrd AUD	Pooled	RE	FE
	(1)	(2)	(3)
education	0.844	0.710	-0.121
	(0.137)	(0.081)	(0.149)
married dummy	5.606	2.365	0.759
	(0.475)	(0.306)	(0.358)
tenure	0.228	0.137	0.100
	(0.052)	(0.034)	(0.035)
$tenure^2$	-0.006	-0.004	-0.003
	(0.002)	(0.001)	(0.001)
experience	-0.018	0.036	0.274
	(0.176)	(0.062)	(0.078)
$experience^2$	0.001	0.003	0.002
	(0.004)	(0.001)	(0.001)
union	0.020	0.880	1.180
	(0.372)	(0.242)	(0.262)
male dummy	0.354	0.782	
	(0.501)	(0.486)	
σ_u		10.671	
σ_e		8.021	
ρ		0.639	
Region	Yes	Yes	No
Number of observations	2,138	2,138	2,138

Table 14: **Estimation results: endogenous income, second stage**

Dep var: Health (1-4)	Pooled oprobit	Pooled oprobit RE in the first stage	RE oprobit RE in the first stage	FE oprobit no correction	FE oprobit Hahn and Newey correcti
	(1)	(2)	(3)	(4)	(5)
real hh income, hndrd AUD	-0.005	-0.019	-0.022	-0.020	-0.029
	(0.021)	(0.020)	(0.023)	(0.027)	(0.026)
correction term	0.008			0.020	0.028
	(0.021)			(0.027)	(0.026)
correction, $\overline{e}_{it}$		0.019	0.021		
		(0.020)	(0.023)		
correction, $\overline{e}_{i}$		0.023	0.028		
		(0.020)	(0.023)		
education	0.038	0.045	0.066	0.010	0.005
	(0.019)	(0.015)	(0.017)	(0.023)	(0.022)
married dummy	0.156	0.158	0.145	0.109	0.123
	(0.124)	(0.063)	(0.066)	(0.057)	(0.055)
age	-0.009	-0.005	-0.003	0.010	0.010
	(0.002)	(0.004)	(0.005)	(0.011)	(0.011)
number of children	-0.011	-0.012	0.029	0.047	0.042
	(0.015)	(0.015)	(0.016)	(0.021)	(0.020)
male dummy	-0.152	-0.136	-0.235	2.292	2.066
	(0.037)	(0.040)	(0.056)	(0.027)	(0.024)
region	Yes	Yes	Yes	No	No
Number of observations	2,138	2,138	2,138	2,138	2,138

Table 15: **Estimation results: regression of FE estimates on independent variables**

VARIABLES	FEhat	FEhat
	(1)	(2)
male dummy	-0.207***	-0.218***
	(0.051)	(0.051)
balance of NSW	-0.195*	-0.112
	(0.100)	(0.099)
Melbourne	-0.044	-0.057
	(0.091)	(0.089)
balance of Victoria	-0.075	0.044
	(0.118)	(0.117)
Brisbane	-0.155	-0.103
	(0.109)	(0.107)
balance of QLD	-0.370***	-0.295***
	(0.103)	(0.102)
Adelaide	-0.228*	-0.156
	(0.124)	(0.122)
balance of SA	-0.325**	-0.255
	(0.161)	(0.159)
Perth	-0.036	-0.014
	(0.118)	(0.116)
Balance of WA	-0.312*	-0.177
	(0.175)	(0.173)
Tasmania	-0.275*	-0.220
	(0.163)	(0.160)
Northern Territory	-0.259	-0.201
	(0.282)	(0.278)
ACT	0.101	0.066
	(0.191)	(0.187)
years of education		0.054***
		(0.010)
married dummy		0.094
		(0.068)
age		-0.017***
		(0.003)
real hh income, hndrd AUD		0.008***
		(0.002)
number of kids		-0.070***
		(0.026)
employment dummy		0.638
		(0.418)
Constant	2.566***	1.774***
	(0.074)	(0.451)
Observations	2,356	2,356
R-squared	0.017	0.056

Note: Standard errors in parentheses, $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.