Mental Health and Labour Supply: Evidence from China

David Lander* & Xueying Zhao†

February 22, 2024

Abstract

This paper studies the effect of mental health on employment and labour-force participation in China using nationally-representative China Family Panel Studies (CFPS) data. Utilising CFPS waves 2012, 2016, and 2018, we construct consistent measures of labour supply and mental health, using the Center for Epidemiological Studies Depression (CES-D) scale. The panel structure enables us to control for unobservable time-invariant individual heterogeneity, in contrast to most prior studies from developing countries. Our estimates reveal that experiencing severe depressive symptoms reduces the likelihood of employment and labour-force participation by, on average, 1.8–1.9 and 1.3–1.4 percentage points, respectively. These estimates are similar in magnitude to those from advanced economies, despite China substantially differing in terms of its development, labour-market structure, and mental-health provisions. Unlike in developed countries, this effect is driven almost exclusively by older male workers. This suggests that adverse mental-health shocks in China may primarily accelerate the transition into retirement for older male workers, rather than affecting the broader prime-age workforce.

JEL codes: I12, J22, J24.
Keywords: CES-D, depression, depressive symptoms, employment, participation.

*Affiliation: Peking University HSBC Business School. Correspondence: lander@phbs.pku.edu.cn.
†Affiliation: Peking University HSBC Business School.

1 Introduction

Empirical studies consistently find that poor mental health, e.g., exhibiting depressive symptoms, adversely impacts workforce engagement and productivity, with effects ranging from increased absenteeism to exiting the labour force (Peng et al., 2016). The majority of these recent findings, however, come from developed/advanced economies; e.g., Australia (Frijters et al., 2014; Bubonya et al., 2019), the UK (Lagomarsino and Spiganti, 2020; Bryan et al., 2022), the Netherlands (Ringdal and Rootjes, 2022), and the US (Mitra and Jones, 2017; Banerjee et al., 2017; Peng et al., 2016).[1]

In this paper, we revisit this question in a fundamentally different environment: China. *A priori*, it is not obvious whether similar findings would appear in China (or other less developed countries), given its vast differences with the aforementioned countries; e.g., composition of the labour force (largely due to its different stage of development), labour-market institutions/regulations, provision of mental-health services, and possible cultural differences in attitudes toward mental health and work.

China is also a country of particular interest for this topic, due to its scale and significance in the global economy. First, with such a substantial population size, the potential productivity implications are staggering; e.g., our nationally-representative sample suggests there are more depressed working-age individuals in China than the entire populations of Australia, the UK, and the Netherlands combined. Additionally, any impacts on labour productivity will indirectly impact consumers all over the world, given the massive volume of products exported by China (subject to relatively labour-intensive production processes).

To investigate this, we use three waves of data from the China Family Panel Studies (CFPS), from 2012, 2016, and 2018, to estimate the causal effect of depressive symptoms on employment and labour-force participation (LFP) for workers in China. The panel dimension is particularly important for studies of mental health, since unobservable time-invariant factors indeed matter, and we are the first to utilise panel data to study this issue in China.[2,3]

Mental health is measured in the CFPS using survey responses, which can be used to diagnose depression according to two scales: the Kessler (K6) and Center for Epidemiological Studies Depression (CES-D). While both of these scales are a valid measure of depressive symptoms, unfortunately the design of the CFPS makes it challenging to compare the scores over time for any individual. Specifically, the CFPS alternates which scale is used across waves,

[1] Two notable exceptions include Sohn (2018) (Indonesia) and Nwosu (2018) (South Africa).

[2] For a detailed discussion of the differences in estimates obtained using cross-section and panel data, see Bryan et al. (2022) or Ringdal and Rootjes (2022).

[3] To the best of our knowledge, Lu et al. (2009) is the only study on this topic using data from China, and it utilised cross-sectional data.

and both scales are never used in the same wave, making consistent comparisons across waves quite challenging. Furthermore, the survey definition of employment differs from the 2010 wave compared to the 2012 wave onwards. Only by using the 2012, 2016, and 2018 waves, we can construct consistent measures of both mental health and labour supply.

Our main results, which control for unobservable time-invariant factors, indicate that severe depressive symptoms reduce the likelihood of working by 1.8–1.9 (and LFP by 1.3–1.4) percentage points, on average. Note that this range encompasses a set of estimates we present throughout the paper. First, the magnitude of these estimates are within the range of those obtained in past studies (accounting for unobservable time-invariant factors); e.g., 1.6 in the UK (Bryan et al., 2022), 1.4–1.7 in the Netherlands (Ringdal and Rootjes, 2022), 2.4 in the US (Peng et al., 2016), 2.8–3.3 in Australia (Bubonya et al., 2019), and 3.6 in South Africa (Nwosu, 2018). While this effect may appear to be relatively small compared to other countries, its impact is orders of magnitude larger due to the enormous size of China's workforce.

Interestingly, we find that this reduction in labour supply, both in terms of employment and LFP, is almost exclusively driven by male workers. This stands in contrast with most previous studies, since they usually find either no difference in the response of men and women (Nwosu, 2018; Sohn, 2018; Bryan et al., 2022), or that both respond, but the magnitude of the response is larger for men (Peng et al., 2016; Bubonya et al., 2019).[4] The only case we found of women in China responding to adverse mental-health shocks were those aged close to retirement in rural areas, which represents a tiny fraction of prime-age female workers.[5]

In addition, we uncover strong evidence indicating that it is older male workers that primarily reduce labour supply in response to severe depressive symptoms. Specifically, for men over the age of 40 and 45, the likelihood of working reduces by as much as 3.8 and 5.2 percentage points, respectively. It is challenging to disentangle this age-effect from a cohort-effect in China, particularly since there are substantial differences in observables across groups (e.g., education/human capital), which is unsurprising given China's rapid development. Nonetheless, our findings suggest that this phenomenon is indeed driven by age, and, since it also similarly impacts LFP, we believe it indicates how the onset of severe depressive symptoms may accelerate the transition to retirement for older workers.

Finally, while recent estimates in the literature are a substantially smaller magnitude than those obtained without controlling for unobservable time-invariant factors, as noted by Peng

[4]Ringdal and Rootjes (2022), using Netherlands data, also observes several significant differences across gender. In their study, adverse mental health generally does not affect LFP or working full-time for women, but it does affect paid employment, and only in response to the onset of severe depressive symptoms.

[5]This finding arises when we use the same threshold for our binary measure of severe depressive symptoms for both men and women. If we allow this threshold to differ by gender, female workers in China require substantially more intense depressive symptoms to reduce labour supply.

et al. (2016) and Bryan et al. (2022), surprisingly this is not the case in China: our main estimate reduces from 3.5 to 1.9 percentage points after including individual fixed effects; in contrast, e.g., Bryan et al. (2022)'s reduces from 9.7 to 1.6. This suggests that selection into mental health problems based on (time-invariant) unobservables appears to play a relatively small role in China compared with developed countries.

The paper proceeds by presenting the empirical model, describing key variables, defining the sample selection criteria, and reporting summary statistics in Section 2. Section 3 reports the main results, heterogeneity analysis, and discusses the robustness of our estimates. Finally, Section 4 offers concluding remarks.

2 Empirics

2.1 Model

The empirical model we use is a standard linear two-way fixed effects model, i.e., a linear regression with individual fixed effects (FEs) and year/wave dummies. Specifically, for individual $i = 1, \dots, N$ and waves $t = 1, \dots, T$:

$$Y_{i,t} = \theta + \beta X_{i,t} + \delta \mathcal{D}_{i,t} + \gamma d_t + \mu_i + \epsilon_{i,t}, \tag{1}$$

where $Y_{i,t}$ is a binary indicator for labour supply (i.e., either employment or LFP) in wave t, $\mathcal{D}_{i,t}$ is a measure of (adverse) mental health in wave t, $X_{i,t}$ vector of observed (time-varying) controls, d_t is a vector of $T-1$ wave dummies, μ_i are individual FEs, $\epsilon_{i,t}$ is the idiosyncratic error term, and θ is a constant. This approach has been used to study the causal effect of mental health on labour supply by, e.g., Bryan et al. (2022) and Ringdal and Rootjes (2022).

An important decision to note is whether the mental-health measure is contemporaneous or lagged, i.e., whether $Y_{i,t}$ is regressed on $\mathcal{D}_{i,t}$ (as above) or $\mathcal{D}_{i,t-1}$. We utilise a contemporaneous measure of mental health primarily for reasons of practicality: we have a relatively long gap between waves of the CFPS (at least two years), and the length of the gap varies across waves (four years between 2012 and 2016, and two years between 2016 and 2018). This type of measure is utilised in Peng et al. (2016) (US survey data with a very short period between waves), Lagomarsino and Spiganti (2020) (annual UK survey data, but restricted to every two years to construct an instrument based on the respondent's social network), Nwosu (2018) (biennial South African survey data), and Sohn (2018) (two-wave Indonesian data with a seven-year gap between waves).

Alternatively, one may argue that, conceptually, a lagged measure is preferable, since it can mitigate potential reverse causality. This type of measure is utilised in, e.g., Bryan et al. (2022) (annual UK survey data), Ringdal and Rootjes (2022) (annual Netherlands survey data), and Bubonya et al. (2019) (annual Australian survey data). If one has relatively frequent surveys, with consistent gaps between surveys, a lagged measure indeed seems more appropriate. However, such high-quality survey data is scarcely available outside advanced economies, particularly relating to mental health, which explains the relative dearth of related studies from developing countries.

2.2 Variables

The variables from the CFPS that we use in our empirical model from Equation (1) include measures of labour-supply outcomes (i.e., the dependent variable), mental health, and time-varying controls. We proceed to describe each in turn.

Dependent variables ($Y_{i,t}$): Working (binary), and labour-force participation (binary).

Mental-health measures ($\mathcal{D}_{i,t}$): We utilise three measures throughout the main analysis, all derived from the CES-D scale (Radloff, 1977).[6] The first is the CES-D8 score (which is calculated using only eight questions from the original twenty-question CES-D scale), since we can directly calculate this for every CFPS wave in our sample (i.e., 2012, 2016, 2018). The CFPS did not utilise the CES-D scale in the 2014 wave, thus we do not include it.[7]

Our preferred measures, which are derived from the CES-D8 score, are binary variables which equal one whenever a respondent exhibits 'moderate' or 'severe' depressive symptoms.[8] Constructing this simply requires us to choose a 'cutoff' value for each level of symptoms, i.e., a different, higher cutoff for severe symptoms vs moderate. Thus, for our two binary measures, $\mathcal{D}=1$ corresponds to an individual's CES-D8 score being greater than (or equal to) the cutoff.

We choose a cutoff value of 7 to indicate moderate or severe symptoms, following Luo and Zhao (2021), and a cutoff of 9 to indicate severe symptoms. Reporting results for two cutoff levels enables us to see how labour-supply changes as the severity of depressive symptoms increases. In Section 3.3, we demonstrate how sensitive our results are to the choice of cutoff values defining each level of symptoms.

Time-varying controls ($X_{i,t}$): This includes measures for age, education, marital status (binary), number of adults in the household, presence of (no) children in the household (binary), and a binary indicator for a child in the household aged 0–4, 5–11, or 12–15, and

[6] Note that the CFPS utilises the Chinese-language version of the CES-D scale, which has been validated by Cheung and Bagley (1998).

[7] See Appendix A for a description of how the CES-D scale was included across CFPS waves.

[8] The labels 'moderate' and 'severe' are chosen to convey information about the relative intensity of depressive symptoms experienced by an individual, rather than to imply a specific diagnosis.

self-assessed health status (binary). These controls are commonly utilised in related studies. Following Bryan et al. (2022), we include age and education as a set of binary indicators to control for potential nonlinearities.[9]

We expand this set of time-varying controls to account for additional factors that potentially affect labour-supply decisions of individuals in China. The main idea is to adequately capture regional differences across provinces as well as within provinces (i.e., between urban and rural areas), as well as the multi-generation household structure often observed in China. Thus, we include controls for residence in a urban vs rural location (binary), a set of province dummies, and the presence of an elder (age 60+) in the household (binary).

Additionally, since there may be complex interactions between these factors, in Section 3.3 we obtain estimates using more flexible specifications with interactions between time-varying controls and gender/region (i.e., urban vs rural residence). In particular, controls relating to children and elders may differentially affect labour-supply choices for men and women across urban/rural areas, since the presence of elders in the home has been shown to impact the labour-supply decision of 'prime-age' women with young children in China (Meng et al., 2023), with varying effects observed across urban and rural areas (Maurer-Fazio et al., 2011).

These factors were absent from previous studies on mental health, because the set of countries under consideration almost exclusively comprised advanced economies in 'Western' countries. However, they are relevant outside of China: differences across urban vs rural areas are important to account for in most developing countries, and family-structure considerations apply to countries with relatively traditional cultures, especially throughout Asia.

2.3 Data and sample selection

The CFPS data used in our empirical analysis includes waves 2012, 2016, and 2018. We do not make use of the 2010, 2014, and 2020 waves due to measurement issues. Measurement error will be introduced by combining 2010/14 waves with others due to different measurement of employment and mental health, while 2020 measurement is post-COVID and lockdown policy responses, which likely confounds the relationship between mental health and employment.

Our sample includes working-age men and women in both urban and rural areas, but excludes individuals who report that they are retired, undertaking (full-time) education, or self-employed. Specifically, we include men aged 16–59 and women aged 16–49 in urban areas, and men and women aged 16–64 in rural areas.[10]

[9]Specifically, each indicator covers five years for age, and milestones for education (i.e., completing middle school, high school, or college/university). This specification indeed fits the data better than simply including age and education (in years) either linearly or plus a quadratic term.

[10]The upper limits in urban areas are set to the statutory retirement age, which differ by gender (Zhang et al.,

After imposing these restrictions, and dropping observations with missing or invalid responses (for the variables listed in Table 1), we are left with an unbalanced panel of 38,580 observations (15,042 individuals).[11] The summary statistics for each variable in our dataset are located in Table 1.

Table 1: Summary statistics

	Mean	S.D.	Min.	Max.
Labour-market measures				
Working	0.869	0.337	0	1
In the labour force	0.881	0.323	0	1
Mental-health measures				
CES-D8 score	5.267	3.826	0	24
Moderate symptoms	0.330	0.470	0	1
Severe symptoms	0.187	0.390	0	1
Individual and household characteristics				
Gender (male)	0.531	0.499	0	1
Ethnicity (minority)	0.092	0.289	0	1
Age (years)	41.129	10.415	16	64
Education (years)	8.021	4.402	0	19
Married	0.893	0.309	0	1
Number of adults	3.036	1.000	0	9
No children (present in HH)	0.535	0.499	0	1
Child aged 0-4	0.171	0.376	0	1
Child aged 5-11	0.258	0.438	0	1
Child aged 12-15	0.158	0.365	0	1
Elder (present in HH)	0.498	0.500	0	1
Health (self-assessed)	3.044	1.186	1	5
Reside in rural area	0.606	0.489	0	1
Observations	38,580			
Individuals	15,042			

Note: CFPS data (2012, 2016, 2018).

From Table 1, we observe that 86.9% of sample is working, while 88.1% is in the labour force; the difference is unemployed. The mean CES-D8 score is 5.27 points. According to the cutoff values, defined in Section 2.1, 33.0% and 18.7% of the sample exhibit moderate and severe depressive symptoms, respectively.

Regarding individual characteristics, 53% of the sample is male, 90.8% are Han ethnicity (i.e., 9.2% are from a 'minority' ethnic group), the mean age is 41 years, the mean education level (in years) is 8, and approximately 89.3% of individuals are married. Regarding household characteristics, 60.6% reside in a rural area, and the mean number of adults in a

2018; Zhou et al., 2021), whereas many workers in rural areas work beyond 60 due to limited pension support (Cheng et al., 2018a; Zang, 2020).

[11] See Appendix A for additional details on our sample, including the distribution of CES-D scores across waves.

household is three (53.5% have no child present in the household, and approximately 50% of households have an elder present).

Table 2: Mental health summary statistics, by employment status

	Full sample	Not Working	Working	Difference	t-stat
CES-D8 score	5.267	5.918	5.168	0.750***	(11.92)
Moderate symptoms	0.330	0.390	0.321	0.070***	(9.52)
Severe symptoms	0.187	0.231	0.180	0.052***	(8.19)
Observations	38,580	5,052	33,528		

Note: CFPS data (2012, 2016, 2018). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We also present conditional means, by employment status, in Table 2. It reveals that a greater fraction of individuals not working exhibit poor mental health, relative to those working. Specifically, those not working have a greater CES-D8 score (by 0.75 points, on average) and are more likely to exhibit moderate or severe depressive symptoms (by 7.0% and 5.2%, respectively). The latter difference (5.2%) is quite substantial, considering the proportion exhibiting severe symptoms in the full sample is 18.7%.

3 Analysis

We proceed to report our estimates of the impact of adverse mental health on labour supply, both in terms of employment as well as labour-force participation, in Section 3.1. These estimates come from our baseline model, i.e., Equation (1). In Section 3.2, we undertake heterogeneity analysis, in terms of gender, location, wealth, and age. Finally, in Section 3.3 we consider a variety of alternative specifications aimed at determining how sensitive/robust our coefficient estimates are to model assumptions (e.g., key measures, sample selection criteria).

3.1 Main results

The coefficient estimates from Equation (1) are presented in Table 3 panel A, labelled 'Fixed Effects' (i.e., including individual FEs). Columns (1) to (3) present estimates of the impact of mental health on employment, while columns (4) to (6) estimate the impact on labour-force participation (LFP). Each column varies in terms of its health measure: the CES-D8 score, the binary indicator for moderate depressive symptoms, and the indicator for severe symptoms.

Table 3: Effect of depressive symptoms on employment and labour-force participation

	Working			**In the labour force**		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Fixed effects						
CES-D8 score	−0.0027***			−0.0022***		
	(0.001)			(0.001)		
Moderate symptoms		−0.0108**			−0.0087*	
		(0.005)			(0.005)	
Severe symptoms			−0.0191***			−0.0145**
			(0.006)			(0.006)
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.6061	0.6059	0.6060	0.6002	0.6000	0.6001
R-squared (within)	0.0174	0.0168	0.0171	0.0170	0.0166	0.0168
Panel B: Pooled OLS						
CES-D8 score	−0.0052***			−0.0043***		
	(0.001)			(0.001)		
Moderate symptoms		−0.0285***			−0.0231***	
		(0.004)			(0.004)	
Severe symptoms			−0.0346***			−0.0294***
			(0.005)			(0.005)
Individual FE	No	No	No	No	No	No
R-squared	0.0685	0.0671	0.0671	0.0622	0.0610	0.0611
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	38,580	38,580	38,580	38,580	38,580	38,580

Note: CFPS data (2012, 2016, 2018). This table reports coefficient estimates for the model in Equation (1): panel A includes individual fixed effects, while panel B does not. The dependent variable in each column is specified in the top row of the table. All specifications include wave and province dummies, as well as the controls listed in Table 1, except panel A excludes time-invariant controls. The full set of coefficient estimates is in Appendix Table B7. Standard errors (clustered by individual) in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Our preferred specification includes both individual fixed effects, which controls for unobserved time-invariant individual heterogeneity, and the indicator for severe depressive symptoms (i.e., panel A, columns 3 and 6), since this provides the clearest measure of whether workers in China adjust labour supply in response to an adverse mental-health shock.

For the effect on employment, this specification yields a coefficient estimate of −0.0191 (significant at the 1% level), which implies that exhibiting severe depressive symptoms reduces the likelihood of employment by 1.91 percentage points (pp), on average. This is within the range of estimated reductions reported in recent studies, such as 1.6 (UK; Bryan et al., 2022), 1.4–1.7 (Netherlands; Ringdal and Rootjes, 2022), 2.4 (US; Peng et al., 2016), 2.8–3.3 (Australia; Bubonya et al., 2019), and 3.6 (South Africa; Nwosu, 2018).

We note that our estimate is toward the lower end of the range reported in the literature.

In Section 3.2 we provide evidence that explains this: unlike most of the studies in this group, we do not find evidence that women reduce labour supply in response to adverse mental health. However, our estimates for men are a similar magnitude to several other studies.

In panel B, we report estimates from a linear regression without individual FEs ('pooled OLS'), i.e., excluding μ_i from Equation (1).[12] Comparing across panels in Table 3 reveals that the magnitude of the coefficient estimate (for adverse mental health) decreases *in every specification* when unobserved time-invariant heterogeneity is accounted for. While these qualitative changes are consistent with prior studies, there is a subtle difference worth highlighting: the reduction in the estimated effect (when controlling for individual time-invariant unobservables) is much smaller than related studies; e.g., estimates in Bryan et al. (2022) reduced quite dramatically (from 9.7 to 1.6 pp), however, ours decline by a relatively modest amount (from 3.46 to 1.91 pp).[13] This could indicate that selection into mental-health problems based on (time-invariant) unobservables appears to play a relatively small role in China compared with developed countries.

It should not, however, be taken as evidence that time-invariant unobservables are not an important factor in the context of either labour supply or mental health in China. There are many time-invariant culture-related factors affecting female labour supply in China that are difficult to measure and control for, e.g., work experience of a wife's mother-in-law (Chen and Ge, 2018), family attitudes toward child gender (Bo, 2018; Fan et al., 2018), preferences for privacy and independence—particularly with regards to multi-generational households (Cheng et al., 2018b).[14]

We can delve a bit deeper into how mental health impacts labour supply by separating labour-market outcomes (i.e., employment, in columns 1 to 3) from LFP decisions (columns 4 to 6). Since the coefficient estimates for participation are consistently smaller in magnitude than employment, this indicates that deteriorating mental health is more likely to lead to unemployment spells (i.e., changes in employment status, conditional on being in the labour force) rather than transitions out of the labour force. We do also find, however, that these estimates become closer in magnitude for men in older cohorts, which we interpret as mental-health shocks possibly accelerating their transition into retirement (see Appendix Table B1).

The full table of coefficient estimates (including all controls) is contained in Appendix Table B7. It sheds light on the relative importance of time-varying factors affecting employment, besides changes in mental health. We proceed to briefly discuss a few highlights.

[12] In this specification, we also include time-invariant controls from Table 1, gender and ethnicity.

[13] This phenomenon appears to be consistent across a wide range of CES-D8 cutoff values (Appendix Figure B1).

[14] We note that our sample period (2012–2018) falls entirely after the implementation of the New Rural Pension Scheme in China had concluded (Cheng et al., 2018a).

Age. Our use of binary indicators for age indeed reveals a nonlinear impact on employment and labour-force attachment. In both cases, there is a distinct inverted U-shape between age and the likelihood of employment/LFP. Compared to the reference group of the youngest workers (ages 16–25), the probability of being employed increases steadily, peaking for individuals aged 41–45 who are more likely to be working by approximately 11.8 pp.

Education. The binary indicators for education also capture significant non-linearities. Compared to those that did not finish middle school (i.e., nine years of formal education), middle-school graduates are marginally more likely to be working (around 2 p.p., although not statistically significant). Completing high school increases the probability of working by about 9.5 pp (significant at the 5% level), while obtaining a college (i.e., 3-year) degree or university (i.e., 4-year) degree increases it by almost 10 pp (significant at the 10% level).

Marriage. Getting married reduces the probability of working and participating in the labour force by around 6.8 and 7.4 pp, respectively (both significant at the 1% level). This result may seem counter-intuitive, since marriage often increases stability. Indeed, Bryan et al. (2022) find that marriage is associated with a 4.5-point increase in the likelihood of employment in the UK. Our result, however, is consistent with previous studies in China, e.g., Ma and Shi (2020), which report evidence that spousal labour supply are substitutes, consistent with traditional gender roles (or a specialised division of labour in the household).[15]

Household structure. Having a child aged 0–4 in the household reduces employment likelihood by 4.2 pp (1% significance), which is qualitatively consistent with many prior studies. It is worth noting that the magnitude of this effect seems relatively large in China: approximately a 4-point reduction compared to 1 in the UK (Bryan et al., 2022). Once children are older (aged 5–15), we observe positive employment effects of having children, and the same applies to more adults in the household; both of which are consistent with decreasing caregiving burden, allowing household members to increase their labour supply.

We do not find that having an elder present in the home significantly impacts either employment or LFP, on average. This result seems counterintuitive, given that prior studies have shown they impact labour supply for women in China (Meng et al., 2023; Maurer-Fazio et al., 2011). Upon further investigation, it appears this effect may be captured by the age dummies: if instead we include quadratic age terms, having an elder in the household (aged 60+) significantly reduces employment by around 2 pp (1% significance)—likely reflecting caregiving demands toward older household members or income effects from elder pensions. With age dummies, however, this effect is reflected by the reduction in coefficients across cohorts aged 46–50 (relative to those aged 41–45).

[15] With individual FEs, this may also be indicative of an increase in labour supply post divorce/widowhood.

Location. Since we include individual FEs the coefficient estimate on rural residence reflects 'within' changes, i.e., moving from urban to rural (or vice versa). Hence, it reveals that moving from an urban to rural residence increases employment by 2.4 pp (5% significance), or, conversely, moving from rural to urban areas decreases employment by the same amount. Interestingly, the effect is not statistically significant for LFP. Together, we interpret these as indicating that migrant workers may experience temporary unemployment spells when moving from rural areas (where they were employed) to urban areas (where they may experience temporary spells of unemployment). To provide clearer insights into differences across regions, we run separate regressions for urban vs rural residents in Section 3.2.

3.2 Heterogeneity analysis

Since we have established that deteriorations in mental health indeed lead to measurable reductions in labour supply for workers in China, we now proceed to investigate which subgroups of the population are most responsive to adverse mental-health shocks. Our analysis focuses on four dimensions: gender (male vs female), region (urban vs rural), family wealth (below vs above median), and age (below vs above 40). The estimates, for each subgroup, are reported in Table 4.

Gender. We find no compelling evidence of a relationship between mental health and employment status for women, yet there is clearly a strong relationship for men (Table 4 panel A). Specifically, a 1.70 and 3.11 pp reduction in the likelihood of employment following the onset of moderate and severe depressive symptoms, respectively (both significant at a 1% level). This latter estimate is remarkably close to those obtained for men in several other prior studies, e.g., 3.2 in South Africa (Nwosu, 2018), and 3.3 in the US (Peng et al., 2016) and Australia (Bubonya et al., 2019).

On the other hand, finding that women do not respond to mental-health shocks is generally inconsistent with prior studies. Specifically, most recent studies that utilise longitudinal data to control for time-invariant unobservables find either no difference in the response of men and women (Nwosu, 2018; Sohn, 2018; Bryan et al., 2022), or that both respond, but the magnitude of the response is larger for men (Peng et al., 2016; Bubonya et al., 2019). This asymmetry in responses by gender is not entirely unprecedented, however: a similar finding was observed in the Netherlands by Ringdal and Rootjes (2022).[16]

[16] They found no evidence that women respond to mental-health shocks in terms of either employment or labour-force participation—but they did find evidence that women reduce *paid* employment.

Table 4: Effect of depressive symptoms on employment status, by subgroup

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Gender	**Female**			**Male**		
CES-D8 score	−0.0016			−0.0033***		
	(0.001)			(0.001)		
Moderate symptoms		−0.0018			−0.0170***	
		(0.008)			(0.006)	
Severe symptoms			−0.0075			−0.0311***
			(0.009)			(0.007)
R-squared (within)	0.0368	0.0366	0.0367	0.0156	0.0146	0.0157
Observations	18,037	18,037	18,037	20,459	20,459	20,459
Panel B: Region	**Urban**			**Rural**		
CES-D8 score	−0.0030**			−0.0026***		
	(0.001)			(0.001)		
Moderate symptoms		−0.0091			−0.0161***	
		(0.010)			(0.006)	
Severe symptoms			−0.0354***			−0.0113
			(0.012)			(0.007)
R-squared (within)	0.0235	0.0229	0.0240	0.0179	0.0176	0.0173
Observations	13,261	13,261	13,261	21,471	21,471	21,471
Panel C: Wealth	**High wealth**			**Low wealth**		
CES-D8 score	−0.0015			−0.0035***		
	(0.001)			(0.001)		
Moderate symptoms		−0.0040			−0.0176***	
		(0.007)			(0.007)	
Severe symptoms			−0.0161*			−0.0197**
			(0.009)			(0.008)
R-squared (within)	0.0240	0.0238	0.0240	0.0174	0.0166	0.0166
Observations	18,818	18,818	18,818	18,981	18,981	18,981
Panel D: Age	**Age < 40**			**Age ≥ 40**		
CES-D8 score	−0.0006			−0.0045***		
	(0.001)			(0.001)		
Moderate symptoms		−0.0000			−0.0222***	
		(0.007)			(0.006)	
Severe symptoms			−0.0093			−0.0270***
			(0.009)			(0.007)
R-squared (within)	0.0238	0.0238	0.0239	0.0195	0.0178	0.0179
Observations	19,406	19,406	19,406	19,174	19,174	19,174
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: CFPS data (2012, 2016, 2018). This table reports coefficient estimates for the model in Equation (1), for a subset of individuals belonging to each group. The dependent variable is a binary indicator for employment status ("Working", from Table 1). All specifications include individual fixed effects, wave and province dummies, and the time-varying controls listed in Table 1. In panels C and D, individuals are separated by wealth (above/below median) and age (above/below 40), respectively, according to their values measured in the first wave observed. Standard errors (clustered by individual) in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To further investigate this seeming non-response by female workers, we plot the coefficient estimate for varying CES-D8 cutoff values in Figure 1. It reveals that we consistently measure a reduction in employment (and LFP) when male workers' CES-D8 scores increase beyond 5. However, for women, a score above 10 is required to measure a response (at a 5% level of significance). This is in line with prior studies finding a higher optimal cutoff value for women (compared to men) with the CES-D scale (Henry et al., 2018). Such a striking gender difference implies a strong asymmetry in labour-supply responses to mental-health shocks, indicating women's employment is less affected, possibly due to lower LFP, household constraints, or cultural factors. This is, to the best of our knowledge, a new dimension differentially affecting labour supply between men and women in China.[17]

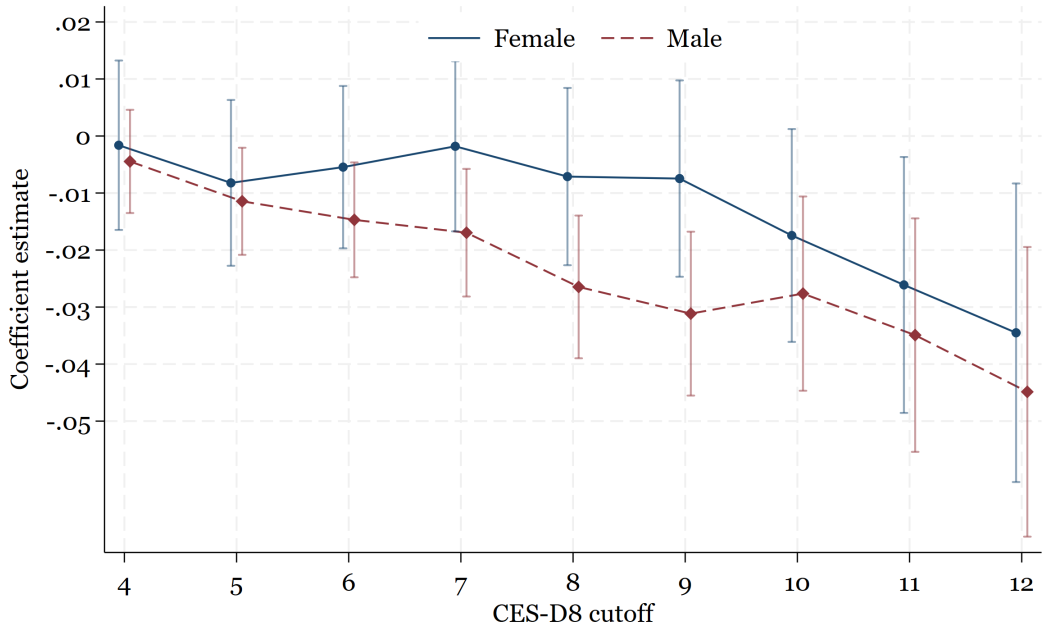
Region. The estimates in Table 4 panel B show that workers in urban and rural areas both reduce labour supply when experiencing adverse mental health, however, those in urban areas are relatively resilient (i.e., less responsive to depressive symptoms). While similar magnitude coefficient estimates on the CES-D8 score may lead us to conclude that there are no regional differences in the relationship between mental health and employment in China, our binary measures uncover nuanced patterns across urban vs rural areas. Specifically, in rural areas, moderate symptoms significantly reduce employment by 1.61 pp (1% significance), on average, but not severe symptoms; while, in urban areas, severe symptoms significantly reduce employment by 3.54 pp (1% significance), on average, but not moderate symptoms.

The greater responsiveness of workers in rural areas to moderate depressive symptoms, relative to urban areas, appears to be consistent with arguments made by Das et al. (2007): that individuals in developing countries experiencing a mental-health shock may be 'insured' against poverty due to having a larger family or village social support systems. In other words, if individuals in rural areas have access to this type of support, they may generally be more flexible to adjust labour supply in response to mental-health shocks.

Wealth. Table 4 panel C reveals that individuals from low-wealth families are generally more responsive to depressive symptoms. Specifically, individuals from low-wealth families exhibiting moderate or severe depressive symptoms reduce employment by, on average, 1.76 or 1.97 pp (significant at the 1% and 5% level), respectively. In contrast, individuals from high-wealth families do not appear to respond to moderate symptoms, but there is some evidence that they reduce employment when experiencing severe depressive symptoms (by 1.61%, on average, significant at the 10% level). This phenomenon is broadly consistent with Bryan et al. (2022), which found that relative poverty exacerbates the impact of adverse mental health on employment in the UK.

[17] See Li and Zax (2003) or Chen et al. (2014) for an analysis of factors affecting female labour supply in China.

Figure 1: Estimated effect on employment for varying CES-D8 cutoffs, by gender

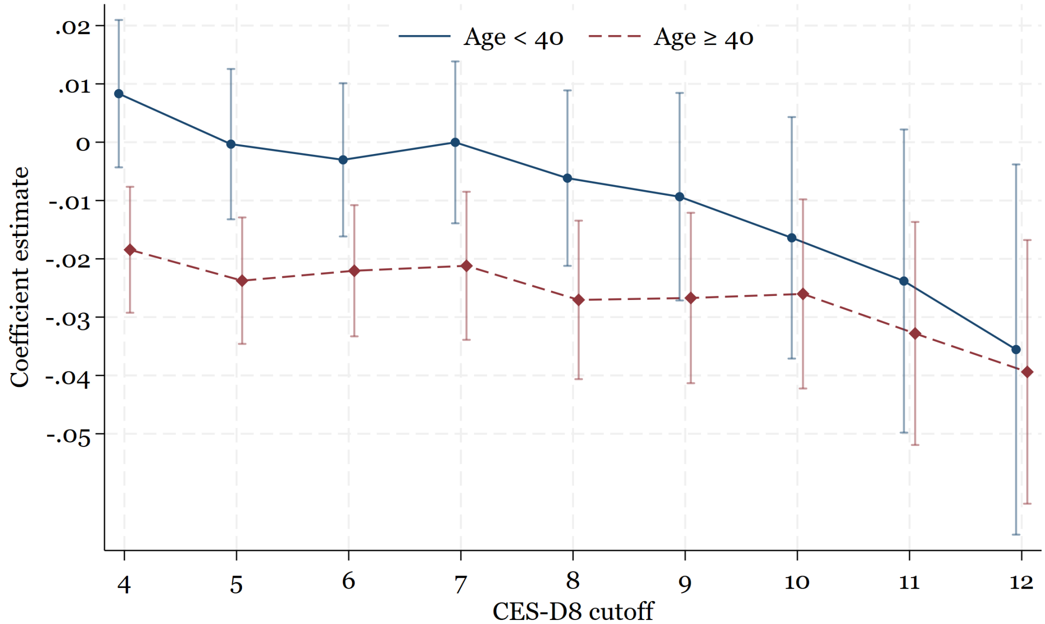


Note: Each series plots coefficient estimates of the indicator for depressive symptoms for varying cutoff values (horizontal axis) from Equation (1) with employment as the dependent variable and including individual fixed effects, separately by gender. The vertical line above/below each point estimate is its 95% confidence interval.

Age. Finally, we explore heterogeneity by age group in Table 4 panel D. Our estimates indicate that workers aged 40 years and older exhibit substantial responsiveness to adverse mental-health shocks. Specifically, the presence of moderate depressive symptoms significantly reduces the likelihood of employment by 2.22 pp, whereas severe depressive symptoms lead to a 2.70-point reduction (both significant at the 1% level). Conversely, younger workers (below age 40) show negligible and statistically insignificant responses.

To investigate this relation in further detail, we proceed to plot coefficient estimate for varying CES-D8 cutoff values in Figure 2. It provides two main insights. First, older workers are particularly vulnerable to adverse mental-health shocks: we consistently measure a reduction in employment (and LFP) when older workers' CES-D8 scores increase beyond 3. In fact, the effect for older workers with a cutoff of 4 (-0.018, significant at the 1% level) is almost as large as our aggregate effect with a cutoff of 7 (-0.019, Table 3). Second, for younger workers, a substantially higher CES-D8 score is required to measure a response: at least 12, at a 5% level of significance. Hence, on average, younger workers, only respond to particularly severe depressive symptoms, similar to female workers. This suggests that both female and younger workers in China are more resilient or face different labour-market constraints that buffer them against the adverse employment effects of poor mental health.

Figure 2: Estimated effect on employment for varying CES-D8 cutoffs, by age cohort



Note: Each series plots coefficient estimates of the indicator for depressive symptoms for varying cutoff values (horizontal axis) from Equation (1) with employment as the dependent variable and including individual fixed effects, separately by age cohort. The vertical line above/below each point estimate is its 95% confidence interval.

We interpret the sensitivity of older workers' labour supply to rising depressive symptoms as evidence that they may be accelerating their transition to retirement. In addition to reducing labour supply when facing less severe depressive symptoms (Figure 2), in Appendix Table B1 we show that men are increasingly likely to stop working (and exit the labour force) as they age from 40 to 50 years (for a fixed cutoff at our baseline of 7). This shows that, as male workers in China approach the statutory retirement age, they become more likely to exit the labour force (which we interpret as early retirement) following a rise in depressive symptoms.

Before proceeding to the next section, we will briefly mention an important caveat to this interpretation. Due to the fact we are not using a lagged measure of mental health, it is possible that this pattern may alternatively be explained by reverse causality, i.e., a reduction in mental health following job loss (as measured by employment) or 'involuntary' early retirement (as measured by LFP). While we cannot investigate this due to the aforementioned limitations in the CFPS survey, we do note that Bubonya et al. (2019) found no evidence that poor employment outcomes contributed to subsequent depressive symptoms in Australia.

3.3 Robustness

Our main findings are robust to a variety of alternative specifications, in terms of sample selection criteria and modelling choices. In this section, we briefly motivate these alternatives and summarise how our inference is affected, but we defer all estimates to Appendix B.

(i) **Gender:** To address concerns there may be gender-specific differences in how some of our time-varying controls (e.g., education) impact labour supply, we consider a specification saturated with gender interaction terms ala Bryan et al. (2022) (Appendix Table B2 panel A).[18] This is arguably more critical in our setting, since China is a relatively traditional culture with well-established gender roles in the household, e.g., the burden of helping children or elders in the home may disproportionately fall on women (Chen and Ge, 2018). In each case the estimates are marginally smaller than our baseline (e.g., 1.87 for the effect of severe symptoms on employment).

(ii) **Region:** To further investigate whether there are important regional differences not controlled for in our baseline specification, we added interaction terms between the rural dummy and family-related variables, i.e., in the same spirit as the exercise in (i), but for the rural dummy instead of the gender dummy. These estimates, located in Appendix Table B2 panel B are almost identical to our baseline estimates.

Additionally, we considered specifications with more complex time-varying regional dummies to control for possible differences in time trends across provinces, as well as across rural and urban areas. Specifically, we added wave-province dummies and wave-province-rural dummies (Appendix Table B3). Estimates remain almost identical.

(iii) **Mental-health scale:** In Appendix Figures B1 and B2 we demonstrate how our baseline estimates vary when we change the cutoff determining our binary indicator for poor mental health. The first figure compares estimates for the Pooled OLS vs Fixed Effects (FE) models, while the second compares estimates for employment and LFP, for the FE model.

In general, we consistently obtain statistically-significant coefficient estimates of at least a 1-point employment reduction for cutoffs above 5, with estimates slightly greater than 2 pp for a cutoff of 10. Furthermore, a cutoff of 7 leads to almost identical estimates (albeit slightly smaller in magnitude) than cutoffs 5 or 6. A similar argument applies to a cutoff of 8 vs 9. This demonstrates that our main findings are not driven by our choice of cutoff values.

We also report estimates using the CES-D20 scale, rather than CES-D8, in Appendix Table B4. This requires converting CES-D8 scores to CES-D20 for many individuals in 2016 and all individuals in 2018.[19] Similar estimates are obtained (marginally smaller for employment,

[18] Gender was interacted with child-related variables, as in Bryan et al. (2022), as well as elder-related variables.

[19] The conversion is based on a mapping determined by the CFPS (see Appendix A for further discussion about

and marginally larger for LFP), using a CES-D20 cutoff of 18 and 22 for moderate and severe depressive symptoms, respectively.[20]

(iv) **Attrition:** Appendix Table B5 investigates the impact of possible non-random attrition on our baseline estimates. When we omit individuals not reporting information on all covariates at each survey wave (i.e., we use a balanced panel), the estimates are generally quite robust, despite the number of observations reducing from 38,580 to 25,488. We observe a marginal reduction in magnitude of the effect of severe depression on employment (1.76) and marginal increase for LFP (1.51), and both remain precisely estimated. On the other hand, the effects for moderate depressive symptoms are more sensitive: there is both a sizeable reduction in magnitude and statistical significance.

(v) **Selection on unobservables:** To consider the possibility of bias driven by unobservable confounders, we employ the method of Oster (2019), adapted to panel data by Bryan et al. (2022). The estimates are reported in Appendix Table B6. For severe depressive symptoms, we find marginal reductions in the magnitude and precision of our estimate on employment (1.80–1.89) and LFP (1.34–1.42). While, for moderate symptoms, we arrive at the same conclusion as our exercise above for attrition: estimates fall in both magnitude and precision to such an extent that we cannot reject there is no effect.

Taken together, we interpret these findings as indicating that non-random sample attrition and omitted-variable bias may lead our baseline estimates to overestimate the impact of moderate depressive symptoms on labour supply. Furthermore, by taking our most conservative bias-adjusted estimates for work and LFP (1.80 and 1.34 pp, Appendix Table B6) with our baseline estimates (1.91 and 1.45 pp), we conclude that exhibiting severe depressive symptoms reduces the likelihood of working by 1.8–1.9 pp and LFP by 1.3–1.4 pp, on average.

4 Concluding remarks

This paper estimates the effect of adverse mental health on labour supply for workers in China using CFPS data from 2012, 2016, and 2018. These waves contain the CES-D scale, which we used to construct binary indicators of moderate and depressive symptoms. By combining our baseline estimates (which include individual fixed effects to account for time-invariant unobservables) with our most conservative estimates (which account for selection on unobservables) we report a range of estimates for both employment and LFP.

the relationship between the CES-D8 and CES-D20 scales). This approach is not our preferred measure, since it may introduce measurement error into our binary measures of mental health across waves.

[20] While 16 and 20 are often considered 'traditional' cutoffs for the CES-D20 scale, following Weissman et al. (1977) and Comstock and Helsing (1977), higher cutoffs yield greater sensitivity and specificity for Chinese (Cheng and Chan, 2005; Zhang et al., 2015).

We find that severe depression reduces the likelihood of working by between 1.8–1.9 percentage points (and LFP by 1.3–1.4), on average. This range is broadly similar to those from prior studies, although these were set in countries with, *inter alia*, remarkably different labour markets, healthcare systems, and cultures, e.g., Australia, USA, UK, Netherlands.

Although these estimates are not substantially different from those found in the literature, the scale of this effect is remarkable in China, due to the enormous size of its workforce. With a working-age population in 2022 of 974,838,887 people, our representative sample implies approximately 847 million employed workers and 182 million people exhibiting severe depressive symptoms. A conservative back-of-the-envelope calculation suggests that a 1.4-point reduction in LFP, given the onset of severe depressive symptoms, would result in upwards of 1 million people ceasing work due to poor mental health if the transition probability from good to poor mental health is at least 0.10—which is lower than the unconditional sample mean of severe depressive symptoms, 0.187 (Table 2).[21]

The panel structure of the CFPS enables us to control for time-invariant unobservables, which is a marked improvement upon related studies outside developed countries, but it does not address the possibility of reverse causality; i.e., the impact of job loss on mental health status (Tefft, 2011; Marcus, 2013). While it is reassuring that Bubonya et al. (2019) found no evidence that poor employment outcomes contributed to subsequent depressive symptoms in Australia, it is not clear whether this is also the case in China. Furthermore, the biennial nature of the CFPS survey data, together with inconsistent mental-health measures across waves, makes it challenging to employ a lagged measure of mental health (to address possible reverse causality), which is typical in studies set in developing countries with annual survey waves; e.g., Bubonya et al. (2019), Bryan et al. (2022), and Ringdal and Rootjes (2022). We view this as the most significant limitation of our study, which unfortunately is a common drawback shared by studies using data from developing countries, e.g., Nwosu (2018), Sohn (2018). Future empirical studies using the CFPS data will be able to take advantage of consistent mental health measures every wave, since the CFPS plans to consistently use the CES-D8 scale in each wave (beginning from 2016). This will reduce the potential for measurement error and broaden the range of possible studies relating to mental health in China.

Finally, while the Chinese government has been proactive in recent years regarding initiatives to raise awareness and national literacy regarding mental health, as well as improving the provision of psychological-related health services, our findings suggest that they may wish to narrow their focus to older workers approaching retirement.[22]

[21] For a more comprehensive analysis of the economic costs of depression in China, see Hu et al. (2007).
[22] For a detailed review of mental-health policies in China from 2009 to 2020, see Chen et al. (2023).

References

BANERJEE, S., P. CHATTERJI, AND K. LAHIRI (2017): "Effects of psychiatric disorders on labor market outcomes: a latent variable approach using multiple clinical indicators," *Health Economics*, 26, 184–205.

BO, S. (2018): "Son preference, children's gender and parents' time allocation: Evidence from China," *Applied Economics*, 50, 4830–4846.

BRYAN, M. L., N. RICE, J. ROBERTS, AND C. SECHEL (2022): "Mental health and employment: a bounding approach using panel data," *Oxford Bulletin of Economics and Statistics*, 84, 1018–1051.

BUBONYA, M., D. A. COBB-CLARK, AND D. C. RIBAR (2019): "The reciprocal relationship between depressive symptoms and employment status," *Economics & Human Biology*, 35, 96–106.

CHEN, J., X. SHAO, G. MURTAZA, AND Z. ZHAO (2014): "Factors that influence female labor force supply in China," *Economic Modelling*, 37, 485–491.

CHEN, R., W. ZHANG, AND X. WU (2023): "Mental health policy and implementation from 2009 to 2020 in China," *SSM-Mental Health*, 100244.

CHEN, X. (2022): "The impact of spousal and own retirement on health: Evidence from urban China," *World Development*, 159, 106025.

CHEN, X. AND S. GE (2018): "Social norms and female labor force participation in urban China," *Journal of Comparative Economics*, 46, 966–987.

CHENG, L., H. LIU, Y. ZHANG, AND Z. ZHAO (2018a): "The health implications of social pensions: evidence from China's new rural pension scheme," *Journal of Comparative Economics*, 46, 53–77.

——— (2018b): "The heterogeneous impact of pension income on elderly living arrangements: evidence from China's new rural pension scheme," *Journal of Population Economics*, 31, 155–192.

CHENG, S.-T. AND A. C. CHAN (2005): "The center for epidemiologic studies depression scale in older Chinese: thresholds for long and short forms," *International Journal of Geriatric Psychiatry*, 20, 465–470.

CHEUNG, C.-K. AND C. BAGLEY (1998): "Validating an American scale in Hong Kong: the center for epidemiological studies depression scale (CES-D)," *The Journal of Psychology*, 132, 169–186.

COMSTOCK, G. W. AND K. J. HELSING (1977): "Symptoms of depression in two communities," *Psychological Medicine*, 6, 551–563.

DAS, J., Q.-T. DO, J. FRIEDMAN, D. MCKENZIE, AND K. SCOTT (2007): "Mental health and poverty in developing countries: Revisiting the relationship," *Social Science & Medicine*, 65, 467–480.

FAN, Y., J. YI, Y. YUAN, AND J. ZHANG (2018): "The glorified mothers of sons: Evidence from child sex composition and parental time allocation in rural China," *Journal of Economic Behavior & Organization*, 145, 249–260.

FRIJTERS, P., D. W. JOHNSTON, AND M. A. SHIELDS (2014): "The effect of mental health on employment: evidence from Australian panel data," *Health Economics*, 23, 1058–1071.

HENRY, S. K., M. M. GRANT, AND K. L. CROPSEY (2018): "Determining the optimal clinical cutoff on the CES-D for depression in a community corrections sample," *Journal of Affective Disorders*, 234, 270–275.

HU, T.-W., Y. HE, M. ZHANG, AND N. CHEN (2007): "Economic costs of depression in China," *Social Psychiatry and Psychiatric Epidemiology*, 42, 110–116.

LAGOMARSINO, E. AND A. SPIGANTI (2020): "No gain in pain: psychological well-being, participation, and wages in the BHPS," *The European Journal of Health Economics*, 21, 1375–1389.

LI, H. AND J. S. ZAX (2003): "Labor supply in urban China," *Journal of Comparative Economics*, 31, 795–817.

LU, C., R. G. FRANK, Y. LIU, AND J. SHEN (2009): "The impact of mental health on labour market outcomes in China," *Journal of Mental Health Policy and Economics*, 12, 157.

LUO, W. AND M. ZHAO (2021): "Trends and socioeconomic differentials in depression prevalence in China, 2010–2018," *Chinese Journal of Sociology*, 7, 535–556.

MA, Y. AND X. SHI (2020): "Are spousal labor supplies substitutes? evidence from the workweek reduction policy in China," *Journal of Development Economics*, 145, 102472.

MARCUS, J. (2013): "The effect of unemployment on the mental health of spouses–Evidence from plant closures in Germany," *Journal of Health Economics*, 32, 546–558.

MAURER-FAZIO, M., R. CONNELLY, L. CHEN, AND L. TANG (2011): "Childcare, Eldercare, and Labor Force Participation of Married Women in Urban China, 1982-2000." *Journal of Human Resources*, 46, 261–294.

MENG, L., Y. ZHANG, AND B. ZOU (2023): "The motherhood penalty in China: Magnitudes, trends, and the role of grandparenting," *Journal of Comparative Economics*, 51, 105–132.

MITRA, S. AND K. JONES (2017): "The impact of recent mental health changes on employment: new evidence from longitudinal data," *Applied Economics*, 49, 96–109.

NWOSU, C. O. (2018): "The relationship between employment and mental and physical health in South Africa," *Development Southern Africa*, 35, 145–162.

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

PENG, L., C. D. MEYERHOEFER, AND S. H. ZUVEKAS (2016): "The Short-Term Effect of Depressive Symptoms on Labor Market Outcomes," *Health Economics*, 25, 1223–1238.

RADLOFF, L. S. (1977): "The CES-D scale: A self-report depression scale for research in the general population," *Applied Psychological Measurement*, 1, 385–401.

RINGDAL, C. AND F. ROOTJES (2022): "Depression and labor supply: Evidence from the Netherlands," *Economics & Human Biology*, 45, 101103.

SOHN, K. (2018): "Depressive symptoms are not related to labor market outcomes in Indonesia," *Hitotsubashi Journal of Economics*, 45–60.

TEFFT, N. (2011): "Insights on unemployment, unemployment insurance, and mental health," *Journal of Health Economics*, 30, 258–264.

WEISSMAN, M. M., D. SHOLOMSKAS, M. POTTENGER, B. A. PRUSOFF, AND B. Z. LOCKE (1977): "Assessing depressive symptoms in five psychiatric populations: a validation study," *American Journal of Epidemiology*, 106, 203–214.

ZANG, E. (2020): "Spillover effects of a husband's retirement on a woman's health: Evidence from urban China," *Social Science & Medicine*, 245, 112684.

ZHANG, Y., M. SALM, AND A. VAN SOEST (2018): "The effect of retirement on healthcare utilization: Evidence from China," *Journal of Health Economics*, 62, 165–177.

ZHANG, Y., R. Z. TING, M. H. LAM, S.-P. LAM, R. O. YEUNG, H. NAN, R. OZAKI, A. O. LUK, A. P. KONG, Y.-K. WING, ET AL. (2015): "Measuring depression with CES-D in Chinese patients with type 2 diabetes: the validity and its comparison to PHQ-9," *BMC Psychiatry*, 15, 198.

ZHOU, Q., K. EGGLESTON, AND G. G. LIU (2021): "Healthcare utilization at retirement in China," *Health Economics*, 30, 2618–2636.

APPENDIX

This appendix contain the supplementary materials for *'Mental Health and Labour Supply: Evidence from China'*. It has the following structure:

A. CFPS details

- Figure A1: Distribution of CES-D8 scores, by CFPS wave
- Figure A2: Distribution of CES-D8 scores, by subgroup

B. Additional results

- Figure B1: Estimates for varying CES-D8 cutoffs, by estimator
- Figure B2: Estimates for varying CES-D8 cutoffs, by labour-supply response
- Table B1: Coefficient estimates for men, by age group
- Table B2: Coefficient estimates, with gender and rural interactions
- Table B3: Coefficient estimates, with time-varying regional dummies
- Table B4: Coefficient estimates, using the CES-D20 scale
- Table B5: Coefficient estimates, with a balanced panel
- Table B6: Coefficient estimates, with bounds
- Table B7: Baseline model estimates (full table)

Appendix A CFPS details

A.1 Mental health scales

This paper uses data from the China Family Panel Studies (CFPS) database, which is designed and constructed by the China Social Science Research Center of Peking University. The sample covers 25 provinces/municipalities/autonomous regions and is nationally representative. The survey focuses on the family relationships and economic activities of the participating households, as well as the education and health information of the participating individuals. It has collected and published six complete waves of data, up to 2020.

Since its inception in 2010, the CFPS has used three different scales to measure mental health: two based on the Center for Epidemiological Studies Depression (CES-D) scale, and one based on the Kessler scale (K6). Specifically, the K6 was used in 2010 and 2014, the CES-D20 was used in 2012 and 2016, and the CES-D8 was used in 2016, 2018, and 2020.

In 2016, the CFPS began its transition away from the 'long-form' CES-D20 to the 'short-form' CES-D8 scale, by asking 20% of respondents the full twenty questions (i.e., the CES-D20) and the remaining 80% only eight representative questions (i.e., the CES-D8). Since the eight CES-D8 questions are a subset of the twenty CES-D20 questions, the CFPS used the results from the 2016 wave to determine a mapping from CES-D8 scores to CES-D20 scores (via 'equipercentile equating').[23]

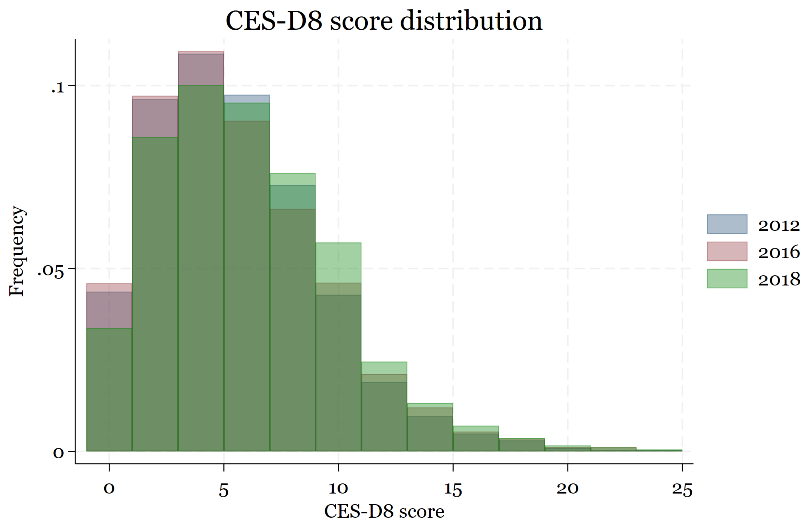
Throughout this paper we use both the CES-D20 and the CES-D8 scale. Whenever we use the CES-D20 scale, we obtain the exact CES-D20 score for all respondents in the 2012 wave and for 20% of respondents in the 2016 wave; but, for the remaining 80% of respondents in wave 2016 and all respondents in wave 2018, we must convert their CES-D8 score to a CES-D20 score using the mapping provided by the CFPS. Alternatively, whenever we use the CES-D8 scale, it is calculated directly for all respondents for waves 2012, 2016, and 2018.

Thus, we face a trade-off when choosing between the two scales in the CFPS. When using the CES-D20, we must use the mapping to convert some (but not all) scores of respondents, hence there may be some measurement error due to inconsistent measures. On the other hand, whenever we use the CES-D8 scale it is consistent across respondents, but we conceivably discard useful information from the longer-form twenty-question scale (available to all respondents in 2012, and some in 2016).

[23] For further details, see: http://www.isss.pku.edu.cn/cfps/docs/20201201085335172101.pdf (in Chinese).

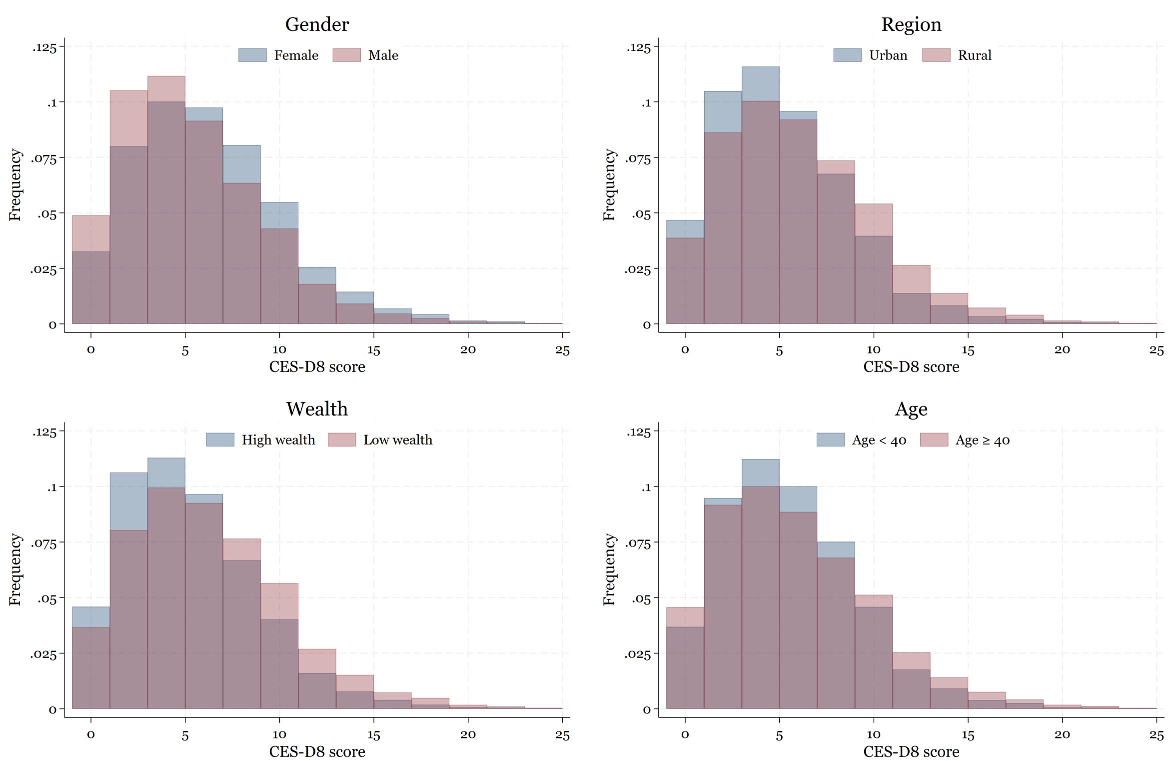
A.2 Distribution of CES-D scores

Figure A1: Distribution of CES-D8 scores, by CFPS wave



Note: CFPS data (2012, 2016, 2018).

Figure A2: Distribution of CES-D8 scores, by subgroup

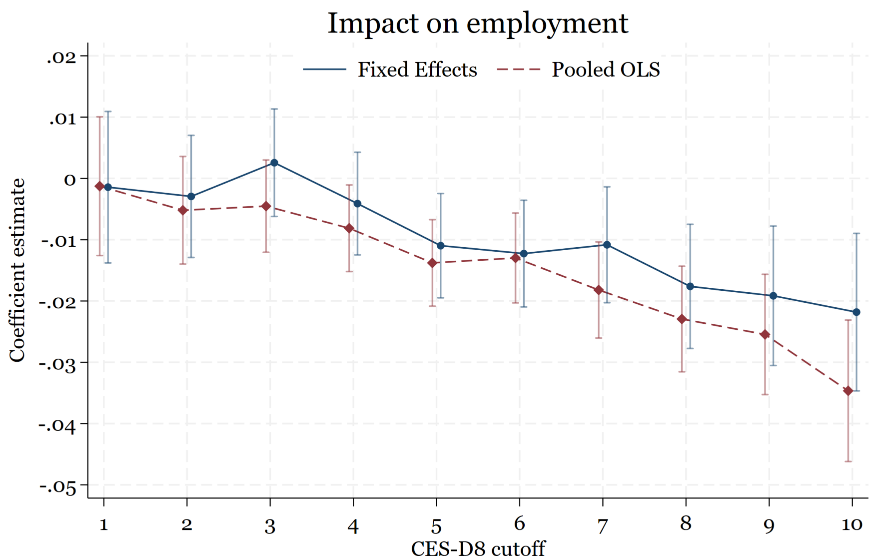


Note: CFPS data (2012, 2016, 2018). Individuals are separated by wealth (above/below median) and age (above/below 40), respectively, according to their values measured in the first wave observed.

Appendix B Additional results

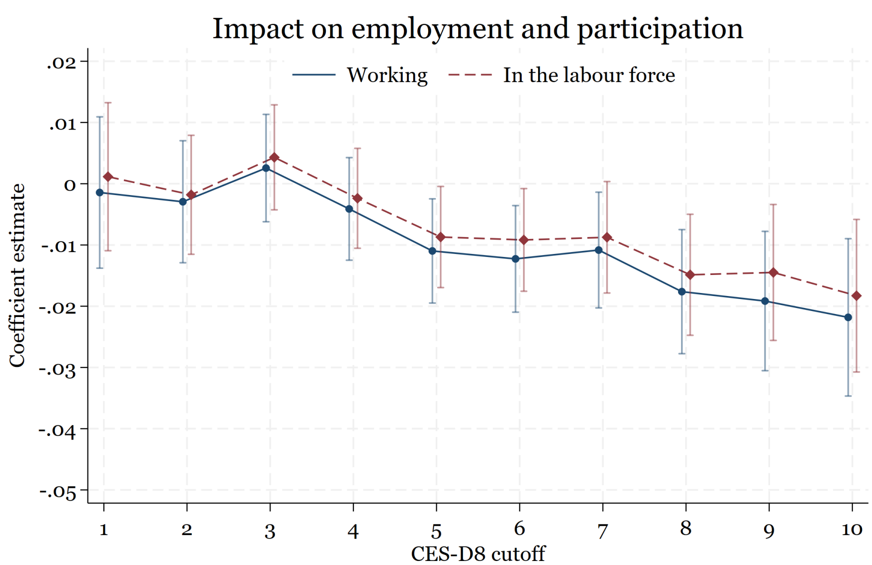
B.1 Figures

Figure B1: Estimated effect on employment for varying CES-D8 cutoffs, by estimator



Note: The 'Fixed Effects' and 'Pooled OLS' series plot the coefficient estimates of the indicator for depressive symptoms for varying cutoff values (horizontal axis) from Equation (1), with and without individual FEs, respectively. The vertical line above/below each point estimate is its 95% confidence interval.

Figure B2: Estimated effect on labour-supply responses for varying CES-D8 cutoffs



Note: The 'Working' and 'In the labour force' series plot the coefficient estimates of the binary indicator for depressive symptoms for varying cutoff values (horizontal axis) from Equation (1), with each label referring to the dependent variable used in Equation (1). Both models include individual FEs. The vertical line above/below each point estimate is its 95% confidence interval.

B.2 Tables

Table B1: Coefficient estimates for men, by age group

	Age ≥ 40		**Age** ≥ 45		**Age** ≥ 50	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Working						
Moderate symptoms	-0.0340^{***}		-0.0369^{***}		-0.0504^{***}	
	(0.008)		(0.009)		(0.013)	
Severe symptoms		-0.0385^{***}		-0.0515^{***}		-0.0528^{***}
		(0.010)		(0.011)		(0.015)
R-squared	0.5642	0.5641	0.5840	0.5847	0.5985	0.5979
R-squared (within)	0.0215	0.0213	0.0261	0.0276	0.0405	0.0392
Panel B: In labour force						
Moderate symptoms	-0.0301^{***}		-0.0341^{***}		-0.0469^{***}	
	(0.007)		(0.009)		(0.013)	
Severe symptoms		-0.0321^{***}		-0.0457^{***}		-0.0468^{***}
		(0.009)		(0.011)		(0.015)
R-squared	0.5454	0.5451	0.5666	0.5671	0.5888	0.5881
R-squared (within)	0.0173	0.0168	0.0227	0.0237	0.0351	0.0335
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,337	11,337	7,699	7,699	3,841	3,841

Note: CFPS data (2012, 2016, 2018). This table reports coefficient estimates for the model in Equation (1), for a subset of observations: male workers meeting the age criteria in the top row. An observation is included in the age-cohort subsample if it meets the age criteria in its first wave observed. The dependent variable in panel A and B is a binary indicator for employment ("Working", from Table 1) and labour-force participation ("In the labour force", from Table 1), respectively. All specifications include individual fixed effects, wave and province dummies, and the time-varying controls listed in Table 1. Standard errors (clustered by individual) in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B2: Coefficient estimates, with gender and rural interactions

	Working			In the labour force		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Gender interactions						
CES-D8 score	−0.0026***			−0.0021***		
	(0.001)			(0.001)		
Moderate symptoms		−0.0100**			−0.0079*	
		(0.005)			(0.005)	
Severe symptoms			−0.0187***			−0.0141**
			(0.006)			(0.006)
R-squared	0.6102	0.6099	0.6101	0.6046	0.6044	0.6045
R-squared (within)	0.0274	0.0269	0.0272	0.0278	0.0274	0.0275
Panel B: Rural interactions						
CES-D8 score	−0.0027***			−0.0023***		
	(0.001)			(0.001)		
Moderate symptoms		−0.0112**			−0.0090*	
		(0.005)			(0.005)	
Severe symptoms			−0.0192***			−0.0145**
			(0.006)			(0.006)
R-squared	0.6065	0.6063	0.6064	0.6006	0.6004	0.6005
R-squared (within)	0.0184	0.0178	0.0181	0.0180	0.0176	0.0178
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	38,580	38,580	38,580	38,580	38,580	38,580

Note: CFPS data (2012, 2016, 2018). This table reports coefficient estimates for the model in Equation (1), amended to incorporate gender and regional interaction terms. Panel A interacts the gender dummy with all variables relating to age, education, marriage, children, elders, and region. Panel B interacts the rural dummy with all variables relating to age, education, marriage, children, elders, and gender. The dependent variable in each column is specified in the top row of the table. All specifications include individual fixed effects, wave and province dummies, and the time-varying controls listed in Table 1. Standard errors (clustered by individual) in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B3: Coefficient estimates, with time-varying regional dummies

	Working			**In the labour force**		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Wave × province dummies						
CES-D8 score	−0.0028***			−0.0023***		
	(0.001)			(0.001)		
Moderate symptoms		−0.0110**			−0.0086*	
		(0.005)			(0.005)	
Severe symptoms			−0.0189***			−0.0142**
			(0.006)			(0.006)
R-squared	0.6083	0.6081	0.6082	0.6025	0.6023	0.6023
R-squared (within)	0.0229	0.0223	0.0225	0.0226	0.0221	0.0223
Panel B: Wave × province × rural dummies						
CES-D8 score	−0.0028***			−0.0023***		
	(0.001)			(0.001)		
Moderate symptoms		−0.0107**			−0.0085*	
		(0.005)			(0.005)	
Severe symptoms			−0.0190***			−0.0148***
			(0.006)			(0.006)
R-squared	0.6106	0.6104	0.6105	0.6049	0.6047	0.6048
R-squared (within)	0.0286	0.0281	0.0283	0.0286	0.0281	0.0283
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	38,580	38,580	38,580	38,580	38,580	38,580

Note: CFPS data (2012, 2016, 2018). This table reports coefficient estimates for the model in Equation (1), with different sets of time-varying regional dummies: wave-by-province dummies, i.e., W×P (panel A), and wave-by-province-by-rural dummies, i.e., W×P×R (panel B). The dependent variable in each column is specified in the top row of the table. All specifications include individual fixed effects and the time-varying controls listed in Table 1. Standard errors (clustered by individual) in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B4: Coefficient estimates, using the CES-D20 scale

	Working			In the labour force		
	(1)	(2)	(3)	(4)	(5)	(6)
CES-D20 score	−0.0012***			−0.0010***		
	(0.000)			(0.000)		
Moderate symptoms		−0.0112**			−0.0098*	
		(0.005)			(0.005)	
Severe symptoms			−0.0184***			−0.0160**
			(0.006)			(0.006)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.6070	0.6059	0.6060	0.6011	0.6000	0.6001
R-squared (within)	0.0173	0.0168	0.0170	0.0169	0.0166	0.0168
Observations	38,413	38,413	38,413	38,413	38,413	38,413

Note: CFPS data (2012, 2016, 2018). This table reports coefficient estimates for the model in Equation (1), using measures of mental health derived from the CES-D20 score, rather than the CES-D8 (as in our baseline). The dependent variable in each column is specified in the top row of the table. All specifications include individual fixed effects, wave and province dummies, and the time-varying controls listed in Table 1. The mental health measures are (i) the CES-D20 score, and binary indicators of (ii) moderate depressive symptoms (i.e., $D^{20} \geq 18$) and (iii) severe depressive symptoms (i.e., $D^{20} \geq 22$). Standard errors (clustered by individual) in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B5: Coefficient estimates, with a balanced panel

	Working			In the labour force		
	(1)	(2)	(3)	(4)	(5)	(6)
CES-D8 score	−0.0028***			−0.0024***		
	(0.001)			(0.001)		
Moderate symptoms		−0.0060			−0.0055	
		(0.005)			(0.005)	
Severe symptoms			−0.0176***			−0.0151**
			(0.007)			(0.006)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.5509	0.5505	0.5506	0.5453	0.5450	0.5451
R-squared (within)	0.0209	0.0201	0.0205	0.0200	0.0193	0.0196
Observations	25,488	25,488	25,488	25,488	25,488	25,488

Note: CFPS data (2012, 2016, 2018). This table restricts the sample to a balanced panel, including only individuals observed in all three waves. The dependent variable in each column is specified in the top row of the table. All specifications include individual fixed effects, wave and province dummies, and the time-varying controls listed in Table 1. Standard errors (clustered by individual) in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B6: Coefficient estimates, with bounds

	$\delta = 0(\tilde{\beta})$	$\delta = 1(\beta^*)$	
		$R_{max} = 1.3\tilde{R}$	$R_{max} = 2.2\tilde{R}$
Panel A: Working			
Moderate symptoms	−0.0108**	−0.0104*	−0.0092
	(0.005)	[0.006]	[0.006]
Severe symptoms	−0.0191***	−0.0189***	−0.0180**
	(0.006)	[0.007]	[0.007]
Panel B: In the labour force			
Moderate symptoms	−0.0087*	−0.0084	−0.0074
	(0.005)	[0.006]	[0.006]
Severe symptoms	−0.0145**	−0.0142**	−0.0134*
	(0.006)	[0.007]	[0.007]
Individual FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	38,580	38,580	38,580

Note: CFPS data (2012, 2016, 2018). Bounds calculated using the method of Oster (2019). Bootstrapped standard errors in square brackets (2,000 reps). The specification includes wave dummies in the baseline regression, following Bryan et al. (2022). Standard errors (clustered by individual) in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B7: Effect of depressive symptoms on employment and labour-force participation

	Working			**In the labour force**		
	(1)	(2)	(3)	(4)	(5)	(6)
CES-D8 score	−0.0027***			−0.0022***		
	(0.001)			(0.001)		
Moderate symptoms		−0.0108**			−0.0087*	
		(0.005)			(0.005)	
Severe symptoms			−0.0191***			−0.0145**
			(0.006)			(0.006)
Married	−0.0698***	−0.0680***	−0.0683***	−0.0753***	−0.0737***	−0.0739***
	(0.014)	(0.014)	(0.014)	(0.013)	(0.013)	(0.013)
Health (self-assessed)	0.0115***	0.0123***	0.0121***	0.0109***	0.0116***	0.0115***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of adults	0.0059	0.0058	0.0059	0.0056	0.0055	0.0056
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
No children (in HH)	−0.0109	−0.0105	−0.0106	−0.0075	−0.0072	−0.0073
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Child aged 0–4	−0.0416***	−0.0415***	−0.0416***	−0.0467***	−0.0467***	−0.0467***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Child aged 5–11	−0.0012	−0.0010	−0.0009	0.0001	0.0004	0.0004
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
Child aged 12–15	0.0002	0.0004	0.0006	0.0019	0.0020	0.0022
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
Elder (present in HH)	−0.0072	−0.0071	−0.0072	−0.0073	−0.0072	−0.0073
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Reside in rural area	0.0244**	0.0243**	0.0245**	0.0137	0.0136	0.0138
	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)
Education (Middle)	0.0213	0.0211	0.0207	0.0092	0.0091	0.0088
	(0.025)	(0.025)	(0.025)	(0.023)	(0.023)	(0.023)
Education (High)	0.0961**	0.0958**	0.0949**	0.0926**	0.0924**	0.0916**
	(0.040)	(0.040)	(0.040)	(0.037)	(0.037)	(0.037)
Education (Degree)	0.0999*	0.0994*	0.0981*	0.0859*	0.0854*	0.0843*
	(0.052)	(0.052)	(0.052)	(0.049)	(0.049)	(0.049)
Age (26–30)	0.0728***	0.0722***	0.0721***	0.0636***	0.0631***	0.0629***
	(0.015)	(0.015)	(0.015)	(0.014)	(0.014)	(0.014)
Age (31–35)	0.0992***	0.0982***	0.0979***	0.0952***	0.0944***	0.0941***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Age (36–40)	0.1069***	0.1058***	0.1053***	0.1086***	0.1076***	0.1072***
	(0.024)	(0.024)	(0.024)	(0.023)	(0.023)	(0.023)
Age (41–45)	0.1192***	0.1180***	0.1176***	0.1216***	0.1206***	0.1203***
	(0.028)	(0.027)	(0.027)	(0.026)	(0.026)	(0.026)
Age (46–50)	0.0964***	0.0954***	0.0949***	0.1020***	0.1012***	0.1007***
	(0.032)	(0.032)	(0.032)	(0.030)	(0.030)	(0.030)
Age (51–55)	0.0683*	0.0673*	0.0667*	0.0815**	0.0806**	0.0801**
	(0.036)	(0.036)	(0.036)	(0.035)	(0.035)	(0.035)
Age (56+)	0.0099	0.0091	0.0083	0.0283	0.0276	0.0269
	(0.041)	(0.041)	(0.041)	(0.039)	(0.039)	(0.039)
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.6061	0.6059	0.6060	0.6002	0.6000	0.6001
R-squared (within)	0.0174	0.0168	0.0171	0.0170	0.0166	0.0168
Observations	38,580	38,580	38,580	38,580	38,580	38,580

Note: CFPS data (2012, 2016, 2018). This table reports coefficient estimates for the model in Equation (1). This table corresponds to Table 3 Panel A. Standard errors (clustered by individual) in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.