UNIVERSITY of York

# HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

WP 19/14

# The Effect of Mental Health on Employment: Accounting for Selection Bias

Mark Bryan; Jennifer Roberts and Cristina Sechel

July 2019

# The Effect of Mental Health on Employment: Accounting for Selection Bias

**Mark Bryan, Jennifer Roberts, Cristina Sechel***

Department of Economics, University of Sheffield, UK

***corresponding author: c.sechel@sheffield.ac.uk**

**WORK IN PROGRESS**

**DO NOT CITE OR CIRCULATE WITHOUTAUTHOR'S PERMISSION**

July 2019

## Abstract

This paper estimates the influence of mental health on the probability of being in employment for prime age workers in England and Wales. We use longitudinal data and fixed effects models, and employ techniques generalised by Oster (2013, 2017) to estimate an unbiased effect of changes in mental health in the presence of unobserved confounders. Our results suggest that selection into mental health is almost entirely based on time-invariant characteristics, and hence fixed effects estimates are unbiased in this context. Our preferred specifications indicate that transitioning into poor mental health leads to a reduction of 1.4 percentage points in the probability of employment. The relatively small effect is comparable to estimates from studies around the world that use similar methods. However, it is substantially smaller than the typical instrumental variable estimates, which dominate the literature, and often provide very specific estimates of a local average treatment effect based on an arbitrary exogenous shock that is unlikely to be a policy target. These findings should provide some reassurance to practitioners using fixed effects methods to investigate the impacts of health on work. They should also be useful to policy makers as the average effect of mental health on employment for those whose mental health changes is a highly relevant policy parameter.

**JEL codes: I12, J14, J24.**

**Key words:** mental health, employment, fixed effects, UKHLS.

**Acknowledgements**

We are grateful to the Health Foundation for funding this research under their Social and Economic Value of Health programme. The UKHLS data (University of Essex, 2018) was accessed via the UK Data Service. Some of the data used are provided to the authors under Special Licence arrangements. We are also very grateful to Nigel Rice and Maarten Lindeboom for very useful comments on an earlier version.

# 1 Introduction

An individual's relationship to the labour market is a key determinant of their financial security and a source of broader wellbeing (Black, 2009). In the UK people with health problems have a much lower employment rate than the rest of the population (Oakley, 2016; WPC, 2017). Every year 300,000 people stop work and become reliant on health-related benefits, costing the government £13bn and employers another £9bn (Black and Frost, 2011). Adverse mental health (MH) seems to be particularly pernicious in its labour market effects. The employment rate for people with MH problems is only 35% (Oakley, 2016), and the disability employment gap between those with and without a MH problem is around 40 percentage points (Munford et al., 2016). Common MH problems, like anxiety and depression, account for over 40% of Employment Support Allowance 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 Europe 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 et al. (2014) summarise 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 on the effect of MH on work relies on instrumental variable (IV) estimation and/or longitudinal data with fixed effects (FE) in an attempt to deal with endogenous selection.[1] 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, these studies often provide very specific estimates of a local average treatment effect (LATE) which in most cases is derived from an arbitrary exogenous shock (for example the death of a close friend) that would not be a policy target. The inclusion of FE eliminates

[1] Technically FE is also an IV estimator, with deviations from the means used as the instruments (see Verbeek, 2012: pp 387-8).

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. This is now possible by exploiting the methods developed by Altonji et al. (2011; 2005), who use selection on observable characteristics to provide information on selection along unobservable factors; and also Oster (2013, 2017) who extends and generalises this method to enable the estimation of an unbiased treatment effect in the presence of unobserved confounders.

We make the following contributions to the literature. Firstly, ours is the first study to employ the Oster method in the context of the health and work relationship, and in particular, we are the first to use the method with individual longitudinal data incorporating FE.[2] We employ the methods developed by Oster (2013, 2017) to estimate the bias that arises from 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, and estimate upper and lower bounds for the impact of MH on employment. Secondly, 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 in England and Wales. The vast majority of existing evidence on the relationship between health and work considers physical health or uses a general measure of overall self-assessed of health (see Ghatak (2010) for a review). Instead, we use two measures of MH derived from validated psychometric instruments; the General Health Questionnaire and the Mental Component Summary (MCS) score from the Short Form-12 health survey. We argue that these measures are good proxies for the true MH stock; they are designed to provide information on all aspects of MH, and they are less likely to suffer from the reporting biases that are present in simple overall evaluative measures (Bound, 1991; Bound et al., 1999). Further, almost all of the existing evidence on the effect of MH on employment comes from countries from outside of the UK, which have very different health care systems to the universal coverage of the National Health Service (and, more generally, different welfare systems); indeed the vast majority of evidence comes from cross-sectional studies in the US. Our estimates for England and Wales contribute to a very small pool of UK evidence, and will be valuable to decision makers given the current policy priority to increase the number of disabled people in work by one million over ten years (DWP, 2017)[3]. Unlike cross sectional studies, where it is very difficult to control for individual unobserved factors that confound the relationship between labour supply and health, our study employs longitudinal data and explores any remain biases that are not removed by the inclusion of FE. Also, much of the evidence on the impact of

[2] In the only panel data applications of the Oster method of which we are aware, Hener et al. (2016) and Cattan et al. (2017) use individual level data with sibling FE, and Black et al. (2014) use firm-level data.
[3] This goal replaces a previous commitment to halve the disability employment gap (DWP, 2016).

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. Finally, 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, varies 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;[4] for example, the Improving Access to Psychological Therapies initiative that has been rolled out in England and Wales from 2008 to help people who suffer from anxiety and depression. As well as providing evidence, we also hope that our application of the Oster method will be useful resource for practitioners who may wish to use this 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 FE 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 from other countries, 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 some limited evidence that MH has larger effects on employment for those without degrees and those who are in poverty. The paper is structured as follows. In Section 2 we explore the background to the FE models that are common in this literature, explaining the econometric estimation problems they are designed to solve and reviewing some of the key evidence. Section 3 describes our estimation method and the Oster (2017) approach. The data and variables are described in Section 4 and the results and sub-group analyses are presented in Section 5. Section 6 includes the discussion and conclusion.

# 2 Background

It is well known in the literature that MH and work are related and that the relationship between them is complex (see for example Currie and Madrian, 1999; Frijters et al., 2014; Steele et al., 2013). However, there is very little quantitative evidence available for the UK on the effect of MH disorders on work, and, across the breadth of evidence from other countries, there is no consensus around the size of the effects. At the same time, policy makers who wish to reduce the MH disability employment gap need reliable quantitative estimates of the effects of health status on the probability of being in employment in order to estimate the real costs to the

[4] This is an important area for health policy; for example Public Health England have a long standing interest in the relationship between work and health, and 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/health-matters-health-and-work

economy and to formulate appropriate policy tools to increase the employment rate of people with MH problems.

The effect of an adverse health event on labour supply is theoretically ambiguous. The Grossman (1972) health investment model shows that deterioration in health can reduce time available for work because of increased time spent being ill, an increased preference for leisure time and/or increased time needed to maintain health; further poorer health can also directly reduce productivity. However, worsening health can also increase labour supply, especially in privatised health care markets like the US. 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. However, even with this type of health care provision, poor health can still increase household costs.

The vast majority of 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. Common instruments used in the IV literature on health and work include: parental history of MH (Banerjee et al., 2017; Ettner et al., 1997; Marcotte et al., 2000); childhood psychiatric disorders (Banerjee et al., 2017; Chatterji et al., 2007; Ettner et al., 1997); participation in religious services and religious beliefs (Alexandre and French, 2001; Chatterji et al., 2007); and perceived social support (Alexandre and French, 2001; Hamilton et al., 1997; Ojeda et al., 2010).[5] The general consensus from these studies is that MH has a negative influence on the probability of being in employment. However, as Chatterji et al. (2011) point out, the chosen instruments are often "hard to justify based on economic theory" (p. 859). Indeed, in their own study, Chatterji et al. (2007) admit that it is a 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 methods proposed by Altonji et al. (2005), which use selection on observable traits to provide information on selection along unobservable factors, 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.

A further problem, which has received little or no attention in the health and work literature, is that the vast majority of IV studies provide a very specific estimate of the local average treatment effect (LATE)

[5] 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 for example DeSimone, 2002; Mullahy and Sindelar, 1996).

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 by providing a more relevant policy parameter. 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 utilise longitudinal data. We know of only one such study for the UK. Garcia-Gomez et al. (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, for working individuals. In a discrete-time hazard framework, they find that worsening MH increases the exit hazard for workers, with the magnitude being greater for men than for women. However, they also find that worsening MH in non-workers increases the hazard of becoming employed for both men and women. This is a difficult finding to explain, 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.[6] Their framework includes both initial and lagged health, but does not allow for FE. Given the shortage of UK evidence it is useful to look to countries, such as Canada, that have similar universal healthcare coverage. One study, by Hamilton et al. (1997), considers data from a relatively small sample of less than 800 Montreal residents. They use a two-equation model for MH (measured by the Psychiatric Symptom Index) and employment, and find that better MH increases employability and vice versa. They also find that unobserved factors correlated with higher employability are also correlated with MH.

For the US, Mitra and Jones (2017) use data from 2 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 i.e. those who are not employed initially and have a health problem do not see an increased probability of employment upon recovery. Also in the US, Peng et al. (2015) use data from 5 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 reduces the likelihood of employment, and that the effect is larger for men than women.

Three studies use data from the Household, Income and Labour Dynamics in Australia (HILDA) survey. Olesen et al. (2013) use path analysis to explore lagged and contemporaneous relationships between unemployment and MH measures of common mental disorders (measured using the Mental Health Inventory,

[6] 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).

MHI-5). Despite using longitudinal data, the 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 et al. (2017) use the same measure of MH to define transitions into and out of depressive episodes, and estimate how these transitions influence employment in a linear probability model with FE. They find that for men the probability of being unemployed rises with the onset of depressive symptoms, while for women unemployment is increased by protracted depressive symptoms. 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 9 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 which 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 to dropping out of high school, and it is 4 times the size of the OLS estimate. The bereavement instrument is shown to be a strong determinant of MH 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 omit 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.

In this paper we investigate the direction and magnitude of the potential bias in FE equations using the method of Oster (2013, 2017). The method formalises arguments that researchers often make when testing the robustness of their results. If a regression coefficient only changes a little when new controls are added, this is taken as evidence that any remaining bias is likely to be small; whereas if the coefficient changes by a lot, it is thought there could still be a substantial bias, undermining confidence in the coefficient estimate. The Oster method (which we explain in more detail in Section 3) allows us to bound the bias by comparing ‘uncontrolled’ and ‘controlled’ regressions under a set of assumptions about the relationship between observable and unobservable selection. As the variables included in the controlled regressions are the standard characteristics from the literature, 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 et al. (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 FE are added to equations for labour market outcomes (for example, Peng et al., 2015). One might then conclude that any remaining bias is in the same direction, although there is no theoretical reason why it should be (Peng et al., 2015). Indeed, our results suggest that the picture is somewhat more complicated.

## 3 Estimation approach

We start by estimating a Linear Probability Model (LPM) with individual FE $\mu_i$ where the dependent variable $Y_{it}$ is a binary indicator for whether individual *i* is employed or not in wave *t*:

$$Y_{it} = \alpha + M_{it-1}\beta + Z_{it}\theta + \mu_i + \gamma_t + \tilde{\varepsilon}_{it} \quad (1)$$

$M_{it-1}$ is a measure of the MH of individual $i$ in wave $t-1$; $Z_{it}$ is a vector of controls, including (time-varying) individual, household and area characteristics, with associated parameter vector $\theta$; $\gamma_t$ are wave FE , and $\tilde{\varepsilon}_{it}$ is the error term. The parameter of interest is $\beta$, the effect of MH in the previous wave on the probability of being employed in the current wave; we denote the estimate of $\beta$ from this regression as $\tilde{\beta}$. 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 (Bubonya et al., 2017; Greve and Nielsen, 2013). It is also used in the wider literature to model binary labour market outcomes. For example, Agüero and Marks (2008) use a 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 a LPM to model the choice to work unpaid overtime. The combination of individual FE and lagged MH is an attempt to minimise 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.[7] However, it is reasonable to assume that MH changes will take some time to feed through to labour market outcomes.

While Equation (1) fully controls for time-invariant heterogeneity, via $\mu_i$, there could still be time-varying heterogeneity not included in $Z_{it}$, which would bias the estimate of $\beta$. To assess the amount of bias, we apply the Oster (2017) approach, which starts with a specification of a complete regression, including observed and unobserved factors as well as any additional measurement error in the outcome variable. Rewriting our equation in Oster form leads to the following model (with subscripts suppressed for clarity):

$$Y = \alpha + M\beta + W_1 + W_2 + \epsilon \quad (2)$$

[7] 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. Further, the Oster bounds also show no bias, as is the case with lagged MH.

where $Y$ and $M$ are the employment and health measures as before; $W_1$ is an index that is a linear combination of observed variables and their corresponding coefficients (which in our case includes $Z$, $\mu$, and $\gamma$); $W_2$ is a similar index of variables that are correlated with both $Y$ and $M$, but which are not observed; and $\epsilon$ is measurement error in $Y$, uncorrelated with $M$, $W_1$ and $W_2$. In general, the Oster method uses information about the correlation between the observables and $M$ to compute the correlation between the unobservables and $M$, in order to estimate the degree of bias in the estimate of $\beta$ arising from omitted variables. Oster (2017) is critical of the argument that is often made in the existing literature, that if a coefficient is stable after the inclusion of the observed controls, then omitted variable bias must be limited.[8] This intuitive argument rests on the idea that bias arising from the observed controls is informative of bias arising from the unobserved factors. Oster (2017) shows that this also depends on how much of the variance in the outcome is explained by the control's inclusion; so we need also to observe how much the R-squared changes once controls are added.

The Oster approach requires a controlled regression, which includes all observable factors, and an uncontrolled regression, which includes only covariates that are not informative of selection on unobservables. Our controlled regression is Equation (1) and we specify our uncontrolled regression as:

$$Y_{it} = \alpha + M_{it-1}\beta + \mu_i + \gamma_t + \dot{\varepsilon}_{it} \tag{3}$$

The individual and wave FE are included in Equation (3) because they capture both the observed and unobserved components of their respective dimensions of variation. As there is no remaining unobserved component, any change in the MH coefficient when they are added does not tell us what would happen if further controls (varying both over time and across individuals) were added.[9] In contrast, the covariates $Z_{it}$ in Equation (1) are assumed to imperfectly capture the relevant time-varying factors that influence the relationship between MH ($M$) and employment ($Y$); and $W_2$ in Equation (2) are the unobserved counterparts to $Z_{it}$.

Two key parameters specify the relationship between observable and unobservable selection and the maximum amount of variation which can be explained by the model. The first parameter, $\delta$, is the coefficient of proportionality in the proportionality of selection equation $\delta\frac{\sigma_{1M}}{\sigma_1^2} = \frac{\sigma_{2M}}{\sigma_2^2}$, where $\sigma_{jM} = cov(W_j, M)$ and $\sigma_j^2 = var(W_j)$ for $j \in \{1,2\}$. $\delta$ 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 $\beta$ in the same direction; when $0 < \delta < 1$ the unobserved factors are less important than the observed factors (and the opposite holds when $\delta > 1$). The second parameter, $R_{max}$, is the maximum R-squared under the full model in Equation (2) where all (observed and unobserved) variables are included. This can be as high as 1 if $Y$ is measured without

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

[9] See the discussion in Oster (2013; p10) of fully observed variables such as sex. Another way to motivate the problem is to recast the analysis in the form of changes – the impact of a change in MH on a change in employment status. Rewriting the regressions as deviations from means would deliver the same coefficients but would remove the individual FE from the analysis.

error ($\epsilon = 0$), this but cannot be smaller than the R-squared obtained from the controlled regression. Both $\delta$ and $R_{max}$ are unknown parameters to be chosen given the particular context of the problem and econometric model. There are no standard values but Oster (2017) argues that an appropriate upper bound for $\delta$ is 1 because the observed variables are usually chosen based on the fact that they are the most important controls; conceptually we can think of the omitted variables as having been stripped of the portion related to the included variables (Oster, 2017, page 10). The range 0 to 1 for δ seems reasonable in our context, as we observe the key control variables that have been identified in the literature on health and work. The bound when $\delta = 0$ is $\tilde{\beta}$ (the estimate from the controlled regression). The other bound, $\beta^*$, can be approximated by the expression:

$$\beta^* \approx \tilde{\beta} - \delta\big[\dot{\beta} - \tilde{\beta}\big]\frac{R_{max}-\tilde{R}}{\tilde{R}-\dot{R}} \tag{4}$$

where $\dot{\beta}$ is the estimate of $\beta$ from the uncontrolled regression in Equation (3). $\tilde{R}$ and $\dot{R}$ are the R-squared values from Equations (1) and (3) respectively. Equation (4) is based on more restrictive assumptions than the full Oster estimator, but the expression is useful for developing the intuition behind the potential bias, which depends on the unknown parameters $\delta$ and $R_{max}$ and the estimated values $\tilde{\beta}$, $\dot{\beta}$, $\tilde{R}$, and $\dot{R}$. In particular, the size of the bias ($\beta^* - \tilde{\beta}$) depends not only on the effect of the observables on $\beta$ (i.e. the difference between $\tilde{\beta}$ and $\dot{\beta}$), but also on how much of the variation in Y the observables explain (the difference between $\tilde{R}$ and $\dot{R}$) relative to how much of this variation we expect the unobservables to explain (the difference between $R_{max}$ and $\tilde{R}$). It is therefore possible to have large bias even when $\beta$ is relatively stable (i.e. $\dot{\beta} - \tilde{\beta}$ is small) if $R_{max} - \tilde{R}$ is large compared to $\tilde{R} - \dot{R}$. Conversely, it is also possible to have little or no bias when $\beta$ is not stable if $\frac{R_{max}-\tilde{R}}{\tilde{R}-\dot{R}}$ is very small.

It is reasonable to assume that $R_{max}$ is less than 1 if idiosyncratic measurement error in $Y$ exists such that there are components of the variation in Y that are orthogonal to $M$, $W_1$ and $W_2$. It seems appropriate to assume $R_{max} < 1$ when using a LPM to model the binary employment outcome. Furthermore, we assume $R_{max}$ cannot be close to 1 in a FE model since the pertinent R-squared in the controlled regression is the within R-squared, which is much lower than the overall R-squared. Oster suggests a heuristic approach setting $R_{max} = 1.3\tilde{R}$ based on a sample of randomized trials (Oster, 2017: pp 3). At least 90% of the trial results she considers are robust to this value, but only 45% of results in a sample of non-randomized studies survive. Cattan et al. (2017) and Hener et al. (2016) use $R_{max} = min\{2.2\tilde{R}, 1\}$. An alternative approach would be to set $R_{max} = \tilde{R} + \big(\tilde{R} - \dot{R}\big)$(see for example Black et al., 2014); this assumes that the unobservables explain as much of the variation in employment as the observed controls (Oster, 2017: pp 15). We consider both $R_{max} = 1.3\tilde{R}$ and $R_{max} = 2.2\tilde{R}$ (generally $2.2\tilde{R} > \tilde{R} + (\tilde{R} - \dot{R})$ in our specifications).

This approach allows us to compute a bounding set $\Delta_s$ with the following bounds on $\beta$: (1) $\tilde{\beta}$ which is the estimate of $\beta$ in the controlled regression Equation (1), and (2) $\beta^*$ which is the effect of MH on employment corrected for omitted variable bias given the specified values of $R_{max}$ and $\delta$. Whether $\tilde{\beta}$ is the upper or lower bound of $\Delta_s$ 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.

We first estimate the FE regression from Equation (1) to obtain $\tilde{\beta}$, and then use the Stata user-written program PSACALC (Oster, 2016) to estimate $\beta^*$. This method gives two solutions when $\delta = 1$. The default is to choose the one that minimizes the squared difference to $\tilde{\beta}$ and does not change the direction of the bias. When $\delta$=0, the solution is $\tilde{\beta}$ (from the controlled regression) regardless of the value of $R_{max}$ or the uncontrolled regression being used.

# 4 Data

We use the first eight waves of the UK Household Longitudinal Study (UKHLS, 2018), with wave 1 data being collected in 2009/2010, wave 2 collected in 2010/2011, and so on until wave 8 which was collected in 2016/2017. Our analysis sample includes 21-55 year olds from England and Wales; we limit the sample to this age range to retain a focus on prime age workers. Table 1 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) [10]; 0 if the individual is unemployed, retired, looking after family/home, or long-term sick/disabled. We exclude individuals who are 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').

We use three alternative measures of MH; two derived from the 12-item General Health Questionnaire (GHQ-12), and one from the Short-Form 12 item health questionnaire (SF-12). The GHQ-12 is a widely recognized instrument that has been adopted by the World Health Organization as a screening tool for psychological disorders and has been validated in a number of international studies (Goldberg et al., 1997; Sartorius and Ustün, 1995; Schmitz et al., 1999). This measure is used as a measure of psychological health in an increasing number of economic studies (see for example Cornaglia et al., 2015; Gardner and Oswald, 2007; Roberts et al., 2011); including studies of the relationship between MH and work (see for exampleGarcia-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 to 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 cut-off is currently used by the NHS

[10] 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.

to monitor the percentage of people who suffer from poor MH in the general population.[11] Our second measure, also from the GHQ-12, is a cardinal measure based on the original 4 point scoring for each question, which ranges from 0-36 (henceforth GHQ36). In the original scoring a higher value signifies worse MH, but for ease of interpretation we recode so that a higher value corresponds to better 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 it brevity and psychometric performance.[12] 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 (Gill et al., 2007; Ware et al., 2002). The 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. The MCS has been used to analyse the MH effects of learning intensity (Hofmann and Mühlenweg, 2018), working-time mismatch (Otterbach et al., 2016), and work schedules of sole-parents (Dockery et al., 2016). Mitra and Jones (2017) use it to estimate the impact of MH changes on labour market outcomes in the US; 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 to 5 (see Jones et al., 2010 for a discussion). 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 rationalise 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.[13]

---

[11] For further details see https://files.digital.nhs.uk/BA/46AF8E/Spec_03J_321VSP2_10_V1.pdf. See also Goldberg et al. (1998) for a discussion of GHQ thresholds around the world.

[12] 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).

[13] 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 or not the respondent had ever received a diagnosis of a major MH disorder. We considered using a self-reported binary indicator of diagnosed depression in our modelling. However, the UKHLS data do not allow for reliable measurement of the incidence of depression.

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[14], 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 better 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.[15] Further, in sensitivity analysis we replace the PCS with a variable derived from questions on Activities of Daily Living; these record whether or not 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 2 and 3 show descriptive statistics for our estimation sample split by the dichotomous GHQ measure of MH. In total, there are 88,143 observations covering 11,263 men and 14,439 women[16]. Approximately a fifth are identified as having poor MH (GHQ12D=1) and these respondents accordingly have lower GHQ36 and MCS scores; they also have worse PH as shown by the PCS scores and problems with ADL. They are also less likely to be employed (69% employed vs. 86% for those who do not have a MH problem), be married, or have higher education. 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 3 shows that among the non-employed, a similar proportion of those with a MH problem are unemployed compared to those without a problem (31.3% and 29.2% respectively), but the largest group of those with poor MH are long-term sick or disabled (34.5%), while the majority of those who do not have poor MH are involved in family/home care (54.4%). Out of the total 25,702 respondents, 10.9% change their employment status at least once over the period of analysis, and 29.2% change their MH status as measured by the dichotomous GHQ variable (not shown in tables). Figure 1 plots the employment gap by age between individuals with good and poor MH (as measured by the GHQ12D binary indicator) in all 8 waves of our UKHLS sample; this gap is substantial and appears to widen with age.

[14] Although we have exact age, we use seven 5-year age groups in our analysis to allow for possible non-linear effects (21-25, 26-30, 31-35, 36-40,41-45, 46-50, 51-55).

[15] 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.

[16] There are slightly fewer observations available for the ADL measures.

## 5 Results

Table 4 contains point estimates for models using the GHQ12D dichotomous measure for MH. For comparison with the FE estimates, the first two columns show the results for the pooled OLS models without controls (column 1) and with controls, $Z$ (column 2). The next two columns show the FE estimates: from the uncontrolled regression, Equation (3), in column 3 and from the full controlled model, Equation (1), in column 4. We have also estimated these models for each gender separately, but we find no significant differential gender effect on employment.[17] However, we find significant gender differences for being married, having children aged 0-4, and having children 12-15. We therefore include in our controls gender interactions with being married and all the children variables.[18]

The OLS coefficient in column 1 shows that poor MH is associated with a 15.6 percentage point lower probability of employment (controlling only for wave dummies). When the main controls are added, the MH coefficient shrinks to -9.7 percentage points, and when we include FE in the specification, the coefficient falls still further to -1.4 percentage points. There is thus quite strong selection into mental health problems based on observed characteristics but especially strong selection based on time-invariant characteristics as a whole (both observed and unobserved). Indeed once FE are included, it makes little difference whether or not we include the additional controls. This provides some tentative evidence (to be investigated formally using the Oster method) 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 1.4 percentage points[19].

The control variables in column 4 all appear to have the expected effects on the employment probability. PH is positively associated with employment, and while the effect is much smaller in the FE model compared to the OLS, it is still statistically significant. In the FE model, the effect of age is significantly larger for all age groups compared to 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 to men, a stronger negative effect of having children aged 0-4, and a larger positive effect of having secondary school aged children (aged 12-15). The education gradient is as expected; those who gained 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

[17] This is in line with the findings of Ettner et al. (1997), who also find no significant gender differences

[18] For conciseness, these interaction effects are not reported in the tables.

[19] The effect is -1.3 percentage points for women and -1.6 for men in separate gender regressions, but as mentioned previously, this difference is not statistically significant

statistically significant, which may reflect the fact that although they vary a lot spatially, they exhibit only limited variation over time.

In Table 5 we re-estimate the specifications from Table 4 using the continuous GHQ36 measure and the SF12 MCS; for both of these measures, higher values represent better MH. As expected, we find a positive relationship between these measures and the probability of being employed.[20] Again, in the OLS regressions the addition of control variables reduces the estimated effect of MH for both measures, and the inclusion of FE reduces both estimates still further. In the FE models, while the GHQ36 coefficient is not changed by the addition of controls, the MCS coefficient is increased (from 0.0009 to 0.0012). The MCS result suggests that, contrary to the cross-sectional selection effects, whereby characteristics that are positively associated 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 using the Oster method.

In Table 5 we also report the standardised coefficients on GHQ36 and MCS. Our preferred specification (column 4) suggests that a one standard deviation increase in GHQ36 (MCS) leads to a 1.0 (1.2) percentage point increase 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 they have equivalently sized effects on employment.

We present Oster bounds for our FE models with full controls in Table 6. For ease of comparison, the first column repeats the estimates from the controlled regression in Equation (1) (i.e. $\tilde{\beta}$ when $\delta = 0$). The bias adjusted Oster estimates ($\beta^*$) 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 both GHQ measures the bias adjusted estimates are the same as the FE estimates ($\tilde{\beta}$) under both $R_{max}$ assumptions. For MCS, as was suggested by the coefficient change noted above, some bias is exhibited; $\tilde{\beta}$ is the lower bound and $\beta^*$ the upper bound. However, this result should be interpreted with caution as these upper bounds are very close to the estimated FE coefficient. For all three MH measures, their 95% confidence intervals overlap with those of $\tilde{\beta}$. Also $\beta^*$ itself is within the 95% confidence intervals of $\tilde{\beta}$ in all cases except for MCS when $R_{max} = 2.2\tilde{R}$).[21] This suggests there is little concern regarding omitted variable bias in these FE models for the effect of MH on employment.

In order to explore the heterogeneity of effects in different subgroups we focus on the preferred model in Equation (1) (column 4 in Tables 4 and 5) and split the sample by age groups (Table 7), education (Table

[20] For conciseness, we do not report the results for the other control variable in Table 5; they are very similar to the results in Table 4.
[21] We have found non-overlapping confidence intervals between $\tilde{\beta}$ and $\beta^*$ in the MCS model only when we increase $R_{max}$ to $3\tilde{R}$, which we do not consider to be a plausible assumption.

8), PH terciles (Table 9) and relative poverty[22] (Table 10). Differences across groups were tested for statistical significance and this is noted in the rightmost column where relevant.[23] The relationship between the GHQ12 dummy and employment remains negative and statistically significant with insignificant differences across most subsamples, with two notable exceptions. The effect of GHQ12D is significantly smaller in magnitude for those with a degree (-0.0063) than for those without a degree (-0.0205) (Table 8), and significantly smaller for those households living above the relative poverty line (-0.0091) than for those below (-0.0221) (Table 10). This suggests that higher education moderates the effect of MH disorders on employment, while relative poverty exacerbates it. However, these group differences are weaker when we use other measures of MH. The effects of the continuous GHQ36 and MCS on employment are also significantly lower for those with a degree, but the latter difference is not statistically significant (Table 8). Similarly, the difference in the effects of GHQ36 and MCS is not significant between those living above and below the relative poverty line (Table 10).

We note that the effects of PH are more heterogeneous. For example, they are significantly larger for those aged 46-55 compared to 31-45 year olds in the model using MCS (Table 7). They are also larger for those in the bottom PCS tercile compared to those in the middle and top terciles across all models (Table 9), as well as for those below the poverty line compared to those above the poverty line (Table 10). However, unlike GHQ12D and GHQ36, the effects of PH are not statistically different between those with a degree and those without (Table 8). It is also worth noting that, in addition to the results presented here, we also explored differences across sub-groups defined by gender, household income and whether or not there is another employed person in the household. We found no significant differences in the effect of MH on employment between these groups.

Table 11 presents Oster bound estimates for select split sample FE models with full controls[24]. Similar to the pooled results, the bias adjusted effects of MH ($\beta^*$) 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 is outside of the 95% confidence interval of $\tilde{\beta}$ in the MCS model for each of the subsamples.

<u>*Sensitivity Analysis*</u>

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 to the coefficient in our benchmark model from Table 4 (-0.01408), the effect of GHQ12D is

[22] The poverty line is 60% of the median net equivalised HH income (before housing costs) in the UK adjusted for inflation using the Consumer Price Index (data source: 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.
[23] Tests of significance were carried out in Stata Version 15.1 using the ***suest*** and ***test*** commands (this required us to estimate the FE model using OLS after first demeaning the data with the user written ***center*** command).
[24] 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.

smaller for the lower cut-off and stronger for the higher cut-offs (see Panel 1 in Table 12) We also consider an alternative measure of PH based on the Activities of Daily Living (ADL) questions. We classify individuals into 4 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 12).

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 to those employed.[25] 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 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)[26] in the neighbourhood[27] where the household is located and local labour market tightness.[28] We find no significant differences in the effect of MH between these sub-groups, across all three measures of MH.

# 6 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. This is a highly relevant parameter for policymakers who wish to understand how deteriorations or improvements in health may affect employment levels in the population. In contrast, IV methods deliver a LATE that typically applies to a narrow subgroup only, for instance those who have suffered a bereavement, and who are probably not the specific target of policy.

[25] 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 generalise about their MH and work relationship.
[26] 2015 IMD data obtained from the Ministry of Housing, Communities & Local Government.
[27] Neighbourhoods are defined as 'lower layer super output areas' (LSOA); these are very small geographic areas with an average population size of 1500.
[28] Labour market conditions may moderate the relationship between MH and employment status (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 2009-2012 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.

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 an employment equation, there is little additional change after FE are included. The observed time-varying characteristics in a FE equation have high explanatory power in an employment equation, but adding them barely changes the estimated effect of MH. There could of course still be a substantial bias if the included controls represent only a small subset of all possible controls. We allow for this in the Oster method 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 Oster 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 minimise the possible influence of reverse causality by using lagged MH, there could still be some residual bias.

Our preferred specifications indicate that transitioning into poor MH (as measured by GHQ) leads to a reduction of 1.4 percentage points in the probability of employment, and that a one standard deviation change in the continuous measures of MH causes a 1.0-1.2 percentage point change in the probability of employment. Comparisons of these effects with previous studies are not straightforward because of differences in the MH measures used and the way the effects are reported (e.g. as discrete or continuous changes in MH). Moreover, there are no directly comparable studies for the UK. However, our effects appear to be considerably smaller than estimates from other countries using 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 pecentage 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, IV estimates deliver a LATE which is probably not relevant to policy. Studies which are more comparable to ours and use FE methods find effects in a similar ballpark to us (albeit again using different MH measures). Estimates of the effects of MH episodes, summarised across specifications and types of transition, are: 1.6-8.0 percentage points depending on the severity of symptoms (US; Peng at al., 2015); 0.0-8.2 (US; MCS measure; Mitra and Jones, 2017); and 0.0-2.9 percentage points (Australia; Bubonya et al., 2017). These more modest effect sizes are arguably more relevant for policy than large IV estimates. The MH effect does not differ across gender but we find tentative evidence that MH has a bigger effect on employment for those is 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 percentage points for people without degree, compared with just 0.6 percentage points for those with degree. Thus there is a case for policy to prioritise 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 to policymakers than the LATE delivered by IV methods. Given the widespread availability of longitudinal data, these findings should provide some reassurance to practitioners using FE methods to investigate the impacts of health on work. We also hope they will be inspired to investigate the reliability of FE models in other contexts. Our two alternative measures of MH gave very similar results, suggesting that either GHQ or MCS can be used as a basis for analysis.

The results also imply that research users can have a good degree of confidence in the reliability of FE results, whereas more scepticism might be warranted about the applicability of IV estimates. At the same time it is 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 mental health (16 percentage points) is much larger that the effect of changing between MH states (1.4 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 and educational attainment. Longer-term structural changes, which impact on all of these factors, will almost certainly be required to eliminate the gap completely.

## References

Agüero, J.M., Marks, M.S., 2008. Motherhood and Female Labor Force Participation: Evidence from Infertility Shocks. American Economic Review 98(2), 500-504.

Alexandre, P.K., 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 4(4), 161-173.

Altonji, J., Conley, T., Elder, T., Taber, C., 2011. Methods for Using Selection on Observed Variables to Address Selection on Unobserved Variables. Mimeo, Yale University.

Altonji, J., Elder, T., Taber, C., 2005. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. Journal of Political Economy 113(1), 151-184.

Andersen, M., 2015. Heterogeneity and the Effect of Mental Health Parity Mandates on the Labour Market. Journal of Health Economics 43, 74-84.

Banerjee, S., Chatterji, P., Lahiri, K., 2017. Effects of Psychiatric Disorders on Labour Market Outcomes: A Latent Variable Approach Using Pultiple Clinical Indicators. Health Economics 26(2), 184-205.

Black, B., de Carvalho, A.G., Khanna, V., Kim, W., Yurtoglu, B., 2014. Methods of Multicountry Studies of Corporate Governance: Evidence From the BRIKT countries. Journal of Econometrics 183(2), 230-240.

Black, C., 2009. Working for a Healthier Tomorrow. Review of the Health of Britain's Working Age Population. Stationery Office 66(1), 1-2.

Black, C., Frost, D., 2011. Health at Work - an Independent Review of Sickness Absence.

Blundell, R., Britton, J., Costa Dias, M., French, E., 2017. The Impact of Health on Labour Supply near Retirement. IFS Working Paper No. W17/18.

Bound, J., 1991. Self-Reported Versus Objective Measures of Health in Retirement Models. Journal of Human Resources 26(1), 106-138.

Bound, J., Schoenbaum, M., Stinebrickner, T.R., Waidmann, T., 1999. The Dynamic Effects of Health on the Labour Force Transitions of Older Workers. Labour Economics 6, 179-202.

Bubonya, M., Cobb-Clark, D.A., Ribar, D., 2017. The Bilateral Relationship between Depressive Symptoms and Employment Status. IZA Discussion Paper No. 10653.

Cattan, S., Kamhofer, D.A., Karlsson, M., Nilsson, T., 2017. The Short and Long-Term Effects of Student Absence: Evidence from Sweden. IFS Working Paper W17/21.

Chatterji, P., Alegria, M., Takeuchi, D., 2007. Psychiatric Disorders and Labor Market Outcomes: Evidence from the National Latino and Asian American Study. Health Economics 16(10), 1069-1090.

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

Chatterji, P., Alegria, M., Takeuchi, D., 2011. Psychiatric Disorders and Labour Market Outcomes: Evidence from the National Comorbidity Survey-Replication. Journal of Health Economics 30, 858-868.

Cornaglia, F., Crivellaro, E., McNally, S., 2015. Mental Health and Education Decisions. Labour Economics 3, 1-12.

Currie, J., Madrian, B., 1999. Health, Health Insurance and the Labour Market. Handbook of Labor Economics. Elsevier, pp. 3309-3416.

DeSimone, J., 2002. Illegal Drug Use and Employment. Journal of Labor Economics 20(4), 952-977.

Dockery, A.M., Li, J., Kendall, G., 2016. Sole-Parent Work Schedules and Adolescent Wellbeing: Evidence from Australia. Social Science and Medicine 168, 167-174.

DWP, 2016. Work, Health and Disability: Improving Lives. Department for Work and Pensions and Department of Health, Green Paper.

DWP, 2017. Improving Lives: The Future of Work, Health and Disability. Department for Work and Pensions and Department of Health, Policy Paper.

Ettner, S.L., Frank, R.G., Kessler, R.C., 1997. The Impact of Psychiatric Disorders on Labor Market Outcomes. Industrial and Labor Relations Review 51(1), 64-81.

Francesconi, M., Van der Klaauw, W., 2007. The Socioeconomic Consequences of "In-Work" Benefit Reform for British Lone Mothers. Journal of Human Resources XLII(1), 1-31.

Frank, R., Gertler, P., 1991. An Assessment of Measurement Error Bias for Estimating the Effect of Mental Distress on Income. Journal of Human Resources 26(1), 154-164.

Frijters, P., Johnston, D.W., Shields, M.A., 2014. The Effect of Mental Health on Employment: Evidence from Australian Panel Data. Health Economics 23, 1058-1071.

Garcia-Gomez, P., Jones, A.M., Rice, N., 2010. Health Effects on Labour Market Exits and Entries. Labour Economics 17, 62-76.

Gardner, J., Oswald, A.J., 2007. Money and Mental Wellbeing: A Longitudinal Study of Medium-sized Lottery Wins. Journal of Health Economics 26(1), 49-60.

Ghatak, A., 2010. Health, Labour Supply and Wages: A Critical Review of the Literature. Bangalore University Institute for Social and Economic Change Working Paper 244.

Gill, S.C., Butterworth, P., Rodgers, B., 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 152, 63-71.

Goldberg, D.P., Gater, R., Sartorius, N., Ustün, T.B., Piccinelli, M., Gureje, O., Rutter, C., 1997. The Validity of two Versions of the GHQ in the WHO Study of Mental Illness in General Health Care. Psychological Medicine 27(1), 191-197.

Goldberg, D.P., Oldehinkel, T., Ormel, J., 1998. Why GHQ Threshold Varies from one Place to another. Psychological Medicine 28(4), 915-921.

Gregg, P., Grout, P.A., Ratcliffe, A., Windmeijer, F., 2011. How Important is Pro-Social Behaviour in the Delivery of Public Services? Journal of Public Economics 95, 758-766.

Greve, J., Nielsen, L.H., 2013. Useful Beautiful Minds - An Analysis of the Relationship Between Schizophrenia and Employment. Journal of Health Economics 32, 1066-1076.

Grossman, M., 1972. On the Concept of Health Capital and the Demand for Health. Journal of Political Economy 80(2), 223-255.

Hamilton, V.H., Merrigan, P., Dufresne, E., 1997. Down and Out: Estimating the Relationship Between Mental Health and Unemployment. Econometrics and Health Economics 6, 397-406.

Hener, T., Rainer, H., 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 179(3), 633-656.

Hofmann, S., Mühlenweg, A., 2018. Learning Intensity Effects in Students' Mental and Physical Health - Evidence from a Large Scale Natural Experiment in Germany. Economics of Education Review 67, 216-234.

Houssemand, C., Meyers, R., 2011. Unemployment and Mental Health in a Favourable Labour Market. International Journal of Psychology 46(5), 377-385.

Jones, A.M., Rice, N., 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 27, 866-880.

Kessler, R.C., Berglund, P.A., Demler, O., Jin, R., Walters, 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 62(6), 593-602.

Kreider, B., Pepper, J.V., 2007. Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors. Journal of the American Statistical Association 102(478), 432-441.

Marcotte, D.E., Wilcox-Gök, V., Redmon, D.P., 2000. The Labour Market Effects of Mental Illness: The Case of Effective Disorders. In: Salkever, Sorkin (Eds.). The Economics of Disability. JAI Press Greenwich, CT, pp. 181-210.

Mavridis, D., 2015. The Unhappily Unemployed Return to Work Faster. IZA Journal of Labor Economics 4(1), 1-22.

McInnes, R., 2012. ESA and Incapacity Benefit Statistics. House of Commons Library Report, SN01420.

Mitra, S., Jones, K., 2017. The Impact of Recent Mental Health Changes on Employment: New Evidence from Longitudinal Data. Applied Economics 49(1), 96-109.

Mullahy, J., Sindelar, J., 1996. Employment, Unemployment, and Problem Drinking. Journal of Health Economics 15(4), 409-434.

Munford, L., Rice, N., Roberts, J., 2016. Econometric Modelling of the Employment Gap for the NHS Outcomes Framework. Report for the Department of Health: http://cfe.org.uk/app/uploads/2016/09/Econometric-Modelling-of-the-Employment-Gap_July-16.pdf.

Oakley, M., 2016. Closing the Gap. Social Market Foundation Report.

OECD/EU, 2018. Health at a Glance: Europe 2018: State of Health in the EU Cycle. OECD Publishing, Paris/EU, Brussels: https://doi.org/10.1787/health_glance_eur-2018-en.

Ojeda, V.D., Frank, R.G., McGuire, T.G., Gilmer, T.P., 2010. Mental Illness, Nativity, Gender and Labor Supply. Health Economics 19(4), 396-421.

Olesen, S.C., Butterworth, P., Leach, L.S., Kelaher, M., Pirkis, J., 2013. Mental Health Effects Future Employment as Job Loss Affects Mental Health: Findings from a Longitudinal Population Study. BMC Psychiatry 13(144).

Oster, E., 2013. Unobservable Selection and Coefficient Stability: Theory and Validation. NBER Working Paper No. 19054.

Oster, E., 2016. PSACALC: Stata Module to Calculate Treatment Effects and Relative Degree of Selection under Proportional Selection of Observables and Unobservables, revised 18 Dec 2016 ed. Boston Collage, Department of Economics, Statistical Software Components S457677.

Oster, E., 2017. Unobservable Selection and Coefficient Stability: Theory and Evidence. Journal of Business & Economic Statistics (forthcoming).

Otterbach, S., Wooden, M., Fok, Y.K., 2016. Working-Time Mismatch and Mental Health. IZA Discussion Paper No. 9818.

Peng, L., Meyerhoefer, C.D., Zuvekas, S.H., 2015. The Short-Term Effect of Depressive Symptoms on Labour Market Outcomes. Health Economics 25(10), 1223-1238.

Roberts, J., Hodgson, R., Dolan, P., 2011. It's Driving Her Mad: Gender Differences in the Effects of Commuting on Psychological Health. Journal of Health Economics 30(5), 1064-1076.

Sartorius, N., Ustün, T.B., 1995. Mental Illness in General Health Care: an International Study. New York: John Wiley & Sons.

Schmitz, N., Kruse, J., Tress, W., 1999. Psychometric Properties of the General Health Questionnaire (GHQ-12) in a German Primary Care Sample. Acta Psychiatrica Scandinavica 100(6), 462-468.

Steele, F., French, R., Bartley, M., 2013. Adjusting the Selection Bias in Longitudinal Analyses Using Simultaneous Equations Modeling: The Relationship between Employment Transitions and Mental Health. Epidemiology 24(5), 703-711.

UKHLS, 2018. Understanding Society: Waves 1-8, 2009-2017 and Harmonised BHPS: Waves 1-18, 1991-2009. [data collection]. 11th Edition. University of Essex. Institute for Social and Economic Research, N.S.R., Kantar Public. UK Data Service. SN: 6614.

Verbeek, M., 2012. A Guide to Modern Econometrics, 4th ed. Chichester, England: John Wiley & Sons.

Waddell, G., Burton, A.K., 2006. Is Work Good for your Health and Well-Being? Stationery Office.

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

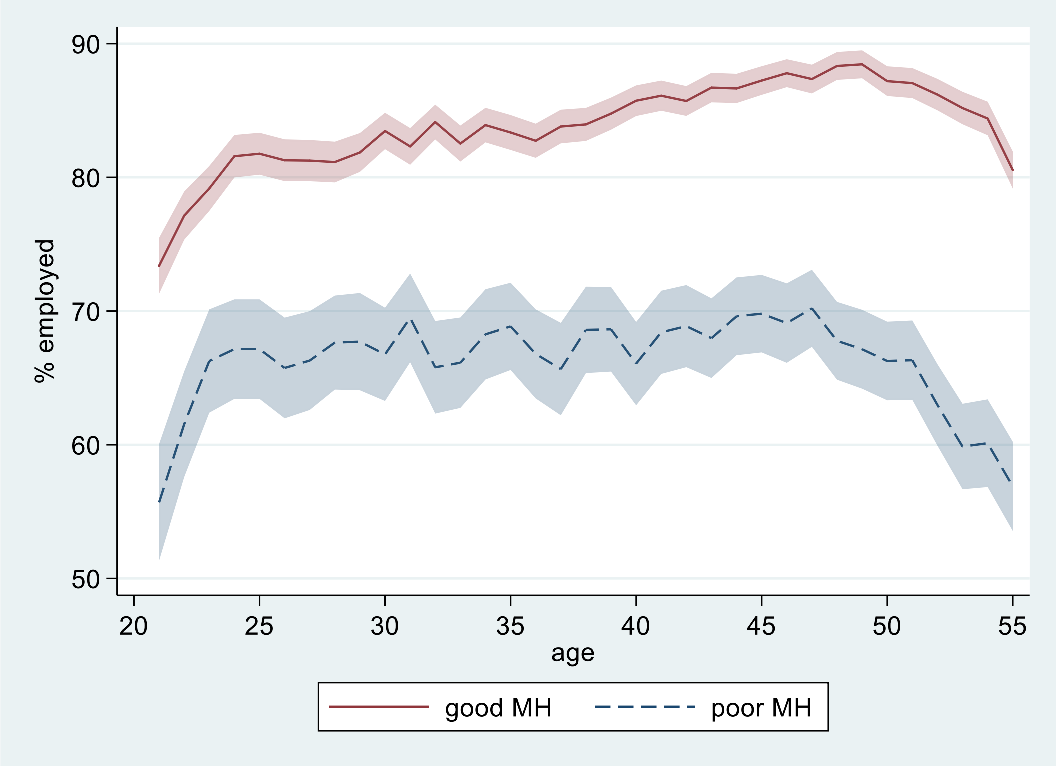
WHO, 2000. Mental Health and Work: Impact, Issues and Good Practices. World Health Organization: https://www.who.int/mental_health/media/en/73.pdf.

WHO, 2008. The Global Burden of Disease: 2004 Update. World Health Organization.

WHO, 2013. Investing in Mental Health: Evidence for Action. World Health Organization.

WPC, 2017. Disability Employment Gap. Seventh Report of Session 2016–17. House of Commons, Work and Pensions Committee.

***Figure 1. Employment across ages by GHQ12D (with 95% confidence intervals).***




Source: UKHLS Waves 1-8 (University of Essex, 2018)

***Table 1. Variable definitions.***

| | 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 (inverted so that 0 represents the most distressed and 36 represents the least distressed). | UKHLS |
| MCS | Mental Component Summary, measured on a 0-100 continuous scale based on the SF-12 questionnaire where 0 denotes low functioning and 100 denotes high functioning. | UKHLS |
| *Individual controls* | | |
| PCS | Physical Component Summary, measured on a 0-100 continuous scale based on the SF-12 questionnaire where 0 denotes low functioning and 100 denotes high 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 |
| 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 (Annual Population Survey) |
| 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 |

***Table 2. Summary statistics.***

| | GHQ12D=0 | | | GHQ12D=1 | | |
|---|---|---|---|---|---|---|
| | obs | mean | (st. dev.) | obs | mean | (st. dev.) |
| Employed | 70,554 | 0.86 | | 17,589 | 0.69 | |
| GHQ12D t-1 | 70,554 | 0.13 | | 17,589 | 0.48 | |
| GHQ36 | 70,554 | 26.90 | (2.88) | 17,589 | 15.59 | (5.19) |
| GHQ36 t-1 | 70,554 | 25.78 | (4.57) | 17,589 | 20.17 | (7.11) |
| MCS | 70,554 | 51.23 | (7.23) | 17,589 | 36.13 | (10.09) |
| MCS t-1 | 70,554 | 50.32 | (8.51) | 17,589 | 41.43 | (11.48) |
| PCS | 70,554 | 52.80 | (8.09) | 17,589 | 48.89 | (13.20) |
| PCS t-1 | 70,554 | 52.98 | (8.27) | 17,589 | 49.08 | (12.17) |
| *ADL problems* | 70,522 | | | 17,566 | | |
| 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 | 70,554 | 40.53 | (9.12) | 17,589 | 40.73 | (9.31) |
| Married | 70,554 | 0.74 | | 17,589 | 0.65 | |
| *Education level* | 70,554 | | | 17,589 | | |
| No education | | 0.04 | | | 0.06 | |
| O-level | | 0.29 | | | 0.32 | |
| A-level | | 0.21 | | | 0.20 | |
| Degree | | 0.46 | | | 0.42 | |
| No child in HH | 70,554 | 0.50 | | 17,589 | 0.53 | |
| Child 0-4 in HH | 70,554 | 0.21 | | 17,589 | 0.19 | |
| Child 5-11 in HH | 70,554 | 0.30 | | 17,589 | 0.28 | |
| Child 12-15 in HH | 70,554 | 0.19 | | 17,589 | 0.19 | |
| Adults in HH | 70,554 | 2.33 | (0.97) | 17,589 | 2.28 | (1.05) |
| Other HH income | 70,554 | 2616 | (2374) | 17,589 | 2441 | (2139) |
| Unemployment rate | 70,554 | 7.14 | (2.90) | 17,589 | 7.40 | (2.95) |
| GVA | 70,554 | 23070 | (13858) | 17,589 | 22908 | (14554) |

***Table 3. Breakdown of observations by employment status (all waves pooled).***

| | GHQ12D=0 | | GHQ12D=1 | |
|---|---|---|---|---|
| | observations | % of non-employed | observations | % of non-employed |
| Self employed | 7,680 | | 1,392 | |
| Paid employment(ft/pt) | 53,263 | | 10,735 | |
| **Total employed** | **60,943** | | **12,127** | |
| | | | | |
| *Non-employed* | | | | |
| Unemployed | 2,804 | (29.2%) | 1,707 | (31.3%) |
| Retired | 388 | (4.0%) | 72 | (1.3%) |
| Family care or home | 5,226 | (54.4%) | 1791 | (32.8%) |
| LT sick or disabled | 1,169 | (12.2%) | 1,883 | (34.5%) |
| On apprenticeship | 24 | (0.2%) | 9 | (0.2%) |
| **Total non-employed** | **9,611** | | **5,462** | |

***Table 4. LPM coefficient estimates for pooled sample (MH = GHQ caseness indicator).***

| | (1) OLS | (2) OLS | (3) FE | (4) FE |
|---|---|---|---|---|
| GHQ12D t-1 | -0.1557 *** | -0.0973 *** | -0.0141 *** | -0.0141 *** |
| | (0.0052) | (0.0042) | (0.0027) | (0.0027) |
| PCS t-1 | | 0.0095 *** | | 0.0008 *** |
| | | (0.0002) | | (0.0002) |
| Female | | 0.0853 *** | | |
| | | (0.0146) | | |
| Age 26-30 | | 0.0150 ** | | 0.0313 *** |
| | | (0.0075) | | (0.0085) |
| Age 31-35 | | 0.0374 *** | | 0.0519 *** |
| | | (0.0082) | | (0.0108) |
| Age 36-40 | | 0.0435 *** | | 0.0558 *** |
| | | (0.0083) | | (0.0125) |
| Age 41-45 | | 0.0454 *** | | 0.0575 *** |
| | | (0.0080) | | (0.0140) |
| Age 46-50 | | 0.0355 *** | | 0.0603 *** |
| | | (0.0077) | | (0.0155) |
| Age 51-55 | | 0.0092 | | 0.0530 *** |
| | | (0.0078) | | (0.0170) |
| Married | | 0.1490 *** | | 0.0484 *** |
| | | (0.0075) | | (0.0078) |
| O-level | | 0.2208 *** | | 0.0553 |
| | | (0.0135) | | (0.0344) |
| A-level | | 0.3086 *** | | 0.0959 *** |
| | | (0.0135) | | (0.0366) |
| Degree | | 0.3392 *** | | 0.1124 *** |
| | | (0.0132) | | (0.0370) |
| No child in HH | | -0.0033 | | -0.0077 |
| | | (0.0081) | | (0.0065) |
| Child 0-4 in HH | | -0.0162 ** | | -0.0092 * |
| | | (0.0069) | | (0.0052) |
| Child 5-11 in HH | | -0.0033 | | 0.0070 |
| | | (0.0059) | | (0.0047) |
| Child 12-15 in HH | | -0.0016 | | 0.0071 |
| | | (0.0070) | | (0.0052) |
| Adults in HH | | 0.0225 *** | | 0.0201 *** |
| | | (0.0024) | | (0.0026) |
| ln(other HH income) | | -0.0283 *** | | -0.0148 *** |
| | | (0.0010) | | (0.0010) |
| Unemployment rate | | -0.0072 *** | | 0.0006 |
| | | (0.0007) | | (0.0005) |
| GVA per head / 10000 | | -0.0043 *** | | 0.0021 |
| | | (0.0014) | | (0.0018) |
| Constant | 0.8324 *** | 0.1796 *** | 0.8168 *** | 0.6861 *** |
| | (0.0037) | (0.0217) | (0.0025) | (0.0377) |
| R-squared | 0.0288 | 0.2325 | 0.8029 | 0.8065 |
| within R-squared | | | 0.0033 | 0.0214 |

Clustered standard errors in parentheses, * p<0.10 ** p<0.05 *** p<0.01. Regressions based on 88,143 observations. All models include wave dummies. Regressions (2) and (4) include gender interaction terms with being married and all the children variables.

***Table 5. LPM coefficient estimates for pooled sample (alternative MH measures).***

| | (1) OLS | (2) OLS | (3) FE | (4) FE |
|---|---|---|---|---|
| ***Panel 1*** | | | | |
| GHQ36 t-1 | 0.0143 *** | 0.0091 *** | 0.0017 *** | 0.0017 *** |
| | (0.0004) | (0.0003) | (0.0002) | (0.0002) |
| | [0.0807] | [0.0515] | [0.0097] | [0.0098] |
| PCS t-1 | | 0.0091 *** | | 0.0008 *** |
| | | (0.0002) | | (0.0002) |
| controls | no | yes | no | yes |
| R-squared | 0.0474 | 0.2398 | 0.8030 | 0.8066 |
| within R-squared | | | 0.0040 | 0.0221 |
| ***Panel 2*** | | | | |
| MCS t-1 | 0.0084 *** | 0.0068 *** | 0.0009 *** | 0.0012 *** |
| | (0.0002) | (0.0002) | (0.0001) | (0.0002) |
| | [0.0823] | [0.0670] | [0.0088] | [0.0119] |
| PCS t-1 | | 0.0102 *** | | 0.0012 *** |
| | | (0.0002) | | (0.0002) |
| controls | no | yes | no | yes |
| R-squared | 0.0490 | 0.2525 | 0.8030 | 0.8067 |
| within R-squared | | | 0.0037 | 0.0224 |

* p<0.10 ** p<0.05 *** p<0.01. Standard errors in parentheses (clustered at individual level). Standardized MH coefficients in brackets. All models include wave dummies. Sample size: 88,143.

***Table 6. Oster bounds for pooled FE models with full controls.***

| | $\delta = 0$ ($\tilde{\beta}$) | $\delta = 1$ ($\beta^*$) $R_{max} = 1.3\tilde{R}$ | $\delta = 1$ ($\beta^*$) $R_{max} = 2.2\tilde{R}$ |
|---|---|---|---|
| GHQ12D t-1 | -0.0141 | -0.0141 | -0.0141 |
| | (0.0027) | [0.0030] | [0.0030] |
| GHQ36 t-1 | 0.0017 | 0.0017 | 0.0017 |
| | (0.0002) | [0.0003] | [0.0003] |
| MCS t-1 | 0.0012 | 0.0014 | **0.0018** |
| | (0.0002) | [0.0002] | [0.0002] |

Bootstrapped standard errors in square brackets (1000 reps). Clustered standard errors in parentheses. Bounds in bold are outside the 95% CI of the coefficient in the controlled regression.

***Table 7. LPM coefficient estimates by age groups.***

| | (1) age 21-30 | (2) age 31-45 | (3) age 46-55 | differences |
|---|---|---|---|---|
| ***Panel 1*** | | | | |
| GHQ12D t-1 | -0.0168 ** | -0.0101 *** | -0.0146 *** | |
| | (0.0079) | (0.0038) | (0.0046) | |
| PCS t-1 | 0.0005 | 0.0004 | 0.0011 *** | |
| | (0.0005) | (0.0003) | (0.0003) | |
| within R-squared | 0.0302 | 0.0222 | 0.0118 | |
| ***Panel 2*** | | | | |
| GHQ36 t-1 | 0.0022 *** | 0.0012 *** | 0.0021 *** | |
| | (0.0006) | (0.0003) | (0.0004) | |
| PCS t-1 | 0.0005 | 0.0004 | 0.0011 *** | |
| | (0.0005) | (0.0003) | (0.0003) | |
| within R-squared | 0.0312 | 0.0226 | 0.0130 | |
| ***Panel 3*** | | | | |
| MCS t-1 | 0.0012 *** | 0.0009 *** | 0.0015 *** | |
| | (0.0004) | (0.0002) | (0.0003) | |
| PCS t-1 | 0.0009 * | 0.0007 *** | 0.0017 *** | (2) ≠ (3) ** |
| | (0.0005) | (0.0003) | (0.0003) | |
| within R-squared | 0.0308 | 0.0228 | 0.0137 | |

* p<0.10 ** p<0.05 *** p<0.01. 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 4 model (4). There are 15,397 observations for 21-30 year olds, 41,639 for 31-45, and 31,107 observations for 46-55.

***Table 8. LPM coefficient estimates by education.***

| | (1) w/o degree | (2) with degree | differences |
|---|---|---|---|
| ***Panel 1*** | | | |
| GHQ12D t-1 | -0.0205 *** | -0.0063 * | (1) ≠ (2) ** |
| | (0.0041) | (0.0036) | |
| PCS t-1 | 0.0010 *** | 0.0006 ** | |
| | (0.0002) | (0.0003) | |
| within R-squared | 0.0278 | 0.0147 | |
| ***Panel 2*** | | | |
| GHQ36 t-1 | 0.0023 *** | 0.0011 *** | (1) ≠ (2) ** |
| | (0.0004) | (0.0003) | |
| PCS t-1 | 0.0010 *** | 0.0006 ** | |
| | (0.0002) | (0.0003) | |
| within R-squared | 0.0286 | 0.0152 | |
| ***Panel 3*** | | | |
| MCS t-1 | 0.0014 *** | 0.0009 *** | |
| | (0.0002) | (0.0002) | |
| PCS t-1 | 0.0015 *** | 0.0009 *** | |
| | (0.0003) | (0.0003) | |
| within R-squared | 0.0287 | 0.0156 | |

* p<0.10 ** p<0.05 *** p<0.01. 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 4 model (4). There are 48,003 observations without a degree and 40,140 with a degree.

***Table 9. LPM coefficient estimates by PCS terciles.***

| | (1)<br>PCS bottom tercile | (2)<br>PCS middle tercile | (3)<br>PCS top tercile | differences |
|---|---|---|---|---|
| ***Panel 1*** | | | | |
| GHQ12D t-1 | -0.0132 ** | -0.0158 ** | -0.0114 ** | |
| | (0.0053) | (0.0066) | (0.0047) | |
| PCS t-1 | 0.0014 *** | -0.0003 | 0.0000 | (1) ≠ (2) *** |
| | (0.0003) | (0.0004) | (0.0004) | (1) ≠ (3) *** |
| within R-squared | 0.0194 | 0.0324 | 0.0209 | |
| ***Panel 2*** | | | | |
| GHQ36 t-1 | 0.0013 *** | 0.0020 *** | 0.0013 *** | |
| | (0.0005) | (0.0006) | (0.0004) | |
| PCS t-1 | 0.0014 *** | -0.0003 | 0.0001 | (1) ≠ (2) *** |
| | (0.0003) | (0.0004) | (0.0004) | (1) ≠ (3) *** |
| within R-squared | 0.0197 | 0.0332 | 0.0213 | |
| ***Panel 3*** | | | | |
| MCS t-1 | 0.0015 *** | 0.0010 *** | 0.0007 *** | |
| | (0.0003) | (0.0004) | (0.0003) | |
| PCS t-1 | 0.0018 *** | 0.0001 | 0.0004 | (1) ≠ (2) *** |
| | (0.0003) | (0.0004) | (0.0004) | (1) ≠ (3) *** |
| within R-squared | 0.0214 | 0.0328 | 0.0210 | |

* p<0.10 ** p<0.05 *** p<0.01. 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 4 model (4). There are 29,251 observations in the bottom PCS tercile, 29,898 in the middle PCS tercile, and 28,994 in the top tercile.

***Table 10. LPM coefficient estimates by relative HH poverty.***

| | (1) above poverty line | (2) below poverty line | differences |
|---|---|---|---|
| ***Panel 1*** | | | |
| GHQ12D t-1 | -0.0091 *** | -0.0221 *** | (1) ≠ (2) ** |
| | (0.0029) | (0.0060) | |
| PCS t-1 | 0.0004 ** | 0.0017 *** | (1) ≠ (2) *** |
| | (0.0002) | (0.0003) | |
| within R-squared | 0.0116 | 0.0614 | |
| ***Panel 2*** | | | |
| GHQ36 t-1 | 0.0014 *** | 0.0023 *** | |
| | (0.0003) | (0.0005) | |
| PCS t-1 | 0.0005 ** | 0.0017 *** | (1) ≠ (2) *** |
| | (0.0002) | (0.0003) | |
| within R-squared | 0.0123 | 0.0620 | |
| ***Panel 3*** | | | |
| MCS t-1 | 0.0010 *** | 0.0015 *** | |
| | (0.0002) | (0.0003) | |
| PCS t-1 | 0.0008 *** | 0.0022 *** | (1) ≠ (2) *** |
| | (0.0002) | (0.0004) | |
| within R-squared | 0.0125 | 0.0622 | |

* p<0.10 ** p<0.05 *** p<0.01. 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 4 model (4). There are 64,188 observations above the poverty line and 23,954 below. The poverty line is 60% of the median net equivalised HH income (before housing costs) in the UK 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.

***Table 11. Oster bounds for split sample FE models with full controls.***

| | $\delta = 0$ ($\tilde{\beta}$) | $\delta = 1$ ($\beta^*$) $R_{max} = 1.3\tilde{R}$ | $\delta = 1$ ($\beta^*$) $R_{max} = 2.2\tilde{R}$ |
|---|---|---|---|
| ***Panel 1 (GHQ12D t-1 coefficients)*** | | | |
| W/o degree | -0.0205 | -0.0204 | -0.0203 |
| | (0.0041) | [0.0045] | [0.0046] |
| With degree | -0.0063 | -0.0062 | -0.0060 |
| | 0.0036 | [0.0038] | [0.0039] |
| Above poverty line | -0.0091 | -0.0091 | -0.0092 |
| | (0.0029) | [0.0031] | [0.0032] |
| Below poverty line | -0.0221 | -0.0212 | -0.0183 |
| | (0.0060) | [0.0065] | [0.0070] |
| ***Panel 2 (GHQ36 t-1 coefficients)*** | | | |
| W/o degree | 0.0023 | 0.0023 | 0.0023 |
| | (0.0004) | [0.0004] | [0.0004] |
| With degree | 0.0011 | 0.0011 | 0.0011 |
| | (0.0003) | [0.0003] | [0.0003] |
| Above poverty line | 0.0014 | 0.0014 | 0.0014 |
| | (0.0003) | [0.0003] | [0.0003] |
| Below poverty line | 0.0023 | 0.0022 | 0.0020 |
| | (0.0005) | [0.0005] | [0.0006] |
| ***Panel 3 (MCS t-1 coefficients)*** | | | |
| W/o degree | 0.0014 | 0.0016 | **0.0021** |
| | (0.0002) | [0.0002] | [0.0003] |
| With degree | 0.0009 | 0.0010 | **0.0014** |
| | (0.0002) | [0.0002] | [0.0003] |
| Above poverty line | 0.0010 | 0.0011 | **0.0014** |
| | (0.0002) | [0.0002] | [0.0003] |
| Below poverty line | 0.0015 | 0.0017 | **0.0025** |
| | (0.0003) | [0.0003] | [0.0004] |

Bootstrapped standard errors in square brackets (1000 reps). Clustered standard errors in parentheses. Bounds in bold are outside the 95% CI of the coefficient in the controlled regression.

***Table 12. Robustness checks.***

| | (1) | (2) | (3) |
|---|---|---|---|
| ***Panel 1*** | | | |
| GHQ12D 2/3 t-1 | -0.0136 *** | | |
| | (0.0025) | | |
| GHQ12D 4/5 t-1 | | -0.0141 *** | |
| | | (0.0030) | |
| GHQ12D 5/6 t-1 | | | -0.0169 *** |
| | | | (0.0033) |
| PCS t-1 | 0.0008 *** | 0.0008 *** | 0.0008 *** |
| | (0.0002) | (0.0002) | (0.0002) |
| observations | 88143 | 88143 | 88143 |
| within R-squared | 0.0215 | 0.0214 | 0.0215 |
| ***Panel 2*** | | | |
| GHQ12D t-1 | -0.0126 *** | | |
| | (0.0027) | | |
| GHQ36 t-1 | | 0.0016 *** | |
| | | (0.0002) | |
| MCS t-1 | | | 0.0009 *** |
| | | | (0.0001) |
| 1-2 ADL t-1 | -0.0111 ** | -0.0103 ** | -0.0116 *** |
| | (0.0045) | (0.0045) | (0.0045) |
| 3-4 ADL t-1 | -0.0510 *** | -0.0491 *** | -0.0515 *** |
| | (0.0092) | (0.0092) | (0.0092) |
| 5+ ADL t-1 | -0.0830 *** | -0.0803 *** | -0.0833 *** |
| | (0.0146) | (0.0145) | (0.0145) |
| observations | 88099 | 88099 | 88099 |
| within R-squared | 0.0222 | 0.0228 | 0.0226 |

* p<0.10 ** p<0.05 *** p<0.01. 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 4 model (4).