The University Of Sheffield.

Department Of Economics.

Sheffield Economic Research Paper Series.

Mental Health and Reporting Bias: Analysis of the GHQ - 12

Sarah Brown, Mark N. Harris, Preety Srivastava and Karl Taylor

ISSN 1749-8368

SERPS no. 2018013

September 2018

www.sheffield.ac.uk/economics

Mental Health and Reporting Bias: Analysis of the $GHQ - 12$ *

Sarah Brown,[1,#] Mark N. Harris,[2] Preety Srivastava[3] and Karl Taylor[1,#]
[1]Department of Economics, University of Sheffield;
[2]School of Economics and Finance, Curtin University;
[3]School of Economics, Finance and Marketing,
Royal Melbourne Institute of Technology;
[#]and IZA Bonn.

September 2018

Abstract

Measures of mental wellbeing are heavily relied upon to identify at-risk individuals. However, self-reported mental health metrics might be unduly affected by mis-reporting (perhaps stemming from stigma effects). In this paper we consider this phenomenon using data from the British Household Panel Survey (*BHPS*) and its successor, Understanding Society, the UK Household Longitudinal Study (*UKHLS*) over the period 1991 to 2016. In particular, in separate analyses of males and females we focus on the $GHQ-12$ measure, and specifically its sub-components, and how inaccurate reporting can adversely affect the distribution of the index. The results suggest that individuals typically over-report pschological wellbeing and that reporting bias is greater for males. The results are then used to adjust the composite $GHQ - 12$ score to take such mis-reporting behaviours into account. To further illustrate the importance of this, we compare the effects of the adjusted and unadjusted $GHQ-12$ index when modelling a number of economic transitions. The results reveal that using the original $GHQ - 12$ score generally leads to an underestimate of the effect of psychological distress on transitions into improved economic states, such as unemployment into employment.

JEL Classification: C3, D1, I1

Keywords: $GHQ - 12$ index, inflated outcomes, mental health and mis-reporting.

*We are grateful to the Data Archive, University of Essex, for supplying the British Household Panel Surveys, waves 1 to 18, and Understanding Society, waves 1 to 7. We are also grateful to Raslan Alzuabi for excellent research assistance and would like to thank Daniel Gray, Arne Rise Hole and Jennifer Roberts for value comments. Funding from the Australian Research Council is kindly acknowledged. The normal disclaimer applies.

1 Introduction

Five of the fifteen leading causes of disability worldwide are psychiatric conditions (Mathers et al., 2008). Mental disorders have become a global public health concern with the World Health Organization (WHO) predicting that one out of four people will endure some kind of mental illness during their life (WHO, 2001), and that the global economic burden of such mental disorders will be of the order of US$16 trillion between 2011 and 2030 (Bloom et al., 2011). Mental illness thus represents an immense psychological, social and economic burden to society and additionally, increases the risk of physical illnesses such as heart disease and diabetes (Stein et al., 2006).

The availability of psychological health alongside physical health information in large-scale individual- and household-level datasets, such as the British Household Panel Survey (*BHPS*) and its successor, Understanding Society, the UK Household Longitudinal Study (*UKHLS*), has enabled researchers to investigate a broad spectrum of policy areas such as employment, education and crime that are impacted by mental illness and *vice versa*. Information on psychiatric health has been collected in such datasets using different diagnostic measurement instruments such as the Kessler Score, the Composite International Diagnostic Interview ($CIDI$), the General Health Questionnaire (GHQ) and the Mental Health Inventory (MHI), which is a subscale of the Short Form-36 ($SF-36$).

The GHQ, for example, covers various dimensions, including depression, anxiety, somatic symptoms, feelings of incompetence, difficulty in coping and sleep disturbance, which are either self- or interviewer-administered, with each item measured using a 4-point Likert-type scale. The accuracy of the information is dependent on respondents providing reliable and accurate responses. It is very likely the case, however, that because of the social stigma associated with adverse mental health (for example, Hinshaw, 2009), respondents have a perceived incentive to mis-report the true status of their health. For example, Bharadwaj et al. (2017) find that survey respondents are significantly more likely to under-report mental illnesses (compared to other health conditions) because of the fear of being stigmatised, socially sanctioned or disgraced.

Mis-reporting leads to information being mis-classified in survey data, which can mask the incidence of such behaviours and lead to biased and inconsistent estimates in statistical analyses (Hausman et al., 1998). Although very little work has been undertaken in analysing the possible extent and consequences of inaccurate reporting in empirical models, its presence has been well-established in the psychology and related literatures. For example, social desirability has been found to be significantly associated with the over-reporting of physical activity and height, and the under-reporting of weight among women (Adams et al., 2005; Ezzati et al., 2006; Hebert et al., 2002).

There is, however, a distinct shortage of research exploring the implications in empirical analyses of utilising such mis-reported survey information, despite the widespread use of such data across the social sciences. That is, very little is known about the reliability of the *GHQ* measure, in identifying at-risk individuals; or, how reliable the effects of mental wellbeing are on a wide range of economic, health and social outcomes, given the presence of mis-reporting. Given the vast amount of resources invested in secondary data collection via large scale surveys, such as the *BHPS* and *UKHLS*, it is obvious that any analysis into the potential implications of reporting behaviour of sensitive information, such as mental wellbeing considered here, will be of importance to both current and future research right across the social and health sciences. It is imperative that the potential for, and implications of, mis-reporting in such data and empirical analyses be fully recognised. Otherwise policy prescriptions based on erroneous conclusions are likely, which may temper effectiveness as a result.

Take, as an example, mental health disorders which are typically present in around 30% of outpatients in primary care settings and are becoming more prevalent and consequently likely to contribute to growing healthcare costs over time; see Schmitz et al. (1999). The $GHQ-12$ is often used in primary care settings to screen patients for non-psychotic and minor psychiatric disorders (for example, Banks and Jackson, 1982; Cooper et al., 1988; Picardi et al., 2000) and was adopted as a screening tool in an international study undertaken by the WHO (Goldberg et al., 1997). Hence, given the above arguments (primarily related to self-validation and stigma effects) mis-reporting is also very likely to occur in such an environment where the $GHQ-12$ is used as a screening tool, and this will have knock-on effects not only in under-estimating the extent of mental health problems but also ultimately in the associated healthcare costs.

The objective of the current paper is to develop a latent-class type modelling approach to analyse the extent of mis-reporting in health instruments that are self-reported. In this study, we focus on the $GHQ-12$ (and its sub-components) which has been widely used to explore a range of important areas such as: education (Cornaglia et al., 2015; Gardner and Oswald, 2002); employment (Boyce and Oswald, 2012; Thomas et al., 2005); financial behaviour (Brown et al., 2005); gambling (Gardner and Oswald, 2007); housing (Ratcliffe, 2015); stock prices (Ratcliffe and Taylor, 2015); transport (Roberts et al., 2011); mortality (Russ et al., 2012; Gardner and Oswald, 2004); crime (Dustmann and Fasani, 2015); and income inequality (Wildman, 2003).

Although survey mis-classification, or mis-reporting, is known to be pervasive and to potentially bias statistical or econometric analyses, there is a limited body of research that has explicitly modelled such behaviours. A key study on mis-classified dependent variables is by Hausman et al. (1998). They consider a binary choice model with two types of mis-classification: the probability that the true 0 is recorded as a 1; and the probability that the true 1 is recorded as a 0, implying that the mis-classification errors are conditionally independent of covariates. A number of other studies have followed, extending this research in terms of semi-parametric estimation (Abrevaya and Hausman, 1999; Lewbel, 2000), the use of ordered data (Dustmann and Van Soest, 2001) and modelling mis-classification as a function of both observables and unobservables (Meyer and Mittag, 2017).

More recently, researchers such as Mahajan (2006), Hu (2008) and Molinari (2008) have attempted to model mis-classification in discrete dependent variables using a secondary measurement or an instrument to identify a nonlinear model. Their approach is based on the assumption that in the presence of classification errors, the relationship between the true variable and its mis-classified representation is given by a linear system of simultaneous equations in which the coefficient matrix is the matrix of mis-classification probabilities. Mis-classification resulting from anchoring, focal point answers and crude rounding in surveys has also increasingly been a subject of interest to researchers (for example, Van Soest and Hurd, 2008; Manski and Molinari, 2010; Kleinjans and Van Soest, 2014). For example, using a random effects multinomial logit model, Kleinjans and Van Soest (2014) explicitly account for such reporting behaviour including non-response where respondents decide not to report any value. Lastly, anchoring vignettes have also been used to measure discrepancies in reporting behaviours, particularly in the case of self-reported health and life satisfaction (Kristensen and Johansson, 2008; Van Soest et al., 2011); however, vignettes are very rare in most large scale data sets.

Our methodology ties in with the literature on latent class type models. Our basic starting hypothesis is that there are inherently two types of individuals in the population with regard to how they respond to particular survey questions of interest: "accurate" and "inaccurate". However, we will never directly observe to which type, or (latent) class, a respondent belongs. Thus, the broad approach we follow is that of latent class or finite mixture models (for a comprehensive review, see McLachlan and Peel (2004)), where our hypothesised classes correspond to these two types of individuals. In latent class modelling, the researcher aims to split the population according to high/low (healthcare, say) users, for example, even with observationally equivalent usage levels. Therefore, a novelty of our approach is to adopt these widely used and accepted techniques to help us identify and quantify any potential inaccurate reporting.

Explicitly, we offer researchers some generic tools with which to account for, and quantify, the effect of any mis-reporting behaviour in large scale surveys. We show how these can be applied to the important area of mental health, and in particular, the commonly used $GHQ-12$ instrument. We then show how these results can be used to identify potential questions of interest that may be particularly subject to mis-reporting, and to also adjust the index so as to obtain a more realistic distribution of the population's mental health over time. Finally, we illustrate how the use of these corrected indices can affect inference regarding the effects of mental wellbeing on several important individual economic outcomes, such that one could draw erroneous policy implications by ignoring these mis-reporting behaviours.

2 The 12-item General Health Questionnaire

The General Health Questionnaire (GHQ) is a self-administered psychometric screening tool that was developed with the aim to detect and assess individuals with a diagnosable psychiatric disorder (Goldberg and Hillier, 1979; Goldberg and Williams, 1988; McDowell, 2006). It was designed to cover four identifiable elements of distress: depression; anxiety; social impairment; and hypochondria. The original questionnaire had 60 items ($GHQ-60$) from which shorter versions of 30 items ($GHQ-30$), 28 items ($GHQ-28$), 20 items ($GHQ-20$) and 12 items ($GHQ-12$) have been constructed.

The 12-item General Health Questionnaire ($GHQ-12$) is the most widely used screening instrument for common mental disorders, in addition to being a more general measure of psychiatric wellbeing. Its brevity makes it attractive for use and its psychometric properties have been studied in various countries (Werneke et al., 2000) and with various types of population, for example, elderly people (Costa et al., 2006).

The $GHQ-12$ has twelve items stemming from the following questions: *'Here are some questions regarding the way you have been feeling over the past few weeks. Have you recently...'; (1) 'been able to concentrate on whatever you're doing?'; (2) 'lost much sleep over worry?'; (3) 'felt that you were playing a useful part in things'; (4) 'felt capable of making decisions about things?'; (5) 'felt constantly under strain?'; (6) 'felt you couldn't overcome difficulties?'; (7) 'been able to enjoy your normal day-to-day activities'; (8) 'been able to face up to problems?'; (9) 'been feeling unhappy or depressed'; (10) 'been losing confidence in yourself?'; (11) 'been thinking of yourself as a worthless person?'; and (12) 'been feeling reasonably happy, all things considered'.* The questions comprise six *"positively"* worded ($GHQs$ 1, 3, 4, 7, 8 & 12) and six *"negatively"* worded ($GHQs$ 2, 5, 6, 9, 10 & 11) sub-items to describe different mood states, see Hu et al. (2007). The responses to each of the twelve questions lie on a four-point Likert scale ranging from 0 to 3. The likert scale of each sub-component is scored so that higher values indicate decreased levels of psychological wellbeing. The $GHQ-12$ score converts valid answers to the 12 items to a single scale by recoding 0 and 1 values (more than usual and same as usual, respectively) on individual sub-items to 0, and 2 and 3 values (less than usual and much less than usual, respectively) to 1, and then summing, giving a scale running from 0 (the least distressed) to 12 (the most distressed).[1]

3 Econometric framework

The main purpose of this study is to determine if there is any bias in the composite $GHQ-12$ measure, which is increasingly being used as a wellbeing measure in economics studies; see, for example, Clark (2003), Roberts et al. (2011), and Cornaglia et al. (2015). Since this is a simple construct from the 12 underlying items or components (by summing the 12 individual 0/1 scores), obvious (related) questions are: *are any of these 12 questions in particular, subject to mis-reporting bias?* And *what is the extent of any mis-reporting bias across these 12 questions?* Hence, any bias in the overall index, must arise from mis-reporting or bias in some, or all, of the composite $GHQ-12$ components. Explicitly, the hypothesis is that, due to stigma and related effects, a proportion of individuals will erroneously over report zero scores in the components (corresponding to an original value

[1] See https://www.gl-assessment.co.uk/products/general-health-questionnaire-ghq/.

of 0 or 1, in the *likert* index; see above). Then, due to the summary composition of the overall index, this hypothesised behaviour in the components will lead to item inflation in the composite, and most likely at 0 in the 0-12 score.

Accordingly, the econometric framework developed here consists of modelling the individual components that, in sum, describe the composite measure. The aim here is to model any potential mis-reporting in these individual components. In doing this, it will be possible to identify particular questions that are more likely to be adversely affected by mis-reporting behaviours. It will also allow us, post-estimation, to construct a new composite $GHQ-12$ index by systematically correcting the 12 individual components where we find the probability of mis-reporting to be high.

A casual inspection of the distribution of the composite $GHQ-12$ measure (see Figure 1) clearly illustrates, as expected, a marked spike at zero; and indeed at a magnitude (apparently) completely at odds with the remainder of the distribution. Indeed, zero values in this composite instrument are important as "a score of zero on the $GHQ-12$ questionnaire can, in contrast (*to a score of more than 4*), be considered to be an indicator of psychological wellbeing" (Scottish-Government, 2013). It is our contention that such a large relative representation of psychological wellbeing, may be an over-representation of the true state of affairs. As noted, the hypothesis is that there is a subset of the population who erroneously identify themselves into this (favourable) category by reporting a 0 or 1 score (on the Likert scale) in all 12 individual components. The reasons for this will presumably be wide and varied across this sub-population, but may result from a desire to appear aligned with social norms and to avoid any associated stigma effects of being identified as having either an actual, potential, or perceived, psychiatric disorder. Thus, we require an econometric model that allows for an "inflation" of the zero outcome in each sub-component. That is, we wish to distinguish "true" zero responses from the "false" ones; or equivalently, to allow for two different types of zero observations, following the latent class literature.

In a similar context, Brown et al. (2018) consider the modelling of illicit drug participation. There, the basic hypothesis is that, due to very similar reasons to those considered in the current paper, a subset of true illicit drug participants, will actually mis-report their true behaviour. Thus, the participation rates often reported in sample survey data, are likely to be an under-estimation of the true ones. Their set-up appears, once more, very well-suited to the problem at hand: We wish to model binary outcome variables for each of the separate 12 questions, where 1 relates to a score of 2 or 3 on the likert scale and a value of 0 relates to a score of 0 or 1 on the likert scale, whereby we believe that, in some cases at least, an excess of zeros is recorded (which is directly analogous to the excess of zeros in drug non-participation in the Brown et al. (2018) approach). Another relevant contribution is Greene et al. (2015), who consider responses to self-assessed health questions based on a 5-point *likert*-scale (1-5), which, on first inspection, appear to be "inflated" in the (binary) *good* and *very good* categories. As with the current paper, their hypothesis is that a subset of the population mis-reports into these favourable categories, again arguably predominantly for reasons of social norms, stigma and opportunity costs of time.

In such a set-up, there are two equations driving the eventual observed outcome. Firstly, a latent variable, $\widetilde{y}_q^*$, is specified that represents the true health status related to the q^{th} question for each of the $q = 1, \ldots, 12$ questions. $\widetilde{y}_q^*$ is a function of variables $\mathbf{z}$ with unknown weights $\boldsymbol{\gamma}_q$, and a standard-normally distributed error term (as is commonly assumed in the literature), u_q, such that

$$\widetilde{y}_q^* = \mathbf{z}'\boldsymbol{\gamma}_q + u_q. \tag{1}$$

This translates into a discrete variable $\widetilde{y}_q$ where $\widetilde{y}_q = 1$ for $\widetilde{y}_q^* > 0$ and $\widetilde{y}_q = 0$ for $\widetilde{y}_q^* \leq 0$. Secondly, and as above, there is a equation which relates to the individual's propensity to report accurately, represented by r_q^* (where $q = 1, 2, \ldots, 12$). Again, this is specified as a function of variables $\mathbf{x}$ with unknown weights $\boldsymbol{\beta}_q$, where there may be some overlap between $\mathbf{x}$ and $\mathbf{z}$, and an error term ε_q, such that

$$r_q^* = \mathbf{x}'\boldsymbol{\beta}_q + \varepsilon_q. \tag{2}$$

The observability criterion for observed y_q is now

$$y_q = \widetilde{y}_q \times r_q. \tag{3}$$

Allowing for the likely correlation between ε_q and u_q $\left(\rho_q\right)$, full probabilities are given by

$$\Pr(y_q) = \begin{cases} \Pr\left(y_q = 0 \left| \mathbf{x} \right.\right) = \left[1 - \Phi\left(\mathbf{z}'\boldsymbol{\gamma}_q\right)\right] + \Phi_2\left(\mathbf{z}'\boldsymbol{\gamma}_q, -\mathbf{x}'\boldsymbol{\beta}_q; -\rho_q\right) \\ \Pr\left(y_q = 1 \left| \mathbf{x} \right.\right) = \Phi_2\left(\mathbf{x}'\boldsymbol{\beta}_q, \mathbf{z}'\boldsymbol{\gamma}_q; \rho_q\right). \end{cases} \tag{4}$$

So here, the probability of a zero observation has been "inflated" as it is a combination of the probability of a "true" 0-score from the health equation plus the probability of an "inaccurately" reported one from the splitting probit model (which we refer to as an *inflated probit* model). Again, once the assumed form of the probabilities is known and observations on $y_{i,q}$ are available in an *i.i.d.* sample of size N from the population, the parameters of the full model $\boldsymbol{\theta}_q = \left(\boldsymbol{\beta}'_q, \boldsymbol{\gamma}'_q, \rho_q\right)'$ can be consistently and efficiently estimated using maximum likelihood (ML) techniques. Note that in theory, it would be possible to also allow the equations across the q questions to be correlated. However, that would require evaluation of multivariate normal distributions of some 2×12 dimensions, therefore we only allow correlations within each q.[2] The likelihood function for a single component (q) is therefore

$$
\begin{aligned}
L_i(\boldsymbol{\theta}) &= \prod_{j=0}^{j=1} \Pr\left(y_i = j \left| \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}\right.\right), \ j = 0, 1 & (5)\\
&= P_i & (6)
\end{aligned}
$$

As argued in Brown et al. (2018), it is generally preferable to have exclusion restrictions across both $\mathbf{x}$ and $\mathbf{z}$, although this is not strictly necessary for formal identification. However, Greene et al. (2015) use a set of potential variables that may be able to more strongly identify the "mis-reporters". Essentially, these fall into two separate blocks: variables that uniquely identify the health equation (equation 1); and variables that identify the mis-reporting one (equation 2). With no other strong priors concerning the remainder of the variables available to the researcher, Greene et al. (2015) suggest entering these in both equations. This is the broad approach we adopt in the following analysis, although we do include some identifying covariates in ($\mathbf{x}$), as discussed below.

In the analysis that follows, we have panel data to hand: that is, for each individual i, we have repeated observations over time periods $t = 1, \ldots, T_i$. Formulating the above model in this context allows one to easily account for unobserved individual heterogeneity in both underlying equations, $\boldsymbol{\alpha}$ (in each of the q components). As is standard in the literature, it is assumed that $\boldsymbol{\alpha} \sim N(0, \Sigma)$; and we denote the individual elements of Σ by $\widetilde{y}_q^*$ and r_q^*, respectively. Since the presence of such unobserved effects complicates evaluation of the resulting likelihood function, we utilise the method of maximum simulated likelihood. Dropping the q subscript for ease of notation, we can define $\mathbf{v}_i$ as a vector of standard normal random variates, which enter the model generically as $\boldsymbol{\Gamma}\mathbf{v}_i$, such that for a single draw of $\mathbf{v}_i$, $\boldsymbol{\Gamma}\mathbf{v}_i = (\alpha_{i,\widetilde{y}^*}, \alpha_{i,r^*})$. $\boldsymbol{\Gamma}$ is the *chol* (Σ) such that $\Sigma = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$. Conditioned

[2]Moreover, this would only yield, arguably minor, efficiency gains if such correlations existed.

on $\mathbf{v}_i$, the sequence of T_i outcomes for individual i are independent, such that the contribution to the likelihood function for a group of t observations is defined as the product of the sequence P_{it} - see equation (6) - which we denote e_i, corresponding to the observed outcome of y_i, $e_i \mid \mathbf{v}_i$,

$$e_i \mid \mathbf{v}_i = \prod_{t=1}^{T_i} \left(P_{it} \mid \mathbf{v}_i\right). \tag{7}$$

The unconditional log-likelihood function is found by integrating out the $\mathbf{v}_i$ as

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \int_{v_i} \prod_{t=1}^{T_i} \left(P_{it} \mid \boldsymbol{\Gamma}\mathbf{v}_i\right) f(\mathbf{v}_i) d\mathbf{v}_i, \tag{8}$$

where all parameters of the model are contained in $\boldsymbol{\theta}$. Using the usual assumption of multivariate normality for $\mathbf{v}_i$ yields

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \int_{v_i} \prod_{t=1}^{T_i} \left(P_{it} \mid \boldsymbol{\Gamma}\mathbf{v}_i\right) \prod_{k=1}^{K} \phi(\mathbf{v}_{ik}) d\mathbf{v}_{ik}, \tag{9}$$

where k indexes the different unobserved effects in the model (so here, $K = 2$ per q). The expected values in the integrals can be evaluated by simulation by drawing R observations on $\mathbf{v}_i$ from the multivariate standard normal population. The following is the resulting simulated log-likelihood function

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} \left(P_{it} \mid \boldsymbol{\Gamma}\mathbf{v}_i\right). \tag{10}$$

Halton sequences of length $R = 100$ were used (see Train, 2009), and this now feasible function is maximized with respect to $\boldsymbol{\theta}$.

As is common in the non-linear panel data literature, given that these unobserved heterogeneity terms are (potentially) correlated with observed heterogeneity terms, the correction proposed by Mundlak (1978) is applied. Consequently we include averages of the continuous covariates of individual i in the set of explanatory variables, $\overline{x_i} = \frac{1}{T_i} \sum_{t=1}^{T_i} x_{it}$.

4 Data

We use the British Household Panel Survey (*BHPS*), a survey conducted by the Institute for Social and Economic Research, which is a large scale representative longitudinal study collecting data on individuals over the period 1991 to 2008. It is household-based and interviews every adult member of sampled households. In 1991 the sample comprised around 5,500 households and over 10,000 individuals living in 250 areas of Great Britain. We also employ the successor to the *BHPS*, Understanding Society - the UK Household Longitudinal Study (*UKHLS*) - which is a nationally representative longitudinal study of the UK population which started in 2009. In the first wave of the *UKHLS*, over 50,000 individuals were interviewed over the period 2009 to 2011 and correspondingly in the latest wave available, wave 7, around 45,000 individuals were interviewed between 2015 and 2017 (hereafter referred as 2016). Both the *BHPS* and *UKHLS* contain detailed information on economic and socio-demographic characteristics. It is possible to track individuals from the *BHPS* into the *UKHLS* hence making a relatively long panel dataset.

We focus upon two unbalanced panels over the period 1991 to 2016 split by gender, where the total number of observations for males is 115,976 comprising 14,378 individuals aged 18 or above, and the respective figures for females are 140,263 observations comprising 16,240 individuals. Males are observed, on average, 14 times over a quarter of a century whilst the corresponding figure for females is 15 times. The percentage of individuals, by gender, present in all periods (*i.e.*, across the 25 years) is 6.7% (7,776 males) and 7.6% (10,704 females). In part of the interview, respondents are asked to complete the self-completion $GHQ-12$ questionnaire. This measure of mental wellbeing is available in both the *BHPS* and the *UKHLS* and has been used to examine a range of policy areas such as education, employment and crime (as noted above).

Figure 1 shows the distribution of the $GHQ-12$ for all individuals and split by gender and Table 1 provides summary statistics for the $GHQ-12$ and its sub-components, by gender. From Figure 1 it is clear that there is around a 10 percentage point differential across gender in reporting complete psychological wellbeing (i.e a score of 0), with it being lower for females. It is also apparent from Figure 1 that over 50% of the sample report none of the above (component) problems, whilst Table 1 reveals that the average number of problems is 1.5 for males compared with 2 for females. Around 13% of males and 19% of females in the sample report in excess of four problems over the period 1991-2016. Considering the elements of the $GHQ-12$, the most common problem faced by individuals is feeling constantly under strain, *i.e.*, 23% for males and just under 30% for females, followed by around 18% of males and 25% of females feeling unhappy or depressed. Interestingly, Table 1 reveals that, across each of the $GHQ-12$ sub-components, problems are more prevalent for females.

In terms of the explanatory variables in both **x** and **z**, we control for: the age of the individual (entered as a quadratic); married or cohabiting (other states constitute the reference group); white; highest educational attainment, specifically a degree, teaching or nursing qualification, A levels, GCSE (or O level), other qualifications (no education is the omitted group); the natural logarithm of labour income last month; the natural logarithm of non-labour income last month; employment status (employed, self-employed or unemployed; other states make up the reference group); housing tenure, specifically whether the home is owned outright, via a mortgage, or rented (other tenure states form the reference category); region of residence; and year of interview. The variables used to model the sub-components of the $GHQ - 12$, given in the vector **z**, follow the existing literature (e.g. Metcalfe et al., 2011).

In addition, we also control for the general health of the individual in **z**. The *BHPS* and *UKHLS* both contain a question on self-assessed health (SAH): '*Please think back over the last 12 months about how your health has been. Compared to people of your own age, would you say that your health has on the whole been excellent/good/fair/poor/very poor?*' However, due to reporting bias and measurement error, the reported SAH may be endogenous in the subsequent analyses. To accommodate this possibility, we follow Stern (1989) and Bound (1991), for example, by conditioning the SAH on a set of instruments namely whether the individual reports specific health problems.[3] The logic here is that more objective measures are used to instrument the endogenous and potentially error ridden subjective health measure. Following the literature, we estimate the health stock of an individual by employing a Generalised Ordered Probit (GOP) model, which allows for the fact that people with the same underlying level of health may apply different thresholds when reporting SAH and hence different ordered categories for similar positions on the assumed underlying continuous scale (Rice et al., 2010; Lindeboom and Van Doorslaer, 2004; Kerkhofs and Lindeboom, 1995). We then take the linear prediction from the GOP model as a measure of an individual's health stock, where higher values denote worse health.

In the vector **x**, we include two additional covariates to identify mis-reporting. Firstly,

[3] Individuals are asked whether they have any of the following health problems: arms, legs or hands; sight; hearing; skin conditions or allergies; chest or breathing; heart or blood pressure; stomach or digestion; diabetes; anxiety or depression; alcohol or drugs; epilepsy or migraine; any other problem.

the percentage of compulsory questions (*i.e.*, those asked to everyone completing the survey) not answered in the individual questionnaire. The idea here is that those individuals who complete a smaller proportion of questions, perhaps because they have a lower level of trust in the survey, will a priori be more likely to answer less accurately. This is consistent with the approach of Brown et al. (2018) and is based on existing literature which suggests that the longer a respondent spends time with the interviewer the more trusting they are of both him/her and the survey in general; see, for example, Corbin and Morse (2003). Secondly, we also condition on whether there is a change in interviewer over time (*i.e.*, between waves) following Niccoletti and Peracchi (2005) and Jenkins et al. (2008). The logic behind the use of this control is similar to the above, in that interviewer continuation is associated with respondent trust, interviewer reputation and rapport with the respondent, and hence continued survey participation over time (see Schrapler (2004) and Vassallo et al. (2015)).[4]

Table 2 provides summary statistics for the control variables used in the empirical analysis, split by gender. Approximately 25% of males and females have either a degree or A-level qualifications as their highest educational attainment, 55% of males are in paid employment compared to 50% of females, around 11% of males are self-employed relative to 4% of females, males have higher earnings from employment but, on average, have lower non-labour income (benefits, child support, *etc.*) than females, 28% own their home outright, whilst around 45% own a home via a mortgage. In terms of the controls in $\mathbf{x}$ used to identify the mis-reporting equation, approximately 30% of respondents experienced a change in interviewer over time, whilst for both males and females roughly 2.5% of the compulsory questions are not answered on the individual questionnaire (the maximum is a third uncompleted for males and over half for females).

5 Results

We estimate the random-effects inflated probit models for each sub-component of the $GHQ-12$, separately for males and females. The estimated coefficients are reported in Tables 3 to 6. The results for males for sub-components $GHQ1$ to $GHQ6$ are shown in Table 3 and for sub-components $GHQ7$ to $GHQ12$ in Table 4. The corresponding results

[4]Note that interviewers in the *BHPS* and *UKHLS* are randomly allocated to respondents the first time that a household appears in the survey and are hence independent of respondent characteristics.

for females are shown in Table 5 for $GHQ1$ to $GHQ6$, and in Table 6 for $GHQ7$ to $GHQ12$. The upper panel of each table reports the coefficients (with corresponding robust standard errors) for the likelihood of reporting a specific problem and the lower panel shows the coefficients (with corresponding standard errors) from the inaccurate reporting equation, where a positive sign denotes a higher probability of accurate reporting.

We find that a number of socio-demographic covariates are associated with the twelve sub-components in terms of the likelihood of reporting a problem and also mis-reporting. For example, older individuals, regardless of gender, are more likely to report problems of: sleep loss; capability of decision making; facing up to problems; and feeling reasonably happy. There also subtle differences in the effect of age between males and females, with older females being more likely to report a problem (where statistically significant). The influence of education is mixed, but typically any significant influence upon problems occurs at higher levels of attainment. For both males and females, education appears to be important for reporting problems of being unhappy or depressed, although the direction of the effect differs across gender with problems increasing in educational attainment (relative to having no qualifications) in the former whilst decreasing in the latter. Income effects are apparent for a number of the sub-elements of the $GHQ-12$ and it is noticeable that problems faced by males are more likely to be influenced by labour income, perhaps signifying greater attachment to the labour market. For males, both higher labour and non-labour income are associated with a lower likelihood of reporting problems with usefulness, capability and losing confidence. Employees (the unemployed) are generally less (more) likely to report a problem compared to those individuals who are out of the labour market. Higher values of the health stock measure denote worse health and, hence, not surprisingly for both males and females a worse health stock is associated with a higher likelihood of reporting each type of problem. Turning to the instruments used to identify inaccurate reporting behaviour, we find that the percentage of (compulsory) questions left unanswered in the questionnaire increases the respondent's propensity to report inaccurately, which is also generally true of changes in the interviewer over time (where significant), which is consistent with our a priori expectations. Moreover, in general, the correlation between the mental health and mis-reporting equations, ρ_q, is statistically significant for each sub-component ($q = 1, \ldots, 12$).[5]

[5] We also estimate a range of partial effects on the marginal probabilty of reporting a problem (e.g. each of the 12 sub-components) and on the respective marginal probability of inaccurate reporting. Due

Of particular importance to the current study, Tables 7 and 8 present summary probabilities for males and females, respectively. These provide insights on the extent of mis-reporting (reporting bias). Column 1 presents the sample proportion of reported psychiatric distress as indicated by survey responses. Using the estimated models, the predicted rates of psychological distress are presented in Column 2 and the resulting reporting bias in Column 3. To be specific, the elements in Column 2, are obtained by evaluating the expression $\Phi\left(\mathbf{z}_i'\widehat{\gamma}_q\right)$ in the first line of equation (4), which corresponds to the "true" probability of psychological distress, in the absence of any reporting bias effects, and averaged over individuals. It should be noted that the standard errors of all of these probability estimates are very small, giving us confidence in their estimated magnitudes.[6] Comparing the numbers in Columns 1 to 2 provides the reporting bias numbers in Column 3, reported as a percentage. These numbers generally indicate statistically significant under-reporting in most of the 12 sub-components of the $GHQ-12$, with the most significant bias of 148% and nearly 185% estimated for $GHQ3$ (*usefulness*), for males and females, respectively. The predicted rates more than doubled for several other sub-components amongst males, such as $GHQ1$ (*concentration*), $GHQ5$ (*strain*), $GHQ10$ (*confidence*) and $GHQ11$ (*worthless*). In comparison, there were generally lower reporting biases among females.

In Column 4, the marginal probabilities of mis-reporting are presented for the 12 elements which largely reflect the results in Column 3, with the highest probability of mis-reporting for $GHQ3$ (*usefulness*) and $GHQ11$ (*worthless*) for males, and $GHQ3$ (*usefulness*) and $GHQ7$ *(enjoying activities)* for females. Lastly we present two sets of posterior probabilities in Columns 5 and 6. As noted above, zero observations come from two sources: mis-reporters; and accurate reporters with a true 0-score. Using posterior probabilities that are conditional on knowing what outcome the individual chooses (we re-visit this below), we can also make a prediction on what percentage of the zeros come from mis-reporters and accurate reporters with a true 0-score respectively, using all the information we have on the individual. All the posterior probabilities again appear to be accurately estimated (with respect to their very small standard errors), with the sub-elements $GHQ1$ *(concentration)*, $GHQ5$ *(strain)* and $GHQ7$ (*enjoyment*) being subject to the greatest amount of mis-reporting in males, and the sub-elements $GHQ1$

to space constraints the results are not reported herein but are available from the authors on request.

[6] Standard errors of all secondary quantities are estimated using the Delta method.

(concentration), $GHQ3$ *(usefulness)* and $GHQ7$ *(enjoyment)*, in females.[7]

6 Adjusting the $GHQ-12$ index

As a natural extension of the above analyses, in this section we show how the results can be used to adjust the $GHQ-12$ index in light of the estimated amount of mis-reporting. We do this on the basis of the estimated posterior probabilities. We favour these as opposed to prior probabilities as they use all the information available on an individual, and should therefore provide more accurate predictions.

On the basis of these posterior probabilities, as noted above, we can make a prediction on what percentage of the reported zeros are related to a true zero-outcome and to mis-reporting, respectively. These are similar to probabilities estimated in latent class models (Greene, 2012) and essentially attempt to answer the question: *given that an individual recorded a zero, what is the probability that they are a mis-reporter versus an accurate reporter with a genuine* 0*-score (given their observed characteristics)?* The posterior probabilities for the two types of zeros for each sub-component q ($q = 1, \ldots, 12$) are given as

$$Pr\left(\widetilde{y}_q = 0|\mathbf{x}, y_q = 0\right) = \frac{f(\widetilde{y}_q = 0|\mathbf{x})}{f(y_q = 0|\mathbf{x})} \tag{11}$$
$$= \frac{1 - \Phi\left(\mathbf{z}'\boldsymbol{\gamma}_q\right)}{\left[1 - \Phi\left(\mathbf{z}'\boldsymbol{\gamma}_q\right)\right] + \Phi_2\left(\mathbf{z}'\boldsymbol{\gamma}_q, -\mathbf{x}'\boldsymbol{\beta}_q; -\rho\right)}$$

$$Pr\left(\widetilde{y}_q = 1, r_q = 0|\mathbf{x}, y_q = 0\right) = \frac{f(\widetilde{y}_q = 1, r_q = 0|\mathbf{x})}{f(y_q = 0|\mathbf{x})} \tag{12}$$
$$= \frac{\Phi_2\left(\mathbf{z}'\boldsymbol{\gamma}_q, -\mathbf{x}'\boldsymbol{\beta}_q; -\rho\right)}{\left[1 - \Phi\left(\mathbf{z}'\boldsymbol{\gamma}_q\right)\right] + \Phi_2\left(\mathbf{z}'\boldsymbol{\gamma}_q, -\mathbf{x}'\boldsymbol{\beta}_q; -\rho\right)}.$$

We estimate the posterior probability of mis-reporting (at an individual level) for each of the twelve components of the $GHQ-12$ (that is, evaluating equation 12). Next, we assign

[7] Our findings relate to existing literature which has found that the $GHQ-12$ sub-components measure both positive and negative mental health dimensions: Hu et al. (2007) explore whether interdependence exists between the two domains. Indeed, our results suggest that mis-reporting bias is generally larger in the case of positively worded questions (namely $GHQs$ 1, 3, 4, 7, 8 & 12). Considering the third column of Tables 7 and 8, the average reporting bias for males (females) across positively worded questions is 89% (77%) compared to 78% (18%) for negatively worded components. Hence, the contrast in the bias between positively and negatively worded sub-components is unambiguous in the case of females. This implies that the phrasing of questions is potentially an important factor in determining the extent of the reporting bias.

the estimated probabilities to individuals who reported a zero to the respective questions and were estimated to have a high posterior probability of mis-reporting. Following the convention with predicted success and failure in empirical work, we use the 0.5 cutoff rule. For example, if individual i's posterior probability of mis-reporting for a sub-component (say, $GHQ5$) is 0.61 (which is > 0.5), we contend that there is a (high) 61% chance that the zero recorded by individual i is mis-reported as against a 39% chance that it is a genuine zero-outcome. Thus, we adjust the zero in $GHQ5$ to 0.61 for individual i. Instead, if we estimate a (low) posterior probability of mis-reporting of 0.29 (which is $\leqslant 0.5$) for individual i, we treat the reported zero as a genuine outcome that does not require any adjustment. After so-adjusting the observed zeros, we then sum all of the 12 sub-components to construct an adjusted $GHQ - 12$ index. To make this adjusted measure comparable to the original index, we simply round the adjusted sum to the nearest integer.

The resulting indices for males and females are illustrated in Figures 2 and 3 respectively, in Panel A labelled *"adjusted"*. While the adjusted $GHQ-12$ indices clearly mimic the overall shape of the original indices, we can see a significant reduction in the frequency of the zeros, which have been predominantly reallocated to the neighbouring outcomes of $1, 2$ and 3. We next explore the robustness of our adjusted index with a slightly different approach. Using the same rule as before, here, where appropriate, we replace the zeros with a 1 instead of the respective posterior probabilities. We notice quite similar patterns in the respective adjusted $GHQ - 12$ index, lending confidence to our approach (shown in Panel B labelled *"robust"* in Figures 2 and 3). As a final exercise, we use the observed sample proportions of the respective sub-components as the cutoff rule to adjust the index. This could be regarded as an upper bound of the adjusted index and is shown in Panel C in Figures 2 and 3 (labelled *"upper bound"*), where for both males and females this measure clearly mimics the original $GHQ - 12$, and so would appear to be the least effective approach out of the three alternatives discussed at correcting for mis-reporting.

As highlighted above, scores in excess of 4 on the $GHQ - 12$ scale are taken to be possibly symptomatic of a mental health issue, as these are states distanced from psychological wellbeing, in contrast to a score of 4 or below (Scottish-Government, 2013). For the original $GHQ - 12$ composite measures, 12.7% in the males sample and 18.9% in the females sample reported a score greater than 4. Hence, females appear to be more psychologically distressed, which is also evident after conditioning upon covariates.

The comparable figures once the composite index has been adjusted using the posterior probabilities are 15.5% and 20.8%, respectively. Thus, the resulting distribution of the composite psychological wellbeing metric has a larger tail reporting states in excess of 4, with a clear gender difference.

6.1 Alternative adjustment approaches

There are a number of alternative strategies that could be used to adjust the $GHQ-12$. These could be more sample specific than the general exercises described above. Three such strategies could be conducted broadly as follows. Firstly, again based on posterior probabilities, it is possible to identify "problem" questions that comprise the 12 items of the composite $GHQ-12$. That is, it is possible to identify questions that appear to be unduly adversely affected by high posterior probabilities of mis-reporting, on average. These individual questions could then be removed from the construction of the overall index. Although an interesting line of research, especially for policy-makers and healthcare professionals, in being able to identify such problematic questions (that could potentially be re-worded, adjusted, or removed in future calculations and surveys), as opposed to the above approach, it is somewhat subjective in defining what level of probability would define "high" for any particular question to identify it as being "problematic". In addition, unlike the above approach, this adjusted index would no longer run from $0-12$, and so would not be directly comparable (unless scaled) with the current version.

Other possible approaches involve identifying "serial offender" respondents (as opposed to questions) and to remove these from the composite index, so as to reduce the zero-inflated bias in this. Firstly, one could look at the average posterior probability of mis-reporting (averaged over the 12 component questions). Then one could remove, say, the top 20% of individuals with the highest "average" mis-reporting probability. Secondly, one could compare individuals' posterior probabilities of mis-reporting in each component to the average across the sample. Those individuals estimated to be "greater" mis-reporters in the bulk of the components, could again be deemed to be "serial offenders" and therefore removed from the calculation of the index. However, as with the first approach, these two strategies rely on rather arbitrary rules for the splitting of the sample.

We have experimented with all such approaches and in general, the broad results were in line with those presented above. However, for the reasons outlined above, we prefer the approaches detailed in Section 6.

7 Applications using the adjusted metrics

In this section we consider applications of the adjusted $GHQ-12$ index to modelling transitions in some key economic outcomes, by focusing on how the mental health instrument is associated with changes in education, labour market status and savings between time $t-1$ and t. Specifically we examine increases in educational attainment ($t-1$ to t); changes from being unemployed or out of the labour force ($t-1$) to paid employment or self-employment (t), for individuals of working age; and changes in the incidence of saving ($t-1$ to t).

The change in the state of each outcome (s_{it}) between $t-1$ to t (Δ) is modelled as a binary outcome, equal to unity if the state improves over time, *i.e.,* an increase in educational attainment, moving out of unemployment into employment, switching from a non-saver to saver. Each outcome is conditioned on a quadratic in age, marital status, total income, housing tenure, year of interview and region of residence, given in vector $\mathbf{z}_{it-1}$.[8] We also control for whether the individual gave a wellbeing score different to zero at $t-1$. That is, for each economic outcome, we compare the effect of not reporting a zero, *i.e.*, exhibiting some psychiatric distress, for the composite $GHQ-12$ and the three alternative adjusted metrics detailed above. This is included as a binary variable, $g_{it-1}=1$, if $GHQ-12 \neq 0$. Each dependent variable is estimated as a panel probit model where μ_i is the individual specific random effect as follows.

$$\Delta s_{it} = \mathbf{1}\left[\mathbf{z}'_{it-1}\boldsymbol{\pi}+\lambda g_{it-1}+\mu_i+\varepsilon_{it-1}>0\right]. \tag{13}$$

The results are shown in Table 9 for males and Table 10 for females, where the first four columns focus on transitions in educational attainment, the next four consider labour market status and the final four columns focus upon transitions in financial behaviour, *i.e.*, whether the individual becomes a saver.[9] Each table provides specifications employing: (A) the original $GHQ-12$; (B) the *adjusted* index (labelled as "Adj. 1"); (C) the *robust* method (labelled as "Adj. 2"); and (D) the *upper bound* measure (labelled as "Adj. 3"),

[8] Following the literature (e.g. Cornaglia et al. (2015), Boyce and Oswald (2012)) additional controls are incorporated in $\mathbf{z}_{it-1}$, specifically: (i) the educational attainment models also condition on labour market status; (ii) the labour market status models include highest educational attainment; and (iii) the savings model includes both labour market status and highest educational attainment.

[9] When the results are based upon the adjusted metrics, *i.e.*, Columns 2 through to 4 for each outcome, given that these wellbeing measures are constructed from model estimates, the standard errors are bootstrapped using 200 replications.

as described above.[10] For brevity we only report the estimate of λ.[11]

The results show that, in general, individuals who report psychiatric distress, *i.e.*, a non-zero score derived from either the original $GHQ-12$ or one of the alternative adjusted measures, *i.e.* $g_{it-1} = 1$, have a lower likelihood of increasing educational attainment (which is consistent with Cornaglia et al. (2015)), moving into employment as previously reported in the literature (for example, Boyce and Oswald, 2012) and, finally, in line with existing literature, becoming savers (for example, Guven, 2012; Frey and Stutzer, 2002). Furthermore, what is particularly noticeable is that both males and females, who exhibit psychiatric distress based upon the adjusted metrics, have an even lower probability of increasing educational attainment. This is also evident for labour market transitions from unemployment into employment, *i.e.*, the negative effect of being in psychiatric distress is more pronounced using the adjusted index relative to using the unadjusted index for both males and females. Interestingly, being in psychiatric distress has similar effects upon the probability of becoming a saver for males across each alternative index with the magnitude of each coefficient being only marginally larger than that associated with the original unadjusted index. Indeed, the adjusted $GHQ - 12$ across the three methods is not significantly different from the original index. Conversely, for females, there are significant differences upon saving behaviour between the original $GHQ - 12$ index and the alternative approaches which allow for mis-reporting (see Table 10).

Moreover, what is also apparent is that the difference in the estimated parameters between the effects of psychiatric distress based upon the original $GHQ - 12$ and the alternative measures are generally statistically significant at the 5% level, as shown by the χ^2 statistics, with the magnitude of the coefficients typically being larger based upon the adjusted measures.[12] This is perhaps not surprising given the inflation observed at the left hand extreme of the subjective mental health distribution observed for both males and females. The results from these applications suggest that the over-reporting of the absence of psychological distress results in an under-estimate of the effect of psychiatric distress on transitions into improved economic states, such as employment and higher educational attainment. Such findings highlight the importance of allowing for potential

[10] Note that the number of observations is reduced as the focus is on the change in state of each outcome over time. Furthermore, the analysis of labour market status is based upon individuals of working age only (males up to 65 and females up to 60), whilst information on savings is not collected in waves 3, 5 and 7 of the *UKHLS*; subsequently this reduces the sample size when examining financial behaviour.

[11] Full results are available upon request.

[12] With the exception of male saving behaviour.

mis-reporting in mental health from a policy perspective.[13]

8 Conclusions

We have analysed the extent and implications of potential mis-reporting of mental health in the 12 sub-components of the $GHQ-12$, a very common and widely used measure of psychological wellbeing. Using data from the British Household Panel Survey and Understanding Society over the period 1991 to 2016, we have employed inflated (latent-class type) models to account for a preponderance of zeros reported in the 12-item questionnaire. We then used the posterior probabilities to adjust the $GHQ-12$ instrument. Importantly, the suggested approach is applicable to any health measure that is self-reported. The analysis shifts the distribution away from high mental wellbeing. In our applications based upon using the adjusted measures, we find that over-reporting no psychiatric distress is generally associated with under-estimating the effect of mental wellbeing on a number of economic transitions relating to educational attainment, employment and financial vulnerability, three areas of particular policy concern.

Furthermore, the $GHQ-12$ index was developed to screen for general (non-psychotic) psychiatric morbidity (Goldberg and Williams, 1988), and the finding that mis-reporting bias is associated with individuals over-estimating their state of mental wellbeing is of policy concern. Interestingly, our results show that older males (females) are more likely to mis-report on 7 (3) out of 12 sub-components of the $GHQ-12$, which given an ageing population is worrying, especially when such metrics are employed as screening tools in primary health care meaning that ultimately long-term health costs may be under-estimated.[14] The technique we use in this study can be used not only to correct the reported $GHQ-12$ but it can also help to: (i) identify those questions in the 12-item

[13] Alternatively, for each transition, we have conditioned on a binary variable, $g_{it-1}=1$, indicting whether the $GHQ-12$ score is in excess of 4 and re-estimated equation 13 for both males and females. For both genders, the results show that having a $GHQ-12$ score above 4 decreases the probability of moving to an improved economic state. However, in contrast to the results reported in Tables 9 and 10, there are no statistically significant differences between the binary variable based upon the original $GHQ-12$ index and the alternative definitions. This in part reflects the fact that the majority of the shift in the distribution of GHQ-12 based upon the alternative metrics occurs between values of 0 to 4, see Figures 2 and 3.

[14] Such costs are potentially not trivial, with a recent independent review for the UK government showing that the cost of poor mental health to the economy is between £74 and £99 billion per year. See https://www.gov.uk/government/publications/thriving-at-work-a-review-of-mental-health-and-employers

questionnaire of the $GHQ-12$ that are significantly mis-reported; and (ii) identify those respondents who systematically mis-report on most or all of the items. Hence it provides a range of measures to address the inherent measurement problem in the $GHQ-12$.

Countries such as the UK are collecting information at a national level on subjective wellbeing. Since 2011, the UK Office for National Statistics has routinely collected measures of subjective wellbeing in the large scale Integrated Household Survey (IHS). This has become particularly pertinent following the Commission on the Measurement of Economic Performance and Social Progress, (Stiglitz et al., 2009), and stems from concerns that traditional measures of living standards, for example, GDP per capita, do not adequately reflect economic and social progress. Hence, investigating mis-reporting of mental wellbeing and seeking alternative ways to take this into account is an important line of future enquiry, given the increasing prominence of wellbeing as an economic indicator.

References

Abrevaya, J. and J. A. Hausman (1999). Semiparametric estimation with mismeasured dependent variables: An application to duration models for unemployment spells. *Annales d'Economie et de Statistique 55–56*, 243–275.

Adams, S. A., C. E. Matthews, C. B. Ebbeling, C. G. Moore, J. E. Cunningham, J. Fulton, and J. R. Hebert (2005). The effect of social desirability and social approval on self-reports of physical activity. *American Journal of Epidemiology 161*(4), 389–398.

Banks, M. H. and P. R. Jackson (1982). Unemployment and risk of minor psychiatric disorder in young people: Cross-sectional and longitudinal evidence. *Psychological Medicine 12*(4), 789–798.

Bharadwaj, P., M. M. Pai, and A. Suziedelyte (2017). Mental health stigma. *Economic Letters 159*(1), 57–60.

Bloom, D. E., E. Cafiero, E. Jané-Llopis, S. Abrahams-Gessel, L. R. Bloom, S. Fathima, A. B. Feigl, T. Gaziano, A. Hamandi, M. Mowafi, A. Pandya, K. Prettner, L. Rosenberg, B. Seligman, A. Z. Stern, and C. Weinstein (2011). The global economic burden of noncommunicable diseases. Technical report, Geneva: World Economic Forum.

Bound, J. (1991). Self-reported versus objective measures of health in retirement models. *Journal of Human Resources 26*(1), 106–138.

Boyce, C. J. and A. J. Oswald (2012). Do people become healthier after being promoted? *Health Economics 21*(5), 580–596.

Brown, S., M. Harris, P. P. Srivastava, and X. Zhang (2018). Modelling illegal drug participation. *Journal of the Royal Statistical Society: Series A 181*(1), 133–154.

Brown, S., K. Taylor, and S. Wheatley-Price (2005). Debt and distress: Evaluating the psychological cost of credit. *Journal of Economic Psychology 26*(3), 642–663.

Clark, A. (2003). Unemployment as a social norm: Psychological evidence from panel data. *Journal of Labor Economics 21*(2), 323–352.

Cooper, P. J., E. A. Campbell, A. Day, H. Kennerley, and A. Bond (1988). Non-psychotic psychiatric disorder after childbirth: a prospective study of prevalence, incidence, course and nature. *The British Journal of Psychiatry 152*(6), 799–806.

Corbin, J. and J. Morse (2003). The unstructured interactive interview: Issues of reciprocity and risks when dealing with sensitive topics. *Qualitative Inquiry 9*(3), 335–354.

Cornaglia, F., E. Crivellaro, and S. McNally (2015). Mental health and education decisions. *Labour Economics 33*, 1–12.

Costa, E., S. M. Barreto, E. Uchoa, J. O. Firmo, M. F. Lima-Costa, and M. Prince (2006). Is the GDS-30 better than the GHQ-12 for screening depression in elderly people in the community? The Bambui Health Aging Study (BHAS). *International Psychogeriatrics 18*(3), 493–503.

Dustmann, C. and F. Fasani (2015). The effect of local area crime on mental health. *The Economic Journal 126*(593), 978–1017.

Dustmann, C. and A. Van Soest (2001). Language fluency and earnings: Estimation with misclassified language indicators. *Review of Economics and Statistics 83*(4), 663–674.

Ezzati, M., H. Martin, S. Skjold, S. Vander Hoorn, and C. J. Murray (2006). Trends in national and state-level obesity in the USA after correction for self-report bias: Analysis of health surveys. *Journal of the Royal Society of Medicine 99*(5), 250–257.

Frey, B. and A. Stutzer (2002). What can economists learn from happiness research? *Journal of Economic Literature 40*(2), 402–435.

Gardner, J. and A. Oswald (2002). How does education affect mental well-being and job satisfaction. *Economics Department Warwick University*.

Gardner, J. and A. Oswald (2004). How is mortality affected by money, marriage, and stress? *Journal of Health Economics 23*(6), 1181–1207.

Gardner, J. and A. J. Oswald (2007). Money and mental wellbeing: A longitudinal study of medium-sized lottery wins. *Journal of Health Economics 26*(1), 49–60.

Goldberg, D., R. Gater, N. Sartorius, T. Ustun, M. Piccinelli, O. Gureje, and C. Rutter (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine 27*(1), 221–230.

Goldberg, D. and P. Williams (1988). *A user's guide to the General Health Questionnaire*. NFER-Nelson.

Goldberg, D. P. and V. F. Hillier (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine 9*(1), 139–145.

Greene, W. (2012). *Econometric Analysis 7e* (seventh ed.). New Jersey, USA: Prentice Hall.

Greene, W., M. Harris, and B. Hollingsworth (2015). Inflated responses in measures of self-assessed health. *American Journal of Health Economics 1*(4), 461–493.

Guven, C. (2012). Reversing the question: Does happiness affect consumption and savings behaviour? *Journal of Economic Psychology 33*(4), 701–717.

Hausman, J. A., J. Abrevaya, and F. M. Scott-Morton (1998). Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics 87*(2), 239–269.

Hebert, J. R., C. B. Ebbeling, C. E. Matthews, T. G. Hurley, M. Yunsheng, S. Druker, and L. Clemow (2002). Systematic errors in middle-aged women's estimates of energy intake: Comparing three self-report measures to total energy expenditure from doubly labeled water. *Annals of Epidemiology 12*(8), 577–586.

Hinshaw, S. P. (2009). *The mark of shame: Stigma of mental illness and an agenda for change.* Oxford University Press.

Hu, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics 144*(1), 27–61.

Hu, Y., S. Stewart-Brown, L. Twigg, and S. Weich (2007). Can the 12-item General Health Questionnaire be used to measure positive mental health? *Psychological Medicine 37*(7), 1005–1013.

Jenkins, R., D. Bhugra, P. Bebbington, T. Brugha, M. Farrell, J. Coid, T. Fryers, S. Weich, N. Singleton, and H. Meltzer (2008). Debt income and mental disorder in the general population. *Psychological Medicine 38*(10), 1485–1493.

Kerkhofs, M. and M. Lindeboom (1995). Subjective health measures and state dependent reporting errors. *Health Economics 4*(3), 221–235.

Kleinjans, K. J. and A. Van Soest (2014). Rounding, focal point answers and nonresponse to subjective probability questions. *Journal of Applied Econometrics 29*(4), 567–585.

Kristensen, N. and E. Johansson (2008). New evidence on cross-country differences in job satisfaction using anchoring vignettes. *Labour Economics 15*(1), 96–117.

Lewbel, A. (2000). Identification of the binary choice model with misclassification. *Econometric Theory 16*(4), 603–609.

Lindeboom, M. and E. Van Doorslaer (2004). Cut-point shift and index shift in self-reported health. *Journal of Health Economics 23*(6), 1083–1099.

Mahajan, A. (2006). Identification and estimation of regression models with misclassification. *Econometrica 74*(3), 631–665.

Manski, C. F. and F. Molinari (2010). Rounding probabilistic expectations in surveys. *Journal of Business & Economic Statistics 28*(2), 219–231.

Mathers, C., D. Fat, and J. Boerma (2008). *The Global Burden of Disease: 2004 Update.* World Health Organization.

McDowell, I. (2006). *Measuring health: A guide to rating scales and questionnaires.* Oxford university press.

McLachlan, G. and D. Peel (2004). *Finite Mixture Models.* John Wiley & Sons.

Metcalfe, R., N. Powdthavee, and P. Dolan (2011). Destruction and distress: Using a quasi-experiment to show the effects of the September 11 attacks on mental well-being in the United Kingdom. *The Economic Journal 121*(550), 81–103.

Meyer, B. and N. Mittag (2017). Misclassification in binary choice models. *Journal of Econometrics 200*(2), 295–311.

Molinari, F. (2008). Partial identification of probability distributions with misclassified data. *Journal of Econometrics 144*(1), 81–117.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica 46*(1), 69–85.

Niccoletti, C. and F. Peracchi (2005). Survey response and survey characteristics: Micro level evidence from the European Community Household Panel. *Journal of the Royal Statistical Society: Series A 168*(4), 763–781.

Picardi, A., D. Abeni, C. Melchi, P. Puddu, and P. Pasquini (2000). Psychiatric morbidity in dermatological outpatients: An issue to be recognized. *British Journal of dermatology 143*(5), 983–991.

Ratcliffe, A. (2015). Wealth effects, local area attributes, and economic prospects: On the relationship between house prices and mental wellbeing. *Review of Income and Wealth 61*(1), 75–92.

Ratcliffe, A. and K. Taylor (2015). Who cares about stock market booms and busts? Evidence from data on mental wellbeing. *Oxford Economic Papers 67*(3), 816–845.

Rice, N., J. Roberts, and A. Jones (2010). Sick of work or too sick to work? Evidence on self-reported health shocks and early retirement from the BHPS. *Economic Modelling 27*(4), 866–880.

Roberts, J., R. Hodgson, and P. Dolan (2011). It's driving her mad: Gender differences in the effects of commuting on psychological health. *Journal of Health Economics 30*(5), 1064–1076.

Russ, T. C., E. Stamatakis, M. Hamer, J. M. Starr, M. Kivimäki, and G. D. Batty (2012). Association between psychological distress and mortality: Individual participant pooled analysis of 10 prospective cohort studies. *British Medical Journal 345*, 1–14.

Schmitz, N., C. Heckrath, L. Alberti, and W. Tress (1999). Diagnosing mental disorders in primary care: The General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments. *Social Psychiatry and Psychiatric Epidemiology 34*(7), 360–366.

Schrapler, J. P. (2004). Respondent behaviour in panel studies: A case study for income-nonresponse by means of the German-Socio Economic Panel (SOEP). *Sociological Methods and Research 33*(1), 118–156.

Scottish-Government (2013). Scottish health survey 2012 - volume 1 main report. Technical report, see http://www.scotland.gov.uk/Publications/2013/09/3684/5.

Stein, M. B., B. J. Cox, T. O. Afifi, S.-L. Belik, and J. Sareen (2006). Does co-morbid depressive illness magnify the impact of chronic physical illness? A population-based perspective. *Psychological Medicine 36*(5), 587–596.

Stern, S. (1989). Measuring the effect of disability on labor force participation. *Journal of Human Resources 24*(3), 361–395.

Stiglitz, J., A. Sen, and J.-P. Fitoussi (2009). Report by the commission on the measurement of economic performance and social progress. Technical report, www.stiglitz-sen-fitoussi.fr/.

Thomas, C., M. Benzeval, and S. A. Stansfeld (2005). Employment transitions and mental health: An analysis from the British Household Panel Survey. *Journal of Epidemiology and Community Health 59*(3), 243–249.

Train, K. E. (2009). *Discrete choice methods with simulation.* Cambridge University Press.

Van Soest, A., L. Delaney, C. Harmon, A. Kapteyn, and J. P. Smith (2011). Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions. *Journal of the Royal Statistical Society: Series A 174*(3), 575–595.

Van Soest, A. and M. Hurd (2008). A test for anchoring and yea-saying in experimental consumption data. *Journal of the American Statistical Association 103*(481), 126–136.

Vassallo, R., G. B. Durrant, and P. W. F. Smith (2015). Interviewer effects on non-response propensity in longitudinal surveys: A multilevel modelling approach. *Journal of the Royal Statistical Society: Series A 178*(1), 83–99.

Werneke, U., D. P. Goldberg, I. Yalcin, and B. Üstün (2000). The stability of the factor structure of the general health questionnaire. *Psychological Medicine 30*(4), 823–829.

WHO (2001). The World Health Report 2001: Mental Health: New Understanding, New Hope. http://www.who.int/whr/.

Wildman, J. (2003). Income related inequalities in mental health in Great Britain: Analysing the causes of health inequality over time. *Journal of Health Economics 22*(2), 295–312.

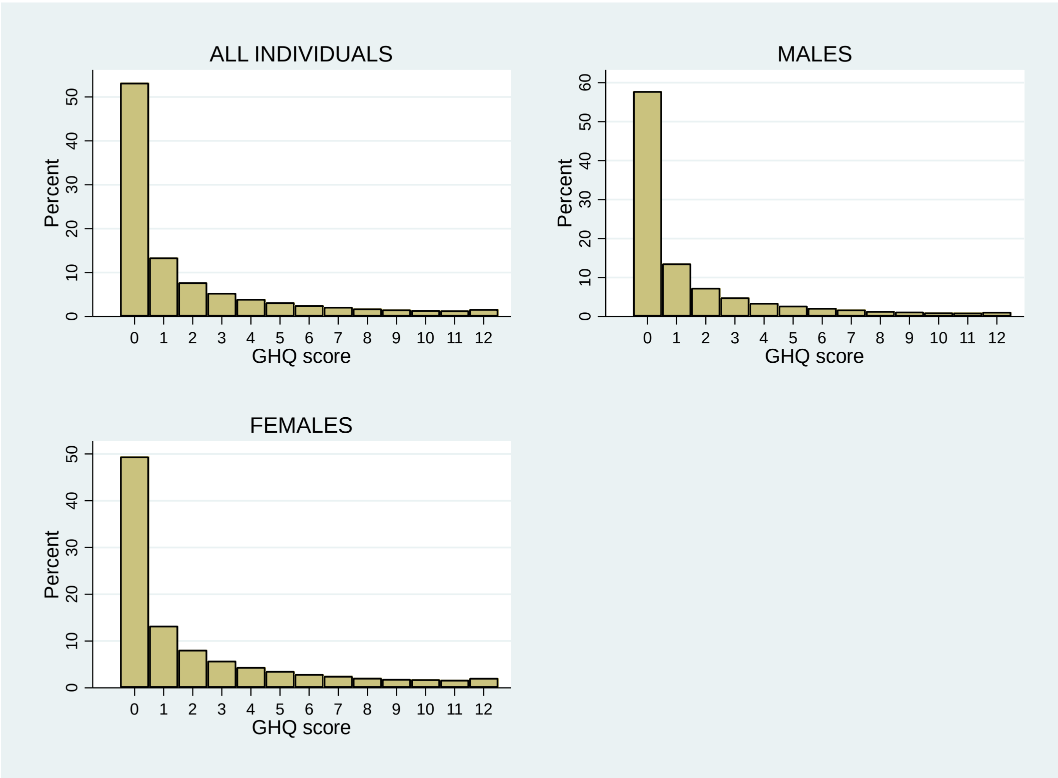

Figure 1: Distribution of the Composite $GHQ - 12$

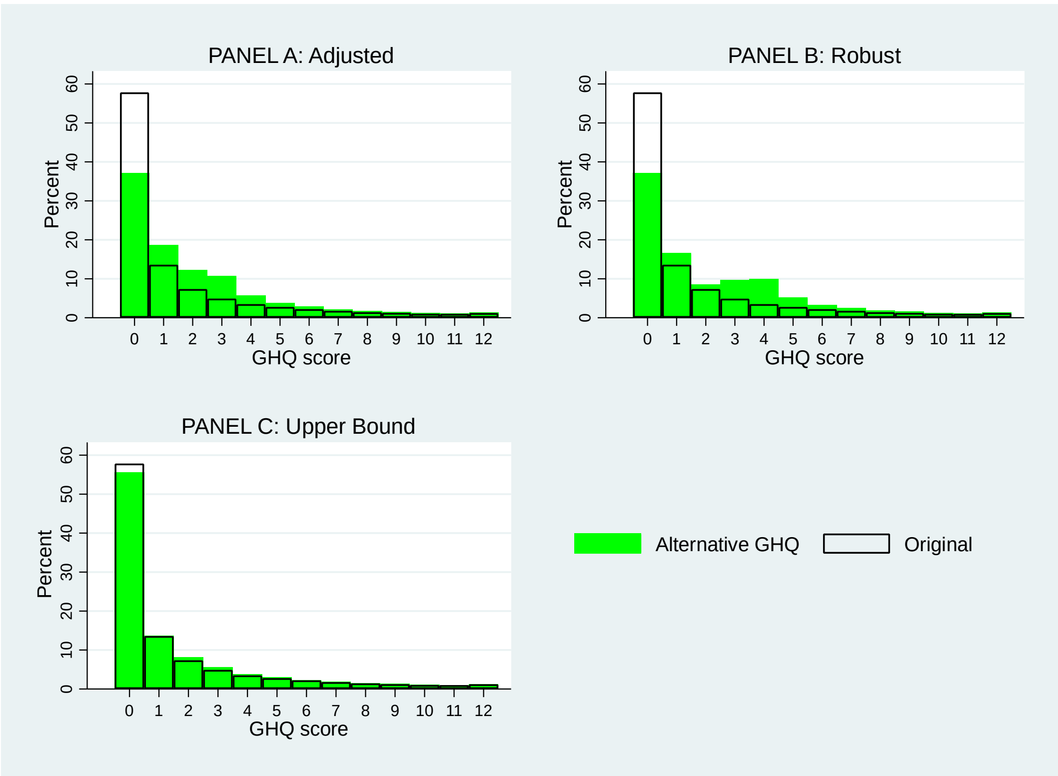

Figure 2: Males - Alternative $GHQ - 12$ index adjusted using Posterior Probabilities

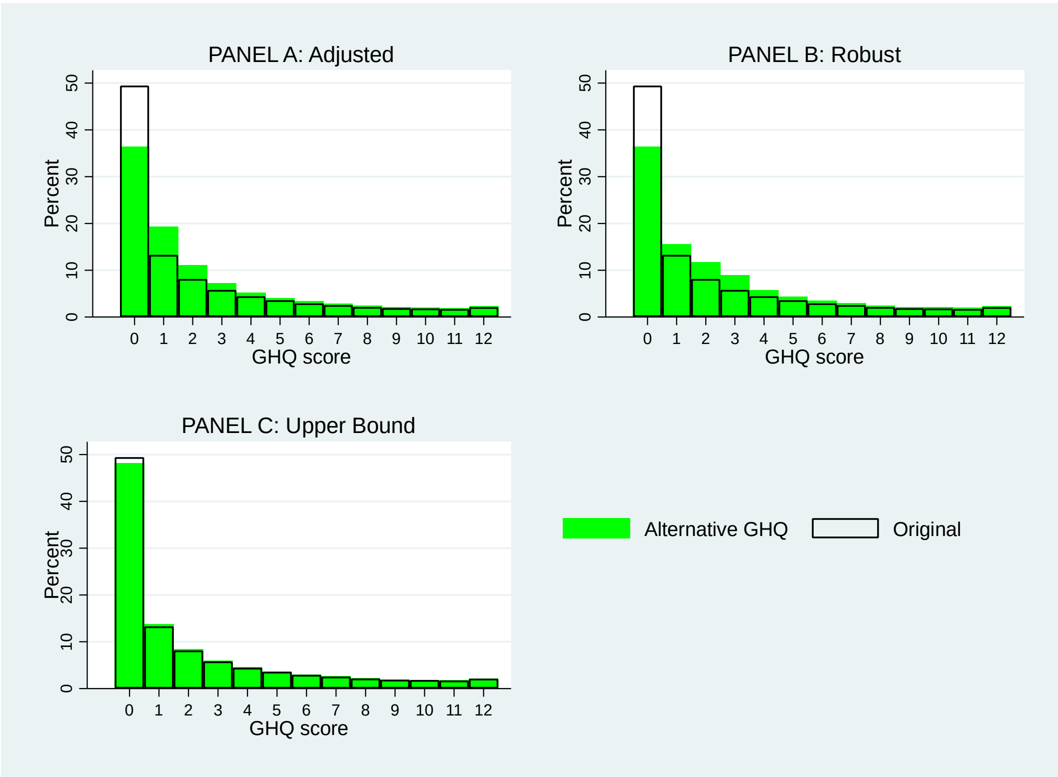

Figure 3: Females - Alternative $GHQ - 12$ index adjusted using Posterior Probabilities

Table 1: Summary Statistics: $GHQ-12$ and Binary Sub-components

	MALES				FEMALES			
	Mean	Standard Deviation	Minimum	Maximum	Mean	Standard Deviation	Minimum	Maximum
Overall $GHQ-12$ index	1.557	2.706	0	12	2.156	3.189	0	12
Sub-components of $GHQ-12$								
GHQ1 – concentration	0.155	0.362	0	1	0.217	0.412	0	1
GHQ2 – sleep loss	0.147	0.354	0	1	0.220	0.414	0	1
GHQ3 – usefulness	0.118	0.323	0	1	0.142	0.349	0	1
GHQ4 – capability	0.076	0.264	0	1	0.116	0.320	0	1
GHQ5 – strain	0.233	0.423	0	1	0.295	0.456	0	1
GHQ6 – overcoming difficulties	0.115	0.319	0	1	0.164	0.370	0	1
GHQ7 – enjoy activities	0.169	0.375	0	1	0.206	0.405	0	1
GHQ8 – face up to problems	0.086	0.280	0	1	0.136	0.343	0	1
GHQ9 – unhappy or depressed	0.175	0.380	0	1	0.241	0.428	0	1
GHQ10 – losing confidence	0.108	0.311	0	1	0.172	0.377	0	1
GHQ11 – worthless person	0.061	0.239	0	1	0.094	0.291	0	1
GHQ12 – feeling reasonably happy	0.114	0.318	0	1	0.154	0.361	0	1
Individuals (N)	14,378				16,240			
Observations ($N \times T$)	115,976				140,263			

Table 2: Summary Statistics: Explanatory Variables

	MALES				FEMALES			
	Mean	Standard Deviation	Minimum	Maximum	Mean	Standard Deviation	Minimum	Maximum
Age	46.388	17.697	18	100	47.187	18.191	18	100
Married or cohabiting	0.683	0.465	0	1	0.616	0.486	0	1
White	0.806	0.395	0	1	0.819	0.385	0	1
Degree	0.149	0.356	0	1	0.128	0.334	0	1
Teaching or nursing	0.281	0.449	0	1	0.242	0.428	0	1
A levels	0.132	0.338	0	1	0.113	0.317	0	1
O levels	0.152	0.359	0	1	0.180	0.384	0	1
Other qualifications	0.080	0.271	0	1	0.082	0.275	0	1
Log of monthly labour income	5.088	3.617	0	11.375	3.923	3.560	0	11.521
Log of monthly non labour income	3.501	3.057	0	12.694	4.522	2.628	0	12.513
Employed	0.546	0.498	0	1	0.500	0.500	0	1
Self-employed	0.113	0.317	0	1	0.038	0.191	0	1
Unemployed	0.051	0.219	0	1	0.027	0.161	0	1
Home owned outright	0.282	0.450	0	1	0.286	0.452	0	1
Home owned on a mortgage	0.471	0.499	0	1	0.435	0.496	0	1
Home rented	0.114	0.318	0	1	0.139	0.346	0	1
Health stock (linear prediction)	0.328	0.535	0	3.572	0.412	0.596	0	3.628
Change in interviewer *t-1* to t	0.318	0.466	0	1	0.320	0.467	0	1
% (compulsory) questions not answered in survey	2.657	3.410	0	32.934	2.745	3.436	0	56.581
Observations ($N \times T$)	115,976				140,263			

Table 3: Males - Estimated Coefficients of the Inflated Probit Model: GHQ1 to GHQ6

	GHQ1		GHQ2		GHQ3		GHQ4		GHQ5		GHQ6	
Index Function for Probit												
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
constant	3.552	(0.265)***	-1.578	(0.099)***	-0.808	(0.207)***	-2.865	(0.138)***	0.057	(0.016)***	-0.876	(0.056)***
Age10 (age divided by 10)	-0.162	(0.010)***	0.070	(0.006)***	-0.010	(0.002)***	0.016	(0.008)**	-0.015	(0.003)***	-0.004	(0.008)
Square of Age10	0.107	(0.008)***	-0.069	(0.005)***	0.011	(0.005)**	0.018	(0.008)***	0.014	(0.004)***	0.006	(0.004)
Married or cohabitating	-0.060	(0.042)	-0.138	(0.025)***	-0.164	(0.039)***	-0.170	(0.036)***	0.018	(0.012)	-0.085	(0.021)***
White	-0.033	(0.060)	-0.103	(0.033)***	-0.019	(0.032)	-0.112	(0.042)***	0.012	(0.031)	-0.076	(0.039)*
Degree	0.087	(0.171)	0.128	(0.101)	0.261	(0.091)***	-0.065	(0.152)	-0.259	(0.306)	-0.265	(0.147)*
Teaching or nursing	0.180	(0.123)	0.115	(0.081)	0.085	(0.145)	-0.026	(0.135)	-0.283	(0.063)***	-0.270	(0.080)***
A levels	0.273	(0.149)*	0.028	(0.085)	0.018	(0.052)	-0.335	(0.140)**	-0.393	(0.090)***	-0.413	(0.087)***
O levels	0.094	(0.136)	0.172	(0.091)*	-0.284	(0.159)*	-0.254	(0.148)*	-0.280	(0.221)	-0.373	(0.125)***
Other qualifications	0.085	(0.208)	0.092	(0.119)	-0.454	(0.126)***	-0.299	(0.190)	-0.353	(0057)***	-0.549	(0.214)***
Log of monthly labour income	0.001	(0.009)	-0.021	(0.006)***	-0.013	(0.001)***	-0.027	(0.008)*	0.002	(0.015)	0.001	(0.013)
Log of monthly non-labour income	-0.020	(0.007)***	-0.002	(0.004)	-0.003	(0.009)	-0.005	(0.005)	-0.006	(0.011)	-0.005	(0.007)
Employed	0.034	(0.068)	-0.169	(0.046)***	0.830	(0.690)	-0.166	(0.064)***	0.995	(0.157)***	0.411	(0.108)***
Unemployed	0.258	(0.063)***	0.159	(0.038)***	1.604	(0.811)*	0.163	(0.050)***	0.921	(0.225)***	0.832	(0.353)**
Self-employed	0.087	(0.076)	-0.047	(0.051)	0.624	(0.464)	-0.101	(0.073)	0.828	(0.357)**	0.665	(0.462)
Home owned outright	-0.023	(0.060)	-0.034	(0.038)	-0.028	(0.073)	-0.020	(0.051)	-0.075	(0.122)	-0.051	(0.103)
Home owned on a mortgage	0.118	(0.057)**	-0.057	(0.030)*	0.069	(0.095)	0.037	(0.041)	0.044	(0.092)	0.095	(0.034)**
Home rented	-0.083	(0.071)	0.007	(0.042)	-0.041	(0.062)	-0.083	(0.057)	-0.032	(0.138)	-0.017	(0.096)
Health stock (linear prediction)	0.513	(0.017)***	0.368	(0.016)***	0.448	(0.026)***	0.422	(0.019)***	0.338	(0.028)***	0.392	(0.018)***
Index Function for Accurate Reporting												
	GHQ1		GHQ2		GHQ3		GHQ4		GHQ5		GHQ6	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
constant	-3.073	(0.223)***	11.616	(1.495)***	4.994	(0.446)***	14.983	(1.618)***	4.5711	(0.401)***	4.996	(0.429)***
Age10 (age divided by 10)	0.118	(0.012)***	-0.870	(0.106)***	0.054	(0.002)***	-0.410	(0.048)***	0.061	(0.004)***	0.075	(0.012)***
Square of Age10	-0.021	(0.014)	0.689	(0.083)***	-0.067	(0.004)***	0.225	(0.035)***	-0.078	(0.006)***	-0.097	(0.019)***
Married or cohabitating	-0.055	(0.052)	-0.006	(0.175)	0.030	(0.011)***	0.049	(0.122)	0.013	(0.005)***	-0.104	(0.142)
White	-0.025	(0.065)	0.584	(0.261)**	-0.014	(0.016)	0.203	(0.170)	-0.081	(0.035)**	-0.017	(0.006)***
Degree	-0.247	(0.196)	-2.186	(1.002)**	-0.360	(0.203)*	-0.013	(0.695)	0.254	(0.276)	0.225	(0.112)**
Teaching or nursing	-0.191	(0.164)	-1.084	(0.764)	0.023	(0.167)	0.231	(0.471)	0.390	(0.189)**	0.558	(0.113)***
A levels	-0.360	(0.177)**	0.123	(0.894)	0.027	(0.438)	0.913	(0.577)	0.446	(0.506)	0.673	(0.196)***
O levels	-0.114	(0.176)	-1.572	(0.882)*	0.259	(0.113)**	0.853	(0.621)	0.332	(0.370)	0.685	(0.138)***
Other qualifications	-0.157	(0.250)	-1.140	(0.943)	0.439	(0.111)***	0.660	(0.773)	0.471	(0.233)**	1.252	(0.085)***
Log of monthly labour income	-0.023	(0.012)*	0.062	(0.038)	-0.031	(0.010)***	0.061	(0.027)**	-0.011	(0.004)***	-0.028	(0.005)***
Log of monthly non-labourincome	0.021	(0.009)**	-0.097	(0.030)***	-0.010	(0.007)	-0.021	(0.021)	-0.001	(0.003)	-0.003	(0.005)
Employed	-0.220	(0.087)**	0.620	(0.315)*	-1.840	(0.182)***	0.078	(0.218)	-1.616	(0.306)***	-1.508	(0.235)***
Unemployed	-0.167	(0.077)**	0.317	(0.283)	-1.831	(0.251)***	0.353	(0.189)*	-1.355	(0.317)***	-1.676	(0.206)***
Self-employed	-0.309	(0.106)***	-0.075	(0.327)	-1.670	(0.241)***	0.061	(0.227)	-1.406	(0.637)***	-1.848	(0.946)***
Home owned outright	0.029	(0.076)	-0.147	(0.270)	0.013	(0.075)	-0.181	(0.160)	-0.004	(0.097)	0.050	(0.034)
Home owned on a mortgage	-0.091	(0.064)	0.480	(0.246)*	-0.031	(0.202)	-0.255	(0.151)*	-0.062	(0.167)	-0.135	(0.047)***
Home rented	0.056	(0.089)	-0.117	(0.309)	-0.019	(0.146)	0.057	(0.181)	-0.031	(0.004)***	-0.111	(0.102)
Change in interviewer *t-1* to *t*	0.001	(0.017)	0.046	(0.074)	-0.019	(0.026)	-0.004	(0.048)	-0.018	(0.039)	-0.033	(0.012)***
% questionnaire not answered	0.041	(0.003)***	0.112	(0.054)**	0.064	(0.046)	0.085	(0.034)**	-0.009	(0.001)***	0.013	(0.006)**
ρ	-0.785	(0.041)***	-0.783	(0.119)***	-0.386	(0.142)***	-0.827	(0.062)***	-0.782	(0.093)***	-0.337	(0.131)***

Note: GHQ1 – concentration; GHQ2 – sleep loss; GHQ3 – usefulness; GHQ4 – capability; GHQ5 – strain; GHQ6 – overcoming difficulties. Standard errors are given in parentheses.* significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 4: Males - Estimated Coefficients of the Inflated Probit Model: GHQ7 to GHQ12

	GHQ7		GHQ8		GHQ9		GHQ10		GHQ11		GHQ12	
Index Function for Probit												
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
constant	3.154	(0.252)***	-2.575	(0.133)***	-1.758	(0.119)***	3.504	(0.312)***	-1.244	(0.969)	-2.242	(0.144)***
Age10 (age divided by 10)	-0.112	(0.010)***	0.015	(0.008)**	0.028	(0.006)***	-0.195	(0.012)***	-0.008	(0.002)***	0.047	(0.009)***
Square of Age10	0.057	(0.008)***	0.018	(0.008)**	0.013	(0.007)*	0.136	(0.008)***	-0.001	(0.012)	-0.001	(0.010)
Married or cohabitating	0.068	(0.045)	-0.159	(0.032)***	-0.153	(0.029)***	-0.184	(0.038)***	-0.060	(0.119)	-0.181	(0.032)***
White	-0.008	(0.061)	-0.077	(0.039)*	-0.031	(0.036)	-0.038	(0.056)	-0.060	(0.028)**	-0.023	(0.037)
Degree	-0.105	(0.188)	0.102	(0.138)	0.325	(0.117)***	-0.108	(0.201)	0.234	(0.171)	0.430	(0.139)***
Teaching or nursing	0.309	(0.142)**	0.050	(0.117)	0.294	(0.102)***	0.222	(0.137)*	-0.402	(0.156)***	0.363	(0.121)***
A levels	0.487	(0.169)***	0.046	(0.125)	0.267	(0.105)**	0.182	(0.154)	-0.282	(0.143)**	0.235	(0.126)*
O levels	0.433	(0.172)**	-0.020	(0.128)	0.272	(0.109)**	0.406	(0.159)**	-0.402	(0.541)	0.210	(0.132)
Other qualifications	0.426	(0.224)*	0.191	(0.159)	0.320	(0.150)**	0.205	(0.221)	-0.916	(2.096)	0.361	(0.169)*
Log of monthly labour income	-0.009	(0.010)	-0.030	(0.007)***	-0.026	(0.006)***	-0.014	(0.008)*	-0.003	(0.048)	-0.033	(0.007)***
Log of monthly non-labour income	-0.024	(0.007)***	-0.003	(0.005)	-0.004	(0.004)	-0.015	(0.065)**	0.003	(0.017)	-0.002	(0.005)
Employed	-0.044	(0.077)	-0.159	(0.056)***	0.001	(0.050)	-0.108	(0.065)*	-0.466	(0.697)	-0.072	(0.056)
Unemployed	0.171	(0.066)***	0.095	(0.044)**	0.334	(0.043)***	0.380	(0.054)***	0.952	(2.240)	0.253	(0.047)***
Self-employed	-0.115	(0.082)	-0.181	(0.064)***	0.086	(0.058)	-0.178	(0.069)**	0.492	(0.391)	-0.012	(0.066)
Home owned outright	-0.131	(0.065)**	0.029	(0.049)	-0.060	(0.043)	-0.037	(0.053)	-0.068	(0.387)	-0.052	(0.048)
Home owned on a mortgage	-0.023	(0.057)	0.042	(0.038)	0.042	(0.033)	0.111	(0.050)**	0.124	(0.260)	-0.002	(0.037)
Home rented	-0.094	(0.076)	-0.103	(0.050)**	-0.085	(0.047)*	0.001	(0.062)	-0.051	(0.058)	0.033	(0.033)
Health stock (linear prediction)	0.633	(0.020)***	0.496	(0.019)***	0.484	(0.017)***	0.484	(0.018)***	0.417	(0.067)***	0.457	(0.019)***
Index Function for Accurate Reporting												
	GHQ7		GHQ8		GHQ9		GHQ10		GHQ11		GHQ12	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
constant	-3.504	(0.255)***	17.223	(1.813)***	7.248	(0.651)***	-2.442	(0.274)***	5.218	(0.723)***	8.724	(0.632)***
Age10 (age divided by 10)	0.066	(0.011)***	-0.412	(0.050)***	-0.275	(0.026)***	0.095	(0.019)***	0.076	(0.014)***	-0.318	(0.023)***
Square of Age10	0.020	(0.011)*	0.225	(0.037)***	0.120	(0.020)***	0.061	(0.028)**	-0.084	(0.021)***	0.187	(0.018)***
Married or cohabitating	-0.158	(0.059)***	-0.008	(0.120)	-0.270	(0.086)***	-0.046	(0.059)	-0.083	(0.242)	-0.149	(0.070)**
White	-0.038	(0.071)	0.113	(0.158)	0.105	(0.122)	0.015	(0.071)	0.075	(0.627)	0.083	(0.097)
Degree	0.070	(0.235)	-1.054	(0.826)	-1.259	(0.504)***	0.336	(0.251)	-0.528	(1.022)	-1.548	(0.458)***
Teaching or nursing	-0.347	(0.203)*	0.305	(0.412)	-0.379	(0.356)	-0.173	(0.214)	0.607	(0.536)	-0.433	(0.318)
A levels	-0.632	(0.228)***	-0.108	(0.497)	-0.246	(0.390)	-0.366	(0.222)	0.552	(0.373)	-0.138	(0.369)
O levels	-0.638	(0.228)***	0.691	(0.548)	-0.110	(0.411)	-0.289	(0.332)	0.625	(1.636)	0.183	(0.411)
Other qualifications	-0.509	(0.286)*	0.667	(0.708)	-0.001	(0.561)	-0.289	(0.306)	1.396	(1.283)	-0.166	(0.411)
Log of monthly labour income	-0.010	(0.013)	0.031	(0.028)	0.043	(0.019)**	-0.015	(0.012)	-0.055	(0.301)	0.038	(0.017)**
Log of monthly non-labour income	0.027	(0.009)***	-0.041	(0.021)**	-0.011	(0.013)	0.020	(0.009)**	-0.010	(0.064)	-0.022	(0.012)*
Employed	-0.044	(0.096)	-0.009	(0.217)	-0.208	(0.152)	-0.147	(0.094)	-1.805	(0.219)***	-0.050	(0.132)
Unemployed	0.057	(0.081)	0.334	(0.200)*	-0.038	(0.131)	-0.014	(0.081)	-1.830	(0.162)***	0.182	(0.118)
Self-employed	0.001	(0.117)	0.282	(0.244)	-0.407	(0.170)**	-0.056	(0.119)	-1.817	(0.169)***	-0.248	(0.146)*
Home owned outright	0.043	(0.084)	0.051	(0.172)	-0.126	(0.128)	0.040	(0.088)	0.027	(0.072)	-0.008	(0.099)
Home owned on a mortgage	-0.031	(0.067)	0.139	(0.160)	-0.055	(0.118)	-0.087	(0.071)	-0.198	(0.865)	0.087	(0.098)
Home rented	0.027	(0.096)	0.123	(0.165)	0.183	(0.131)	-0.014	(0.095)	0.044	(0.029)	-0.037	(0.109)
Change in interviewer *t-1* to *t*	-0.027	(0.017)	-0.038	(0.048)	-0.039	(0.033)	0.023	(0.016)**	0.016	(0.067)	0.011	(0.027)
% questionnaire not answered	0.029	(0.013)**	0.060	(0.026)***	0.039	(0.020)**	0.051	(0.018)***	0.081	(0.189)	0.068	(0.019)***
ρ	-0.753	(0.042)***	-0.761	(0.070)***	-0.753	(0.056)***	-0.675	(0.062)***	-0.291	(0.446)	-0.079	(0.039)***

Note: GHQ7 – enjoy activities; GHQ8 – face up to problems; GHQ9 – unhappy or depressed; GHQ10 – losing confidence; GHQ11 – worthless person;

GHQ12 – feeling reasonably happy. Standard errors are given in parentheses.* significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 5: Females - Estimated Coefficients of the Inflated Probit Model: GHQ1 to GHQ6

	GHQ1		GHQ2		GHQ3		GHQ4		GHQ5		GHQ6	
Index Function for Probit												
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
constant	1.835	(0.483)***	-1.056	(0.098)***	2.297	(0.325)***	-1.958	(0.106)***	-0.689	(0.097)***	-1.063	(0.108)***
Age10 (age divided by 10)	-0.069	(0.004)***	0.025	(0.004)***	-0.082	(0.010)***	0.060	(0.006)***	0.028	(0.003)***	0.063	(0.005)***
Square of Age10	0.060	(0.005)***	-0.006	(0.004)	0.019	(0.008)**	-0.038	(0.006)***	-0.031	(0.004)***	-0.058	(0.005)***
Married or cohabitating	-0.180	(0.030)***	-0.180	(0.020)***	-0.196	(0.053)***	-0.071	(0.026)***	-0.025	(0.008)***	-0.101	(0.022)***
White	0.015	(0.060)	-0.063	(0.032)*	0.061	(0.088)	-0.092	(0.034)***	-0.033	(0.025)	-0.086	(0.035)**
Degree	0.172	(0.103)*	0.113	(0.081)	0.962	(0.239)***	0.180	(0.113)	0.031	(0.162)	0.122	(0.091)
Teaching or nursing	0.085	(0.162)	0.072	(0.067)	0.620	(0.205)***	-0.076	(0.098)	-0.012	(0.104)	0.047	(0.076)
A levels	0.413	(0.190)**	0.000	(0.072)	1.058	(0.225)***	-0.023	(0.104)	-0.027	(0.055)	0.025	(0.082)
O levels	0.025	(0.083)	0.069	(0.070)	0.913	(0.220)***	-0.129	(0.100)	-0.026	(0.023)	0.035	(0.082)
Other qualifications	0.134	(0.375)	0.031	(0.092)	0.235	(0.290)	-0.041	(0.131)	-0.005	(0.179)	0.332	(0.109)***
Log of monthly labour income	0.012	(0.015)	-0.007	(0.005)	0.002	(0.013)	-0.002	(0.006)	0.009	(0.012)	-0.017	(0.006)***
Log of monthly non-labour income	-0.009	(0.019)	-0.004	(0.003)	-0.006	(0.009)	-0.007	(0.004)	-0.001	(0.006)	-0.003	(0.004)
Employed	-0.059	(0.130)	-0.042	(0.032)	-0.064	(0.089)	-0.219	(0.038)***	-0.246	(0.053)***	-0.071	(0.036)**
Unemployed	0.121	(0.074)	0.169	(0.035)***	0.414	(0.084)***	0.177	(0.046)***	0.036	(0.068)	0.156	(0.037)***
Self-employed	-0.071	(0.103)	-0.006	(0.045)	-0.068	(0.122)	-0.178	(0.068)***	-0.255	(0.058)***	-0.100	(0.053)*
Home owned outright	0.042	(0.097)	-0.092	(0.032)***	-0.248	(0.084)***	0.086	(0.042)**	-0.086	(0.042)**	0.005	(0.037)
Home owned on a mortgage	0.072	(0.123)	-0.069	(0.027)***	-0.194	(0.078)**	-0.009	(0.033)	-0.055	(0.043)	-0.035	(0.031)
Home rented	0.022	(0.049)	-0.012	(0.035)	-0.149	(0.087)*	-0.023	(0.043)	-0.046	(0.033)	-0.014	(0.036)
Health stock (linear prediction)	0.310	(0.010)***	0.296	(0.012)***	0.600	(0.024)***	0.428	(0.015)***	0.290	(0.024)***	0.385	(0.012)***
Index Function for Accurate Reporting												
	GHQ1		GHQ2		GHQ3		GHQ4		GHQ5		GHQ6	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
constant	-1.168	(0.219)***	5.437	(0.759)***	-1.998	(0.304)***	13.257	(1.439)***	2.771	(0.032)***	4.857	(0.776)***
Age10 (age divided by 10)	0.088	(0.010)***	-0.254	(0.033)***	0.069	(0.010)***	-0.569	(0.054)***	-0.071	(0.005)***	-0.541	(0.048)***
Square of Age10	-0.089	(0.007)***	0.064	(0.023)***	0.022	(0.009)**	0.404	(0.041)***	0.062	(0.008)***	0.470	(0.041)***
Married or cohabitating	0.120	(0.058)**	0.125	(0.118)	-0.039	(0.058)	-0.717	(0.137)***	0.014	(0.044)	-0.276	(0.128)**
White	-0.023	(0.052)	0.273	(0.176)	-0.085	(0.095)	0.516	(0.175)***	0.074	(0.022)***	0.282	(0.196)
Degree	-0.103	(0.137)	-0.199	(0.543)	-0.653	(0.258)**	-1.026	(0.796)	-0.059	(0.910)	-0.957	(0.560)*
Teaching or nursing	-0.050	(0.149)	0.029	(0.445)	-0.602	(0.233)***	0.176	(0.502)	-0.017	(0.243)	-0.963	(0.424)**
A levels	-0.348	(0.165)**	0.485	(0.513)	-1.119	(0.252)***	0.345	(0.620)	-0.192	(0.408)	-0.428	(0.500)
O levels	0.026	(0.049)	0.113	(0.506)	-0.956	(0.246)***	0.730	(0.579)	-0.045	(0.238)	-0.085	(0.503)
Other qualifications	-0.109	(0.376)	0.183	(0.660)	-0.374	(0.315)	0.871	(0.668)	0.119	(0.637)	-3.946	(0.741)***
Log of monthly labour income	0.003	(0.018)	0.025	(0.029)	-0.034	(0.015)**	0.037	(0.032)	0.004	(0.034)	0.119	(0.035)***
Log of monthly non-labour income	0.007	(0.022)	0.003	(0.020)	-0.001	(0.010)	-0.012	(0.022)	0.022	(0.018)	0.018	(0.022)
Employed	-0.139	(0.159)	-0.022	(0.196)	-0.134	(0.098)	0.013	(0.219)	3.799	(2.295)*	-0.077	(0.230)
Unemployed	-0.062	(0.079)	0.150	(0.246)	0.101	(0.089)	0.238	(0.258)	0.434	(0.288)	0.494	(0.262)*
Self-employed	-0.120	(0.122)	-0.184	(0.264)	-0.173	(0.148)	-0.151	(0.316)	4.124	(54.683)	0.210	(0.313)
Home owned outright	-0.024	(0.128)	0.288	(0.172)*	0.299	(0.095)***	-0.204	(0.188)	0.130	(0.110)	-0.263	(0.226)
Home owned on a mortgage	-0.071	(0.140)	0.225	(0.161)	0.299	(0.083)***	0.148	(0.183)	0.116	(0.146)	0.274	(0.193)
Home rented	-0.019	(0.056)	-0.246	(0.188)	0.145	(0.090)	-0.176	(0.197)	0.062	(0.092)	-0.149	(0.186)
Change in interviewer *t-1* to *t*	-0.015	(0.008)*	-0.027	(0.050)	0.017	(0.016)	-0.061	(0.052)	0.005	(0.026)	-0.056	(0.055)
% questionnaire not answered	0.006	(0.010)	-0.033	(0.037)	0.066	(0.013)***	0.056	(0.037)	0.038	(0.007)***	0.107	(0.041)***
ρ	-0.993	(0.010)***	-0.668	(0.125)***	-0.573	(0.063)***	-0.610	(0.108)***	-0.462	(0.218)**	-0.861	(0.068)***

Note: GHQ1 – concentration; GHQ2 – sleep loss; GHQ3 – usefulness; GHQ4 – capability; GHQ5 – strain; GHQ6 – overcoming difficulties. Standard errors are given in parentheses.* significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 6: Females - Estimated Coefficients of the Inflated Probit Model: GHQ7 to GHQ12

	GHQ7		GHQ8		GHQ9		GHQ10		GHQ11		GHQ12	
Index Function for Probit												
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
constant	0.429	(0.243)*	-1.411	(0.077)***	-0.642	(0.099)***	-0.694	(0.065)***	-1.307	(0.130)***	-2.104	(0.114)***
Age10 (age divided by 10)	-0.144	(0.012)***	0.025	(0.005)***	0.000	(0.004)	0.024	(0.003)***	0.017	(0.002)***	0.020	(0.006)***
Square of Age10	0.134	(0.010)***	-0.024	(0.006)***	0.025	(0.004)***	-0.034	(0.003)***	-0.026	(0.005)***	0.028	(0.006)***
Married or cohabitating	-0.212	(0.044)***	-0.127	(0.034)***	-0.198	(0.020)***	-0.100	(0.011)***	-0.121	(0.023)***	-0.089	(0.026)***
White	-0.031	(0.061)	-0.032	(0.017)*	-0.088	(0.034)**	0.018	(0.010)*	0.018	(0.009)**	-0.052	(0.038)
Degree	0.012	(0.193)	0.034	(0.241)	0.021	(0.083)	0.151	(0.081)*	0.113	(0.098)	-0.042	(0.114)
Teaching or nursing	-0.091	(0.154)	-0.059	(0.113)	-0.135	(0.067)**	0.016	(0.012)	-0.137	(0.039)***	-0.147	(0.097)
A levels	0.024	(0.170)	-0.064	(0.195)	-0.159	(0.075)**	-0.015	(0.041)	-0.131	(0.063)**	-0.213	(0.104)**
O levels	-0.212	(0.169)	-0.064	(0.121)	-0.140	(0.071)**	0.141	(0.048)***	-0.005	(0.049)	-0.179	(0.100)*
Other qualifications	-0.190	(0.216)	-0.037	(0.120)	-0.047	(0.097)	-0.079	(0.087)	0.062	(0.116)	-0.027	(0.134)
Log of monthly labour income	0.024	(0.010)**	-0.004	(0.006)	-0.010	(0.005)**	-0.006	(0.016)	-0.018	(0.006)***	-0.021	(0.006)***
Log of monthly non-labour income	-0.009	(0.007)	-0.001	(0.007)	0.001	(0.003)	0.003	(0.002)	-0.003	(0.003)	-0.001	(0.005)
Employed	-0.217	(0.067)***	-0.069	(0.086)	-0.005	(0.032)	-0.127	(0.053)**	-0.027	(0.015)*	0.082	(0.041)**
Unemployed	0.199	(0.079)**	0.167	(0.085)**	0.236	(0.035)***	0.203	(0.052)***	0.222	(0.043)***	0.240	(0.044)***
Self-employed	-0.245	(0.094)***	-0.105	(0.086)	-0.068	(0.047)	-0.192	(0.053)***	-0.138	(0.045)***	0.056	(0.067)
Home owned outright	-0.011	(0.067)	-0.032	(0.045)	0.011	(0.033)	-0.030	(0.036)	-0.003	(0.045)	0.026	(0.044)
Home owned on a mortgage	0.121	(0.061)**	0.008	(0.036)	0.054	(0.027)**	-0.033	(0.018)*	-0.014	(0.025)	-0.020	(0.034)
Home rented	-0.109	(0.075)	-0.045	(0.038)	-0.026	(0.034)	-0.038	(0.027)	-0.032	(0.028)	0.049	(0.044)
Health stock (linear prediction)	0.879	(0.030)***	0.374	(0.018)***	0.408	(0.012)***	0.332	(0.013)***	0.342	(0.030)***	0.425	(0.015)***
Index Function for Accurate Reporting												
	GHQ7		GHQ8		GHQ9		GHQ10		GHQ11		GHQ12	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
constant	-0.868	(0.222)***	9.194	(0.235)***	4.956	(0.630)***	6.073	(0.832)***	7.209	(0.150)***	6.153	(0.595)***
Age10 (age divided by 10)	0.148	(0.010)***	-0.230	(0.019)***	-0.167	(0.026)***	-0.188	(0.009)***	-0.148	(0.001)***	-0.192	(0.020)***
Square of Age10	-0.128	(0.009)***	0.191	(0.021)***	-0.002	(0.020)	0.168	(0.007)***	0.115	(0.006)***	0.050	(0.016)***
Married or cohabitating	0.157	(0.049)***	-0.124	(0.220)	0.268	(0.105)**	-0.226	(0.072)***	-0.129	(0.076)*	-0.335	(0.078)***
White	0.052	(0.068)	0.222	(0.271)	0.568	(0.171)***	-0.006	(0.003)**	0.068	(0.026)***	0.251	(0.109)**
Degree	0.315	(0.238)	-1.054	(5.094)	-0.210	(0.479)	-0.537	(0.544)	-2.031	(0.117)***	0.182	(0.382)
Teaching or nursing	0.167	(0.207)	0.451	(0.831)	0.256	(0.359)	0.189	(0.085)**	0.093	(0.093)	0.405	(0.275)
A levels	0.067	(0.220)	1.508	(0.430)***	0.400	(0.399)	0.605	(0.412)	0.431	(0.206)**	0.684	(0.328)**
O levels	0.278	(0.222)	-0.584	(0.130)***	0.559	(0.419)	-0.837	(0.108)***	-0.640	(0.090)***	0.662	(0.298)**
Other qualifications	0.262	(0.275)	0.465	(0.825)	-0.696	(0.515)	0.501	(0.667)	-0.985	(0.340)***	-0.187	(0.418)
Log of monthly labour income	0.010	(0.012)	0.038	(0.041)	0.031	(0.026)	0.029	(0.063)	0.102	(0.019)***	0.038	(0.020)*
Log of monthly non-labour income	0.001	(0.009)	0.003	(0.012)	0.002	(0.017)	-0.037	(0.016)**	-0.027	(0.006)***	0.002	(0.013)
Employed	-0.197	(0.076)**	-0.115	(0.079)	-0.092	(0.190)	0.068	(0.156)	-0.665	(0.089)***	-0.375	(0.139)***
Unemployed	0.111	(0.088)	0.058	(0.534)	-0.122	(0.228)	0.090	(0.230)	-0.064	(0.142)	0.208	(0.156)
Self-employed	-0.115	(0.131)	5.191	(3.396)	-0.051	(0.229)	5.257	(10.062)	4.497	(1.365)***	-0.496	(0.194)**
Home owned outright	0.082	(0.079)	0.253	(0.163)	-0.170	(0.163)	0.181	(0.027)***	-0.138	(0.120)	0.059	(0.119)
Home owned on a mortgage	-0.029	(0.069)	0.152	(0.033)***	-0.081	(0.154)	0.229	(0.108)**	0.071	(0.085)	0.298	(0.113)***
Home rented	0.055	(0.082)	0.039	(0.231)	-0.076	(0.158)	0.108	(0.083)	0.018	(0.024)	-0.093	(0.117)
Change in interviewer *t-1* to *t*	0.037	(0.018)**	-0.001	(0.066)	-0.082	(0.045)*	0.021	(0.021)	-0.007	(0.034)	0.012	(0.029)
% questionnaire not answered	0.058	(0.014)***	0.178	(0.090)**	-0.018	(0.032)	0.033	(0.022)	0.102	(0.081)	0.073	(0.023)***
ρ	0.045	(0.146)	-0.364	(0.117)***	-0.733	(0.088)***	-0.649	(0.073)***	-0.535	(0.072)***	-0.624	(0.068)***

Note: GHQ7 – enjoy activities; GHQ8 – face up to problems; GHQ9 – unhappy or depressed; GHQ10 – losing confidence; GHQ11 – worthless person;

GHQ12 – feeling reasonably happy. Standard errors are given in parentheses.* significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 7: Males - Predicted Probabilities and Reporting Bias for Individual $GHQ-12$ Components

	Proportion of reported psychological distress	Predicted rate of psychological distress	Reporting bias	Predicted marginal probability of misreporting zeros	Posterior Probabilities	
					0-Score	mis-reporting
		$\mathbf{Pr}(\widetilde{y}=1\|\mathbf{x})$	%	$\mathbf{Pr}(r=0\|\mathbf{x})$	$\mathbf{Pr}(\widetilde{y}=0\|\mathbf{x},\mathbf{y}=\mathbf{0})$	$\mathbf{Pr}(\widetilde{y}=1,r=0\|\mathbf{x},\mathbf{y}=\mathbf{0})$
	(1)	(2)	(3)	(4)	(5)	(6)
GHQ1	0.155	0.354	-129%	0.245	0.751	0.249
		(0.015)***		(0.017)***	(0.018)***	(0.018)***
GHQ2	0.147	0.152	-4%	0.023	0.986	0.014
		(0.004)***		(0.007)***	(0.003)***	(0.003)***
GHQ3	0.118	0.293	-148%	0.409	0.802	0.198
		(0.135)**		(0.123)***	(0.151)***	(0.151)
GHQ4	0.076	0.097	-28%	0.081	0.952	0.048
		(0.006)***		(0.017)***	(0.007)***	(0.007)***
GHQ5	0.233	0.483	-107%	0.277	0.665	0.335
		(0.081)***		(0.068)***	(0.108)***	(0.108)***
GHQ6	0.115	0.213	-85%	0.280	0.888	0.112
		(0.018)***		(0.055)***	(0.019)***	(0.019)***
GHQ7	0.169	0.397	-134%	0.258	0.730	0.270
		(0.016)***		(0.019)***	(0.020)***	(0.020)***
GHQ8	0.086	0.099	-15%	0.077	0.957	0.043
		(0.006)***		(0.017)***	(0.007)***	(0.007)***
GHQ9	0.175	0.234	-34%	0.103	0.914	0.086
		(0.008)***		(0.016)***	(0.011)***	(0.011)***
GHQ10	0.108	0.226	-108%	0.221	0.833	0.167
		(0.013)***		(0.017)***	(0.014)***	(0.014)***
GHQ11	0.061	0.139	-130%	0.365	0.915	0.085
		(0.181)		(0.235)	(0.195)***	(0.195)
GHQ12	0.114	0.201	-77%	0.178	0.876	0.124
		(0.010)***		(0.016)***	(0.012)***	(0.012)***

Note: GHQ1 – concentration; GHQ2 – sleep loss; GHQ3 – usefulness; GHQ4 – capability; GHQ5 – strain; GHQ6 – overcoming difficulties; GHQ7 – enjoy activities; GHQ8 – face up to problems; GHQ9 – unhappy or depressed; GHQ10 – losing confidence; GHQ11 – worthless person; GHQ12 – feeling reasonably happy. Standard errors are given in parentheses.* significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 8: Females - Predicted Probabilities and Reporting Bias for Individual $GHQ - 12$ Components

	Proportion of reported psychological distress	Predicted rate of psychological distress	Reporting bias	Predicted marginal probability of misreporting zeros	Posterior Probabilities	
					0-Score	mis-reporting
		$\mathbf{Pr}(\widetilde{y}=1\|\mathbf{x})$	%	$\mathbf{Pr}(r=0\|\mathbf{x})$	$\mathbf{Pr}(\widetilde{y}=0\|\mathbf{x},\mathbf{y}=\mathbf{0})$	$\mathbf{Pr}(\widetilde{y}=1,r=0\|\mathbf{x},\mathbf{y}=\mathbf{0})$
	(1)	(2)	(3)	(4)	(5)	(6)
GHQ1	0.217	0.463	-113%	0.246	0.683	0.317
		(0.031)***		(0.031)***	(0.040)***	(0.040)***
GHQ2	0.220	0.244	-11%	0.019	0.983	0.017
		(0.005)***		(0.006)***	(0.004)***	(0.004)***
GHQ3	0.142	0.405	-185%	0.370	0.698	0.302
		(0.020)***		(0.026)***	(0.025)***	(0.025)***
GHQ4	0.116	0.140	-21%	0.069	0.958	0.042
		(0.005)***		(0.018)***	(0.007)***	(0.007)***
GHQ5	0.295	0.341	-16%	0.074	0.931	0.069
		(0.012)***		(0.008)***	(0.017)***	(0.017)***
GHQ6	0.164	0.175	-6%	0.005	0.994	0.006
		(0.004)***		(0.002)***	(0.002)***	(0.002)***
GHQ7	0.206	0.368	-78%	0.370	0.774	0.226
		(0.023)***		(0.020)***	(0.033)***	(0.033)***
GHQ8	0.136	0.153	-13%	0.044	0.978	0.022
		(0.004)***		(0.013)***	(0.005)***	(0.005)***
GHQ9	0.241	0.269	-12%	0.018	0.980	0.020
		(0.004)***		(0.004)***	(0.004)***	(0.004)***
GHQ10	0.172	0.213	-24%	0.065	0.947	0.053
		(0.004)***		(0.009)***	(0.005)***	(0.005)***
GHQ11	0.094	0.130	-39%	0.100	0.957	0.043
		(0.005)***		(0.013)***	(0.006)***	(0.006)***
GHQ12	0.154	0.231	-50%	0.142	0.899	0.101
		(0.008)***		(0.018)***	(0.010)***	(0.010)***

Note: GHQ1 – concentration; GHQ2 – sleep loss; GHQ3 – usefulness; GHQ4 – capability; GHQ5 – strain; GHQ6 – overcoming difficulties; GHQ7 – enjoy activities; GHQ8 – face up to problems; GHQ9 – unhappy or depressed; GHQ10 – losing confidence; GHQ11 – worthless person; GHQ12 – feeling reasonably happy. Standard errors are given in parentheses.* significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 9: Males - Application of the Adjusted $GHQ-12$ to Modelling Transitions in Economic Outcomes

	Educational attainment *(increase in highest qualification obtained)*				Labour market status *(unemployed to employee)*				Savings *(non-saver to saver)*			
	λ (1)	λ (2)	λ (3)	λ (4)	λ (1)	λ (2)	λ (3)	λ (4)	λ (1)	λ (2)	λ (3)	λ (4)
Definition of $GHQ-12$:												
A: Original	0.007				-0.067***				-0.067***			
	(0.021)				(0.024)				(0.012)			
B: Adj. 1		-0.047**				-0.126***				-0.068***		
		(0.023)				(0.015)				(0.011)		
C: Adj. 2			-0.047**				-0.129***				-0.068***	
			(0.022)				(0.014)				(0.011)	
D: Adj. 3				0.011				-0.067***				-0.072***
				(0.020)				(0.019)				(0.011)
Obs. $(N \times T)$		91,309				74,727				91,210		
χ^2 equality		5.68				6.14				0.01		
$\lambda(1) = \lambda(2)$		*p=0.017*				*p=0.013*				*p=0.994*		
χ^2 equality		5.68				6.31				0.01		
$\lambda(1) = \lambda(3)$		*p=0.017*				*p=0.012*				*p=0.994*		
χ^2 equality		0.04				0.04				0.13		
$\lambda(1) = \lambda(4)$		*p=0.845*				*p=0.840*				*p=0.721*		

Note: results in each column are based upon random effects probit estimates conditioning on a quadratic in age, marital status, total income, housing tenure, year of interview and region of residence. Additional controls in the educational attainment models are labour market status. Additional controls in the labour market status models are highest educational attainment. The savings model includes both labour market status and highest educational attainment. Coefficients are reported with associated standard errors given in parentheses. The label "Adj. 1" refers to the *adjusted* method, "Adj. 2" refers to the *robust* method and "Adj. 3" refers to the *upper bound*, as described in section 6. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 10: Females - Application of the Adjusted $GHQ-12$ to Modelling Transitions in Economic Outcomes

	Educational attainment *(increase in highest qualification obtained)*				Labour market status *(unemployed to employee)*				Savings *(non-saver to saver)*			
	λ (1)	λ (2)	λ (3)	λ (4)	λ (1)	λ (2)	λ (3)	λ (4)	λ (1)	λ (2)	λ (3)	λ (4)
Definition of *$GHQ-12$* :												
A: Original	-0.038**				-0.068***				-0.037***			
	(0.018)				(0.015)				(0.011)			
B: Adj. 1		-0.060***				-0.089***				-0.050***		
		(0.021)				(0.020)				(0.009)		
C: Adj. 2			-0.059***				-0.088***				-0.048***	
			(0.022)				(0.019)				(0.010)	
D: Adj. 3				-0.039**				-0.069***				-0.037***
				(0.018)				(0.015)				(0.011)
Obs. ($N \times T$)		111,687				82,414				110,819		
χ^2 equality		10.72				5.52				6.24		
$\lambda(1)=\lambda(2)$		*p=0.000*				*p=0.019*				*p=0.008*		
χ^2 equality		10.01				5.68				6.04		
$\lambda(1)=\lambda(3)$		*p=0.000*				*p=0.017*				*p=0.014*		
χ^2 equality		0.01				0.01				0.06		
$\lambda(1)=\lambda(4)$		*p=0.927*				*p=0.920*				*p=0.808*		

Note: results in each column are based upon random effects probit estimates conditioning on a quadratic in age, marital status, total income, housing tenure, year of interview and region of residence. Additional controls in the educational attainment models are labour market status. Additional controls in the labour market status models are highest educational attainment. The savings model includes both labour market status and highest educational attainment. Coefficients are reported with associated standard errors given in parentheses. The label "Adj. 1" refers to the *adjusted* method, "Adj. 2" refers to the *robust* method and "Adj. 3" refers to the *upper bound*, as described in section 6. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.