HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

THE UNIVERSITY of York

WP 20/12

Mental health and employment: a bounding approach using panel data

Mark Bryan; Nigel Rice; Jennifer Roberts and

Cristina Sechel

June 2020

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

Mental health and employment: a bounding approach using panel data

Mark Bryan[a], Nigel Rice[b,*], Jennifer Roberts[a], Cristina Sechel[a]

[a] *Department of Economics, University of Sheffield, UK*
[b] *Centre for Health Economics and Department of Economics and Related Studies, University of York, UK*

Abstract

The disability employment gap is an issue of concern in most Western developed economies. This paper provides important empirical evidence on the influence of mental health on the probability of being in employment for prime age workers. We use longitudinal data and recently developed techniques, which use selection on observable characteristics to provide information on selection along unobservable factors, to estimate an unbiased effect of changes in mental health. 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 results indicate that transitioning into poor mental health leads to a reduction of 1.6 percentage points in the probability of employment. This is approximately 10 per cent of the raw employment gap. This effect 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. 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.

Keywords: Mental health, employment, fixed effects, Oster bounds, UKHLS.

JEL: I12, J14, J24.

*Corresponding author

Email addresses: m.l.bryan@sheffield.ac.uk (Mark Bryan), nigel.rice@york.ac.uk (Nigel Rice), j.r.roberts@sheffield.ac.uk (Jennifer Roberts), c.sechel@sheffield.ac.uk (Cristina Sechel)

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 most countries people with health problems have a much lower employment rate than the rest of the population.[1] In the UK, every year 300,000 people stop work and become reliant on health-related benefits, costing the government £13bn and employers another £9bn (Black & Frost 2011). Recent work by Jones et al. (2020) shows that acute health shocks substantially increase the probability of exiting the labour market and reduce hours and earnings. Adverse mental health (MH) seems to be particularly pernicious in its labour market effects. The employment rate for people with a MH problem 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 UK disability insurance claims (McInnes 2012). MH is neglected in terms of health spending, and often hidden in the workplace due to stigma and discrimination (WHO 2013). Internationally, the World Health Organization (WHO 2008) estimate that MH disorders comprise around 13% of the global burden of disease; and the OECD estimates that MH problems affect more than one in six people across the 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 & 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

[1] https://ilostat.ilo.org/2019/12/03/how-do-people-with-disabilities-fare-in-the-labour-market

an individual); and endogenous selection (since unobserved characteristics and circumstances that affect health outcomes are also likely to be related to labour market outcomes). Our study focuses on the latter problem, but we also employ methods that aim to reduce the biases arising from the first two issues.

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

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

et al. (2005, 2011), who use selection on observable characteristics to provide information on selection along unobservable factors; and in particular Oster (2013*b*, 2019) who extends and generalises this method to enable the estimation of an unbiased treatment effect in the presence of unobserved confounders.

We make two key contributions to the literature. Firstly, we fill a number of important gaps in the evidence base by providing quantitative estimates of the effect of MH on the employment of prime age adults. This is important evidence for social and economic policy across all countries. The vast majority of existing evidence on the relationship between health and work considers either physical health, or general measures of overall self-assessed health (see Ghatak (2010) for a review). In contrast, we use measures of MH derived from two psychometric instruments; the General Health Questionnaire and the Short Form-12 health survey. These measures are good proxies for the true MH stock; they are designed to provide information on all aspects of MH, and are less likely to suffer from the reporting biases that are present in simple overall evaluative measures (Bound 1991, Bound et al. 1999, Lindeboom & Kerkhofs 2009). To date almost all of the existing evidence comes from the US; a country that has very different health and welfare systems to many other countries; and in particular to the universal health care coverage of the UK National Health Service. Our estimates for England and Wales contribute to a very small pool of UK evidence, and 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). In addition, much of the evidence on the impact of health on labour market outcomes is for older workers, since this is where the burden of most physical ill-health is felt. In contrast, MH disorders are particularly prevalent in prime age workers (Kessler et al. 2005), so evidence is needed for this key group. As well as estimating average effects for our sample of prime age individuals, we also explore how both the health-employment relationship, and any bias in the estimates, varies across a number of sub-groups differentiated by sex, age, ed-

with the existing literature, we reserve the term LATE for explicit IV methods.

ucation, physical health and household income. These results will also make a valuable contribution to the economic analysis of the cost-effectiveness of health care interventions that are expected to have important labour market effects[3]; for example, the Improving Access to Psychological Therapies initiative that has been rolled out in England and Wales from 2008 to help people who suffer from anxiety and depression.

Secondly, the vast majority of evidence comes from cross-sectional studies and our longitudinal analysis is more useful for policy formulation, because we can control for individual unobserved factors that confound the relationship between employment and health. Further, we explore any remaining biases that are not removed by the inclusion of FE, by employing Oster's method (Oster 2013*b*, 2019) to deal with unobservable selection. Ours is the first study to use this method with individual longitudinal data incorporating FE.[4] We estimate the bias that arises from omitting important influences on both health and employment in a FE framework that has no exclusion restrictions. We also calculate a consistent estimate of the biased-adjusted treatment effect, under certain assumptions. We discuss the interpretation of the FE treatment effect and contrast this with the narrow LATEs that are often estimated from instrumental variable studies. This application will be a useful resource for practitioners who may wish to use the method in other contexts; and for the policy community who wish to judge the quality of evidence from econometric studies.

Our results show that while there is strong evidence of cross-sectional selection in pooled OLS estimates of the effect of MH on employment, there is little or no additional selection bias once FE are included. Even under weak assumptions, we cannot reject that the bias-corrected estimates are the same as

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

[4]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.

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 evidence that MH has larger effects on employment for those without higher education and those who are in poverty. The paper is structured as follows. In Section 2 we explore the background to the FE models, explaining the estimation problems they are designed to solve and reviewing some of the key evidence. Section 3 describes our estimation method, and the data and variables are described in Section 4. The results and sub-group analyses are presented in Section 5. Section 6 includes the discussion and conclusion.

2. Background

It is well known that MH and work are related and that the relationship between them is complex (see for example, Currie & Madrian (1999), Frijters et al. (2014), Steele et al. (2013)). However, there is very little quantitative evidence available on the effect of MH disorders on work. In particular, there is virtually no evidence for the UK, and in the 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 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. Grossman's health investment model (Grossman 1972) 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, espe-

cially 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. Commonly used instruments 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 & French 2001, Chatterji et al. 2007); and perceived social support (Alexandre & 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, p. 859) point out, the chosen instruments are often "hard to justify based on economic theory". Indeed, in their own study, Chatterji et al. (2007) admit that it is difficult to make a strong case for the exogeneity of their instrument; childhood psychiatric disorders, for example, can be argued to be underlying individual traits that can manifest later in life. Further, Chatterji et al. (2011)

[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 & Sindelar (1996), Terza (2002)).

use the methods proposed by Altonji et al. (2005) to show the sensitivity of IV estimates to the extent of unobserved selection bias, and recommend that longitudinal data be used to explore selection based on unobserved personal characteristics.

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) 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] 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.

[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).

(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 & 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, 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; further, it is difficult to state the direction of this bias with any certainty.

In this paper we investigate the direction and magnitude of the potential bias in FE equations using information on selection on unobservables. Oster (2013*b*, 2019) is critical of the intuitive argument, 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.[7] This rests on the assumption that bias arising from observed controls is informative of the bias arising from omitted unobserved factors. But this is not sufficient because it is also necessary to know how much of the variance in the outcome is explained by the inclusion of the controls. The approach we use 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; so one might then conclude that any remaining bias is in the same direction, although there is no theoretical reason why this should be the case (Peng et al. 2015).

3. Estimation approach

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

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

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

where N is the total number of individuals and T is the total number of waves; M_{it-1} is a measure of the MH of individual i in wave $t-1$; Z_{it} is a vector of observed controls, including (time-varying) individual, household and area characteristics, with associated parameter vector θ; d_t is a vector of $T-1$ wave dummy variables that control for time effects; μ_i are individual FE; and ϵ_{it} is the idiosyncratic error term. The parameter of interest is β, the effect of MH in the previous wave on the probability of being employed in the current wave. The LPM allows us to control for individual-specific effects that are correlated with the covariates, and is often the preferred choice to model health and work with FE (Bubonya et al. 2017, Greve & Nielsen 2013). It is also used in the wider literature to model binary labour market outcomes. For example, Agüero & Marks (2008) use a LPM to investigate the relationship between children and female labour force participation in Latin America; Francesconi & 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.[8] However, it is reasonable to assume that MH changes will take some time to feed through to labour market outcomes.

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

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

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

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

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

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

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

where $\ddot{\nu}_i = \nu_{it} - \bar{\nu}_i$, and ν_{it} is the error term from the FE model $Y_{it} = \alpha + \beta M_{it-1} + d_t\gamma + \mu_i + \nu_{it}$. $\bar{\nu}_i$ is the within-individual mean of ν_{it}. The demeaned time dummies, $\ddot{d}_t$, are included in Equation (3) because they capture time trends

that may be correlated with M. As these time trends are fully captured by $\ddot{d}_t$, any change in the coefficient on M when they are added does not tell us what would happen if further time varying controls were added.[9] In contrast, the covariates $\ddot{Z}_{it}$ in Equation (2) are assumed to imperfectly capture the relevant time-varying factors that influence the relationship between M and Y; thus there are unobserved counterparts to $\ddot{Z}_{it}$.

Two key parameters specify the relationship between observable and unobservable selection and the maximum amount of variation that can be explained by the model. The first parameter, δ, defines the importance of the unobservables relative to the observables in influencing M. When $\delta = 1$ the observables and the unobservables are equally important and affect β in the same direction; when $0 < \delta < 1$ the unobserved factors are less important than the observed factors (and the opposite holds when $\delta > 1$).[10] The second parameter, R_{max}, is the (theoretical) maximum R-squared from the full model where all observed and unobserved variables are included. This can be as high as 1 if Y is measured without error, but cannot be smaller than the R-squared obtained from the controlled regression. Both δ and R_{max} are unknown parameters to be chosen given the particular context of the problem. It is generally argued that an appropriate upper limit for δ is 1 because the observed variables are usually chosen based on the fact that they are the most important controls (based on theory and/or previous empirical evidence). The range 0 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. It is reasonable to assume that R_{max} is less than 1 if idiosyncratic measurement error in Y exists. It also seems appropriate

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

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

to assume $R_{max} < 1$ when modelling a discrete employment outcome using a linear equation.[11] In our analysis we consider a range of values suggested by Oster's (Oster 2019) empirical survey of randomised studies, $R_{max} = 1.3\tilde{R}$ and $R_{max} = 2.2\tilde{R}$; where $\tilde{R}$ is the R-squared value from equation (2). While the assumptions around δ and R_{max} are not testable in our analysis we apply the method essentially as a robustness check on our results rather than as a method to correct for any bias that may arise. This allows us to compute a bounding set Δ_s with the following bounds on β: (i) $\tilde{\beta}$ which is the estimate of β in the controlled regression (Equation (2)), and (ii) β^* which is the effect of MH on employment corrected for omitted variable bias given the specified values of R_{max} and δ.[12] Whether $\tilde{\beta}$ is the upper or lower bound of Δ_s will depend on the direction of the MH effect and the direction of the bias. For a positive MH effect, $\tilde{\beta}$ is a lower bound in the presence of downward bias, and an upper bound in the presence of upward bias. The opposite is true if the MH effect is negative.

4. Data

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

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

[12] These bounds can be estimated using Stata (Oster 2013*a*).

time)[13]; 0 if the individual is unemployed, retired, looking after family/home, or long-term sick/disabled.

We use three alternative measures of MH for our key explanatory variable; 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 & 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 & Oswald (2007), Roberts et al. (2011)); including studies of the relationship between MH and work (see for example, Garcia-Gomez et al. (2010), Mavridis (2015)). Our primary measure of MH status is a binary indicator that identifies individuals with a possible psychiatric disorder. This measure is derived from the GHQ-12 caseness score. The original GHQ scale permits responses of 0 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 to monitor the percentage of people who suffer from poor MH in the general population.[14] 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) where a higher value corresponds to worse MH.

Our third measure of MH is the Mental Component Summary (MCS) derived from the SF-12. The SF-12 is a multidimensional generic measure of health-

[13] 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.

[14] 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.

related quality of life that is widely used in clinical trials and routine outcome assessment because of it brevity and psychometric performance.[15] 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 original score ranges from 0 to 100 where higher values denote better MH and the scoring method is based on an algorithm developed by Ware et al. (2002); this uses population norm based scoring so that the measure has a mean of 50 and a standard deviation of 10. For consistency with our other two MH measures, we recode the MCS so that higher values denote worse MH. The MCS has been used to analyse the MH effects of learning intensity (Hofmann & Mühlenweg 2019), working-time mismatch (Otterbach et al. 2016), and work schedules of sole-parents (Dockery et al. 2016). Mitra & 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 & 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 & Pepper 2007). In addition, our measures

[15] 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).

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.[16]

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[17], marital status, highest level of education achieved, presence of children in household (by age groups), number of adults in household, and other household income. We also control for the physical health (PH) of the individual using the SF-12 Physical Component Summary (PCS); this is the PH equivalent of the MCS, with the score ranging from 0 to 100 where higher values denote worse health (Ware et al. 2002). In some specifications we also allow for comorbidity between MH and PH by including an interaction term between the two measures.[18] 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.

[16] In contrast, Frank & 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.

[17] 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).

[18] 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.

Table 1: Summary statistics

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

Tables 1 and 2 show descriptive statistics for our estimation sample split by the dichotomous GHQ measure of MH. In total, there are 98,435 observations covering 11,683 men and 14,851 women.[19] Approximately a fifth are identified as having poor MH (GHQ12D=1) and these respondents accordingly have higher GHQ36 and MCS scores; they also have worse PH as shown by the PCS scores and problems with ADL. They are also less likely to be employed (70% employed vs. 87% 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

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

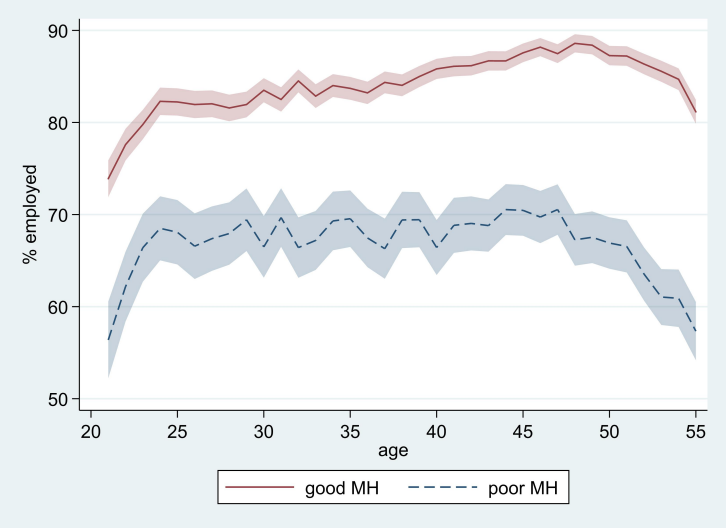
Table 2: Observations by employment status (all waves pooled)

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

household income is lower in the households of people with poor MH; and they also live in areas with a higher unemployment rate and lower GVA. Table 2 shows that among the non-employed, a similar proportion of those with a MH problem are unemployed compared to those without a problem (31.1% and 29.2% respectively), but the largest group of those with poor MH are long-term sick or disabled (34.9%), while the majority of those who do not have poor MH are involved in family/home care (54.1%). Out of the total 26,534 respondents, 11.7% change their employment status over the period of analysis, and 30.8% change their MH status as measured by the dichotomous GHQ variable (not shown in tables). Of those who change employment status, most experience only one transition (4.6% move from non-employment to employment and 3.2% move from employment to non-employment), while only 3.9% change employment status multiple times. In contrast, a large proportion of respondents experience multiple changes in MH status (17.1%), while 6.7% change MH once from good to poor and 7.1% change once from poor to good. Figure 1 plots the employment gap by age between individuals with good and poor MH (as measured by the GHQ12D binary indicator) in all 9 waves of our UKHLS sample; this gap is substantial and appears to widen with age.[20]

[20]Note that the employment gap (of around 15 percentage points) in our data is narrower

Figure 1: Employment by age and MH (95% confidence intervals)



Source: UKHLS Waves 1-9 (UKHLS, 2019)

5. Results

Table 3 contains point estimates for models using the GHQ12D dichotomous measure for MH. For comparison purposes, the first two columns show the results when the LPM is estimated without FE by applying OLS to the raw untransformed data (we refer to these as pooled OLS models). We report estimates without controls (column 1) and with controls, $\ddot{Z}$, (column 2).[21] The next two columns show the estimates for the LPM with FE, obtained using OLS on the demeaned data (we will refer to these as FE models): from the uncontrolled regression, Equation (3), in column 3 and from the full controlled model, Equation (2), in column 4. We have also estimated these models for each gender separately, but we find no significant differential gender effect on employment.[22] However, we find significant gender differences for being mar-

than that commonly cited in recent reports, and this is largely due to our focus on prime age workers.

[21] All equations include wave dummy variables.

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

ried, having children aged 0-4, 5-11 and 12-15. We therefore include in our controls gender interactions with being married and all the children variables.[23]

The pooled OLS coefficient in column 1 shows that poor MH is associated with a 15.5 percentage point lower probability of employment (controlling only for wave dummies). When the main controls are added, the absolute size of the effect is reduced to -9.7 percentage points, and when we include FE in the specification, the effect falls still further to -1.6 percentage points. There is thus quite strong selection into 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 (which we investigate formally below) that once cross-sectional selection is removed, there is little remaining time-varying selection bias. In the preferred specification (column 4), having poor MH lowers the probability of being employed by approximately 1.6 percentage points, which suggests that the causal effect of MH accounts for about 10% of the raw employment gap (column 1).[24]

The control variables in column 4 all appear to have the expected effects on the employment probability. Poor PH is associated with reduced employment, and while the effect is much smaller in the FE model compared to the pooled OLS coefficient in column 2, it is still statistically significant. In the FE model, the effect of age is significantly larger for all age groups compared 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

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

[24] The effect is very similar in separate gender regressions (-1.60 percentage points for women and -1.62 for men), and as mentioned previously, this difference is not statistically significant. We also considered asymmetric effects by including separate MH variables for positive differences from the mean and for negative differences from the mean (an approach similar to Allison (2019) who uses first differences instead of demeaning). We find no evidence of asymmetric effects for our two GHQ measures (i.e. the effect of a positive change in MH is not statistically different from the opposite effect of a negative MH change) but the effect of a positive change in MH is significantly larger in magnitude than the opposite effect of a negative change in MH for our MCS measure.

Table 3: LPM: MH = GHQ caseness indicators

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

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

Table 4: LPM: alternative MH measures

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

Where $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors in parentheses (clustered at individual level). Standardized MH coefficients in brackets. All models include wave dummies. Sample size (NT): 98,435. Regressions (1) and (2) are LPM without individual fixed-effects using OLS on the untransformed data; regressions (3) and (4) are LPM estimated using OLS on the demeaned data.

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 or children aged 5-11, and a larger positive effect of having secondary school aged children (aged 12-15). The education gradient is as expected; those who gained qualifications, especially A-levels or a degree, have a higher likelihood of being employed than those with no formal qualifications. A higher number of adults living in the household also increases the probability of being employed, while a higher level of other household income lowers it. Neither of the area level controls (unemployment rate and GVA per head) are statistically significant, which may reflect the fact that although they vary a lot spatially, they exhibit only limited variation over time.

In Table 4 we re-estimate the specifications from Table 3 using the continuous GHQ36 measure and the SF12 MCS; for both of these measures, higher values represent worse MH. As expected, we find a negative relationship between these measures and the probability of being employed.[25] Again, in the pooled OLS regressions the addition of control variables reduces the magnitude of the estimated effect of MH for both measures, and the inclusion of FE reduces both estimates still further. In the FE models, the addition of controls increases the magnitude of both the GHQ36 and the MCS coefficients slightly. This is consistent with the GHQ12D models from Table 3 and suggests that, contrary to the cross-sectional selection effects, whereby characteristics that are positively associated with employment are also positively related to MH, the opposite is true for time-varying selection - that is, changes in characteristics which increase the probability of employment also lead to reduced MH. Again, we investigate this formally below.

In Table 4 we also report the standardised coefficients on GHQ36 and MCS (reported in brackets). Our preferred specification (column 4) suggests that a one standard deviation increase in GHQ36 (MCS) leads to a 1.0 (1.3) percentage point decrease in the probability of employment. In the MCS model, the MH and PH measures are directly comparable since they both use SF-12 summary scores; the results suggest that changes in physical health have a slightly larger effect on employment compared to mental health.

Table 5 reports the bounds of the value of β from the FE models with full controls. For ease of comparison, the first column repeats the estimates from the controlled regression in Equation (2) (i.e. $\tilde{\beta}$ when $\delta = 0$). The bias adjusted estimates (β^*) under the assumption that $\delta = 1$ are shown in column 2 (setting $R_{max} = 1.3\tilde{R}$) and column 3 (setting $R_{max} = 2.2\tilde{R}$) with bootstrapped standard errors in brackets. We find that for the GHQ36 measure the bias adjusted estimates are the same as the FE estimates ($\tilde{\beta}$) under both R_{max} assumptions.

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

Table 5: Oster bounds for FE models with full controls

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

Bootstrapped standard errors in square brackets (1000 reps). Clustered standard errors in parentheses. Estimates in bold denote non-overlapping 95% confidence intervals between the bounds and original estimates.

For GHQ12D and MCS, as was suggested by the coefficient changes noted above, some bias is exhibited; $\tilde{\beta}$ is the lower bound (in terms of magnitude) and β^* the upper bound. However, this result should be interpreted with caution as these upper bounds are very close to the estimated coefficient. The bounds fall within the 95% confidence intervals of $\tilde{\beta}$ in all cases with the exception of MCS when $R_{max} = 2.2\tilde{R}$.[26] This suggests there is little concern regarding omitted variable bias in these FE models for the effect of MH on employment.

5.1. Heterogeneity Analysis

In order to explore the heterogeneity of effects in different subgroups we focus on the preferred model (Equation (2); column 4 in Tables 3 and 4) and split the sample by education, PH terciles and relative poverty. Respective results are reported in the Appendix as Tables A2 to A4. Differences across groups were tested for statistical significance and this is noted in the rightmost column where relevant. The relationship between the GHQ12 dummy and employment remains negative and statistically significant. The effect of MH is significantly

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

Table 6: Oster bounds for split sample FE models with full controls

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

Bootstrapped standard errors in square brackets (1000 reps). Clustered standard errors in parentheses. Estimates in bold denote non-overlapping 95% confidence intervals between the bounds and original estimates.

smaller in magnitude across all three MH measures for those with a degree than for those without a degree (Table A2), and significantly smaller for those households living above the relative poverty line than for those below (Table A4). This suggests that higher education moderates the effect of MH disorders on employment, while relative poverty exacerbates it. There is also some evidence that the effect of MH is larger for individuals with worse PH (Table A3), but

this difference is only statistically significant for MCS (comparing those in the top PCS tercile to those in the bottom PCS tercile).

We note that the effects of PH are more heterogeneous across levels of PH. They are larger for those in the bottom PCS tercile compared to those in the middle and top terciles across all models (Table A3). Like MH, the effect of PH is also larger for those below the poverty line compared to those above the poverty line (Table A4). However, unlike MH, the effects of PH are not statistically different between those with a degree and those without in all three models (Table A2). It is also worth noting that, in addition to the results presented here, we also explored differences across sub-groups defined by age, 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 6 presents Oster bound estimates for select split sample FE models with full controls.[27] Similar to the pooled results, the bias adjusted effects of MH (β^*) on employment have the same sign, and are close in magnitude to the estimated coefficients from the controlled FE regressions with overlapping 95% confidence intervals. However, the bias adjusted effect of MH on employment when $R_{max} = 2.2\tilde{R}$ is outside of the 95% confidence interval of $\tilde{\beta}$ in the MCS model for each of the subsamples.

5.2. Sensitivity Analysis

We run a number of sensitivity checks. Our results are robust to different GHQ cut-offs for our binary GHQ12D. For the pooled FE model we consider one lower cut-off at 2/3 and two higher cut-offs at 4/5 and 5/6. Compared to the coefficient in our benchmark model from Table 3 (-0.0160), the effect of GHQ12D is smaller for the cut-offs at 2/3 and 4/5, and larger for cut-off at 5/6 (see Panel 1 in Table 7). We also consider an alternative measure of PH based on the Activities of Daily Living (ADL) questions. We classify individuals into

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

Table 7: Robustness checks (FE models)

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

Where $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. 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 3 model (4).

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 7).

The effect of MH may depend on the nature of employment, particularly on whether the individual is self-employed or not. The self-employed are likely

to have a differential degree of autonomy and control at work, which can lead to different effects of MH on employment compared to those employed.[28] 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)[29] in the neighbourhood where the household is located and local labour market tightness.[30] 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

[28] 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.

[29] 2015 IMD data (for England only) obtained from the Ministry of Housing, Communities & Local Government.

[30] Labour market conditions may moderate the relationship between MH and employment status (see for example, Houssemand & 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.

contrast, IV methods deliver a LATE that typically applies to a narrow subgroup only, for instance those who have suffered a bereavement.

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 by assuming that adding the missing controls would more than double the explained longitudinal variance. Even under this fairly extreme assumption, we cannot reject that the bounds are the same as the FE estimates. The results indicate that selection into MH is almost entirely based on time-invariant characteristics and so we conclude that FE estimates of the effect of MH on employment are unbiased. There is certainly no evidence of upward bias in the size of the MH effect, as may be expected from the intuition that changing circumstances that favour work also favour MH. A caveat to our results is that while we try to 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 1.6 percentage points in the probability of employment (which makes up about 10% of the raw MH employment gap), and that a one standard deviation change in the continuous measures

of MH causes a 1.0-1.3 percentage point change in the probability of employment. Comparisons of these effects with previous studies are not straightforward because of differences in the MH measures used and the way the effects are reported; 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 et al. 2015); 0.0-8.2 (US; MCS measure; Mitra & 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 2.5 percentage points for people without degree, compared with just 0.5 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 (15 percentage points) is much larger that the effect of changing between MH states (1.6 percentage points), improving the MH of those with conditions amenable to treatment may only have a small direct effect on closing the MH employment gap. As well as chronic health problems, much of the raw gap is also due to differences in other factors, such as income and educational attainment. Longer-term structural changes, which impact on all of these factors, will almost certainly be required to eliminate the gap completely.

Acknowledgements

We are grateful to the Health Foundation for funding this research under their Social and Economic Value of Health programme, Award No. 751630. The UKHLS data (University of Essex, 2018) 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 Andrew Jones, Maarten Lindeboom and Arne Risa Hole for very useful comments on earlier versions of this paper.

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.

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

Altonji, J., Conley, T., Elder, T., & Taber, C. (2011), Methods for using selection on observed variables to address selection on unobserved variables, Technical report, 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 multiple clinical indicators', *Health Economics* **26**(2), 184–205.

Black, B., de Carvalho, A. G., Khanna, V. & Kim, W. e. a. (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', *The Stationery Office* **66**(1), 1–2.

Black, C. & Frost, D. (2011), 'Health at work - an independent review of sickness absence', *The Stationery Office* **Cm 8205**.

Blundell, R., Britton, J., Costa Dias, M. & French, E. (2017), The impact of health on labour supply near retirement, Technical Report No. W17/18, Institute for Fiscal Studies Working Paper.

Bound, J. (1991), 'Self-reported versus objective measures of health in retirement models', *Journal of Human Resources* **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, Technical Report No. 10653, IZA Discussion Paper.

Cattan, S., Kamhofer, D. A., Karlsson, M. & Nilsson, T. (2017), The short and long-term effects of student absence: Evidence from Sweden, Technical Report W17/21, Institute for Fiscal Studies Working Paper.

Chatterji, P., Alegria, M. & Takeuchi, D. (2007), 'Psychiatric disorders and labor market outcomes: Evidence from the National Latino and Asian American Study', *Health Economics* **16**(10), 1069–90.

Chatterji, P., Alegria, M. & Takeuchi, D. (2008), Psychiatric disorders and labour market outcomes: Evidence from the National Comorbidity Survey - replication, Technical Report Working Paper 14404, National Bureau of Economic Research (NBER).

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 labor market, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', 1 edn, Vol. 3, Part C, Elsevier, chapter 50, pp. 3309–3416.

DeSimone, J. (2002), 'Illegal drug use and employment', *Journal of Labor Economics* **20**(4), 952–77.

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 (2017), Improving lives: The future of work, health and disability, Technical Report Policy Paper, Department for Work and Pensions and Department of Health.

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), 154–64.

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**(11), 49–60.

Ghatak, A. (2010), Health, labour supply and wages: A critical review of the literature., Technical Report Working Paper 244, Institute for Social and Economic Change.

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., Ustˋun, 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. (2019), '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.

Jones, A. M., Rice, N. & Zantomio, F. (2020), 'Acute health schocks and labour market outcomes: Evidence from the post crash era', *Economics and Human Biology* **36**, 100811.

Kessler, R. C., Berglund, P. A., Demler, O., Jin, R., Merikangas, K. R. & Walterset, E. E. (2005), 'Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication (NCS-R)', *Archives of General Psychiatry* **62**(6), 593–602.

Kreider, B. & Pepper, J. V. (2007), 'Disability and employment: Re-evaluating the evidence in light of reporting errors', *Journal of the American Statistical Association* **102**(478), 432–441.

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

Marcotte, D. E., Wilcox-Gök, V. & Redmon, D. P. (2000), *The Labour Market Effects of Mental Illness: The Case of Effective Disorders*, JAI Press, Greenwich, CT; pp 181-210.

Mavridis, D. (2015), 'The unhappily unemployed return to work faster', *Journal of Labor Economics* **4**(1), 1–22.

McInnes, R. (2012), ESA and Incapacity Benefit statistics, Technical Report Standard Note SN01420, House of Commons Library.

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), 409–34.

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

Munford, L., Rice, N. & Roberts, J. (2016), Econometric modelling of the employment gap for the NHS Outcomes Framework, Technical report, Report for the Department of Health.
URL: *http://cfe.org.uk/app/uploads/2016/09/Econometric-Modelling-of-the-Employment-Gap_July-16.pdf*

Oakley, M. (2016), Closing the gap, Technical report, Social Market Foundation.

OECD/EU (2018), Health at a glance: Europe 2018: State of health in the EU cycle, Technical report, OECD Publishing.
URL: *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. e. a. (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*a*), PSACALC: Stata module to calculate treatment effects and relative degree of selection under proportional selection of observables and

unobservables. revised 18 Dec 2016 ed., Technical Report S457677, Boston Collage Department of Economics.

Oster, E. (2013*b*), Unobservable selection and coefficient stability: Theory and validation, Technical Report No. 19054, NBER Working Paper.

Oster, E. (2019), 'Unobservable selection and coefficient stability: Theory and evidence', *Journal of Business & Economic Statistics* **37**(2), 187–204.

Otterbach, S., Wooden, M. & Fok, Y. K. (2016), Working-time mismatch and mental health, Technical Report Discussion Paper No. 9818, IZA.

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*, John Wiley & Sons, New York.

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), 703–11.

Terza, J. V. (2002), 'Alcohol abuse and employment: A second look', *Journal of Applied Econometrics* **17**, 393–404.

UKHLS (2019), Understanding society: Waves 1-8 2009-2017 and Harmonised BHPS: Waves 1-18 1991-2009. [data collection], Technical report, University

of Essex. Institute for Social and Economic Research NatCen Social Research, Kantar Public. UK Data Service.
URL: *http://doi.org/10.5255/UKDA-SN-6614-13*

Verbeek, M. (2012), *A Guide to Modern Econometrics*, John Wiley & Sons, Chichester England.

Waddell, G. & Burton, A. K. (2006), Is work good for your health and well-being?, Technical report, The 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), Technical report, Quality Metric Incorporated, Lincoln RI.

WHO (2000), Mental health and work: Impact issues and good practices, Technical report, World Health Organization.

WHO (2008), The global burden of disease: 2004 update, Technical report, World Health Organization.

WHO (2013), Investing in mental health: Evidence for action, Technical report, World Health Organization.

Appendix

Table A1: Variable definitions

Variable	Definition	Source
Dependent variable		
Employed	=1 if self-employed or in paid employment (full- or part-time); 0 if individual is unemployed, retired, looking after family/home, or long-term sick/disabled. We exclude individuals out of the labour force (i.e. full-time students, on maternity leave, on a government training scheme or apprenticeship, untrained workers in family business, and those 'doing something else').	UKHLS
Mental health measures		
GHQ12D	Binary measure of Caseness based on the 0-12 scoring method of the 12-item General Health Questionnaire (=1 if score is 4 or higher, which identifies the possible presence of psychiatric morbidity).	UKHLS
GHQ36	Continuous measure based on the 0-36 Likert scale scoring method of the GHQ (0 represents the least distressed and 36 represents the most distressed).	UKHL
MCS	Mental Component Summary, measured on a 0-100 continuous scale based on the SF-12 questionnaire where 0 denotes high functioning and 100 denotes low functioning.	UKHLS
Individual controls		
PCS	Physical Component Summary, measured on a 0-100 continuous scale based on the SF-12 questionnaire where 0 denotes high functioning and 100 denotes low functioning.	UKHLS
ADL	Individuals are classified into one of four groups based on reported Activities of Daily Living: no ADL problems, 1-2 ADL problems, 3-4 problems, and those with 5 or more.	UKHLS
Age	Age of respondent in years.	UKHLS
Education	Highest level of education achieved at the time of the interview: no educational attainment (baseline), O-level or equivalent, A-level or equivalent, and having a degree or equivalent.	UKHLS
Married	=1 if individual is married, in a registered same-sex civil partnership or living as a couple; 0 otherwise.	UKHLS
Household controls		
No child in HH	=1 if no children 0-15 living in household; 0 otherwise	UKHLS
Child 0-4 in HH	=1 if children 0-4 living in household; 0 otherwise	UKHLS
Child 5-11 in HH	=1 if children 5-11 living in household; 0 otherwise	UKHLS
Child 12-15 in HH	=1 if children 12-15 living in household; 0 otherwise	UKHLS
Adults in HH	Number of adults living in household.	UKHLS
Other HH income	Derived by subtracting own gross monthly labour income from total gross household income in the month before interview (real, adjusted using RPI 2013=100).	UKHLS
Area controls		
Unemployment rate	Unemployment rate in the Local Authority District (LAD) where the household is located.	NOMIS†
GVA	Gross Value Added per head of the LAD where the household is located. Calculated using the balanced approach and the resident population of that region.	ONS

† Annual Population Survey

Table A2: LPM: FE models by education

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

Where $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. 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 3 model (4). There are 53,139 observations (NT) without a degree and 45,296 with a degree.

Table A3: LPM: FE models by PCS tertiles

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

Where $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. 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 3 model (4). There are 33,302 observations (NT) in the bottom PCS tercile, 32,582 in the middle PCS tercile, and 32,551 in the top tercile.

Table A4: LPM: FE models by relative HH poverty

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

Where $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. 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 3 model (4). There are 73,652 observations (NT) above the poverty line and 24,782 below. The poverty line is 60% of the median net equivalised HH income (before housing costs) in the 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.

Supplementary Appendix

Mental health and employment: a bounding approach using panel data

Mark Bryan†, Nigel Rice‡, Jennifer Roberts†, Cristina Sechel†

†Department of Economics, University of Sheffield, UK
‡Centre for Health Economics and Department of Economics and Related Studies, University of York, UK

S1. Introduction

Methods for determining the likely bias in regression estimates of a parameter of interest, β, arising from unobservable characteristics correlated with both a 'treatment' variable of interest, and an outcome have been set out by Oster (2019). The approach is useful in defining bounds to estimated treatment effects in the absence of information on the unobservable confounders. The approach is intended for use in linear regressions with cross-sectional data. This appendix provides further detail on the method and a formal exposition of its application to panel-data. The appendix is intended to be read alongside Oster (2019).

S2. Exposition of the Oster approach applied to panel data

This section sets out a formal exposition of the Oster approach applied to the use of panel data. To aid comparison we adopt a similar notation to Oster (2019). Consider the following true population panel regression model:

$$y_{it} = \alpha + \beta x_{it} + \phi\omega_{it}^{0} + W_{2it} + \zeta z_i + \eta_i + \varepsilon_{it}, \tag{S1}$$

$\forall, i = 1, \ldots, N; t = 1, \ldots, T.$ $\varepsilon_{it} \sim N\left(0, \sigma_{\varepsilon}^{2}\right)$ is the usual idiosyncratic error term. y_{it} is an outcome, x_{it} is a variable of key interest and β the corresponding population parameter, ω_{it}^{0} is a vector of time-varying characteristics observed by the researcher, and z_i is a vector of observed time-invariant characteristics. Assume there is an index of time-varying characteristics, W_{2it}, unobservable to the researcher. In addition, η_i is an unobserved time-invariant individual-specific effect, often assumed to be normally distributed. Interest lies in obtaining unbiased estimates of the population parameter β, which we denote $\hat{\beta}$.

The time-invariant observed and unobserved effects (z, η) can be eliminated by taking first-differences or demeaning equation (S1), such that:

$$\ddot{y}_{it} = \beta \ddot{x}_{it} + \phi \ddot{\omega}^0_{it} + \ddot{W}_{2it} + \ddot{\varepsilon}_{it}, \tag{S2}$$

where, for example, $\ddot{y}_{it}$ is the within-individual mean-transformation: $\ddot{y}_{it} = y_{it} - \bar{y}_{it}$, where $\bar{y}_{it} = \sum_{t=1}^{T} \frac{y_{it}}{T}$ for each individual, i.

Assume, due to the index, $\ddot{W}_{2it}$, being unobserved we cannot estimate equation (S2) and instead estimate the following omitted variables alternative,

$$\ddot{y}_{it} = \beta \ddot{x}_{it} + \phi \ddot{\omega}^0_{it} + \ddot{\varepsilon}_{it}. \tag{S3}$$

We refer to (S3) as the controlled model. Due to omitted variable, $\ddot{W}_{2it}$ being correlated with $\ddot{x}_{it}$, estimation of the controlled model can lead to bias in the estimate of the population parameter, β. The direction and size of this bias is unknown. The approach relies on the intuition that information on the effect of included regressors (here, $\ddot{\omega}^0_{it}$) on the estimate of β can inform the likely bias from omitting unobservable variables, W_{2it}.

Under certain conditions, the bias due to omitting both unobservable, $\ddot{W}_2$, and observable characteristics, $\ddot{\omega}^0$, can be recovered from estimation of model (S3) and the following uncontrolled model,

$$\ddot{y}_{it} = \beta \ddot{x}_{it} + \ddot{\varepsilon}_{it}. \tag{S4}$$

S2.1. Definitions and assumptions

Denote the estimate of β obtained from regression of the controlled model (S3) as $\tilde{\beta}$, and the corresponding estimate from regression of model (S4) as $\breve{\beta}$.

As in Oster, to simplify the exposition, assume that all elements of ω^0 are orthogonal to W_2, and that all j elements of ω^0 are orthogonal to each other. The general ideas extend to situations were these conditions do not hold, as shown by Oster.

S2.2. Bias in estimating the uncontrolled model

Using standard procedures for omitted variable bias we can compute a biased estimate for β from the uncontrolled model (S4). Due to the assumption that elements of $\ddot{\omega}^0$ are orthogonal to $\ddot{W}_2$ there are two components of bias, one for the omitted, but observable regressors, $\ddot{\omega}^0$, and one for the omitted but unobserved index, $\ddot{W}_2$. Orthogonality implies the bias terms are independent. The estimate of β, denoted $\breve{\beta}$, from regression of model (S4) is given by,

$$\breve{\beta} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{x}_{it} \ddot{y}_{it}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{x}_{it}^2}. \tag{S5}$$

Substituting for $\ddot{y}_{it}$ from the population model (S2) and simplifying gives,

$$
\begin{aligned}
\breve{\beta} &= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}\left(\hat{\beta}\ddot{x}_{it}+\hat{\phi}\ddot{\omega}_{it}^{0}+\ddot{W}_{2it}+\hat{\ddot{\varepsilon}}_{it}\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}^{2}} \\
&= \frac{1}{\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}^{2}}\left(\hat{\beta}\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}^{2}+\hat{\phi}\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}\ddot{\omega}_{it}^{0}+\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}\ddot{W}_{2it}+\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}\hat{\ddot{\varepsilon}}_{it}\right) \\
&= \frac{1}{\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}^{2}}\left(\hat{\beta}\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}^{2}+\hat{\phi}\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}\ddot{\omega}_{it}^{0}+\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}\ddot{W}_{2it}\right) \\
&= \hat{\beta}+\hat{\phi}\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}\ddot{\omega}_{it}^{0}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}^{2}}+\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}\ddot{W}_{2it}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}^{2}},
\end{aligned}
\tag{S6}
$$

which can be expressed as,

$$
\breve{\beta} = \hat{\beta}+\hat{\phi}\hat{\theta}+\hat{\varphi}^{u},
$$

where $\hat{\theta}$ is the set of coefficients from regressing each of the elements of $\ddot{\omega}^{0}$ on $\ddot{x}$. Similarly, $\hat{\varphi}^{u}$ is the coefficient estimate from regressing the index $\ddot{W}_{2}$ on $\ddot{x}$. Taking probability limits we have,

$$
\breve{\beta} \xrightarrow{p} \beta+\phi\theta+\varphi^{u}. \tag{S7}
$$

S2.3. Bias in estimating the controlled model

Using a similar approach to the above we can compute an estimate for β, denoted $\tilde{\beta}$, in the controlled model (S3), by the following expression:

$$
\tilde{\beta} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}\ddot{y}_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}^{2}}, \tag{S8}
$$

where $\hat{\ddot{v}}_{it}$ is an estimate of the residual; $\hat{\ddot{v}}_{it}=\ddot{x}_{it}-\hat{\ddot{x}}_{it}$, and where $\hat{\ddot{x}}_{it}$ is estimated from the regression of $\ddot{x}_{it}$ on the other covariates, $\ddot{\omega}_{it}^{0}$, of equation (S3)[31],

$$
\ddot{x}_{it} = \mu_{0}+\mu\ddot{\omega}_{it}^{0}+\ddot{v}_{it}, \tag{S9}
$$

with $\ddot{v}_{it}\sim N\left(0,\sigma_{\ddot{v}}^{2}\right)$. Accordingly, $\hat{\ddot{x}}_{it}=\hat{\mu}_{0}+\hat{\mu}\ddot{\omega}_{it}^{0}$. To compute the bias in $\tilde{\beta}$ substitute $\ddot{y}$ from the true population equation (S2) into equation (S8),

[31] That is, we partial out the effect of $\ddot{\omega}^{0}$ on $\ddot{x}$ by running the regression $\ddot{x}_{it}=\omega_{0}^{0}+\omega_{1}^{0}\ddot{\omega}_{it}^{0}+\ddot{v}_{it}$, and calculating the residual, $\hat{\ddot{v}}_{it}=\ddot{x}_{it}-\hat{\ddot{x}}_{it}$

$$
\begin{aligned}
\tilde{\beta} &= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}\ddot{y}_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}^{2}}, \\
&= \frac{1}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}^{2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}\left(\hat{\beta}\ddot{x}_{it}+\hat{\phi}\ddot{\omega}_{it}^{0}+\ddot{W}_{2it}+\hat{\ddot{\varepsilon}}_{it}\right), \\
&= \frac{1}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}^{2}}\left(\hat{\beta}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}\ddot{x}_{it}+\hat{\phi}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}\ddot{\omega}_{it}^{0}+\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}\ddot{W}_{2it}+\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}\hat{\ddot{\varepsilon}}_{it}\right).
\end{aligned}
\tag{S10}
$$

Note that, $\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}\ddot{\omega}_{it}^{0}=0$. In addition, $\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}\hat{\ddot{\varepsilon}}_{it}=0$. The latter follows from observing that in the regression of equation (S2) by construction $\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}\hat{\ddot{\varepsilon}}_{it}=0$. Therefore,

$$
\begin{aligned}
\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}\hat{\ddot{\varepsilon}}_{it} &= \sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\ddot{x}}_{it}+\hat{\ddot{v}}_{it}\right)\hat{\ddot{\varepsilon}}_{it}=0, \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\mu}_{0}+\hat{\mu}\ddot{\omega}_{it}^{0}+\hat{\ddot{v}}_{it}\right)\hat{\ddot{\varepsilon}}_{it}=0, \\
&= \hat{\mu}_{0}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{\varepsilon}}_{it}+\hat{\mu}\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{\omega}_{it}^{0}\hat{\ddot{\varepsilon}}_{it}+\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}\hat{\ddot{\varepsilon}}_{it}=0, \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}\hat{\ddot{\varepsilon}}_{it}=0.
\end{aligned}
$$

The equalities set out above can be used to simplify equation (S10) as follows,

$$
\tilde{\beta} = \frac{1}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}^{2}}\left(\hat{\beta}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}\ddot{x}_{it}+\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}\ddot{W}_{2it}\right).
$$

Note that $\sum_{i=1}^{N}\sum_{i=1}^{T}\hat{\ddot{v}}_{it}\ddot{x}_{it}=\sum_{i=1}^{N}\sum_{i=1}^{T}\hat{\ddot{v}}_{it}\left(\hat{\ddot{x}}_{it}+\hat{\ddot{v}}_{it}\right)=\hat{\mu}_{0}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}+\hat{\mu}\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{\omega}_{it}^{0}\hat{\ddot{v}}_{it}+\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}\hat{\ddot{v}}_{it}=\sum_{i=1}^{N}\sum_{i=1}^{T}\hat{\ddot{v}}_{it}^{2}$. Thus it follows that:

$$
\begin{aligned}
\tilde{\beta} &= \hat{\beta}+\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}\ddot{W}_{2it}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{v}}_{it}^{2}}, \\
&= \hat{\beta}+\hat{\varphi}^{c}.
\end{aligned}
$$

where $\hat{\varphi}$ is the coefficient on $\ddot{x}$ from the regression of the unobserved index $\ddot{W}_{2}$ onto $\hat{\ddot{v}}$. Expressed in terms of probability limits, we have,

$$\tilde{\beta} \xrightarrow{p} \beta + \varphi^c. \tag{S11}$$

S2.4. Restricted model

To develop consistent estimators for the omitted variable bias arising from the unobservable $\ddot{W}_2$ in equation (S3) and (S4), Oster imposes the following two simplifying assumptions.[32] First,

$$\delta = 1, \text{ such that } \frac{\sigma_{\ddot{W}_1,\ddot{x}}}{\sigma^2_{\ddot{W}_1}} = \frac{\sigma_{\ddot{W}_2,\ddot{x}}}{\sigma^2_{\ddot{W}_2}}. \tag{S12}$$

where the above expression represents the proportionality assumption that the bias that occurs from omitted observable variables ($\ddot{W}_1 = \phi\ddot{\omega}^0$) is the same as that arising from omitted unobservables, $\ddot{W}_2$. δ is the coefficient of proportionality for which the restriction imposed here is relaxed at the end of the appendix to provide an approximation to omitted variable bias.

Second is the assumption,

$$\frac{\phi_k}{\phi_l} = \frac{\mu_k}{\mu_l} \forall k, l. \tag{S13}$$

The above assumes that the relative contribution of each of the coefficients $(\mu_j, j = 1, \ldots, J)$ from a regression of $\ddot{x}$ on observables $\ddot{\omega}^0$, is the same as the respective relative contribution of the coefficients on $\ddot{\omega}^0$, $(\phi_j, j = 1, \ldots, J)$, from a regression of $\ddot{y}$ on $\ddot{x}$ and $\ddot{\omega}^0$. This is useful as Oster (in the Appendix to the paper) shows that under these conditions, the regression of $\ddot{y}$ on $\ddot{x}$ and index $\ddot{W}_1 (= \phi\ddot{\omega}^0)$ yields the same coefficient on $\ddot{x}$ as obtained from a regression of $\ddot{y}$ on $\ddot{x}$ and $\ddot{\omega}^0$.[33] This result allows the simplification of the following limits.

$$\begin{aligned}
\hat{\phi}\hat{\theta} &\xrightarrow{p} \phi \frac{cov\left(\ddot{\omega}^0, \ddot{x}\right)}{var(\ddot{x})} = \phi \frac{cov\left(\ddot{W}_1/\phi, \ddot{x}\right)}{var\left(\ddot{x}\right)} = \frac{\sigma_{\ddot{W}_1,\ddot{x}}}{\sigma^2_{\ddot{x}}}, \\
\hat{\varphi}^u &\xrightarrow{p} \frac{cov\left(\ddot{W}_2, \ddot{x}\right)}{var(\ddot{x})} = \frac{\sigma_{\ddot{W}_2,\ddot{x}}}{\sigma^2_{\ddot{x}}}, \\
\hat{\varphi}^c &\xrightarrow{p} \frac{cov\left(\ddot{W}_2, \ddot{v}\right)}{var(\ddot{v})} = \frac{cov\left(\ddot{W}_2, \ddot{x}\right)}{var(\ddot{v})} = \frac{\sigma_{\ddot{W}_2,\ddot{x}}}{\sigma^2_{\ddot{v}}},
\end{aligned} \tag{S14}$$

where $\ddot{v}$ is the residual from the regression of $\ddot{x}$ on $\ddot{\omega}^0$. The final equality holds, since $\ddot{\omega}$ and $\ddot{W}_2$ are orthogonal; hence $cov\left(\ddot{W}_2, \ddot{v}\right) = cov\left(\ddot{W}_2, \ddot{x}\right)$.

[32] These assumptions are for ease of exposition and can be relaxed without loss of generality.

[33] Oster acknowledges that the condition is unlikely to hold exactly in practice, but assumes where deviations are not large, an approximation to a consistent estimator of the bias is achieved. When $\ddot{\omega}^0$ is a single variable then the condition clearly holds.

Noting that,

$$
\begin{aligned}
var\left(\ddot{v}\right) &= var\left(\ddot{x} - \left(\mu_0 + \mu\ddot{\omega}^0\right)\right) = var\left(\ddot{x} - \left(\mu_0 + \nu\ddot{W}_1\right)\right), \\
&= var\left(\ddot{x}\right) + \nu^2 var\left(\ddot{W}_1\right) - 2\nu cov\left(\ddot{x}, \ddot{W}_1\right), \\
&= var\left(\ddot{x}\right) + \left[\frac{cov\left(\ddot{x}, \ddot{W}_1\right)^2}{var\left(\ddot{W}_1\right)^2} var\left(\ddot{W}_1\right)\right] - 2\left[\frac{cov\left(\ddot{x}, \ddot{W}_1\right)}{var\left(\ddot{W}_1\right)} cov\left(\ddot{x}, \ddot{W}_1\right)\right], \\
&= var\left(\ddot{x}\right) - \frac{cov\left(\ddot{x}, \ddot{W}_1\right)^2}{var\left(\ddot{W}_1\right)},
\end{aligned}
\tag{S15}
$$

then in the limit, $var\left(\ddot{v}\right) \xrightarrow{p} \left(\sigma^2_{\ddot{x}} - \frac{\sigma^2_{\ddot{W}_1,\ddot{x}}}{\sigma^2_{\ddot{W}_1}}\right)$.

Hence, applying the proportionality assumption to the limits of (S14) and simplifying gives,

$$
\begin{aligned}
\hat{\varphi}^u &\xrightarrow{p} \frac{\sigma_{\ddot{W}_2,\ddot{x}}}{\sigma^2_{\ddot{x}}} = \frac{\sigma^2_{\ddot{W}_2}\sigma_{\ddot{W}_1,\ddot{x}}}{\sigma^2_{\ddot{W}_1}\sigma^2_{\ddot{x}}}, \\
\hat{\varphi}^c &\xrightarrow{p} \frac{\sigma_{\ddot{W}_2,\ddot{x}}}{\sigma^2_{\ddot{v}}} = \frac{\sigma^2_{\ddot{W}_2}\sigma_{\ddot{W}_1,\ddot{x}}}{\sigma^2_{\ddot{W}_1}\left(\sigma^2_{\ddot{x}} - \frac{\sigma^2_{\ddot{W}_1,\ddot{x}}}{\sigma^2_{\ddot{W}_1}}\right)}.
\end{aligned}
\tag{S16}
$$

Denoting the latter term in the above as Π (the bias inherent in the controlled model), then we have,

$$
\begin{aligned}
\breve{\beta} &\xrightarrow{p} \beta + \frac{\sigma_{\ddot{W}_1,\ddot{x}}}{\sigma^2_{\ddot{x}}} + \Pi\frac{\left(\sigma^2_{\ddot{x}} - \frac{\sigma^2_{\ddot{W}_1\ddot{x}}}{\sigma^2_{\ddot{W}_1}}\right)}{\sigma^2_{\ddot{x}}}, \\
\tilde{\beta} &\xrightarrow{p} \beta + \Pi.
\end{aligned}
\tag{S17}
$$

Here, the bias arising from omitted variables, $\ddot{W}_1$, and denoted, Π, is the panel data analogous expression for bias in the cross-sectional model.

Oster (2019) shows how the above expressions for bias arising from the controlled and uncontrolled models can be used to derive a bias-adjusted parameter β^* which in the limit approximates β. Defining, $\breve{R}$ as the R-squared from the uncontrolled regression (S4); $\tilde{R}$ as the corresponding R-squared from the controlled regression (S3); and $R_{\max}$ as the (unknown) R-squared from the hypothetical regression (S2), Oster derives the bias-adjusted parameter, β^*, as,

$$\beta^* = \tilde{\beta} - \left[\breve{\beta} - \tilde{\beta}\right] \frac{R_{\max} - \tilde{R}}{\tilde{R} - \breve{R}}. \tag{S18}$$

Relaxing the assumption of equal selection and allowing for proportional selection $\left(\delta \frac{\sigma_{\ddot{W}_1,\ddot{x}}}{\sigma^2_{\ddot{W}_1}} = \frac{\sigma_{\ddot{W}_2,\ddot{x}}}{\sigma^2_{\ddot{W}_2}}\right)$, where δ is the coefficient of proportionality, an approximation to β^* can be derived by extending (S18) to,

$$\beta^* \approx \tilde{\beta} - \delta \left[\breve{\beta} - \tilde{\beta}\right] \frac{R_{\max} - \tilde{R}}{\tilde{R} - \breve{R}}. \tag{S19}$$

The above illustrates that Oster's approach for quantifying omitted variable bias from unobservable characteristics in cross-sectional models can be extended in a straightforward way to a panel data context. Once the model (for example, model (S1)) has been demeaned or first-differenced to remove time-invariant effects, including unobservables, then the approach follows naturally. All other formulae, assumptions and proofs inherent in the approach follow as provided in Oster (2019).

References

Oster, E. (2019) Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics*, 37:2, 187-204, DOI: 10.1080/07350015.2016.1227711.

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