Article:
Bryan, M., Rice, N., Roberts, J. et al. (2022) Mental health and employment: a bounding approach using panel data. Oxford Bulletin of Economics and Statistics, 84 (5). pp. 1018-1051. ISSN: 0305-9049

https://doi.org/10.1111/obes.12489

OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 0305-9049
doi: 10.1111/obes.12489

Mental Health and Employment: A Bounding Approach Using Panel Data*

MARK L BRYAN,† NIGEL RICE,‡ JENNIFER ROBERTS† and CRISTINA SECHEL†

†*Department of Economics, University of Sheffield, Sheffield, UK (e-mail: m.l.bryan@sheffield.ac.uk)*
‡*Centre for Health Economics and Department of Economics and Related Studies, University of York, York, UK (e-mail: nigel.rice@york.ac.uk)*

Abstract

The effect of mental health on employment is a key policy question, but reliable causal estimates are elusive. Exploiting panel data and extending recent techniques using selection on observables to provide information on selection along unobservables, we estimate that transitioning into poor mental health leads to a 1.6% point reduction in the probability of employment; approximately 10% of the raw employment gap. Selection into mental health is almost entirely based on time-invariant characteristics, rendering fixed effects estimates unbiased in this context, meaning researchers no longer have to rely on the narrow local average treatment effects of most health/work IV studies.

I. Introduction

An individual's relationship to the labour market is a key determinant of their financial security and a source of broader well-being (Black, 2009). In most countries, people with health problems have a much lower employment rate than the rest of the population.[1] In the United Kingdom, every year 300,000 people stop work and become reliant on health-related benefits, costing the government £13 bn and employers another £9 bn (Black and Frost, 2011). Recent work by Jones, Rice, and Zantomio (2020) shows that acute health shocks substantially increase the probability of exiting the labour market and reduce hours and earnings. Adverse mental health (MH) seems to be particularly pernicious in its labour market effects. The employment rate for people with an MH problem is only 35% (Oakley, 2016), and the disability employment gap between those with and without an MH problem is around 40 percentage points (Munford, Rice, and

JEL Classification numbers: I12, J14, J24.

*We are grateful to the Health Foundation for funding this research under their Social and Economic Value of Health programme, Award No. 751630. The UKHLS data (University of Essex, 2018) were accessed via the UK Data Service. Some of the data used are provided to the authors under Special Licence arrangements. We are also grateful to Andrew Jones, Maarten Lindeboom, Arne Risa Hole, and two anonymous reviewers for helpful comments on earlier versions of this article.

[1] https://ilostat.ilo.org/how-do-people-with-disabilities-fare-in-the-labour-market/

Roberts, 2016). Common MH problems, like anxiety and depression, account for over 40% of UK disability insurance claims (McInnes, 2012). MH is neglected in terms of health spending, and often hidden in the workplace due to stigma and discrimination (WHO, 2013). Internationally, the World Health Organization (WHO, 2008) estimate that MH disorders comprise around 13% of the global burden of disease; and the OECD estimates that MH problems affect more than one in six people across the European Union in any one year (OECD/EU, 2018).

There is a complex relationship between MH and work. Work is generally good for MH (Waddell and Burton, 2006), but there can also be adverse effects from long hours, stress, and job insecurity (WHO, 2000). MH is also an important determinant of an individual's labour market situation, affecting the chances of obtaining employment, 'good work', and adequate reward. This complex relationship poses a number of problems for the estimation of causal effects. Frijters, Johnston, and Shields (2014) summarize these as: reverse causality (since health affects work and vice versa); measurement error (as we do not observe the true health stock of an individual); and endogenous selection (since unobserved characteristics and circumstances that affect health outcomes are also likely to be related to labour market outcomes). Our study focuses on the latter problem, but we also employ methods that aim to reduce the biases arising from the first two issues.

Causal estimation of the effect of an individual's MH status on their chance of being in employment requires independent variation in MH. However, many of the tools that are often used to create a pseudo-experimental framework for estimation of causal effects (such as exogenous policy changes or other 'shocks') are not valid, or have only weak validity, in the context of MH and work. Most of the recent econometric evidence relies on instrumental variable (IV) estimation and/or longitudinal data with fixed effects (FEs) in an attempt to deal with endogenous selection. Few of the IV studies are satisfactory; the instruments used have little theoretical support and virtually none of the studies provide convincing empirical evidence on instrument validity. Further, where the identification strategy is arguably stronger, the results often provide very specific estimates of a local average treatment effect (LATE), which in most cases is derived from an arbitrary exogenous shock (e.g. the death of a close friend used by Frijters *et al.*, 2014). The inclusion of FE eliminates endogenous selection bias arising from time-invariant unobserved variables (such as childhood circumstances) that influence both health and work outcomes. Also, FE may give a more relevant policy parameter, because these models estimate the average effect on work outcomes for those whose MH changes, rather than a more narrowly defined LATE. However, these models cannot deal with unobserved effects that vary over time (such as changes in work relationships); if these are present, they will bias the estimated effect of health on work, providing a misleading basis for policy formulation. Practitioners face a dilemma given the difficulty of finding suitable instruments for MH and the need for reliable quantitative evidence. In this context, the use of FE models without instrumentation warrants deeper scrutiny.[2] This is now possible by exploiting the methods developed by (Altonji, Elder, and Taber, 2005; Altonji

[2]Technically FE is also an IV estimator, with deviations from the means used as the instruments (see Verbeek, 2012, p. 387-8). Thus, the effect identified (the average affect for the subgroup whose MH changes) is also a LATE. However, for clarity, when comparing our work with the existing literature, we reserve the term LATE for explicit IV methods.

et al., 2011), who use selection on observable characteristics to provide information on selection along unobservable factors; and in particular Oster (2013b, 2019) who extends and generalizes this method to enable the estimation of an unbiased treatment effect in the presence of unobserved confounders.

We make two key contributions to the literature. Firstly, we fill a number of important gaps in the evidence base by providing quantitative estimates of the effect of MH on the employment of prime age adults. This is important evidence for social and economic policy across all countries. The vast majority of existing evidence on the relationship between health and work considers either physical health or general measures of overall self-assessed health (see Ghatak, 2010 for a review). Where MH is considered, the estimates often pertain only to severe health problems. In contrast, we use measures of general MH derived from two psychometric instruments; the General Health Questionnaire and the Short Form-12 health survey. These measures are good proxies for the true MH stock: they are designed to provide information on all aspects of MH, and are less likely to suffer from the reporting biases that are present in simple overall evaluative measures (Bound, 1991; Bound *et al.*, 1999; Lindeboom and Kerkhofs, 2009). To date, almost all of the existing evidence comes from the United States: a country that has very different health and welfare systems to many other countries, and in particular to the universal health care coverage of the UK National Health Service. Our estimates for England and Wales contribute to a very small pool of UK evidence, and provide important context to the current policy priority to increase the number of disabled people in work by one million over 10 years (DWP, 2017). In addition, much of the evidence on the impact of health on labour market outcomes is for older workers, since this is where the burden of most physical ill-health is felt. In contrast, MH disorders are particularly prevalent in prime age workers (Kessler *et al.*, 2005), so evidence is needed for this key group. As well as estimating average effects for our sample of prime age individuals, we also explore how both the health−employment relationship and any bias in the estimates vary across a number of sub-groups differentiated by sex, age, education, physical health, and household income. These results will also make a valuable contribution to the economic analysis of the cost-effectiveness of health care interventions that are expected to have important labour market effects[3]; for example, the Improving Access to Psychological Therapies (IAPT) initiative that has been rolled out in England and Wales from 2008 to help people who suffer from anxiety and depression.

Secondly, the vast majority of evidence comes from cross-sectional studies, whereas our longitudinal analysis can control for individual unobserved factors that confound the relationship between employment and health. Further, we explore any remaining biases that are not removed by the inclusion of FE, by employing Oster's method (Oster, 2013b; Oster, 2019) to deal with unobservable selection. Ours is the first study to use this method with individual longitudinal data incorporating FE.[4] We estimate the bias that arises from

[3]This is an important area for health policy; for example, Public Health England have recently commissioned a model to estimate the cost-effectiveness of health interventions that are expected to have significant labour market effects https://www.gov.uk/government/publications/health-matters-health-and-work

[4]In the only panel data applications of the Oster method of which we are aware, Hener, Rainer, and Siedler (2016) and Cattan *et al.* (2017) use individual-level data with sibling FE, and Black *et al.* (2014) use firm-level data. By assuming that the FEs capture all time-invariant variables and have no unobserved counterparts, we apply Oster's

omitting important influences on both health and employment in a FE framework that has no exclusion restrictions. We also calculate a consistent estimate of the biased-adjusted treatment effect under certain assumptions. We discuss the interpretation of the FE treatment effect and contrast this with the narrow LATEs that are often estimated from IV studies. This application will be a useful resource for practitioners who may wish to use the method in other contexts; and for the policy community who wish to judge the quality of evidence from econometric studies.

Our results show that while there is strong evidence of cross-sectional selection in pooled OLS estimates of the effect of MH on employment, there is little or no additional selection bias once FEs are included. Even under weak assumptions, we cannot reject that the bias-corrected estimates are the same as the FE coefficients. Our preferred estimates are reasonably similar to the small amount of comparable longitudinal evidence, but they are substantially smaller than typical IV estimates in the literature, suggesting that much existing evidence may overestimate the average effect of MH on employment. We find evidence that MH has larger effects on employment for those without higher education and those who are in poverty. The article is structured as follows. In section II we explore the background to the FE models, explaining the estimation problems they are designed to solve and reviewing some of the key evidence. Section III describes our estimation method, and the data and variables are described in section IV. The results and sub-group analyses are presented in section V. Section VI includes the discussion and conclusion.

II. Background

It is well known that MH and work are related and that the relationship between them is complex (see, e.g. Currie and Madrian, 1999; Steele, French, and Bartley, 2013; Frijters *et al.*, 2014). However, the existing evidence falls short in a number of respects, and our contribution lies in attempting to correct for these shortcomings. Firstly, in contrast to the literature that explores the effects of physical health problems on work, MH studies are in relatively short supply. In particular, there is virtually no evidence for the United Kingdom, and in the available evidence for other countries, there is no consensus around the size of the effects. Further, it is hard to compare the estimates from the various studies due to the use of different data and methods, and the different ways in which MH is measured. Secondly, the majority of this evidence considers only severe MH problems (e.g. Greve and Nielsen, 2013 on schizophrenia, and Hakulinen *et al.*, 2019; Hakulinen *et al.*, 2020 on severe mental disorders requiring hospitalization) and there is little to inform the employment effects of common mental disorders (such as anxiety and depression); and this despite the growing prevalence of these problems among prime age workers. Thirdly, some of the studies with the strongest identification strategies use data from very specific samples with limited external validity, such as Andreeva *et al.* (2015) who study the link between depression and unemployment in relation to organizational downsizing in Sweden, and Hamilton, Merrigan, and Dufresne (1997) who estimate the relationship between psychiatric symptoms and employment using a relatively small sample of less

approach to individual-level panel data. Under this assumption and taking first-differences, we compute the bounds as outlined in the Supplementary Appendix.

than 800 Montreal residents. Finally, the majority of studies use cross-sectional data combined with IV strategies and we return to this issue below.

Our theoretical framework follows Grossman (1972), which in turn draws on Becker (1964); here health is a form of human capital, analogous to education. Health is valued both directly and also because poor health detracts from time available for both market and non-market work. Grossman's health investment model can be solved to derive a labour supply function that depends on an endogenous health variable (see Currie and Madrian, 1999 for an excellent exposition). The effect of an adverse health event on labour supply is theoretically ambiguous. A deterioration in health can reduce time available for work and reduce productivity; further it can lead to an increased preference for leisure time and/or increased time needed to maintain health. However, worsening health can also increase labour supply, especially in privatized health care markets like the United States. In these systems, for prime age adults, health insurance is generally provided with employment, and thus adverse health events can increase the costs of job loss, thus increasing the opportunity cost of non-work time; further, more work may be needed to cover the costs of health care that are not included in insurance coverage. Given that we are studying England and Wales, which provide universal health care coverage under the National Health Service, we would expect the negative impacts of worsening health on labour supply to dominate. This is particularly true for the effects of common mental disorders like anxiety and depression, which will be by far the most prevalent problems reflected in the MH measures we use in our empirical work. These disorders are largely treated in primary care by ‘talking therapies’ or pharmaceutical interventions, and they rarely require hospitalization.[5]

In some of the earliest econometric work on MH and employment, Bartel and Taubman (1979); Bartel and Taubman (1986) studied a sample of older aged male twins in the United States; in common with other early studies, they assume that MH is exogenous (e.g. Frank and Gertler, 1991). They find that physician-diagnosed psychoses and neuroses (which are relatively severe MH problems) were associated with a lower likelihood of employment. The vast majority of later evidence on the relationship between MH and employment comes from US cross-section studies that use IV in an attempt to deal with endogenous selection. Endogenous selection occurs because unobserved characteristics (such as motivation or childhood circumstances) and/or circumstances (like work relationships or the local economic environment) are correlated with both health and work outcomes. Commonly used instruments include: parental history of MH (Ettner, Frank, and Kessler, 1997; Marcotte, Wilcox-Gök, and Redmon, 2000; Banerjee, Chatterji, and Lahiri, 2017); childhood psychiatric disorders (Ettner *et al.*, 1997; Chatterji, Alegria, and Takeuchi, 2007; Fletcher, 2014; Banerjee *et al.*, 2017); participation in religious services and religious beliefs (Alexandre and French, 2001; Chatterji *et al.*, 2007); and perceived social support (Hamilton *et al.*, 1997; Alexandre and French, 2001; Ojeda

[5] In addition, treatment for these disorders are readily available in England and Wales, especially since the advent of IAPT, which is widely recognized as the most ambitious programme of talking therapies in the world. For example, in 2019, more than one million people accessed IAPT services and 89% of referrals waited less than 6 weeks to enter treatment. https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-report-on-the-use-of-iapt-services/december-2019-final-including-reports-on-the-iapt-pilots-and-quarter-3-2019-20-data/waiting-times

et al., 2010; Lagomarsino and Spiganti, 2020).[6] The general consensus from these studies is that MH has a negative influence on the probability of being in employment. However, as Chatterji, Alegria, and Takeuchi (2011) point out, the chosen instruments are often 'hard to justify based on economic theory'. Indeed, in their own study, Chatterji *et al.* (2007) admit that it is difficult to make a strong case for the exogeneity of their instrument; childhood psychiatric disorders, for example, can be argued to be underlying individual traits that can manifest later in life. Further, Chatterji *et al.* (2011) use the methods proposed by Altonji *et al.* (2005) to show the sensitivity of IV estimates to the extent of unobserved selection bias, and recommend that longitudinal data be used to explore selection based on unobserved personal characteristics, which is the primary focus of our study.

A further problem, which has received little or no attention in the health and work literature, is that many of the IV studies provide a very specific estimate of the local average treatment effect (LATE) calculated from some arbitrary exogenous shock that, in most cases, would not be an appropriate policy target. For example, causal evidence derived from religiosity does not help current policy makers design tools to tackle the MH disability employment gap. FE models can be useful in this respect because they estimate the average effect on labour market outcomes for those whose MH changes; and while this is not the effect of a particular intervention (which would be another specific LATE), it is easy to interpret and shows the scale of the problem to be tackled.

The most recent studies on MH and employment utilize longitudinal data. We know of only three such studies for the United Kingdom.[7] Using a discrete-time hazard framework Garcia-Gomez, Jones, and Rice (2010) use data from the British Household Panel Survey 1991 to 2002 to estimate the effect of psychological health (measured by the GHQ) on both entries to and exits from the labour market. They find that worsening MH increases the exit hazard for workers, with the magnitude being greater for men than women. However, they also find that worsening MH in non-workers increases the hazard of becoming employed. This is a difficult finding to explain, and they argue that it is because those individuals who are less happy with their current situation (not working) are more likely to return to employment.[8] More recently Lagomarsino and Spiganti (2020) have also used the same data set to estimate the effect of the GHQ score on wages, taking into account selection into employment. Their FE model shows a very small effect of GHQ on the employment probability, which is consistent with our results. However, their identifying variable for selection (perceived social support) is questionable given this can directly affect both employment and wages, as well as MH; and also they do not control for physical health in their analysis. In a recent working paper Jolivet and Postel-Vinay (2020) specify a dynamic structural model allowing for the two-way interaction between MH and work. They use this model to simulate the employment responses to a severe MH shock at age

[6]There is also a related strand of literature on the impact of substance abuse on employment outcomes, which has used instruments based on parental substance abuse problems and regional variation in alcohol and drug policies (see e.g. Mullahy and Sindelar, 1996; DeSimone, 2002; Terza, 2002).

[7]Steele *et al.* (2013) is a UK study, but their focus is the effect of labour market transitions on MH for males only.

[8]This is consistent with recent findings from the subjective well-being literature, that people who suffer a bigger drop in life satisfaction on becoming unemployed seem to search harder for a job and may find one more quickly (Mavridis, 2015).

30, using UKHLS data for men of prime working age. These simulations suggest that severe MH shocks have persistent effects on participation; these are partly a result of the direct health effects, but are also exacerbated by feedback loops where non-participation in one period worsens the probability of finding employment in the future.

For the United States, Mitra and Jones (2017) use data from two waves of the National Survey of Alcohol, Drug, and Mental Health Problems. Their preferred specification is a split first difference model, which they estimate separately for individuals who are initially employed and not employed; they also differentiate between mental illness onset and recovery. They find a positive association between the onset of an MH problem and a transition to non-employment for those who are initially employed with no MH problem; but little evidence for the reverse effect, that is, those who are not employed initially and have a health problem do not see an increased probability of employment upon recovery. Also in the United States, Peng, Meyerhoefer, and Zuvekas (2015) use data from five waves of the Medical Expenditure Panel Survey to explore the effects of depressive symptoms on employment. They use FE and correlated random effects models and find that exhibiting depressive symptoms reduce the likelihood of employment, and that the effect is larger for men than women.

Four studies use data from the Household, Income and Labour Dynamics in Australia (HILDA) survey. Butterworth *et al.* (2012) use the first five waves of the HILDA to track individuals who are not unemployed at baseline across four subsequent years. They find that baseline MH (measured using the Mental Health Inventory sub-scale of the SF-36) is a significant predictor of overall time spent unemployed.[9] Olesen *et al.* (2013) extend this to nine waves of HILDA and use path analysis to explore lagged and contemporaneous relationships between unemployment and MH (measured in the same way). Despite using longitudinal data, this study does not appear to account for unobserved individual effects. MH was shown to be both a risk factor for, and consequence of, unemployment. The strength of these two effects was similar for women, but for men the effect of MH on unemployment dominated. Bubonya, Cobb-Clark, and Ribar (2019) analyse the reciprocal lagged relationship between severe depressive symptoms and employment status by estimating correlated random effects models using 14 waves of HILDA. They find that severe depression contributes to a 26% increase in subsequent non-employment rates, with larger effects for men than women. They also find no evidence for men, and only limited evidence for women, that adverse labour market outcomes increase the incidence of depression. Finally for Australia, Frijters *et al.* (2014) use 10 waves of the HILDA data with an IV-FE model identified using the recent death of a close friend. They create their own measure of MH using nine questions from the SF-36 general health survey and explore its effect on employment. The results suggest that a one-standard-deviation decrease in MH leads to a 30-percentage-point decrease in the probability of being employed; an effect that is stronger for older than younger workers. This is a very large effect; for example, they show that it is roughly twice that of having a degree compared with dropping out of high school, and it is four times the size of the OLS estimate. The bereavement instrument is shown to be a strong determinant of MH

[9]Kaspersen *et al.* (2016) carry out a very similar study using data from a large Norwegian cohort, and find similar results.

and placebo tests suggest that it only affects labour market outcomes through its effect on MH. However, one issue not discussed by the authors is that the large IV estimate may be a result of the fact that it is a LATE showing the effect on employment for people whose MH has been affected by the death of close friend. It is not appropriate to extrapolate this estimate to the wider population of workers.

The advantage of longitudinal data and FE models is that they can be used to remove any bias arising from unobserved time-invariant factors that might influence both health and work; for example, the influence of adverse childhood circumstances (that are predetermined in a model for working age adults). However, there are also likely to be important unobserved factors, relating to both employment and health, which vary over time. For example, people's family circumstances, work relationships, tastes, and the macroeconomic environment are all things that are likely to affect both MH and employment; they vary over time and are rarely completely observed in secondary data. The inclusion of FE does not deal with this issue, and thus estimates of the effect of health on employment from FE models may still be biased; further, it is difficult to state the direction of this bias with any certainty.

In this article, we investigate the direction and magnitude of the potential bias in FE estimation using information on selection on unobservables. Our starting point is the common argument in the empirical literature that if a coefficient is stable after the inclusion of the observed controls, then omitted variable bias must be limited.[10] This reasoning rests on the assumption that bias arising from observed controls is informative of the bias arising from omitted unobserved factors. However, (Oster, 2013b, 2019) notes that it is also necessary to know how much of the variance in the outcome is explained by the inclusion of the controls. For instance, a control may be highly correlated with the variable of interest, but if it only explains a small fraction of the outcome variance, the coefficient will not change much when the control is added. In this example, the estimated bias would need to be scaled up to allow for the small contribution of the control to the variance explained. Oster combines this insight with assumptions about the relationship between observable and unobservable selection in order to derive bounds on the coefficients in OLS regressions. We apply Oster's method in a FE framework in which we compare 'uncontrolled' and 'controlled' employment regressions. The controlled regressions include the standard characteristics from the literature, so it is worth noting that there is already some evidence about selection into health. However, there does not appear to be a consensus about the direction of selection biases. For instance, Chatterji, Alegria, and Takeuchi (2008) provide evidence that people may be selected into psychiatric disorders along characteristics associated with better labour market outcomes (white ethnicity and divorced status for women) or worse outcomes (lack of college education and disadvantaged background). It is typically found that the size of the health effect diminishes when FEs are added to equations for labour market outcomes; so one might then conclude that any remaining bias is in the same direction, although there is no theoretical reason why this should be the case (Peng *et al.*, 2015).

[10]For example Frijters *et al.* (2014) rely on this reasoning to justify omitting certain variables from their model (p. 1063; footnote 4.)

III. Estimation approach

We start with a linear probability model (LPM) where the dependent variable, Y_{it}, is a binary indicator for whether an individual, i, is employed or not in wave t:

$$Y_{it} = \alpha + \beta M_{it-1} + Z'_{it}\theta + d_t\gamma + \mu_i + \varepsilon_{it}; \qquad i = 1, \ldots, N; \quad t = 1, \ldots, T, \tag{1}$$

where N is the total number of individuals and T is the total number of waves; M_{it-1} is a measure of the MH of individual i in wave $t-1$; Z_{it} is a vector of observed controls, including (time-varying) individual, household, and area characteristics, with associated parameter vector θ; d_t is a vector of $T-1$ wave dummy variables that control for time effects; μ_i are individual FE; and ε_{it} is the idiosyncratic error term. The parameter of interest is β, the effect of MH in the previous wave on the probability of being employed in the current wave. The LPM allows us to control for individual-specific effects that are correlated with the covariates, and is often the preferred choice to model health and work with FE (Greve and Nielsen, 2013; Bubonya *et al.*, 2019). It is also used in the wider literature to model binary labour market outcomes. For example, Agüero and Marks (2008) use an LPM to investigate the relationship between children and female labour force participation in Latin America; Francesconi and Van der Klaauw (2007) use it to model employment and other binary outcomes, such as benefit receipt, among lone parents; and Gregg *et al.* (2011) use an LPM to model the choice to work unpaid overtime. The combination of individual FE and lagged MH is an attempt to minimize reverse causality bias from employment to health status. One drawback of this approach is that we do not obtain an estimate for the effect of contemporaneous MH.[11] However, it is reasonable to assume that MH changes will take some time to feed through to labour market outcomes.

While the model in equation (1) fully controls for time-invariant heterogeneity by including individual FE, μ_i, there could still be time-varying heterogeneity that is not fully controlled for by the observed variables in Z_{it}. The method developed by Oster (2019) is useful for assessing the amount of bias that these omitted unobserved variables would cause under certain assumptions. Here we provide a brief description of the method and how we apply this to individual longitudinal data, with more detail provided in the Supplementary Appendix.

The method rests on the specification of two regression equations: a *controlled* regression, which, like equation (1) above, includes the key variable of interest (MH, M, in our case), as well as all observable factors Z; and an *uncontrolled* regression, which includes only M, and any observed covariates whose correlation with the key explanatory variable of interest is not informative about selection bias. Conceptually there is also a set of unobserved variables that are correlated with both M and the outcome Y, but which are necessarily omitted from the controlled regression. In order to estimate the degree of bias in the estimate of β rising from these omitted variables, the method utilizes the correlation between the observables and M, together with information on how much of

[11] In models with contemporaneous health (not reported here), we find that the effect of MH is larger in magnitude but qualitatively the same as in our lagged models.

the R-squared is explained by the observed controls, to compute the correlation between the unobservables and M under certain assumptions.

The original method was developed for cross-sectional models, so in order to apply it to our longitudinal case, we transform our data using within-individual means, and denote the demeaned linear model as:

$$\ddot{Y}_{it} = \beta\ddot{M}_{it-1} + \ddot{Z}_{it}'\theta + \ddot{d}_t\gamma + \ddot{\varepsilon}_{it}; \qquad i = 1, \ldots, N; \quad t = 1, \ldots, T, \tag{2}$$

where $\ddot{Y}_{it} = Y_{it} - \overline{Y}_i$, and $\overline{Y}_i = \sum_t^T Y_{it}/T_i$; and similarly for the other variables. This eliminates the individual FE, μ_i from equation (1), and allows us to estimate β using OLS. We denote the estimate of β from this regression, commonly known as the within estimator, as $\tilde{\beta}$. We show the derivation of the omitted variable bias for this model in the Supplementary Appendix. Our corresponding uncontrolled regression is:

$$\ddot{Y}_{it} = \beta\ddot{M}_{it-1} + \ddot{d}_t\gamma + \ddot{\nu}_{it}; \qquad i = 1, \ldots, N; \quad t = 1, \ldots, T, \tag{3}$$

where $\ddot{\nu}_i = \nu_{it} - \overline{\nu}_i$, and ν_{it} is the error term from the FE model $Y_{it} = \alpha + \beta M_{it-1} + d_t\gamma + \mu_i + \nu_{it}$. $\overline{\nu}_i$ is the within-individual mean of ν_{it}. The demeaned time dummies, $\ddot{d}_t$, are included in equation (3) because they capture time trends that may be correlated with M. As these time trends are fully captured by $\ddot{d}_t$, any change in the coefficient on M when they are added does not tell us what would happen if further time-varying controls were added.[12] In contrast, the covariates $\ddot{Z}_{it}$ in equation (2) are assumed to imperfectly capture the relevant time-varying factors that influence the relationship between M and Y; thus there are unobserved counterparts to $\ddot{Z}_{it}$.

Two key parameters specify the relationship between observable and unobservable selection and the maximum amount of variation that can be explained by the model. The first parameter, δ, defines the importance of the unobservables relative to the observables in influencing M. When $\delta = 1$ the observables and the unobservables are equally important and affect β in the same direction; when $0 < \delta < 1$ the unobserved factors are less important than the observed factors (and the opposite holds when $\delta > 1$).[13] The second parameter, $R_{\max}$, is the (theoretical) maximum R-squared from the full model where all observed and unobserved variables are included. This can be as high as 1 if Y is measured without error, but cannot be smaller than the R-squared obtained from the controlled regression. Both δ and $R_{\max}$ are unknown parameters to be chosen given the particular context of the problem. It is generally argued that an appropriate upper limit for δ is 1 because the observed variables are usually chosen based on the fact that they are the most important controls (based on theory and/or previous empirical evidence). The range 0–1 for δ seems reasonable in our context, as we observe the key control variables that

[12]Oster (2013b) discusses the case of controls that fully capture the relevant explanatory variables and therefore have no unobserved counterpart. Take for example gender: 'since it is fully observed it may be inappropriate to assume that resulting coefficient movements reflect what would happen with additional controls' (Oster, 2013b, p. 10). However, the choice of which controls are informative of selection bias and which are not is more complex and depends on the econometric model being used as well as theoretical considerations (see discussion in section 3.2.1 in Oster, 2013b).

[13]δ can also be negative in theory if the effect of the unobservables on β is in the opposite direction to the observables. However, we do not expect this to be the case in our application.

have been identified in the literature on health and work. It is reasonable to assume that $R_{\max}$ is less than 1 if idiosyncratic measurement error in Y exists. It also seems appropriate to assume $R_{\max} < 1$ when modelling a discrete employment outcome using a linear equation.[14] In our analysis, we consider a range of values suggested by Oster's (Oster, 2019) empirical survey of randomized studies, $R_{\max} = 1.3\tilde{R}$ and $R_{\max} = 2.2\tilde{R}$; where $\tilde{R}$ is the R-squared value from equation (2). While the assumptions around δ and $R_{\max}$ are not testable in our analysis, we apply the method essentially as a robustness check on our results rather than as a method to correct for any bias that may arise. This allows us to compute a bounding set Δ_s with the following bounds on β: (i) $\tilde{\beta}$ which is the estimate of β in the controlled regression (equation (2)), and (ii) β^*, which is the effect of MH on employment corrected for omitted variable bias given the specified values of $R_{\max}$ and δ.[15] Whether $\tilde{\beta}$ is the upper or lower bound of Δ_s will depend on the direction of the MH effect and the direction of the bias. For a positive MH effect, $\tilde{\beta}$ is a lower bound in the presence of downward bias, and an upper bound in the presence of upward bias. The opposite is true if the MH effect is negative.

IV. Data

We use the first nine waves of the UK Household Longitudinal Study (UKHLS, 2019), with wave 1 data being collected in 2009/2010, wave 2 in 2010/2011, and so on until wave 9, which was collected in 2017/2018. We limit our analysis sample to those aged 21–55 years from England and Wales, in order to retain a focus on prime age workers.[16] Table A1 in the Appendix provides detailed definitions for all the variables in our models. The dependent variable (Y) takes the value 1 if the individual is self-employed or in paid employment (full- or part-time);[17] 0 if the individual is unemployed, retired, looking after family/home, or long-term sick/disabled.

We use three main alternative measures of MH for our key explanatory variable (and consider three further measures in our sensitivity analyses); two derived from the 12-item General Health Questionnaire (GHQ-12) and one from the Short-Form 12 health questionnaire (SF-12). The GHQ-12 is a widely recognized instrument that has been adopted by the WHO as a screening tool for psychological disorders and has been validated in a number of international studies (Sartorius and Ustün, 1995; Goldberg *et al.*, 1997; Schmitz, Kruse, and Tress, 1999). The GHQ has also been shown to

[14]The R-squared from a within-individual regression will typically be much smaller than from a cross-sectional or pooled regression. To check that our results were not sensitive to small values of R-squared, we also estimated a correlated random effects specification (Mundlak, 1978); this is specified in levels but includes the individual means of all observed time-varying characteristics to model the unobserved individual effect. The individual means were included in both the uncontrolled and controlled equations and the overall R-squared used in the bias calculation. The results were very close to those from our within-individual specifications (results available on request).

[15]These bounds can be estimated using Stata (Oster, 2013a).

[16]Our data cover the period immediately following the financial crisis of 2007/8. In common with many other countries around the world, the United Kingdom experienced a period of austerity followed by economic recovery. Our arguments are primarily methodological and we do not wish to imply that the quantitative effects we report will hold for other countries or other time periods. Nevertheless, the unemployment rate experienced in the United Kingdom over our analysis period followed a very similar trend to that observed in other EU countries.

[17]Approximately 10% of the observations in our sample are self-employed individuals. We also conduct the analysis excluding this group and the results do not change.

be predictive of face-to-face clinical diagnosis of MH problems (Anjara *et al.*, 2020). This instrument is used as a measure of psychological health in an increasing number of economic studies (see, e.g. Cornaglia, Crivellaro, and McNally, 2015; Gardner and Oswald, 2007; Roberts, Hodgson, and Dolan, 2011); including studies of the relationship between MH and work (see, e.g. Garcia-Gomez *et al.*, 2010; Mavridis, 2015). Our primary measure of MH status is a binary indicator that identifies individuals with a possible psychiatric disorder. This measure is derived from the GHQ-12 caseness score. The original GHQ scale permits responses of 0−3 for each of the 12 questions. The caseness score recodes values of 0 and 1 on individual questions to 0, and values of 2 and 3 to 1; the sum then gives a scale running from 0 (least distressed) to 12 (the most distressed). Our dummy indicator (GHQ12D) is 1 when the GHQ-12 caseness score is between 4 and 12, and 0 when the score is between 0 and 3. This 3/4 cut-off is used by the NHS to monitor the percentage of people who suffer from poor MH in the general population.[18] In section V, we also vary this measure to consider a lower cut-off at 2/3 and two higher cut-offs at 4/5 and 5/6. Our second MH measure, also from the GHQ-12, is a cardinal measure based on the original four-point Likert scoring for each question, which ranges from 0 to 36 (henceforth GHQ36) where a higher value corresponds to worse MH.

Our third measure of MH is the mental component summary (MCS) derived from the SF-12. The SF-12 is a multidimensional generic measure of health-related quality of life that is widely used in clinical trials and routine outcome assessment because of its brevity and psychometric performance.[19] The MCS is designed to have construct validity in that it is able to discriminate between groups of patients who differ in MH condition according to clinically assessed diagnoses (Ware *et al.*, 2002; Gill *et al.*, 2007). The original score ranges from 0 to 100 where higher values denote better MH and the scoring method is based on an algorithm developed by Ware *et al.* (2002); this uses population norm based scoring so that the measure has a mean of 50 and a standard deviation of 10. For consistency with our other two MH measures, we recode the MCS so that higher values denote worse MH. The MCS has been used to analyse the MH effects of learning intensity (Hofmann and Mühlenweg, 2019), working-time mismatch (Otterbach, Wooden, and Fok, 2016), and work schedules of sole-parents (Dockery, Li, and Kendall, 2016). Mitra and Jones (2017) use it to estimate the impact of MH changes on labour market outcomes in the United States; and Andersen (2015) uses it to explore the effects of changes to MH insurance mandates on a number of labour market outcomes.

Previous work on the health and employment relationship has revealed that the estimated effects are quite sensitive to the health measures used (see Currie and Madrian, 1999 for a review). Reporting bias is a concern for the general self-assessed health measures that are often used in economic analysis of the health and work relationship, such as where the respondent is asked to rate their overall health on a scale of 1−5 (see

[18]For further details see https://files.digital.nhs.uk/BA/46AF8E/Spec_03J_321VSP2_10_V1.pdf. See also Goldberg, Oldehinkel, and Ormel (1998) for a discussion of GHQ thresholds around the world.
[19]The SF-12 is itself derived from the longer SF-36 health questionnaire; it was designed to be a briefer survey than the SF-36 with minimal loss of information (Ware *et al.*, 2002).

Jones *et al.*, 2010 for a discussion).[20] However, this type of bias is much less likely to be present in the validated psychometric instruments we use here, which are comprised of sets of relatively objective questions on specific aspects of health and functioning and do not explicitly refer to work capability. These questions are less prone to the potential positive bias that arises where individuals rationalize poor employment outcomes by self-reporting poor MH (Kreider and Pepper, 2007). In addition, our measures are also preferred to the use of specific MH conditions, such as anxiety and depression, since these are unlikely to capture all of the important aspects of the MH stock that influence employment, and they rarely contain any additional information on severity. Blundell *et al.* (2017) show that the use of these narrow objective measures leads to a downward bias in the estimated effect of health on employment. In contrast, Frank and Gertler (1991) find very similar estimates of the effect of MH conditions on wages whether they use assessment based on detailed interviews or a simple self-report of whether the respondent had ever received a diagnosis of a major MH disorder.[21]

For the individual and household level controls (Z in equation (1)), we consider those variables that are commonly used in the existing literature. These include age,[22] marital status, highest level of education achieved, presence of children in household (by age groups), number of adults in household, and other household income. We also control for the physical health (PH) of the individual using the SF-12 physical component summary (PCS); this is the PH equivalent of the MCS, with the score ranging from 0 to 100 where higher values denote worse health (Ware *et al.*, 2002). In some specifications, we also allow for comorbidity between MH and PH by including an interaction term between the two measures.[23] Further, in sensitivity analysis, we replace the PCS with a variable derived from questions on activities of daily living; these record whether the respondent has difficulties with physical functioning, such as mobility, manual dexterity, or hearing. As with MH, the PH measures are also included as lagged values. To take account of the local economic environment, we include two variables at the local authority district (LAD) level, namely the unemployment rate and gross value added (GVA). All other time-invariant characteristics available in the data (such as sex) are captured by the individual FE.

Tables 1 and 2 show descriptive statistics for our estimation sample split by the dichotomous GHQ measure of MH. In total, there are 98,435 observations covering 11,683 men and 14,851 women.[24] Approximately a fifth are identified as having poor MH (GHQ12D = 1) and these respondents accordingly have higher GHQ36 and MCS scores; they also have worse PH as shown by the PCS scores and problems with ADL. They are

[20] Brown *et al.* (2018) argue that there is ‘over-reporting’ of the ‘best’ response categories in the GHQ questions, and show that this can lead to underestimation of the effect of GHQ on economic outcomes, including employment.
[21] We did consider using a self-reported indicator of diagnosed depression in our modelling. However, while respondents can report new diagnoses in each wave of the UKHLS, in subsequent waves there is no mechanism for them to record whether they still have the problems they reported previously. This makes the diagnosis questions an unreliable indicator of the specific health problems that a respondent has in any particular survey wave.
[22] Although we have exact age, we use seven 5-year age groups in our analysis to allow for possible nonlinear effects (21–25, 26–30, 31–35, 36–40, 41–45, 46–50, 51–55).
[23] For conciseness, we do not report these results as the interaction effects between MH and PH were very small and the main effects were largely unchanged by their inclusion.
[24] There are slightly fewer observations available for the ADL measures.

TABLE 1
Summary statistics

	GHQ12D = 0 (good MH)			*GHQ12D = 1 (poor MH)*		
	NT	*Mean*	*SD*	*NT*	*Mean*	*SD*
Employed	78,719	0.87		19,716	0.70	
GHQ12D t-1	78,719	0.13		19,716	0.48	
GHQ36	78,719	9.13	(2.88)	19,716	20.44	(5.19)
GHQ36 t-1	78,719	10.22	(4.57)	19,716	15.81	(7.09)
MCS	78,719	48.85	(7.31)	19,716	64.01	(10.13)
MCS t-1	78,719	49.77	(8.57)	19,716	58.68	(11.49)
PCS	78,719	47.21	(8.07)	19,716	51.06	(13.18)
PCS t-1	78,719	47.01	(8.25)	19,716	50.88	(12.13)
ADL problems	78,675			19,689		
None		0.91			0.74	
1–2		0.07			0.15	
3–4		0.02			0.07	
5 or more		0.01			0.04	
Age	78,719	40.59	(9.14)	19,716	40.71	(9.32)
Married	78,719	0.74		19,716	0.65	
Education level	78,719			19,716		
No education		0.04			0.06	
O-level		0.28			0.31	
A-level		0.21			0.20	
Degree		0.47			0.43	
No child in HH	78,719	0.50		19,716	0.53	
Child 0–4 in HH	78,719	0.20		19,716	0.19	
Child 5–11 in HH	78,719	0.30		19,716	0.27	
Child 12–15 in HH	78,719	0.20		19,716	0.19	
Adults in HH	78,719	2.34	(0.98)	19,716	2.29	(1.05)
Other HH income	78,719	2752	(2411)	19,716	2565	(2263)
Unemployment rate	78,719	6.90	(2.89)	19,716	7.12	(2.95)
GVA	78,719	24,777	(15903)	19,716	24,603	(16737)

Notes: GHQ12D is the GHQ caseness dummy; MCS is the SF12 Mental Component Summary; PCS is the SF12 Physical Component Summary; GVA is Gross Value Added (per capita). With the exception of Age and GVA, T-tests of differences in means reveals all are significantly different to zero at the 1% level, except the difference in A-level, which is significant at the 5% level. Age and GVA are not statistically different at conventional levels.

also less likely to be employed (70% employed vs. 87% for those who do not have an MH problem), be married, or have higher education.[25] However, they are similar in terms of the age distribution. Other household income is lower in the households of people with poor MH; and they also live in areas with a higher unemployment rate and lower GVA. Table 2 shows that among the non-employed, a similar proportion of those with a MH problem are unemployed compared with those without a problem (31.1% and 29.2% respectively), but the largest group of those with poor MH are long-term sick or disabled (34.9%), while the majority of those who do not have poor MH are involved in family/home care (54.1%). Out of the total 26,534 respondents, 11.7% change their employment status over the period

[25]This employment rate (70%) is much higher than that cited for the general population in the introduction (35%), because our analysis sample includes only people aged 21–55 years and excludes full-time students and those on government training schemes.

TABLE 2

Observations by employment status (all waves pooled)

	GHQ12D = 0		GHQ12D = 1	
	NT	% of non-employed	NT	% of non-employed
Self employed	8,593		1,569	
Paid employment (ft/pt)	59,617		12,142	
Total employed	68,210		13,711	
Non-employed				
Unemployed	3,073	29.2%	1,865	31.1%
Retired	424	4.0%	83	1.4%
Family care or home	5,685	54.1%	1952	32.5%
LT sick or disabled	1,297	12.3%	2,094	34.9%
On apprenticeship	30	0.3%	11	0.2%
Total non-employed	10,509		6,005	

of analysis with most experiencing only one transition (4.6% move from non-employment to employment and 3.2% move from employment to non-employment), while only 3.9% change employment status multiple times. Exactly 68% of our sample show no change in MH status (according to the binary indicator GHQ-12D) throughout. Exactly 7% show one change from good to poor MH, and approximately the same proportion show one change in the opposite direction. The vast majority of the rest of the sample (9%) show two changes, going from good MH to poor and then back to good. Exactly 2% have the reverse two-change pattern, and the remaining very small proportion have more than two changes. For the controls, sufficient change is needed to be informative about the bias due to time-varying unobserved factors, so we assess the amount of within-individual variance as a proportion of total variance (Appendix Table A2). For most of the controls, within variance is in the region of 25–40% of total variance. The main exceptions are education, which not surprisingly does not change much in our prime age adult sample (within variance is only 1–2% of total variance), marital status (13%), and the number of adults in the household (17%). The change in GVA is also small (6%), however the local unemployment rate changes considerably (37%).

Figure 1 plots the employment gap by age between individuals with good and poor MH (as measured by the GHQ12D binary indicator) in all nine waves of our UKHLS sample; this gap is substantial and appears to widen with age.[26]

V. Results

Table 3 contains point estimates for models using the GHQ12D dichotomous measure for MH. For comparison purposes, the first two columns show the results when the LPM is estimated without FE by applying OLS to the raw untransformed data (we refer to these as pooled OLS models). We report estimates without controls (column 1) and with controls, $\ddot{Z}$, (column 2).[27] The next two columns show the estimates for the LPM with

[26]Note that the employment gap (of around 15 percentage points) in our data is narrower than that commonly cited in recent reports, and this is largely due to our focus on prime age workers.

[27]All equations include wave dummy variables.

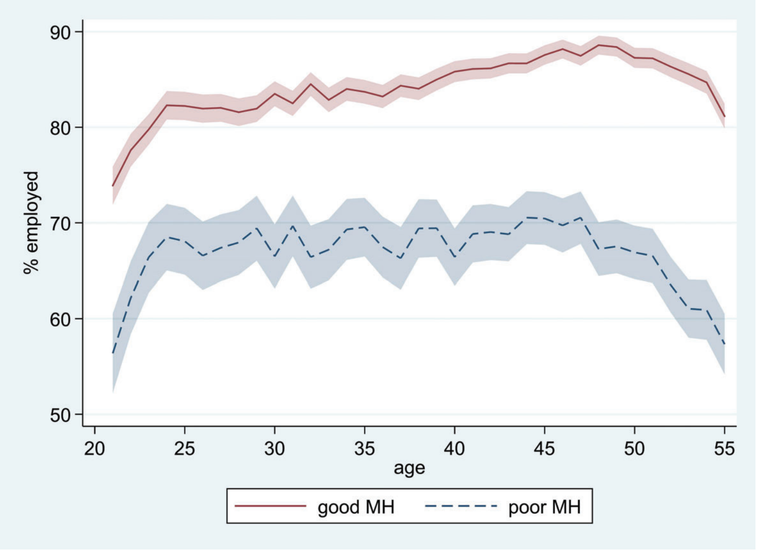

Figure 1. Employment by age and MH (95% confidence intervals)
Source: UKHLS Waves 1–9 (UKHLS, 2019).

FE, obtained using OLS on the demeaned data (we will refer to these as FE models): from the uncontrolled regression, equation (3), in column 3 and from the full controlled model, equation (2), in column 4. We have also estimated these models for each gender separately, but we find no significant differential gender effect on employment.[28] However, we find significant gender differences for being married, having children aged 0–4, 5–11, and 12–15. We therefore include in our controls gender interactions with being married and all the children variables.[29]

The pooled OLS coefficient in column 1 shows that poor MH is associated with a 15.5 percentage point lower probability of employment (controlling only for wave dummies). When the main controls are added, the absolute size of the effect is reduced to −9.7 percentage points, and when we include FE in the specification, the effect falls still further to −1.6 percentage points. There is thus quite strong selection into MH problems based on observed characteristics but especially strong selection based on time-invariant characteristics as a whole (both observed and unobserved). Indeed once FEs are included, it makes little difference whether we include the additional controls. This provides some tentative evidence (which we investigate formally below) that once cross-sectional selection is removed, there is little remaining time-varying selection bias. In the preferred specification (column 4), having poor MH lowers the probability of being employed by approximately 1.6 percentage points, which suggests that the causal effect of MH accounts for about 10% of the raw employment gap (column 1).[30] The effect is very similar in separate gender regressions (−1.60 percentage points for women and −1.62

[28]This is in line with the findings of Ettner *et al.* (1997), who also find no significant gender differences.
[29]For conciseness, these interaction effects are not reported in the tables.
[30]Cseh (2008) finds a similar result from his US study of the relationship between depression and wages. His pooled OLS models show a strong negative correlation between depression and wages, but this largely disappears once

TABLE 3
LPM: MH = GHQ caseness indicators

	(1) Pooled OLS	(2) Pooled OLS	(3) FE	(4) FE
GHQ12D t-1	−0.1547***	−0.0972***	−0.0157***	−0.0160***
	(0.0051)	(0.0040)	(0.0026)	(0.0026)
PCS t-1		−0.0095***		0.0010***
		(0.0002)		(0.0002)
Female		0.0839***		
		(0.0140)		
Age 26–30		0.0124*		0.0346***
		(0.0070)		(0.0079)
Age 31–35		0.0346***		0.0565***
		(0.0078)		(0.0100)
Age 36–40		0.0402***		0.0627***
		(0.0078)		(0.0115)
Age 41–45		0.0418***		0.0681***
		(0.0076)		(0.0130)
Age 46–50		0.0332***		0.0688***
		(0.0073)		(0.0144)
Age 51–55		0.0080		0.0609***
		(0.0074)		(0.0157)
Married		0.1474***		0.0445***
		(0.0073)		(0.0075)
O-level		0.2238***		0.0534*
		(0.0133)		(0.0320)
A-level		0.3098***		0.0962***
		(0.0134)		(0.0355)
Degree		0.3409***		0.1061***
		(0.0131)		(0.0353)
No child in HH		−0.0053		−0.0068
		(0.0077)		(0.0062)
Child 0–4 in HH		−0.0171***		−0.0101**
		(0.0066)		(0.0050)
Child 5–11 in HH		−0.0031		0.0057
		(0.0056)		(0.0045)
Child 12–15 in HH		−0.0021		0.0065
		(0.0066)		(0.0050)
Adults in HH		0.0212***		0.0208***
		(0.0023)		(0.0024)
ln (other HH income)		−0.0276***		−0.0149***
		(0.0009)		(0.0009)
Unemployment rate		−0.0072***		0.0001
		(0.0006)		(0.0005)
GVA per head/10000		−0.0034***		0.0025
		(0.0012)		(0.0015)
R-squared	0.0293	0.2318	0.7915	0.7956
Within R-squared			0.0036	0.0236

Notes: Regressions based on 98,435 observations (NT). All models include wave dummies. Regressions (2) and (4) include gender interaction terms with being married and all the children variables. Regressions (1) and (2) are LPM without individual fixed effects using OLS on the untransformed data; regressions (3) and (4) are LPM estimated using OLS on the demeaned data.
Clustered standard errors in parentheses, $^{*}P < 0.10$, $^{**}P < 0.05$, $^{***}P < 0.01$.

for men), and as mentioned previously, this difference is not statistically significant. We also considered asymmetric effects by including separate MH variables for positive and negative differences from the mean (an approach similar to Allison, 2019). We find no evidence of asymmetric effects for our two GHQ measures (i.e. the effect of a positive change in MH is not statistically different from the opposite effect of a negative MH change), but the effect of a positive change in MH is significantly larger in magnitude than the opposite effect of a negative change in MH for our MCS measure.

The control variables in column 4 all appear to have the expected effects on the employment probability. Poor PH is associated with reduced employment, and while the effect is much smaller in the FE model compared with the pooled OLS coefficient in column 2, it is still statistically significant. In the FE model, the effect of age is significantly larger for all age groups compared with those 21−25 (the youngest group). Being married increases the probability of being employed, but having pre-school aged children (aged 0−4) in the household lowers it. The gender interaction terms (not shown in table) reveal a significantly lower marriage effect on employment for women compared with men, a stronger negative effect of having children aged 0−4 or children aged 5−11, and a larger positive effect of having secondary school aged children (aged 12−15). The education gradient is as expected; those who gained qualifications, especially A-levels or a degree, have a higher likelihood of being employed than those with no formal qualifications. A higher number of adults living in the household also increases the probability of being employed, while a higher level of other household income lowers it. Neither of the area level controls (unemployment rate and GVA per head) are statistically significant in the FE model, which for GVA may reflect the fact that although it varies a lot spatially, it exhibits only limited variation over time. The local unemployment rate does exhibit a large amount of within variation (as shown in Table A2), but it was also found to be statistically insignificant in FE models of the effect of health on employment by Auld (2002) and Webber and Bjelleand (2015) for the United States, and Cai (2021) for the United Kingdom.[31]

In Table 4 we re-estimate the specifications from Table 3 using the continuous GHQ36 measure and the SF12 MCS; for both of these measures, higher values represent worse MH. As expected, we find a negative relationship between these measures and the probability of being employed.[32] Again, in the pooled OLS regressions, the addition of control variables reduces the magnitude of the estimated effect of MH for both measures, and the inclusion of FE reduces both estimates still further. In the FE models, the addition of controls increases the magnitude of both the GHQ36 and the MCS coefficients slightly. This is consistent with the GHQ12D models from Table 3 and suggests that, contrary to the cross-sectional selection effects, whereby characteristics that are positively associated

FEs are introduced. His interpretation is that any correlation between depression and wages reflects differences in personality, rather than MH.

[31]We have also explored the effects of other caring responsibilities, since these may affect both labour supply and MH (Savage and Bailey, 2004; Carmichael *et al.*, 2008). Models including variables that indicate whether the individual provides care for another adult inside or outside the household give very similar results to those presented here. The caring variables are negative and significant in the pooled OLS models but not in the FE models. The corresponding gender interactions are also negative.

[32]For conciseness, we do not report the results for the other control variables in Table 4; they are very similar to those in Table 3.

TABLE 4

LPM: Alternative MH measures

	(1) *Pooled OLS*	*(2)* *Pooled OLS*	*(3)* *FE*	*(4)* *FE*
Panel 1				
GHQ36 t-1	−0.0141***	−0.0090***	−0.0018***	−0.0019***
	(0.0004)	(0.0003)	(0.0002)	(0.0002)
	[−0.0794]	[−0.0506]	[−0.0103]	[−0.0105]
PCS t-1		−0.0091***		0.0010***
		(0.0002)		(0.0002)
Controls	No	Yes	No	Yes
R-squared	0.0471	0.2386	0.7916	0.7958
Within R-squared			0.0042	0.0243
Panel 2				
MCS t-1	−0.0082***	−0.0067***	−0.0009***	−0.0013***
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
	[−0.0809]	[−0.0659]	[−0.0093]	[−0.0131]
PCS t-1		−0.0102***		−0.0015***
		(0.0002)		(0.0002)
Controls	No	Yes	No	Yes
R-squared	0.0486	0.2511	0.7915	0.7959
Within R-squared			0.0039	0.0247

Notes: Standard errors in parentheses (clustered at individual level). Standardized MH coefficients in brackets. All models include wave dummies. Sample size (NT): 98,435. Regressions (1) and (2) are LPM without individual fixed effects using OLS on the untransformed data; regressions (3) and (4) are LPM estimated using OLS on the demeaned data.

Where $^{*}P < 0.10$, $^{**}P < 0.05$, $^{***}P < 0.01$.

with employment are also positively related to MH, the opposite is true for time-varying selection, that is, changes in characteristics, which increase the probability of employment, also lead to reduced MH. Again, we investigate this formally below.

In Table 4 we also report the standardized coefficients on GHQ36 and MCS (reported in brackets). Our preferred specification (column 4) suggests that a one standard deviation increase in GHQ36 (MCS) leads to a 1.0 (1.3) percentage point decrease in the probability of employment. In the MCS model, the MH and PH measures are directly comparable since they both use SF-12 summary scores; the results suggest that changes in physical health have a slightly larger effect on employment compared with MH. In Table 3, the results showed that when comparing the OLS model with no controls to the FE model with controls, the causal effect of MH explained approximately 10% of the raw gap. For the GHQ36 (Table 4 Panel 1), the equivalent statistic (based on relative coefficient sizes) is 13%, and for the MCS (Panel 2) it is 16%.

Table 5 reports the bounds of the value of β from the FE models with full controls. For ease of comparison, the first column repeats the estimates from the controlled regression in equation (2) (i.e. $\tilde{\beta}$ when $\delta = 0$). The bias-adjusted estimates (β^{*}) under the assumption that $\delta = 1$ are shown in column 2 (setting $R_{\max} = 1.3\tilde{R}$) and column 3 (setting $R_{\max} = 2.2\tilde{R}$) with bootstrapped standard errors in brackets. We find that for the GHQ36 measure the bias-adjusted estimates are the same as the FE estimates ($\tilde{\beta}$) under both $R_{\max}$ assumptions. For GHQ12D and MCS, as was suggested by the coefficient changes noted above, some bias is exhibited; $\tilde{\beta}$ is the lower bound (in terms of magnitude)

TABLE 5

Oster bounds for FE models with full controls

		$\delta = 1\,(\beta^*)$	
	$\delta = 0(\tilde{\beta})$	$R_{max} = 1.3\tilde{R}$	$R_{max} = 2.2\tilde{R}$
GHQ12D t-1	−0.0160	−0.0161	−0.0165
	(0.0026)	[0.0028]	[0.0028]
GHQ36 t-1	−0.0019	−0.0019	−0.0019
	(0.0002)	[0.0002]	[0.0002]
MCS t-1	−0.0013	−0.0015	**−0.0020**
	(0.0002)	[0.0002]	[0.0002]

Notes: Bootstrapped standard errors in square brackets (1,000 reps). Clustered standard errors in parentheses. Estimates in bold denote non-overlapping 95% confidence intervals between the bounds and original estimates.

and β^* the upper bound. However, this result should be interpreted with caution as these upper bounds are very close to the estimated coefficient. The bounds fall within the 95% confidence intervals of $\tilde{\beta}$ in all cases with the exception of MCS when $R_{\max} = 2.2\tilde{R}$.[33] This suggests there is little concern regarding omitted variable bias in these FE models for the effect of MH on employment.

Sample attrition

Analyses of longitudinal data create the risk of introducing bias due to survey non-response, either by individuals exiting the survey or failing to respond to particular survey questions.[34] Survey non-response may be directly related to labour market outcomes, for example, people moving for reasons related to employment may be lost to follow-up, and may have different health and socio-economic characteristics than those who remain in the survey. Non-response may also be directly related to health status, with individuals suffering from poor MH being less able or less motivated to respond to survey questionnaires.[35] We explore the presence of non-response bias using tests proposed by Verbeek and Nijman (1992). First, we compare estimates of the parameter for MH ($\hat{\beta}$) derived from the unbalanced sample (the sample used in the main estimation results) to that derived from a balanced sample. Estimates that coincide suggest a lack of item non-response and attrition bias. Secondly, we use a variable addition test where the constructed variable representing whether an individual is observed to be in the next wave

[33]In fact, the confidence intervals between $\tilde{R}$ and β^* (when $R_{\max} = 2.2\tilde{R}$) in the MCS model do not overlap. These findings hold even when we increase $R_{\max}$ to $3\tilde{R}$, which we do not consider to be a plausible assumption, but serve to demonstrate the robustness of the results.

[34]Nicoletti and Peracchi (2005) provide a useful taxonomy of reasons for non-participation in surveys.

[35]Jones, Koolman, and Rice (2006) provide evidence on survey non-response and attrition bias for models of self-reported health in the British Household Panel Study (the precursor to the UKHLS) and find evidence of health-related non-response bias. Individuals with poor initial health were generally more likely to drop out of the survey. However, younger individuals in good health were also more likely to provide non-responses. A comparison of parameter estimates corrected for non-response (using inverse-probability weighting) showed little substantive differences in the magnitudes of average partial effects of interest. This finding appears to corroborate other literature that reports a limited influence of non-response bias, for example, in models of income dynamics and various labour market outcomes (e.g. see Hausman and Wise, 1979; Becketti *et al.*, 1988; Lillard and Panis, 1998; Zabel, 1998; Ziliac and Kniesner, 1998; Jimenez-Martin and Peracchi, 2002).

of the survey is included in an augmented regression on the unbalanced sample. The null hypothesis under the assumption of no item or survey non-response bias is that the parameter estimate is zero.

Estimates from FE on the unbalanced and balanced samples,[36] using the GHQ12D dichotomous measure of MH, are very similar [unbalanced coefficient: −0.016, (SE: 0.0026); balanced coefficient: −0.015, (SE: 0.0056)].[37] Augmenting equation (1) and undertaking the variable addition test results in a coefficient on the variable next wave of 0.0018 (SE: 0.0024).[38] Accordingly, there is insufficient evidence to reject the null of no item or survey non-response bias.[39]

Heterogeneity analysis

In order to explore the heterogeneity of effects in different subgroups, we focus on the preferred model (equation (2); column 4 in Tables 3 and 4) and split the sample by education, PH terciles, and relative poverty. Respective results are reported in the Appendix as Tables A4 to A6. Differences across groups were tested for statistical significance and this is noted in the rightmost column where relevant. The relationship between the GHQ12 dummy and employment remains negative and statistically significant. The effect of MH is significantly smaller in magnitude across all three MH measures for those with a degree than for those without a degree (Table A4), and significantly smaller for those households living above the relative poverty line than for those below (Table A6). This suggests that higher education moderates the effect of MH disorders on employment, while relative poverty exacerbates it. There is also some evidence that the effect of MH is larger for individuals with worse PH (Table A5), but this difference is only statistically significant for MCS (comparing those in the top PCS tercile to those in the bottom PCS tercile).

We note that the effects of PH are more heterogeneous across levels of PH. They are larger for those in the bottom PCS tercile compared with those in the middle and top terciles across all models (Table A5). Like MH, the effect of PH is also larger for those below the poverty line compared with those above the poverty line (Table A6). However, unlike MH, the effects of PH are not statistically different between those with a degree and those without in all three models (Table A4). It is also worth noting that, in addition to the results presented here, we also explored differences across subgroups defined by age, gender, household income, and whether there is another employed person in the household. We found no significant differences in the effect of MH on employment between these groups.

[36]Full results available on request.

[37]The difference in these parameter estimates can be formally tested using a Hausman test (Hausman, 1978). It is clear, however, that they do not diverge significantly across our estimates.

[38]Similar results are found when using the alternative measures of MH (GHQ36 and PCS – unbalanced coefficient: −0.019(0.0002); balanced coefficient: −0.0018(0.0005), next wave – coefficient: 0.00158(0.0024); MCS and PCS – unbalanced coefficient: −0.0013(0.0001); balanced coefficient: −0.0012(0.0003), next wave coefficient: 0.00158(0.0024)).

[39]This result is supported by Table A3, which shows that, of those respondents who report good MH at time t, 22% have missing information on employment status at time t+1, a very similar proportion to those who report poor MH at time t (23%).

TABLE 6

Oster bounds for split sample FE models with full controls

		$\delta = 1\,(\beta^*)$	
	$\delta = 0(\tilde{\beta})$	$R_{\max} = 1.3\tilde{R}$	$R_{\max} = 2.2\tilde{R}$
Panel 1 (GHQ12D t-1 coefficients)			
W/o degree	−0.0249	−0.0249	−0.0247
	(0.0039)	[0.0038]	[0.0038]
With degree	−0.0054	−0.0056	−0.0062
	(0.0033)	[0.0036]	[0.0036]
Above poverty line	−0.0108	−0.0110	−0.0116
	(0.0027)	[0.0030]	[0.0030]
Below poverty line	−0.0256	−0.0247	−0.0219
	(0.0060)	[0.0071]	[0.0078]
Panel 2 (GHQ36 t-1 coefficients)			
W/o degree	−0.0026	−0.0026	−0.0026
	(0.0003)	[0.0004]	[0.0004]
With degree	−0.0010	−0.0010	−0.0010
	(0.0003)	[0.0002]	[0.0003]
Above poverty line	−0.0014	−0.0014	−0.0014
	(0.0002)	[0.0003]	[0.0003]
Below poverty line	−0.0028	−0.0027	−0.0024
	(0.0005)	[0.0004]	[0.0005]
Panel 3 (MCS t-1 coefficients)			
W/o degree	−0.0016	−0.0018	**−0.0024**
	(0.0002)	[0.0002]	[0.0003]
With degree	−0.0009	−0.0011	**−0.0015**
	(0.0002)	[0.0002]	[0.0003]
Above poverty line	−0.0010	−0.0011	**−0.0015**
	(0.0002)	[0.0002]	[0.0002]
Below poverty line	−0.0020	−0.0022	**−0.0031**
	(0.0003)	[0.0003]	[0.0004]

Notes: Bootstrapped standard errors in square brackets (1,000 reps). Clustered standard errors in parentheses. Estimates in bold denote non-overlapping 95% confidence intervals between the bounds and original estimates.

Table 6 presents Oster bound estimates for select split sample FE models with full controls.[40] Similar to the pooled results, the bias-adjusted effects of MH (β^*) on employment have the same sign, and are close in magnitude to the estimated coefficients from the controlled FE regressions with overlapping 95% confidence intervals. However, the bias-adjusted effect of MH on employment when $R_{\max} = 2.2\tilde{R}$ is outside of the 95% confidence interval of $\tilde{\beta}$ in the MCS model for each of the subsamples.

Sensitivity analysis

We run a number of sensitivity checks. Our results are robust to different GHQ cut-offs for our binary GHQ12D. For the pooled FE model, we consider one lower cut-off at 2/3 and two higher cut-offs at 4/5 and 5/6. Compared with the coefficient in our benchmark model from Table 3 (−0.0160), the effect of GHQ12D is smaller for the cut-offs at

[40]We do not report bounds for other split sample estimates because no significant differences were found across these groups, but results are available upon request.

TABLE 7

Robustness checks (FE models)

	(1)	*(2)*	*(3)*
Panel 1			
GHQ12D 2/3 t-1	−0.0156***		
	(0.0023)		
GHQ12D 4/5 t-1		−0.0153***	
		(0.0028)	
GHQ12D 5/6 t-1			−0.0182***
			(0.0031)
PCS t-1	−0.0010***	−0.0010***	−0.0010***
	(0.0002)	(0.0002)	(0.0002)
NT	98435	98435	98435
Within R-squared	0.0236	0.0234	0.0235
Panel 2			
GHQ12D t-1	−0.0145***		
	(0.0026)		
GHQ36 t-1		−0.0017***	
		(0.0002)	
MCS t-1			−0.0009***
			(0.0001)
1-2 ADL t-1	−0.0141***	−0.0133***	−0.0146***
	(0.0042)	(0.0042)	(0.0042)
3-4 ADL t-1	−0.0516***	−0.0496***	−0.0522***
	(0.0086)	(0.0086)	(0.0086)
5+ ADL t-1	−0.0781***	−0.0755***	−0.0785***
	(0.0141)	(0.0141)	(0.0141)
NT	98381	98381	98381
Within R-squared	0.0240	0.0246	0.0244

Notes: Standard errors in parentheses (clustered at individual level). All models are LPM with FE and include a constant, wave dummies, and all controls included in Table 3 model (4).
Where $^{*}P < 0.10$, $^{**}P < 0.05$, $^{***}P < 0.01$.

2/3 and 4/5, and larger for cut-off at 5/6 (see Panel 1 in Table 7). We also consider an alternative measure of PH based on the activities of daily living (ADL) questions. We classify individuals into four categories: those with no ADL problems, those with 1−2 ADL problems, those with 3−4, and those with 5 or more. We re-run the pooled FE model using a categorical variable (with no ADL problems as the baseline) and find there is little change in the effect of MH on the probability of being employed; it is generally smaller for all three MH measures, but remains highly significant (see Panel 2 in Table 7).

The effect of MH may depend on the nature of employment, particularly on whether the individual is self-employed or not. The self-employed are likely to have a differential degree of autonomy and control at work, which can lead to different effects of MH on employment compared with those employed.[41] In our preferred model with individual FE, we find no substantial differences in the effect of MH on employment when we exclude respondents who are self-employed. We also explore sensitivity to geographical

[41]The self-employed are a very heterogeneous group consisting of, for example, highly paid consultants as well as low paid workers in the gig economy; thus it is difficult to generalize about their MH and work relationship.

location by excluding London and the results remain qualitatively the same. We consider additional geographical variation by running separate regressions for households located in urban/rural areas, and households in the north/south of England, and find no significant differences. Lastly, we split the sample by terciles of the Index of Multiple Deprivation (IMD)[42] in the neighbourhood where the household is located and local labour market tightness.[43] We find no significant differences in the effect of MH between these subgroups, across all three measures of MH.

VI. Discussion and conclusion

Given the wide availability of longitudinal data including measures of health status and labour market outcomes, FE models are an attractive method for estimating the effect of health on employment. They are straightforward to estimate and they control for the many time-invariant, but unobserved, characteristics likely to be correlated with both health and employment. They also provide a natural interpretation for the estimated relationship as the average effect of health on employment for those whose health changes.

Despite these advantages, a concern with FE is that, while removing the effects of time invariant heterogeneity, there could still be omitted time-varying characteristics that bias the estimates. In the MH and work context, likely omitted factors are people's changing family circumstances, work relationships and attitudes, as well as unobserved macroeconomic conditions. There is no firm indication from previous literature about which way the bias might go, particularly as much evidence comes from cross-sectional rather than longitudinal data. We have argued that cross-sectional selection provides little guidance about the remaining bias due to time-varying factors. Indeed, we find that while there are large reductions in the size of the MH coefficient when controls are added to a cross-sectional employment equation, adding controls to an FE equation barely changes the MH coefficient (even though the controls are highly significant). There could of course still be a substantial bias if the included controls represent only a small subset of all possible controls and add only a small amount of explained variance. We allow for this by assuming that adding the missing controls would more than double the explained longitudinal variance. Even under this fairly extreme assumption, we cannot reject that the bounds are the same as the FE estimates. The results indicate that selection into MH is almost entirely based on time-invariant characteristics and so we conclude that FE estimates of the effect of MH on employment are unbiased. There is certainly no evidence of upward bias in the size of the MH effect, as may be expected from the intuition that changing circumstances that favour work also favour MH. A caveat to our results is that while we try to minimize the possible influence of reverse causality by using lagged MH, there could still be some residual bias.

[42]2015 IMD data (for England only) obtained from the Ministry of Housing, Communities & Local Government.
[43]Labour market conditions may moderate the relationship between MH and employment status (see, e.g. Houssemand and Meyers, 2011). Labour market tightness is calculated at the LAD level with data obtained from NOMIS as: job vacancies/unemployment count. As job vacancy data are not available after 2012, we use average labour market tightness from 2009 to 2012 in each LAD to split the sample into households in LADs with average labour market tightness in the bottom quartile, top quartile, and middle two quartiles.

Our preferred specifications indicate that transitioning into poor MH (as measured by GHQ) leads to a reduction of 1.6 percentage points in the probability of employment (which makes up about 10% of the raw MH employment gap), and that a one standard deviation change in the continuous measures of MH causes a 1.0−1.3 percentage point change in the probability of employment. Comparisons of these effects with previous studies are not straightforward because of differences in data and methods, and differences in the way in which MH is measured. However, our estimated effects appear to be considerably smaller than most estimates that use IV methods. Across specifications studies report the effect of a one standard deviation change in MH as: 14−33 percentage points (US; Banerjee *et al.*, 2017) and 30 percentage points (Australia; Frijters *et al.*, 2014); while having a psychiatric disorder reduces employment by 13−14 percentage points (US; Ettner *et al.*, 1997). While these studies use different MH measures to us, the effects appear extremely large. However, as we have argued above, it is very difficult to make a strong case for the exogeneity of some of the instruments used (Ettner *et al.*, 1997; Banerjee *et al.*, 2017), so those estimates are still likely to be biased by unobserved effects that influence both MH and labour market outcomes. Further, in the case of Frijters *et al.* (2014), the estimate is of a very specific LATE. Studies that are more comparable to ours and use FE methods find much smaller effects in a similar ballpark to us. Estimates of the effects of MH episodes, summarized across specifications and types of transition, are: 1.6−8.0 percentage points depending on the severity of symptoms (US; Peng *et al.*, 2015); 0.0−8.2 (US; MCS measure; Mitra and Jones, 2017); and 0.0−2.9 percentage points (Australia; Bubonya *et al.*, 2019). Further, a UK study by Lagomarsino and Spiganti (2020) utilizes longitudinal data and IV, and they show that their large IV estimates (30 percentage points) are substantially reduced to 6 percentage points by the introduction of FE.[44] The MH effect does not differ across gender, but we find tentative evidence that MH has a bigger effect on employment for those in less advantaged positions, notably those without higher education and who start off in poverty. For instance, falling into poor MH (GHQ) reduces employment by 2.5 percentage points for people without degree, compared with just 0.5 percentage points for those with a degree. Thus there is a case for policy to prioritize these groups, although further evidence is required.

We have shown that simple FE methods can deliver estimates of the effect of MH on employment, which are both robust, and arguably more relevant for policy than the LATE delivered by many IV studies. Our alternative measures of MH gave very similar results, suggesting that either GHQ or MCS can be used as a basis for analysis. Given the widespread availability of longitudinal data, our findings should provide some reassurance to practitioners using FE methods to investigate the impacts of MH on work. We also hope they will be inspired to investigate the reliability of FE models in other contexts.

One limitation of our study is that the results show that FEs capture much of the effect of MH on employment, and this does not in itself help policy makers to design policies that will increase the employment probabilities of specific groups. However, we are able

[44]While this estimate is slightly larger than ours, this could be explained by the fact that Lagomarsino and Spiganti (2020) omit physical health as a control, and they include contemporaneous MH, whereas our measure is lagged by one period. As we explain in footnote 11, our estimated effect is larger if we do not lag MH.

to give a warning to policy makers to treat with caution the very large IV estimates that have been found in some studies. Reliance on these estimates could lead them to conclude that improving MH will directly lead to quite large increases in employment. This is not the case and could lead to resources being inappropriately directed. It is also important to remember that the FE impacts relate only to individuals whose MH changes. By their nature FE methods cannot identify the impacts of chronic, underlying MH conditions where no change is observed over time. Since the cross-sectional gap between those in good and poor MH (15 percentage points) is much larger than the effect of changing between MH states (1.6 percentage points), improving the MH of those with conditions amenable to treatment may only have a small direct effect on closing the MH employment gap. As well as chronic health problems, much of the raw gap is also due to differences in other factors, such as income, educational attainment, and childhood circumstances. Longer-term structural changes, which impact on all of these factors, will almost certainly be required to eliminate the gap completely.

Appendix

TABLE A1

Variable definitions

Variable	*Definition*	*Source*
Dependent variable		
Employed	=1 if self-employed or in paid employment (full- or part-time); 0 if individual is unemployed, retired, looking after family/home, or long-term sick/disabled. We exclude individuals out of the labour force (i.e. full-time students, on maternity leave, on a government training scheme or apprenticeship, untrained workers in family business, and those ‘doing something else’).	UKHLS
Mental health measures		
GHQ12D	Binary measure of Caseness based on the 0–12 scoring method of the 12-item General Health Questionnaire (=1 if score is 4 or higher, which identifies the possible presence of psychiatric morbidity).	UKHLS
GHQ36	Continuous measure based on the 0–36 Likert scale scoring method of the GHQ (0 represents the least distressed and 36 represents the most distressed).	UKHLS
MCS	Mental component summary, measured on a 0–100 continuous scale based on the SF-12 questionnaire where 0 denotes high functioning and 100 denotes low functioning.	UKHLS
Individual controls		
PCS	Physical Component Summary, measured on a 0–100 continuous scale based on the SF-12 questionnaire where 0 denotes high functioning and 100 denotes low functioning.	UKHLS
ADL	Individuals are classified into one of four groups based on reported activities of daily living: no ADL problems, 1–2 ADL problems, 3–4 problems, and those with 5 or more.	UKHLS
Age	Age of respondent in years.	UKHLS

(Continued)

TABLE A1

(Continued)

Variable	*Definition*	*Source*
Education	Highest level of education achieved at the time of the interview: no educational attainment (baseline), O-level or equivalent, A-level or equivalent, and having a degree or equivalent.	UKHLS
Married	=1 if individual is married, in a registered same-sex civil partnership or living as a couple; 0 otherwise.	UKHLS
Household controls		
No child in HH	=1 if no children 0–15 living in household; 0 otherwise	UKHLS
Child 0–4 in HH	=1 if children 0–4 living in household; 0 otherwise	UKHLS
Child 5–11 in HH	=1 if children 5–11 living in household; 0 otherwise	UKHLS
Child 12–15 in HH	=1 if children 12–15 living in household; 0 otherwise	UKHLS
Adults in HH	Number of adults living in household.	UKHLS
Other HH income	Derived by subtracting own gross monthly labour income from total gross household income in the month before interview (real, adjusted using RPI 2013=100).	UKHLS
Area controls		
Unemployment rate	Unemployment rate in the Local Authority District (LAD) where the household is located.	NOMIS*
GVA	Gross Value Added per head of the LAD where the household is located. Calculated using the balanced approach and the resident population of that region.	ONS

*Annual Population Survey.

TABLE A2

Components of variation

		Variance			
	Mean	*Total*	*Between*	*Within*	*Within as % of total*
Employed	0.832	0.140	0.130	0.029	20.9
GHQ12D	0.200	0.160	0.101	0.080	50.0
GHQ12D t-1	0.199	0.160	0.100	0.081	50.5
GHQ36	11.39	32.53	23.87	12.30	37.8
GHQ36 t-1	11.34	31.78	23.37	12.08	38.0
MCS	51.89	100.03	76.679	34.34	34.3
MCS t-1	51.56	97.90	74.56	34.62	35.4
PCS	47.98	89.30	73.94	23.82	26.7
PCS t-1	47.79	86.27	70.37	24.14	28.0
ADL problems					
None	0.876	0.109	0.083	0.037	33.9
1–2	0.085	0.078	0.050	0.039	49.6
3–4	0.026	0.025	0.016	0.012	48.0
5 or more	0.012	0.012	0.009	0.004	36.5
Age 21–25	0.067	0.062	0.083	0.014	22.9
Age 26–30	0.108	0.096	0.077	0.035	36.7
Age 31–35	0.136	0.118	0.082	0.047	40.2
Age 36–40	0.156	0.132	0.084	0.057	43.6
Age 41–45	0.177	0.146	0.087	0.067	45.8
Age 46–50	0.183	0.150	0.085	0.071	47.2
Age 51–55	0.172	0.143	0.132	0.036	25.3
Married	0.726	0.199	0.192	0.025	12.5

(Continued)

TABLE A2

(Continued)

	Mean	Variance Total	Between	Within	Within as % of total
Education level					
No education	0.040	0.039	0.048	0.001	3.4
O-level	0.289	0.205	0.204	0.004	2.0
A-level	0.211	0.166	0.165	0.004	2.5
Degree	0.460	0.248	0.244	0.003	1.4
No child in HH	0.504	0.250	0.223	0.037	15.0
Child 0–4 in HH	0.200	0.160	0.126	0.048	30.0
Child 5–11 in HH	0.297	0.209	0.164	0.052	24.9
Child 12–15 in HH	0.193	0.156	0.111	0.056	35.9
Adults in HH	2.327	0.988	1.035	0.164	16.6
ln (other HH income)	7.208	3.724	2.992	0.979	26.3
Unemployment rate	6.943	8.415	6.290	3.109	36.9
GVA	2.474	2.584	2.944	0.147	5.7

Notes: GHQ12D is the GHQ caseness dummy; GVA is gross value added (per capita); MCS is the SF12 mental component summary; PCS is the SF12 physical component summary.

TABLE A3

Mental health at time t and employment status information at time t+1

	Employment status (t+1) Not employed	Employed	Missing employment status
Good MH (t)	13,267	77,925	25,441
(%)	(11)	(67)	(22)
Poor MH (t)	7,000	16,038	6,824
(%)	(23)	(54)	(23)

Notes: Calculated from the UKHLS waves 1–8. Missing employment status can arise because the respondent is present but does not answer this question, or because there is no record for that respondent.

TABLE A4

LPM: FE models by education

	(1) Educational attainment Without degree	(2) With degree	Difference
Panel 1			
GHQ12D t-1	−0.0249***	−0.0054	(1) ≠ (2)***
	(0.0039)	(0.0033)	
PCS t-1	−0.0011***	−0.0009***	
	(0.0002)	(0.0003)	
Within R-squared	0.0307	0.0162	
Panel 2			
GHQ36 t-1	−0.0026***	−0.0010***	(1) ≠ (2)***
	(0.0003)	(0.0003)	
PCS t-1	−0.0011***	−0.0009***	
	(0.0002)	(0.0003)	
Within R-squared	0.0316	0.0166	
Panel 3			
MCS t-1	−0.0016***	−0.0009***	(1) ≠ (2)**
	(0.0002)	(0.0002)	
PCS t-1	−0.0017***	−0.0012***	
	(0.0002)	(0.0003)	
Within R-squared	0.0318	0.0171	

Notes: Standard errors in parentheses (clustered at individual level). All models are LPM with FE and include a constant, wave dummies, and all controls included in Table 3 model (4). There are 53,139 observations (NT) without a degree and 45,296 with a degree.
Where $^{*}P < 0.10$, $^{**}P < 0.05$, $^{***}P < 0.01$.

TABLE A5

LPM: FE models by PCS tertiles

	(1)	(2)	(3)	
	Physical component summary (PCS) score			
	Bottom tercile	Middle tercile	Top tercile	Difference
Panel 1				
GHQ12D t-1	−0.0110***	−0.0153**	−0.0158***	
	(0.0043)	(0.0061)	(0.0050)	
PCS t-1	−0.0002	0.0005	−0.0017	(2) ≠ (3)***
	(0.0004)	(0.0004)	(0.0003)	(1) ≠ (3)***
Within R-squared	0.0218	0.0319	0.0223	
Panel 2				
GHQ36 t-1	−0.0012***	−0.0018***	−0.0016***	
	(0.0004)	(0.0005)	(0.0004)	
PCS t-1	−0.0003	0.0005	−0.0017***	(2) ≠ (3)***
	(0.0004)	(0.0004)	(0.0003)	(1) ≠ (3)***
Within R-squared	0.0221	0.0325	0.0227	
Panel 3				
MCS t-1	−0.0008***	−0.0012***	−0.0017***	(1) ≠ (3)**
	(0.0003)	(0.0003)	(0.0003)	
PCS t-1	−0.0006	0.0001	−0.0021***	(2) ≠ (3)***
	(0.0004)	(0.0004)	(0.0003)	(1) ≠ (3)***
Within R-squared	0.0221	0.0326	0.0248	

Notes: Standard errors in parentheses (clustered at individual level). All models are LPM with FE and include a constant, wave dummies, and all controls included in Table 3 model (4). There are 33,302 observations (NT) in the bottom PCS tercile, 32,582 in the middle PCS tercile, and 32,551 in the top tercile.
Where $^{*}P < 0.10$, $^{**}P < 0.05$, $^{***}P < 0.01$.

TABLE A6

LPM: FE models by relative HH poverty

	(1)	(2)	
	Poverty line		
	Above	Below	Difference
Panel 1			
GHQ12D t-1	−0.0108***	−0.0256***	(1) ≠ (2)**
	(0.0027)	(0.0060)	
PCS t-1	−0.0007***	−0.0019***	(1) ≠ (2)***
	(0.0002)	(0.0003)	
Within R-squared	0.0140	0.0650	
Panel 2			
GHQ36 t-1	−0.0014***	−0.0028***	(1) ≠ (2)***
	(0.0002)	(0.0005)	
PCS t-1	−0.0007***	−0.0019***	(1) ≠ (2)***
	(0.0002)	(0.0003)	
Within R-squared	0.0145	0.0661	
Panel 3			
MCS t-1	−0.0010***	−0.0020***	(1) ≠ (2)***
	(0.0002)	(0.0003)	
PCS t-1	−0.0011***	−0.0025***	(1) ≠ (2)***
	(0.0002)	(0.0004)	
Within R-squared	0.0147	0.0670	

Notes: Standard errors in parentheses (clustered at individual level). All models are LPM with FE and include a constant, wave dummies, and all controls included in Table 3 model (4). There are 73,652 observations (NT) above the poverty line and 24,782 below. The poverty line is 60% of the median net equivalised HH income (before housing costs) in the United Kingdom adjusted for inflation using the Consumer Price Index (data available from IFS). Households are classified based on whether they are above or below this relative poverty line in the first wave that they appear in the analysis sample.
Where $^{*}P < 0.10$, $^{**}P < 0.05$, $^{***}P < 0.01$.

Final Manuscript Received: February 2021

References

Agüero, J. M. and Marks, M. S. (2008). 'Motherhood and female labor force participation: evidence from infertility shocks', *American Economic Review*, Vol. 98, pp. 500–504.

Alexandre, P. K. and French, M. T. (2001). 'Labor supply of poor residents in metropolitan Miami, Florida: the role of depression and the co-morbid effects of substance use', *Journal of Mental Health Policy and Economics*, Vol. 4, pp. 161–173.

Allison, P. D. (2019). 'Asymmetric fixed-effects models for panel data', *Sociological Research for a Dynamic World*, Vol. 5, pp. 1–12.

Altonji, J., Elder, T. and Taber, C. (2005). 'Selection on observed and unobserved variables: assessing the effectiveness of catholic schools', *Journal of Political Economy*, Vol. 113, pp. 151–184.

Altonji, J., Conley, T., Elder, T. and Taber, C. (2011). *Methods for Using Selection on Observed Variables to Address Selection on Unobserved Variables*, Technical Report, Yale University, Mimeo.

Andersen, M. (2015). 'Heterogeneity and the effect of mental health parity mandates on the labor market', *Journal of Health Economics*, Vol. 43, pp. 74–84.

Andreeva, E., Magnusson Hanson, L., Westerlund, H., Theorell, T. and Brenner, M. (2015). 'Depressive symptoms as a cause and effect of job loss in men and women: evidence in the context of organisational downsizing from the Swedish longitudinal occupational survey of health', *BMC Public Health*, Vol. 15, pp. 1–11.

Anjara, S., Bonetto, C., Van Bartel, T. and Brayne, C. (2020). 'Using the GHQ-12 to screen for mental health problems among primary care patients: psychometrics and practical considerations', *International Journal of Mental Health Systems*, Vol. 15, p. 62. https://doi.org/10.1186/s13033-020-00397-0

Auld, C. (2002). 'Disentangling the effects of morbidity and life expectancy on labor market outcomes', *Health Economics*, Vol. 11, pp. 471–483.

Banerjee, S., Chatterji, P. and Lahiri, K. (2017). 'Effects of psychiatric disorders on labor market outcomes: a latent variable approach using multiple clinical indicators', *Health Economics*, Vol. 26, pp. 184–205.

Bartel, A. and Taubman, P. (1979). 'Health and labor market success: the role of various diseases', *Review of Economics and Statistics*, Vol. 61, pp. 1–8.

Bartel, A. and Taubman, P. (1986). 'Some economic and demographic consequences of mental illness', *Journal of Labor Economics*, Vol. 4, pp. 243–256.

Becker, G. (1964). *Human Capitol* 2nd ed., Columbia University Press, New York.

Becketti, S., Gould, W., Lillard, L. and Welch, F. (1988). 'The panel study of income dynamics after fourteen years: an evaluation', *Journal of Labor Economics*, Vol. 6, pp. 427–492.

Black, B., de Carvalho, A. G., Khanna, V., Kim, W. and Yurtoglu, B. (2014). 'Methods of multi-country studies of corporate governance: evidence from the BRIKT countries', *Journal of Econometrics*, Vol. 183, pp. 230–240.

Black, C. (2009). 'Working for a healthier tomorrow. Review of the health of Britain's working age population', *The Stationery Office*, Vol. 66, pp. 1–2.

Black, C. and Frost, D. (2011). *Health at Work-An Independent Review of Sickness Absence*, The Stationery Office, London.

Blundell, R., Britton, J., Costa Dias, M. & French, E. (2017). *The Impact of Health on Labor Supply Near Retirement*, Technical Report No. W17/18, Institute for Fiscal Studies Working Paper.

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

Bound, J., Schoenbaum, M., Stinebrickner, T. R. and Waidmann, T. (1999). 'The dynamic effects of health on the labor force transitions of older workers', *Labour Economics*, Vol. 6, pp. 179–202.

Brown, S., Harris, M. N., Srivastava, P. and Taylor, K. B. (2018). *Mental Health and Reporting Bias: Analysis of the GHQ-12*, IZA Discussion Paper No. 11771. https://ssrn.com/abstract=3249885

Bubonya, M., Cobb-Clark, D. A. and Ribar, D. (2019). 'The reciprocal relationship between depressive symptoms and employment status', *Economics and Human Biology*, Vol. 35, pp. 96–106.

Butterworth, P., Leach, L., Pirkis, J. and Kelaher, M. (2012). 'Poor mental health influences risk and duration of unemployment: a prospective study', *Social Psychiatry Psychiatric Epidemiology*, Vol. 47, pp. 1013–1021.

Cai, L. (2021). 'The effects of health on the extensive and intensive margins of labor supply', *Journal of the Royal Statistical Society, Series A*, Vol. 184, pp. 87–117.

Carmichael, F., Hulme, C., Sheppard, S. and Connell, G. (2008). 'Work-life imbalance: informal care and paid employment in the UK', *Feminist Economics*, Vol. 14, pp. 3–35.

Cattan, S., Kamhofer, D. A., Karlsson, M. & Nilsson, T. (2017). *The Short and Long-Term Effects of Student Absence: Evidence from Sweden*, Technical Report W17/21, Institute for Fiscal Studies Working Paper.

Chatterji, P., Alegria, M. and Takeuchi, D. (2007). 'Psychiatric disorders and labor market outcomes: evidence from the National Latino and Asian American Study', *Health Economics*, Vol. 16, pp. 1069–1090.

Chatterji, P., Alegria, M. & Takeuchi, D. (2008). *Psychiatric Disorders and Labor Market Outcomes: Evidence from the National Comorbidity Survey-Replication*, National Bureau of Economic Research (NBER) Technical Report Working Paper 14404.

Chatterji, P., Alegria, M. and Takeuchi, D. (2011). 'Psychiatric disorders and labor market outcomes: evidence from the National Comorbidity Survey-Replication', *Journal of Health Economics*, Vol. 30, pp. 858–868.

Cornaglia, F., Crivellaro, E. and McNally, S. (2015). 'Mental health and education decisions', *Labour Economics*, Vol. 3, pp. 1–12.

Cseh, A. (2008). 'The effects of depressive symptoms on earnings', *Southern Economic Journal*, Vol. 75, pp. 383–409.

Currie, J. and Madrian, B. (1999). 'Health, health insurance and the labor market', in Ashenfelter O. and Card D. (eds), *Handbook of Labor Economics* Vol. 3, 1st ed., Elsevier, Amsterdam, The Netherlands, chapter 50, pp. 3309–3416.

DeSimone, J. (2002). 'Illegal drug use and employment', *Journal of Labor Economics*, Vol. 20, pp. 952–977.

Dockery, A. M., Li, J. and Kendall, G. (2016). 'Sole-parent work schedules and adolescent wellbeing: Evidence from Australia', *Social Science and Medicine*, Vol. 168, pp. 167–174.

DWP (2017). *Improving Lives: The Future of Work, Health and Disability*, Department for Work and Pensions and Department of Health Technical Report Policy Paper.

Ettner, S. L., Frank, R. G. and Kessler, R. C. (1997). 'The impact of psychiatric disorders on labor market outcomes', *Industrial and Labor Relations Review*, Vol. 51, pp. 64–81.

Fletcher, J. (2014). 'The effects of childhood adhd on adult labor market outcomes', *Health Economics*, Vol. 23, pp. 159–181.

Francesconi, M. and Van der Klaauw, W. (2007). 'The socioeconomic consequences of ''In-Work" benefit reform for British lone mothers', *Journal of Human Resources*, Vol. XLII, pp. 1–31.

Frank, R. and Gertler, P. (1991). 'An assessment of measurement error bias for estimating the effect of mental distress on income', *Journal of Human Resources*, Vol. 26, pp. 154–164.

Frijters, P., Johnston, D. W. and Shields, M. A. (2014). 'The effect of mental health on employment: evidence from Australian panel data', *Health Economics*, Vol. 23, pp. 1058–1071.

Garcia-Gomez, P., Jones, A. M. and Rice, N. (2010). 'Health effects on labour market exits and entries', *Labour Economics*, Vol. 17, pp. 62–76.

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

Ghatak, A. (2010). *Health, Labour Supply and Wages: A Critical Review of the Literature*, Institute for Social and Economic Change, Technical Report Working Paper 244.

Gill, S. C., Butterworth, P., Rodgers, B. and Mackinnon, A. (2007). 'Validity of the mental health component scale of the 12-Item Short-Form Health Survey (MCS-12) as measure of common mental disorders in the general population', *Psychiatry Research*, Vol. 152, pp. 63–71.

Goldberg, D. P., Gater, R., Sartorius, N., Us̀tun, T. B., Piccinelli, M., Gureje, O. and Rutter, C. (1997). 'The validity of two versions of the GHQ in the WHO study of mental illness in general health care', *Psychological Medicine*, Vol. 27, pp. 191–197.

Goldberg, D. P., Oldehinkel, T. and Ormel, J. (1998). 'Why GHQ threshold varies from one place to another', *Psychological Medicine*, Vol. 28, pp. 915–921.

Gregg, P., Grout, P. A., Ratcliffe, A. and Windmeijer, F. (2011). 'How important is pro-social behaviour in the delivery of public services?', *Journal of Public Economics*, Vol. 95, pp. 758–766.

Greve, J. and Nielsen, L. H. (2013). 'Useful beautiful minds-An analysis of the relationship between schizophrenia and employment', *Journal of Health Economics*, Vol. 32, pp. 1066–1076.

Grossman, M. (1972). 'On the concept of health capital and the demand for health', *Journal of Political Economy*, Vol. 80, pp. 223–255.

Hakulinen, C., Elovainio, M., Arffman, M., Lumme, S., Pirkola, S., Keskimäki, I., Manderbacka, K. and Böckerman, P. (2019). 'Mental disorders and long-term labour market outcomes: nationwide cohort study of 2,055,720 individuals', *Acta Psychiatrica Scandinavica*, Vol. 140, pp. 371–381.

Hakulinen, C., Elovainio, M., Arffman, M., Lumme, S., Suokas, K., Pirkola, S., Keskimäki, I., Manderbacka, K. and Böckerman, P. (2020). 'Employment status and personal income before and after onset of a severe mental disorder: a case-control study', *Psychiatric Services*, Vol. 71, pp. 250–255.

Hamilton, V. H., Merrigan, P. and Dufresne, E. (1997). 'Down and out: estimating the relationship between mental health and unemployment', *Econometrics and Health Economics*, Vol. 6, pp. 397–406.

Hausman, J. (1978). 'Specification tests in econometrics', *Econometrica*, Vol. 46, pp. 1251–1271.

Hausman, J. and Wise, D. (1979). 'Attrition bias in experimental and panel data: the Gary income maintenance experiment', *Econometrica*, Vol. 47, pp. 455–474.

Hener, T., Rainer, H. and Siedler, T. (2016). 'Political socialization in flux? Linking family non-intactness during childhood to adult civic engagement', *Journal of the Royal Statistical Society Series A*, Vol. 179, pp. 633–656.

Hofmann, S. and Mühlenweg, A. (2019). 'Learning intensity effects in students' mental and physical health-Evidence from a large scale natural experiment in Germany', *Economics of Education Review*, Vol. 67, pp. 216–234.

Houssemand, C. and Meyers, R. (2011). 'Unemployment and mental health in a favourable labour market', *International Journal of Psychology*, Vol. 46, pp. 377–385.

Jimenez-Martin, S. and Peracchi, F. (2002). 'Sample attrition and labor force dynamics: evidence from the Spanish labor force survey', *Spanish Economic Review*, Vol. 4, pp. 79–102.

Jolivet, G. and Postel-Vinay, F. (2020). *A Structural Analysis of Mental Health and Labor Market Trajectories*, IZA Discussion Paper No. 13518.

Jones, A. M., Koolman, X. and Rice, N. (2006). 'Health-related non-response in the BHPS and ECHP: using inverse probability weighted estimators in non-linear models', *Journal of the Royal Statistical Society, Series A*, Vol. 169, pp. 543–569.

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

Jones, A. M., Rice, N. and Zantomio, F. (2020). 'Acute health shocks and labour market outcomes: evidence from the post crash era', *Economics and Human Biology*, Vol. 36, 100811. https://doi.org/10.1016/j.ehb.2019.100811

Kaspersen, S., Pape, K., Vie, G., Ose, S. O., Krokstad, S., Gunnell, D. and Bjørngaard, J. H. (2016). 'Health and unemployment: 14 years of follow-up on job loss in the Norwegian HUNT study', *The European Journal of Public Health*, Vol. 26, pp. 312–317.

Kessler, R. C., Berglund, P. A., Demler, O., Jin, R., Merikangas, K. R. and Walterset, E. E. (2005). 'Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication (NCS-R)', *Archives of General Psychiatry*, Vol. 62, pp. 593–602.

Kreider, B. and Pepper, J. V. (2007). 'Disability and employment: re-evaluating the evidence in light of reporting errors', *Journal of the American Statistical Association*, Vol. 102, pp. 432–441.

Lagomarsino, E. and Spiganti, A. (2020). 'No gain in pain: psychological well-being, participation, and wages in the BHPS', *European Journal of Health Economics*, Vol. 21, pp. 1375–1389.

Lillard, L. and Panis, C. (1998). 'Panel attrition from the panel study of income dynamics: household income, marital status, and mortality', *Journal of Human Resources*, Vol. 33, pp. 437–457.

Lindeboom, M. and Kerkhofs, M. (2009). 'Health and work of the elderly: subjective health measures, reporting errors and endogeneity in the relationship between health and work', *Journal of Applied Econometrics*, Vol. 24, pp. 1024–1046.

Marcotte, D. E., Wilcox-Gök, V. and Redmon, D. P. (2000). *The Labour Market Effects of Mental Illness: The Case of Effective Disorders*, JAI Press, Greenwich, CT, pp. 181–210.
Mavridis, D. (2015). 'The unhappily unemployed return to work faster', *Journal of Labor Economics*, Vol. 4, pp. 1–22.
McInnes, R. (2012). *ESA and Incapacity Benefit Statistics*, House of Commons Library, Technical Report Standard Note SN01420.
Mitra, S. and Jones, K. (2017). 'The impact of recent mental health changes on employment: new evidence from longitudinal data', *Applied Economics*, Vol. 49, pp. 96–109.
Mullahy, J. and Sindelar, J. (1996). 'Employment, unemployment, and problem drinking', *Journal of Health Economics*, Vol. 15, pp. 409–434.
Mundlak, Y. (1978). 'On the pooling of time series and cross section data', *Econometrica*, Vol. 46, pp. 69–85.
Munford, L., Rice, N. and Roberts, J. (2016). *Econometric Modelling of the Employment Gap for the NHS Outcomes Framework*, Technical Report, Report for the Department of Health. http://cfe.org.uk/app/uploads/2016/09/Econometric-Modelling-of-the-Employment-Gap_July-16.pdf
Nicoletti, C. and Peracchi, F. (2005). 'A cross-country comparison of survey nonparticipation in the ECHP', *Journal of the Royal Statistical Society, Series A*, Vol. 168, pp. 763–781.
Oakley, M. (2016). *Closing the Gap*, Social Market Foundation, Technical Report.
OECD/EU (2018). *Health at a Glance: Europe 2018: State of Health in the EU Cycle*, OECD Publishing, Technical Report. https://doi.org/10.1787/health_glance_eur-2018-en
Ojeda, V. D., Frank, R. G., McGuire, T. G. and Gilmer, T. P. (2010). 'Mental illness nativity gender and labor supply', *Health Economics*, Vol. 19, pp. 396–421.
Olesen, S. C., Butterworth, P., Leach, L. S., Kelaher, M. and Pirkis, J. (2013). 'Mental health affects future employment as job loss affects mental health: findings from a longitudinal population study', *BMC Psychiatry*, Vol. 13. https://doi.org/10.1186/1471-244X-13-144
Oster, E. (2013a). *PSACALC: Stata Module to Calculate Treatment Effects and Relative Degree of Selection Under Proportional Selection of Observables and Unobservables. Revised 18 December 2016 ed*, Boston Collage Department of Economics, Technical Report S457677.
Oster, E. (2013b). *Unobservable Selection and Coefficient Stability: Theory and Validation*, NBER Working Paper, Technical Report No. 19054.
Oster, E. (2019). 'Unobservable selection and coefficient stability: theory and evidence', *Journal of Business & Economic Statistics*, Vol. 37, pp. 187–204.
Otterbach, S., Wooden, M. and Fok, Y. K. (2016). *Working-Time Mismatch and Mental Health*, IZA, Technical Report Discussion Paper No. 9818.
Peng, L., Meyerhoefer, C. D. and Zuvekas, S. H. (2015). 'The short-term effect of depressive symptoms on labour market outcomes', *Health Economics*, Vol. 25, pp. 1223–1238.
Roberts, J., Hodgson, R. and Dolan, P. (2011). 'It's driving her mad: gender differences in the effects of commuting on psychological health', *Journal of Health Economics*, Vol. 30, pp. 1064–1076.
Sartorius, N. and Ustün, T. B. (1995). *Mental Illness in General Health Care: An International Study*, John Wiley & Sons, New York.
Savage, S. and Bailey, S. (2004). 'The impact of caring on caregivers' mental health: a review of the literature', *Australian Health Review*, Vol. 27, pp. 103–109.
Schmitz, N., Kruse, J. and Tress, W. (1999). 'Psychometric properties of the general health questionnaire (GHQ-12) in a German primary care sample', *Acta Psychiatrica Scandinavica*, Vol. 100, pp. 462–468.
Steele, F., French, R. and Bartley, M. (2013). 'Adjusting the selection bias in longitudinal analyses using simultaneous equations modeling: the relationship between employment transitions and mental health', *Epidemiology*, Vol. 24, pp. 703–711.
Terza, J. V. (2002). 'Alcohol abuse and employment: a second look', *Journal of Applied Econometrics*, Vol. 17, pp. 393–404.
UKHLS (2019). *Understanding Society: Waves 1-8 2009-2017 and Harmonised BHPS: Waves 1-18 1991-2009. [Data Collection]*, Technical Report, University of Essex. Institute for Social and Economic Research NatCen Social Research, Kantar Public. UK Data Service. https://doi.org/10.5255/UKDA-SN-6614-13
Verbeek, M. (2012). *A Guide to Modern Econometrics*, John Wiley & Sons, Chichester, England.

Verbeek, M. and Nijman, T. (1992). 'Testing for selectivity bias in panel data models', *International Economic Review*, Vol. 33, pp. 681–703.

Waddell, G. and Burton, A. K. (2006). *Is Work Good for Your Health and Well-Being?* The Stationery Office, Technical Report.

Ware, J. E., Kosinski, M., Turner-Bowker, D. M. and Gandek, B. (2002). *How to Score Version 2 of the SF-12 Health Survey (With a Supplement Documenting Version 1)*, Technical Report, Quality Metric Incorporated, Lincoln RI.

Webber, D. and Bjelleand, M. (2015). 'The impact of work-limiting disability on labor force participation', *Health Economics*, Vol. 24, pp. 333–352.

WHO (2000). *Mental Health and Work: Impact Issues and Good Practices*, World Health Organization, Technical Report.

WHO (2008). *The Global Burden of Disease: 2004 Update*, World Health Organization, Technical Report.

WHO (2013). *Investing in Mental Health: Evidence for Action*, World Health Organization, Technical Report.

Zabel, J. (1998). 'An analysis of attrition in the panel study of income dynamics and the survey of income and program participation with an application to a model of labor market behaviour', *Journal of Human Resources*, Vol. 33, pp. 479–506.

Ziliac, J. and Kniesner, T. (1998). 'The importance of sample selection in life cycle labor supply estimation', *Journal of Human Resources*, Vol. 33, pp. 507–530.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Supplementary Appendix