What Can We Learn About the Effect of Mental Health on Labor Market Outcomes Under Weak Assumptions? Evidence from the NLSY79.

Giuseppe Germinario[1]; Vikesh Amin[2]; Carlos A. Flores[3]; Alfonso Flores-Lagunes[4]

[1] Syracuse University
[2] Central Michigan University
[3] California Polytechnic State University
[4] Syracuse University, IZA, and GLO

Abstract
We employ a nonparametric partial identification approach to bound the causal effect of poor mental health on employment and earnings using the National Longitudinal Study of Youth 1979. Our approach allows us to provide bounds on the population average treatment effect based on relatively weak, credible assumptions. We also provide insights into the heterogeneity of the effects on labor market outcomes at different levels of adverse mental health experienced (no-to-mild, moderate, and severe depressive symptoms). We find that (1) being categorized as depressed decreases employment by 10% and earnings by 27% at most, but we cannot statistically rule out a zero effect, and (2) going from having no-to-mild to severe depressive symptoms reduces employment by 2-16% and earnings by 8-37%, with both estimated bounds statistically ruling out a zero effect.

Key Words: depression; mental health; employment; earnings; partial identification; bounds

We acknowledge research funding from NIH grant number 1R01HD094011-01. Vikesh Amin also acknowledges research funding from the College of Business Administration at Central Michigan University.

1. Introduction

It is well-documented that poor mental health is correlated with worse labor market outcomes. A review of early studies by Marcotte and Wilcox-Gök (2001) concluded that 5–6 million workers in the US between the ages of 16 and 54 lose, fail to seek, or cannot find employment due to mental illnesses, and that mental illness decreases annual income by $3,500-$6000, conditional on working. Poor mental health could limit employment and lower earnings because (1) it affects factors such as mood, memory or motivation, (2) it increases absenteeism and presenteeism, (3) of employer taste-based discrimination or (4) of employers' unwillingness to accommodate health problems. There is empirical evidence showing that poor mental health is associated with higher rates of absenteeism and presenteeism (Bubonya et al. 2017) and lower productivity (Oswald et al. 2015).[1,2] Establishing causality though is complicated because of omitted variable bias and reverse causality. Omitted variable bias arises due to "third" factors (e.g., genetic endowments, cognitive ability, childhood circumstances, personality) that are correlated with both mental health and labor market outcomes, while reverse causality occurs when lack of employment or reduced earnings worsen mental health.

To identify causal effects several studies (summarized in section 2) have used instrumental variables (IV) strategies to isolate variation in mental health that is argued to be orthogonal to unobserved factors. Instruments used have pertained to childhood and parental mental health, religiosity, death of a close friend, and availability of social services and social support. Other studies have used panel data and individual fixed-effect regression specifications to control for time invariant individual characteristics correlated with mental health and labor market outcomes. The consensus from prior studies is that poor mental health reduces employment but effects for earnings are mixed.

While IV and panel data approaches are an improvement upon naïve OLS estimates, the assumptions needed for causality may be violated. Valid instruments rely on the exclusion restriction—that the instruments only affect labor market outcomes indirectly through their effect on mental health. One can think of reasons why the exclusion restriction may not be satisfied for each of the instruments used in the literature. Childhood/adolescent mental health may violate the exclusion restriction because poor adolescent mental health is associated with lower educational attainment (Fletcher 2008) and worse labor market outcomes (Fletcher 2013a; Lundberg et al. 2014; Mousteri et al 2019). Parental mental illness has been found to increase the probability of high school dropout of children (Farahati et al. 2003), which could in turn affect labor market outcomes. Religiosity is likely correlated with earnings (Gruber 2005) and could affect labor market outcomes through other indirect channels (e.g. attending religious services could be helpful for career networking). Furthermore, in the presence of heterogeneous effects, IV identifies a local average treatment effect (LATE) for those individuals whose treatment (mental health) is affected by the instrument used—the so-called "compliers" (Imbens and Angrist 1994). LATE estimates derived from random shocks such as death of a close friend are potentially less informative for policy. Individual fixed-effect estimates are identified from individuals who experience changes in mental health and labor market outcomes. If within-person variation in mental health and labor market outcomes are randomly distributed across all individuals, then individual fixed-effect estimates will identify the average treatment effect (ATE) on the population,

[1] Using the first 14 waves of Household, Income and Labour Dynamics in Australia and conditional fixed-effect logit models, Bubonya et al. (2015) find that absence rates are five percent higher among workers who report being in poor mental health. They also find that the odds that workers in poor mental health report diminished productivity due to emotional problems is six times higher than those of similar workers in good mental health.

[2] Oswald et al. (2015) conduct a series of laboratory experiments at an elite university in the UK where students are randomly assigned "happiness" (showing comedy movie clips, providing chocolate, fruit and drinks) before completing a piece rate task. They find treated individuals have 12% higher productivity.

which is arguably more policy relevant. Alternatively, the estimates could reflect an effect for a specific subpopulation if the within-person variation is present in certain subgroups of the population. Moreover, if there is insufficient within-person variation in mental health and labor market outcomes, then estimates may lack statistical power to detect effects of mental health. Individual fixed-effects also do not control for time varying factors correlated with mental health and labor market outcomes, such as stress and health shocks. In addition, attenuation bias due to random measurement error in mental health is typically exacerbated in fixed-effects models.

We estimate the effect of mental health (based on a count of depressive symptoms from the Center for Epidemiologic Studies Depression Scale) in the 1992 round (when individuals are 30 years old) of the National Longitudinal Study of Youth 1979 (NLSY79) on employment and earnings in 1993. Our key contribution is that we provide novel evidence by employing a nonparametric partial identification approach (Manski and Pepper 2000) to bound the causal effect, which has at least three advantages. First, it provides bounds on the population ATE as opposed to a subpopulation. Second, it allows for arbitrary correlations between mental health and unobserved factors that can affect employment and earnings. Third, it relies on relatively weak assumptions, which are, arguably, more credible. These advantages though come at a cost of obtaining a range of possible values for the causal population ATE rather than a point estimate of it. The assumptions we use are three: (1) monotone treatment selection (MTS) which posits that individuals "selected" into worse mental health have lower latent employment probabilities and earnings. (2) Monotone treatment response (MTR) which imposes the restriction that worse mental health does not improve labor market outcomes. (3) We employ adolescent test scores as a monotone instrumental variable (MIV)—a variable that is assumed to have a weakly increasing mean relationship with potential outcomes—to help tighten the bounds under the MTS and MTR assumptions. In our context, the MIV assumption states that individuals with higher test scores have no lower average latent employment probabilities and earnings than those with lower test scores. The MIV assumption is weaker than the exclusion restriction in IV models, since the MIV is allowed to have a direct impact on potential outcomes and may be non-random itself. We discuss and assess these assumptions below. The nonparametric partial identification method that we employ has been used in other contexts but to our knowledge has not been applied to bound effects of mental health on labor market outcomes.[3] Lastly, since mental health is inherently a continuous condition but is typically measured in the literature with an indicator variable for being depressed or having a psychiatric condition, we provide insights into the effects at different levels of adverse mental health experienced (no-to-mild, moderate, and severe depressive symptoms).

Our estimated bounds indicate that being categorized as depressed decreases employment by 10% and earnings by 27% at most, but we cannot statistically exclude a zero effect. When considering different levels of depression, we find that going from having no-to-mild to severe depressive symptoms reduces employment by 2-16% and earnings by 8-37%, with both estimated bounds statistically ruling out zero effects. Therefore, our results provide evidence of a statistically significant average causal effect of going from having no-to-mild depression symptoms to having severe depression symptoms on labor market outcomes. They also allow ruling out relevant magnitudes of the effects of depression on labor market outcomes.

The remainder of the paper is organized as follows. Section 2 provides a review of the literature and section 3 gives a description of the data. Our econometric approach is explained in section 4 and the results are presented in section 5. Finally, section 6 concludes.

---

[3] Examples of previous applications include bounding the causal effect of parents' schooling on children's schooling (De Haan 2011), unemployment on mental health (Cyagn-Rehm et al. 2017), English proficiency on labor market outcomes (Gonzalez 2005), criminal convictions on labor market outcomes (Richey 2015, education on social support (Huang et al. 2012), and social activities on cognition (Christelis and Dobrescu 2020).

2. Literature Review

Most of the IV evidence on the effect of mental health on labor market outcomes for the US comes from analysis of the National Comorbidity Survey (NCS), a nationally representative survey of 15-54 year-olds designed to study the prevalence, causes and consequences of comorbidity between substance abuse disorders and nonsubstance abuse psychiatric disorders. Using the 1990-1992 NCS and instrumenting adult mental health with childhood and parental mental illnesses, Ettner et al. (1997) found that a diagnosis of any psychiatric disorder in the last year reduced the probability of being employed by 11 percentage points. They also found negative effects of a past year psychiatric disorder on earnings, though IV estimates were only statistically significant for women. Macotte and Wilcox-Gök (2003) estimated the effect of depression, dysthymia, anxiety disorder and anti-social personality disorder on earnings in the 1990-1992 NCS using measures of parental mental health as the instruments. They only found a statistically significant negative effect for anxiety disorder on earnings for women, with IV estimates showing that an anxiety disorder in the past year reduced earnings by 49%. Based on IV quantile regressions, they did however find large negative effects of all four disorders at the lower tail of the earnings distribution. Chatterji et al. (2011) estimated the effect of a past year psychiatric disorder on employment and earnings in the 2001-2003 National Comorbidity Survey-Replication, using childhood mental health and religiosity as instruments. Their IV estimates showed that a past year psychiatric disorder reduced the likelihood of employment by 17 percentage points for men and 9 percentage points for women.[4,5] Unlike Ettner et al. (1997) and Marcotte and Wilcox-Gök (2003), Chatterji et al. (2011) did not find any evidence of negative effects for earnings. They conjecture that the difference in findings could be due to the US economy being stronger in 2001-2003 compared to 1990-1992.

Other IV studies for the US not using the NCS are Alexandre and French (2001), Chatterji et al. (2007) and Ojeda et al. (2010). Alexandre and French (2001) use survey data collected from a low-income neighborhood in Miami-Dade County, 1996-1997. Using religiosity and social supports as instruments they found that depression reduces employment by 19 percentage points. Chatterji et al. (2007) estimate labor market effects of a past year psychiatric disorder in the 2002-2003 National Latino and Asian American Study using childhood mental health and religiosity as instruments. They found large effects for Latinos— a past year psychiatric disorder decreased employment by 8 percentage points for men and 26 percentage points for women— and smaller negative effects of 1-3% on employment for Asians. Ojeda et al. (2010) estimate the effect of the K6 Scale of Mental Illness on labor supply for natives and immigrants in the 2002 National Survey on Drug Use and Health using social support as the IV. They found that immigrants' labor supply is less responsive to mental health problems than the labor supply of natives.

Results from US panel data studies also indicate detrimental effects of poor mental health on employment, though the magnitudes are substantially smaller than IV estimates, and there is mixed evidence for earnings. Peng et al. (2015) used the 2004-2009 Medical Expenditure Panel Survey and correlated random effects specifications to control for omitted variable bias. They found that depression reduces the probability of employment by 2.6 percentage points but didn't affect earnings. Cseh (2008) estimated regressions with individual fixed-effects using the 1992,

---

[4] These IV estimates are from appendix tables 2A and 2B in their NBER 2008 working paper (Chatterji et al. 2008). The IV results for employment are not shown in their published paper (Chatterji et al. 2011).

[5] Findings were similar based on bivariate probit models using the methods in Altonji et al. (2005), where identification comes from making assumptions about the correlation between unobserved and observed factors determining the outcome and endogenous variable, and the functional form. The bivariate probit estimates showed that a past year psychiatric disorder decreases employment by 14 percentage points for men and 13 percentage points for women.

1994 and 2004 rounds of the NLSY79, and found that depression didn't affect wages for women, but reduced wages of men by 3.4-4.3%.

Detrimental effects of poor mental health on employment have also been reported for other countries. For example, Frijters et al. (2014) estimate the effect of mental health on employment using 10 waves of the Household, Income and Labour Dynamics in Australia survey. They construct a mental health index based on nine questions from the short-form general health survey. They use death of a close friend as an instrument and estimate IV regressions with individual fixed-effects. They find that a one standard deviation decrease in the mental health index decreases employment by 30 percentage points. Bryan et al. (2020) use nine waves of the UK Household Longitudinal Study and estimate individual fixed-effects regressions which indicate that depression reduces employment by 1.6 percentage points. In a recent study, Biasi, et al. (2021) leverage the approval of lithium as a treatment for bipolar disorder in Denmark in 1976 as a natural experiment to identify the causal effect of access to treatment on career earnings. They compare those with access at age 20—the usual age for onset of bipolar disorder—to those born prior to 1956, using administrative data. They find that access to treatment reduces the observed earnings difference of 38% between those with bipolar disorder and those without by 28%, and reduces the probability of zero earnings by 33%.

In sum, previous studies have consistently found negative effects of poor mental health on employment in the US and internationally, and mixed findings for earnings. IV estimates typically show larger detrimental effects of poor mental health on employment than individual fixed-effect estimates.

3. Data

We use the NLSY79 which is a nationally representative panel study that follows 12,686 individuals from 1979, when they were aged 14–22, through the present. We measure mental health using the Center for Epidemiologic Studies Depression Scale (CES-D). The CES-D was developed to provide a self-reported measure of symptoms of clinical depressive episodes to allow for the epidemiological study of psychiatric problems in the general population (Radloff 1977). The CES-D assesses depressive symptoms as a continuum and measures the frequency of depressive symptoms over one week with 20 items. For example, respondents are asked how often in the last week they were bothered by things not normally bothersome; could not shake the blues; and felt like they were not as good as others. With four response categories (0 = none to 3 = almost every day), the maximum score of the scale is 60. A score of 16 or more on the CES-D indicates that one has a higher risk of clinical depression. Individuals can also be classified as having no-to-mild depressive symptoms (CES-D score of 0-15), moderate depressive symptoms (CES-D score of 16-23) and severe depressive symptoms (CES-D score of 24-60) (Center for Child and Human Development, Georgetown University 2021).[6] The original 20-item CES-D was first administered at the 1992 round, when individuals were 30 years old on average. A total of 9,016 individuals were interviewed in 1992, of which 8,978 have valid data on the CES-D score.

We estimate the effect of mental health measured in 1992 on employment and annual earnings in the following year, 1993, to reduce concerns of reverse causality. Employment is measured by a dummy variable equal to one if the individual reports working positive hours, and zero if the individual reports working zero hours. Individuals who report being employed but work zero hours, and individuals in the armed forces are coded as missing. A useful aspect of the NLSY79 is that it contains Armed Forces Qualification Test (AFQT) scores from the 1981 round when respondents were between 16-21 years old. As we describe in the next section, we use the AFQT score as a monotone instrumental variable. In total we have a sample of 7,665 individuals

[6]We also considered other classifications, such as no to very mild, mild, severe, and very severe depression symptoms. The results are in general similar to those obtained with the 3-group classification we use herein.

with non-missing information on depressive symptoms, employment, earnings and the AFQT score. Throughout our analysis we use the 1993 panel weights provided in the NLSY79.[7]

Summary statistics for our estimation sample are shown in Table 1. In the full sample (column 1) the average age in 1992 is 31 years and 51% are female. White individuals make up 80% of the sample, and there is a higher proportion of Black individuals (14%) than Hispanic individuals (6%).[8] The average CES-D score in 1992 is 9.04 and 21% are classified as depressed (CES-D≥16). Most individuals (82%) have no-to-mild depressive symptoms (0≤CES-D≤15), while 10% have moderate (16≤CES-D≤23) and 8% have severe (CES-D≥24) depressive symptoms. Women have worse mental health than men—22% of women are depressed whereas 15% of men are depressed. Black and Hispanic individuals are also more likely to be depressed than whites. The 1993 employment rate in the full sample is 88%, but there are differences by gender and race. The employment rate of men (94%) is substantially higher than women (82%). Similarly, white individuals have a higher employment rate (89%) than Black (80%) and Hispanic (84%) individuals. A similar pattern is also observed for earnings. Average earnings (in 1993 dollars) are higher for men than women, and for white individuals compared to Black and Hispanic individuals.

4. Methodology and Identifying Assumptions

Let every individual i have a response function $Y\ (\cdot): T \rightarrow Y$ which maps treatments $t \in T$ into potential outcomes $y_i\ (t) \in Y$. In our context, the treatment t is depressive symptoms measured by the CES-D score in the form of either a binary indicator for depression or a discretized version consisting of three levels: no-to-mild symptoms, moderate symptoms, and severe symptoms. The outcome Y is employment or earnings. Let S_i denote the realized treatment received by individual i, so that $Y_i \equiv \sum_{t \in T} 1\{S_i = t\} \cdot Y_i\ (t)$ is the associated observed outcome.

To illustrate how the bounds are obtained, in this section we will focus on the case of three treatment levels $t_1, t_2, \text{ and } t_3$ which respectively correspond to no-to-mild, moderate, and severe depressive symptoms.[9] We are interested in the population ATE of, for example, worsening mental health from t_1 to t_2 on labor market outcomes defined as:

$$(1)\quad \Delta\ (t_1, t_2) = E[Y(t_2)] - E[Y(t_1)]$$

Estimation of the ATE is complicated because the potential outcome $y(t_2)$ is unobserved for individuals with treatment level of t_1, and the potential outcome $y(t_1)$ is unobserved for individuals with treatment level of t_2. This identification problem can be seen by using the law of iterated expectations to write the expected potential outcome $E[Y(t_1)]$ as:

$$(2)\ E[Y(t_1)] = E[Y(t_1)|s = t_1] * P(S = t_1) + E[Y(t_1)|S \neq t_1] * P(S \neq t_1)$$

The data identify the sample analogues of all the right-hand side quantities except of the counterfactual $E[Y(t_1)|S \neq t_1]$. We thus need to impose identifying assumptions about this missing counterfactual. Manski (1989) suggested a bounded support assumption, whereby one

[7]The information on primary sampling units (PSUs) for the NLSY79 is part of the restricted geocode data, which presents a complication for the bootstrapping inside our estimation and inference procedure described below. We use the respondent's Census region as the PSU, which is the finest level of geographical information available in the public use dataset. We also follow Altonji, et al. (2012) and stratify based on race and sex for the bootstrap samples.

[8]Respondents reported their race/ethnicity at baseline, which was mapped into three groups: black, Hispanic, and non-black/non-Hispanic. We refer to the last group as white, although it is a heterogenous group.

[9]A similar discussion would follow for a binary treatment defined by an indicator for depression.

uses the minimum (Y_{min}) and maximum (Y_{max}) of the outcome variable in place of $E[Y(t_1)|S \neq t_1]$. This gives Manski's (1989) "no-assumption" bounds:

$$(3)\ E[Y|S = t_1] * P(S = t_1) + Y_{min} * P(S \neq t_1) \\ \leq E[y(t_1)] \leq \\ E[Y|S = t_1] * P(S = t_1) + Y_{max} * P(S \neq t_1)$$

The no-assumption lower (respectively, upper) bound on the ATE $\Delta\,(t_1, t_2)$ is calculated by subtracting the upper (lower) bound of $E[Y(t_1)]$ from the lower (upper) bound of $E[Y(t_2)]$. Bounds for other treatment effects such as $\Delta\,(t_2, t_3)$ or $\Delta\,(t_1, t_3)$ are computed analogously. In practice, the no-assumption bounds are typically wide and uninformative and they contain zero by construction. To tighten the bounds, we employ three monotonicity assumptions introduced in Manski (1997) and Manski and Pepper (2000): (1) monotone treatment selection; (2) monotone treatment response, and (3) monotone instrumental variable.

4.1 Monotone Treatment Selection (MTS)

The MTS assumption captures the notion that individuals who "selected" into higher CES-D scores have lower latent employment probability and earnings. Formally, the non-positive MTS assumption states that for each $t \in T$ and two treatment levels μ_1 and μ_2

$$(4)\ \mu_2 \geq \mu_1 \Rightarrow E[Y(t) \mid S = \mu_2] \leq E[Y(t) \mid S = \mu_1]$$

In our context, the MTS assumption requires that, on average, individuals with worse mental health have weakly lower potential outcomes than individuals with better mental health. The MTS assumption is untestable since potential outcomes are unobserved. However, there is evidence that poor mental health is associated with traits such as lower measured intelligence and higher neuroticism which are in turn associated with lower employment and earnings (Lin et al., 2018; Fletcher 2013b). Moreover, we can indirectly shed light on the plausibility of the MTS assumption in our sample by computing average characteristics of individuals in different categories of depressive symptoms. Intuitively, we would expect to see that individuals with worse depressive symptoms show traits correlated with worse labor market outcomes.[10] Table 2 provides summary statistics for some observed characteristics by no-to-mild, moderate, and severe depressive symptoms. Average parental education is lower for individuals with moderate and severe depressive symptoms compared to individuals with no-to-mild depressive symptoms. Individuals with moderate and severe depressive symptoms also have statistically significantly lower AFQT scores in adolescence compared to individuals with no-to-mild depressive symptoms. Furthermore, in adolescence they are also statistically significantly less likely to expect to graduate from college.[11] Education attained in adulthood (by 1993) may be influenced by mental health, but it is suggestive that individuals with moderate and severe depressive symptoms have statistically significantly fewer years of education and are less likely to be college graduates. Given the observed differences in those characteristics, the MTS assumption seems plausible because individuals with worse mental health likely come from lower SES families, have lower ability and education, all of which are associated with worse labor market outcomes. Alternatively,

[10] In principle, this exercise should use characteristics that are not themselves affected by the individual's mental health. Some of the characteristics in Table 2 satisfy this condition (e.g., mother's and father's education), while others may not (e.g., years of education). We still present averages for the latter variables for reference.

[11] At the 1981 round in the NLSY respondents were asked "as things now stand, what is the highest grade or year you think you will actually complete?". We classify individuals as expecting to graduate college if they report 4th year of college, or 5th year of college, or 6th+ year of college.

the MTS assumption would be violated if individuals with severe (moderate) depressive symptoms have better potential outcomes than individuals with moderate (no-to-mild) depressive symptoms. This situation seems unlikely given that average parental education, AFQT scores, educational expectations, and educational attainment are all (weakly) lower for individuals with moderate and severe depressive symptoms compared to individuals with no-to-mild depressive symptoms.[12]

To illustrate the derivation of the bounds on the ATE $\Delta\,(t_1, t_2)$ under the MTS assumption, consider bounding the term $E[Y(t_2)]$ in equation (1). Using the law of iterated expectations, we can write $E[Y(t_2)]$ as:

$$E[Y(t_2)] = E[Y(t_2)|S < t_2] * P(S < t_2) + E[Y(t_2)|S = t_2] * P(S = t_2) + E[Y(t_2)|s > t_2] * P(S > t_2)$$

Then, the MTS bounds on $E[Y(t_2)]$ are given by (Manski and Pepper, 2000):

$$\begin{gathered}(5)\ E[Y|S = t_2] * P(S < t_2) + E[Y|S = t_2] * P(S = t_2) + Y_{min} * P(S > t_2) \\ \leq E[Y(t_2)] \leq \\ Y_{max} * P(S < t_2) + E[Y|S = t_2] * P(S = t_2) + E[Y|s = t_2] * P(S > t_2)\end{gathered}$$

Notice how MTS tightens the bounds on $E[Y(t_2)]$ relative to the no-assumption case. First, for the conditional mean potential outcomes $E[\,Y(t_2) \mid S < t_2\,]$, we previously could only conclude that this is bounded below by Y_{min} due to the assumption of bounded support on the outcome. However, under MTS, equation (4) further implies that this cannot be less than $E[\,Y(t_2) \mid S = t_2\,]$, which is identified by the observed mean for those receiving t_2. Thus, the lower bound on $E[Y(t_2)]$ under MTS will exceed the no-assumption lower bound to the extent that $E[Y|S = t_2]$ is larger than Y_{min}.Similarly, for the conditional mean potential outcome $E[\,Y(t_2) \mid S > t_2\,]$, bounded support alone could only yield an upper bound using Y_{max}. Again using equation (4), MTS implies that the unidentified quantity can be no larger than $E[\,Y(t_2) \mid S = t_2\,]$, or the observed mean for $S = t_2,\ E[Y|S = t_2]$. Therefore, the upper bound on $E[Y(t_2)]$ under MTS will be below that from the no-assumption case.

As before, the lower (respectively, upper) bound on the ATE $\Delta\,(t_1, t_2)$ is calculated by subtracting the upper (lower) bound of $E[Y(t_1)]$ from the lower (upper) bound of $E[Y(t_2)]$, and likewise for other comparisons of interest.

4.2 Monotone Treatment Response (MTR)

The non-positive MTR assumption we employ imposes the restriction that worse mental health (higher CES-D score) does not improve labor market outcomes. Formally, for each individual and any treatment levels t_k and t_j:

$$(6) t_j \geq t_k \Longrightarrow Y(t_j) \leq Y(t_k)$$

We believe that MTR is a reasonable assumption in our context because of the well-documented correlations between poor mental health and worse labor market outcomes (e.g., Chatterji et al. 2007, 2011; Ettner et al. 1997; Marcotte and Wilcox-Gök 2001) and evidence that worse mental health is associated with higher rates of absenteeism, “presenteeism” (Bubanya et

[12] Corresponding summary statistics by gender and by race are given in Online Appendix tables A1 and A2. A similar pattern is observed to that in Table 2. The one notable point is that for Hispanic individuals average parental education, AFQT scores, college expectations and educational attainment are higher for individuals with severe depressive symptoms compared to individuals with moderate depressive symptoms. However, those differences are not statistically different from zero, and thus they are consistent with the weak inequality in the MTS assumption.

al. 2017), and lower productivity (Oswald et al. 2015). Theoretical models also imply that poor health leads to worse labor market outcomes. For example, the Grossman (1972) health investment model shows that poor health reduces time available for work because of increased time spent being ill, increased preferences for leisure, or increased time needed to maintain health. Health is also an input into the production function for human capital, and those with worse health have lower human capital investments. Given the positive correlation of human capital investments and labor market outcomes, as well as the economic theory behind the positive effect of human capital accumulation on labor market outcomes (e.g., Card 1999), MTR is consistent with those theories.

A key implication from MTR is that, for example, $E[Y(t_2)|S = t_\ell] \leq E[Y(t_1)|S = t_\ell]$ for any ℓ, given that $t_2 > t_1$. This provides tighter bounds on, say $E[Y(t_2)]$, relative to the no-assumption case in the following way. For any treatment levels $t < t_2$, MTR implies that the conditional mean $E[Y(t_2)|S = t]$ is no greater than $E[Y(t)|s = t]$, or the observed mean of Y at t, $E[Y|S = t]$. This decreases the upper bound on $E[Y(t_2)]$, relative to that obtained from the bounded support assumption alone. Further, for treatment levels $t' > t_2$, MTR implies that the conditional mean $E[Y(t_2)|s = t']$ cannot lie below $E[Y(t')|s = t']$, which is identified by the observed mean of Y at t', $E[Y|s = t']$. This raises the lower bound on the unconditional mean $E[Y(t_2)]$ when compared to the no-assumption lower bound.

The MTR bounds on $E[Y(t_2)]$ are given by (Manski, 1997):

$$(7)\; Y_{min} * P(S < t_2) + E[Y|S = t_2] * P(S = t_2) + E[Y|S > t_2] * P(S > t_2) \\ \leq E[Y(t_2)] \leq \\ E[Y|S < t_2] * P(S < t_2) + E[Y|S = t_2] * P(S = t_2) + Y_{max} * P(S > t_2)$$

The MTR lower (respectively, upper) bound on the ATE $\Delta\,(t_1, t_2)$ is calculated by subtracting the upper (lower) bound of $E[Y(t_1)]$ from the lower (upper) bound of $E[Y(t_2)]$. Under the non-positive MTR assumption the upper bound on $\Delta\,(t_1, t_2)$ is never above zero, because the MTR rules out the possibility that worse mental health improves labor market outcomes.

The MTR and MTS assumptions can be combined to provide tighter bounds on the mean potential outcomes. Taking again the example of $E[Y(t_2)]$, recall that, among treatment levels $t < t_2$, MTS worked to increase the lower bound, while MTR decreases the upper bound. When imposed simultaneously, the bounds are narrowed from both sides, as we are able to conclude that $E[Y(t_2)|s = t]$ lies between the observed mean at t_2 (from below) and the observed mean at t (from above). In a similar way, the conditional mean $E[Y(t_2)|S = t']$, for any $t' > t_2$, lies weakly above the observed mean at t' (due to MTR) and weakly below the observed mean at t_2 (due to MTS).

The MTR-MTS bounds on $E[Y(t_2)]$ are given by (Manski and Pepper, 2000):

$$(8)\; E[Y|S = t_2] * P(S < t_2) + E[Y|S = t_2] * P(S = t_2) + E[Y|S > t_2] * P(S > t_2) \\ \leq E[Y(t_2)] \leq \\ E[Y|S < t_2] * P(S < t_2) + E[Y|S = t_2] * P(S = t_2) + E[Y|S = t_2] * P(S > t_2)$$

As before, the MTR-MTS lower (upper) bound on the ATE $\Delta\,(t_1, t_2)$ is calculated by subtracting the upper (lower) bound of $E[Y(t_1)]$ from the lower (upper) bound of $E[Y(t_2)]$. Note that the MTS and MTR assumptions imposed together yield a testable implication that observed mean labor market outcomes are weakly decreasing in the CES-D score. That is, for any two treatments t_k and t_j, $t_j > t_k$ implies that $E\left[Y\middle|S = t_j\right] \leq E[Y|S = t_k]$. Table 3 shows that average labor market outcomes are decreasing as a function of depressive symptoms, consistent with the implication from combining the MTS and MTR assumptions. The employment rate (respectively, average earnings) for individuals with no-to-mild depressive symptoms is 90% (\$22,830); in contrast, it is

82% ($16,752) for individuals with moderate depressive symptoms and 75% ($12,509) for individuals with severe depressive symptoms.[13]

4.3 Monotone Instrumental Variable (MIV)

The MTR-MTS bounds can be further narrowed by using a monotone instrumental variable (MIV), which is a variable that has a monotone (weakly increasing or weakly decreasing) mean relationship with the potential outcomes $Y(t)$. The MIV assumption is weaker than the exclusion restriction in IV models, which requires the instrument to be mean independent of the outcome. The MIV assumption also does not require that the variable has a causal effect on the outcome. Specifically, a weakly increasing MIV Z satisfies:

$$(9)\quad m_1 \le m \le m_2 \;\Rightarrow\; E[Y(t)|Z = m_1] \le E[Y(t)|Z = m] \le E[Y(t)|Z = m_2]$$

for all treatment levels $t \in T$.

In this paper, we use adolescent AFQT test scores as a MIV. Using this measure of cognitive skills or ability is easy to justify based economic models of human capital (e.g., Ben-Porath 1967). These models imply that higher innate ability is related to higher labor market outcomes both directly and indirectly through education. Those models, as well as our MIV assumption, are consistent with the well-documented positive relationship between the AFQT (and in general adolescent cognitive ability) and better labor market outcomes, as well as on genetic correlations between intelligence and household income.[14]

With a variable Z satisfying the MIV assumption, we can divide the sample into bins defined by the values of Z and compute the MTR-MTS bounds within each bin. In our case of a non-negative MIV, equation (9) implies that the lower bound on $E[Y(t_2)|Z = m]$ is no lower than the lower bound on $E[Y(t_2)|Z = m_1]$, and its upper bound is no higher than the upper bound on $E[Y(t_2)|Z = m_2]$. For the bin where Z has a value of m, we can thus obtain a new lower bound by taking the largest lower bound over all bins where $Z \le m$. Likewise, we can obtain a new upper bound by taking the smallest upper bound over all bins where $Z \ge m$. The MIV-MTR-MTS bounds are then obtained by taking the weighted average over all the conditional bounds (which follows from the law of iterated expectations):

$$
\begin{aligned}
(10)\ \textstyle\sum_{m \in M} P(Z = m) * \left[max_{m_1 \le m}\, LB_{E[Y(t_1)|Z = m_1]}\right] \\
\le E[Y(t_1)] \le \\
\textstyle\sum_{m \in M} P(Z = m) * \left[min_{m_2 \ge m}\, LB_{E[Y(t_1)|Z = m_2]}\right]
\end{aligned}
$$

The MIV-MTR-MTS lower (upper) bound on the ATE $\Delta\,(t_1, t_2)$ is calculated once again by subtracting the upper (lower) bound of $E[Y(t_1)]$ from the lower (upper) bound of $E[Y(t_2)]$.

We make an additional observation regarding the MIV-MTR-MTS bounds which cannot be easily seen from Equation (10), but which can allow for further tightening of the bounds in practice. Consider the ATE in (1): when MTR in (6) is imposed alongside an MIV it implies that, within each cell of the MIV, the lower bound on $E[Y(t_2)]$ must be at least as large as the upper bound on $E[Y(t_1)]$ (with $t_2 > t_1$). This implication of MTR (that bin-specific ATEs are non-

[13] Online Appendix table A3 shows that mean employment and earnings are also decreasing in the CES-D score for men, women, white and Black individuals. Hispanic individuals with severe depressive symptoms have higher employment and earnings on average than individuals with moderate depressive symptoms, but the differences are not statistically significantly different from zero, and thus statistically consistent with the weak inequality in the testable implication.

[14] Using the UK Biobank Hill et al. (2019) find a genetic correlation of 0.69 between intelligence and income.

negative) represents a potential source for tightening the bounds relative to not assuming MTR. This was originally pointed out by (McCarthy et al. 2015).

4.4 Estimation and Inference

The no-assumption, MTS, MTR, and MTR-MTS bounds are all estimated straightforwardly by plugging in sample analogs for the expectations and probabilities in the corresponding bounds' expressions. Inference is undertaken by constructing Imbens and Manski (2004) confidence intervals. Estimation and inference under the MIV-MTR-MTS bounds require that we deal with two issues which have been noted since Manski and Pepper (2000). The first is that the plug-in estimators of Equation (10)—an example of so-called intersection bounds—suffers from bias in finite samples that make them narrower relative to the corresponding true identified set. The bias then carries over to estimated bounds on the average treatment effects of interest. The second, related issue is that the corresponding confidence intervals do not have the expected coverage at the desired level. Both of these issues arise because of the non-concavity and non-convexity, respectively, of the min and max operators in equation (10).

We address both issues in the bounds involving the MIV assumption by employing the estimation and valid-inference procedure in Chernozhukov et al. (2013; hereafter, CLR) for intersection bounds.[15] The CLR procedure allows us to obtain lower and upper bound estimators that satisfy a half-median unbiasedness property, that is, the estimated lower/upper bound will fall below/above the true lower/upper bound with a probability of at least one-half asymptotically. This property is important because Hirano and Porter (2012) showed that there exist no locally asymptotically unbiased estimators of parameters that contain min and max operators, implying that methods aimed at reducing bias (such as the bootstrap) cannot completely eliminate it and reducing bias too much eventually leads the variance of such methods to increase significantly. The details on our implementation of the CLR procedure can be found in Online Appendix B.

5. Results

5.1 Main Results

Table 4 provides results for the effect of being depressed on employment and earnings. The OLS estimates in column 1 indicate that depressed individuals are 10 percentage points less likely to be employed, and their earnings are $7,931 less than non-depressed individuals. OLS estimates are unlikely to be causal because of unobserved factors that are correlated with mental health and labor market outcomes. Columns 2-6 provide bounds and 95% confidence intervals on the ATE of being depressed using the nonparametric partial identification procedure described in section 3. The no assumption bounds in column 2 are wide, which is typical of this kind of bounds. They indicate that the true causal effect of depression on employment ranges from -77 to 23 percentage points, and from -$34,516 to $66,441 for earnings. Still, they rule out, without assumptions, detrimental effects from depression that are larger than 77 percentage points and $34,516 for employment and earnings, respectively. Adding the MTS assumption—that depressed individuals have lower average potential labor market outcomes than individuals who are not depressed—substantially increases the lower bounds. The MTS bounds in column 3 indicate that the effect of depression is to at most reduce employment by 11 percentage points and earnings by $7,935. The MTS bounds are still wide and null effects as well as relatively large positive effects cannot be ruled out. Column 4 shows bounds under the MTR assumption, which states that worse mental health does not improve labor market outcomes. This means that the upper bound under the MTR assumption is zero, while the lower bound is the same as in the no

[15] See also Flores and Flores-Lagunes (2013) for additional discussion on the CLR procedure and an application estimating bounds on local average treatment effects without the exclusion restriction under a different set of monotonicity assumptions.

assumption bounds. The combination of the MTR and MTS assumptions in column 5 provide considerably tighter bounds compared to the previous bounds. The bounds for employment and earnings are [-0.11, 0.00] and [-$7,935, $0.00], respectively. Adding the MIV assumption to the MTR+MTS bounds further narrows the bounds. The MIV+MTS+MTR bounds in column 6 indicate that being depressed decreases employment by at most 9 percentage points and earnings by at most $6,082. These estimates correspond to effects of at most 10% for employment and 27% for earnings relative to average employment and earnings for individuals who are not depressed.[16] These effects are smaller than the effect indicated by the OLS point estimates, though the OLS point estimates are included in the corresponding 95% confidence interval for employment, but not for earnings. The estimated bounds do not exclude zero and are consistent with both relatively large effects reported in IV studies and small effects reported in studies using individual fixed-effects of depression. However, it is important to recall that in the presence of heterogeneous treatment effects, IV and individual fixed-effect methods estimate a parameter that is different from the ATE, which is the one we are bounding.

Table 5 presents bounds on the ATE of going from having (1) no-to-mild to moderate depressive symptoms, (2) moderate to severe depressive symptoms and (3) no-to-mild to severe depressive symptoms. OLS estimates show large labor market disparities between individuals with no-to-mild and moderate depressive symptoms. Individuals with moderate depressive symptoms are 8 percentage points less likely to be employed compared to individuals with no-to-mild depressive symptoms, and their earnings are $6,077 lower. Employment and earnings differences between individuals with moderate and severe depressive symptoms are slightly smaller; the employment gap is 7 percentage points and the earnings differential is $4,243. The largest labor market disparities are observed between individuals with severe and no-to-mild depressive symptoms, with an employment gap of 15 percentage points and an earnings differential of $10,321. As before, the OLS estimates are likely not causal due to omitted variable bias. The pattern of results from sequentially adding the MTS, MTR and MIV assumptions is the same as in table 4. The narrowest bounds are provided by employing the MTS, MTR and MIV assumptions in column 6. These bounds do not rule out a null effect for going from having no-to-mild to moderate depressive symptoms. The lower bounds though suggest slightly smaller effects than the corresponding OLS estimates, although the latter are inside the bounds' 95% confidence intervals. The effect of going from having no-to-mild to moderate depressive symptoms is to decrease employment by at most 7 percentage points and earnings by at most $5,045. These effects correspond to decreases of at most 8% for employment and 22% for earnings, relative to average employment and earnings for individuals with no-to-mild depressive symptoms. The MIV+MTS+MTR bounds on the effect for going from having moderate to severe depressive symptoms also do not exclude zero and the OLS estimates are within the estimated bounds. The lower bound indicates that going from moderate to severe depressive symptoms at most reduces employment by 12 percentage points (15%) and earnings by $7,390 (44%).

Notably, the results in Table 5 show that a null effect is statistically excluded (at the 95% confidence level) in the estimated bounds for severe vs no-to-mild depressive symptoms under the MTS, MTR and MIV assumptions. These estimated bounds imply that the reduction in employment (respectively, earnings) is at least 2 percentage points ($1,785) and at most 14 percentage points ($8,400). In percentage terms, these results imply that the population causal average effect of going from no-to-mild to severe depression symptoms is a decrease of 2-16% for employment, and 8-37% for earnings. These estimated bounds on the effects of interest are meaningful as they are obtained under relatively weak assumptions. To get a sense about the magnitude of these estimated effects, Table 6 provides bounds on the effect of education on

[16] Average employment and earnings for non-depressed individuals are 0.90 and $22,830 respectively.

employment and earnings for our estimation sample.[17] These estimated bounds closely mirror the original application in Manski and Pepper (2000). The MIV+MTS+MTR bounds on high school dropout vs high school graduate/some college do not exclude zero, but indicate that going from being a high school dropout to high school graduate/some college at most increases employment by 11 percentage points (15%) and earnings by $9,148 (87%). The maximum reduction in employment (earnings) in percentage terms from going from no-to-mild to severe depressive symptoms is similar to (half the size of) the increase in employment (earnings) from going from being a high school dropout to high school graduate/some college. While this is a comparison of the upper bounds on the effects, it suggests that the effects of no-to-mild to severe depression symptoms on labor market outcomes are non-negligible. Another point of comparison for the magnitude of our estimated effects is to consider traditional estimates of the returns to schooling. In a survey by Card (1999) of IV estimates of the return on earnings to one year of schooling, he finds that they range from about 6-15%. Our inference on the effect of no-to-mild to severe depressive symptoms on earnings of between 8-37% is *at least*, comparable to the earnings effect of an additional year of schooling, informally reinforcing that the magnitude of our estimated effects is not negligible (as before, we point out that the IV estimates surveyed in Card (1999) apply to a subpopulation defined by each of the instruments used in those studies).

5.2 Results by Gender, and by Race

In this section, we further analyze our effects of interest across different demographic groups. The results are summarized in Figures 1 and 2, which show the MIV+MTR+MTS bounds and 95% confidence intervals, along with the OLS estimates (full results are given in Online Appendix tables A4-A8). While it is difficult to make formal comparisons across groups because the bounds overlap, there are some interesting findings. First, examining the width of the bounds, the range of possible values of the causal effect is clearly narrower for women than for men. Second, the bounds on the ATE of going from no-to-mild to severe depressive symptoms for employment are statistically negative for men, but nor for women. Going from no-to-mild to severe depressive symptoms decreases the probability of being employed by 3-17 percentage points (3-18%) and earnings by $2,500-$11,172 (9-39%) for men. Third, the bounds on “severe vs no-to-mild” depressive symptoms for employment and earnings exclude zero for white, black and Hispanic individuals, but their corresponding 95% confidence intervals include zero. The exception is the 95% confidence interval on earnings for Hispanic individuals, which indicates a reduction in earnings of $2,133 to $8,932 (11-45%).

6. Summary

Credibly identifying the causal effect of poor mental health on labor market outcomes is challenging because of omitted variable bias and reverse causality. To address these issues, previous literature has used childhood and parental mental health, religiosity, death of a close friend, availability of social services and social support as instruments, but these instruments may not satisfy the exclusion restriction. Other studies have used individual fixed-effects, which control for unobserved time invariant characteristics, but they are unable to control for time varying confounders, such as stress. In addition, in the presence of heterogeneous effects, IV and individual fixed-effect methods estimate effects for specific subpopulations that may or may not correspond to the population of interest. We contribute to the literature by using a nonparametric partial identification method to provide bounds on the population average effect of depression on employment and earnings. Although we do not point identify the causal effect, our approach uses relatively weak assumptions and provides bounds on the population average treatment effect. These assumptions are based on weak monotonicity relationships on the selection process

[17] For this exercise, we drop 6 observations from our estimation sample that have missing values on education, resulting in a sample size of 7,659.

(MTS), the treatment response (MTR), and the link between a variable (referred to as a monotone instrument) and the outcome (MIV).

When looking at a binary indicator for depression, our results under the MTS, MTR and MIV assumptions indicate that the effect of depression is at most 10% on employment and 27% on earnings, thus ruling out potentially plausible magnitudes for these effects. When considering the dose-response nature of depression symptoms and labor market outcomes, we find statistically significant and economically non-negligible effects of going from no-to-mild to severe depression symptoms on employment and earnings. In particular, we find that these effects result in decreases of 2-16% for employment, and 8-37% for earnings. Interestingly, these bounds are somewhat comparable in magnitude (in absolute value) to corresponding bounds obtained for going from being a high school dropout to high school graduate/some college on labor market outcomes. Moreover, our estimated bounds rule out, for the population average effect, the magnitude of some of the effects obtained in prior studies, particularly very large detrimental effects on employment, and a zero effect on earnings (at least for going from no-to-mild to severe depressive symptoms). Therefore, our findings, obtained using relatively weak assumptions, point to potentially significant detrimental causal effects of depression symptoms on labor market outcomes.

References

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), 161-174.

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, *113*(1), 151-184.

Altonji, J.G., Bharadwaj, P., & Lange, F. (2012). Changes in the characteristics of American youth: Implications for adult outcomes. *Journal of Labor Economics*, 30, 783-828.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy*, *75*(4, Part 1), 352-365.

Biasi, B., Dahl, M.S. & Moser, P. (2021). Career effects of mental health. NBER Working Paper 29031.

Bubonya, M., Cobb-Clark, D. A., & Wooden, M. (2017). Mental health and productivity at work: Does what you do matter?. *Labour Economics*, *46*, 150-165.

Bryan, M. L., Rice, N., Roberts, J., & Sechel, C. (2020). Mental health and employment: a bounding approach using panel data. The Sheffield Economic Research Paper Series (SERPS), 2020006(2020006).

Card, D. (1999). The causal effect of education on earnings. *Handbook of Labor Economics*, *3*, 1801-1863.

Center for Child and Human Development, Georgetown University. (2021). “Center for Epidemiologic Studies-Depression Scale (CES-D).” The Right Questions are the Key: Best Practices in Screening by Home Visitors. GeorgetownX. Accessed August 10, 2021 https://edge.edx.org/courses/coursev1:GeorgetownX+CCHD+2016/0c2aa1d840ab4ba2982d50f2a6fe44e5/

Chernozhukov, V., Lee, S., & Rosen, A. M. (2013). Intersection bounds: estimation and inference. *Econometrica*, *81*(2), 667-737.

Cseh, A. (2008). The effects of depressive symptoms on earnings. *Southern Economic Journal*, 383-409.

Chatterji, P., Alegria, M., Lu, M., & Takeuchi, D. (2007). Psychiatric disorders and labor market outcomes: evidence from the National Latino and Asian American Study. *Health Economics*, *16*(10), 1069-1090.

Chatterji, P., Alegria, M., & Takeuchi, D. (2008). Psychiatric disorders and labor market outcomes: Evidence from the National Comorbidity Survey-Replication. NBER Working Paper Number 14404.

Chatterji, P., Alegria, M., & Takeuchi, D. (2011). Psychiatric disorders and labor market outcomes: Evidence from the National Comorbidity Survey-Replication. *Journal of Health Economics*, *30*(5), 858-868.

Christelis, D., & Dobrescu, L. (2019). The causal effect of social activities on cognition: Evidence from 20 European countries. *Social Science & Medicine, 247,* 1-9

Cygan-Rehm, K., Kuehnle, D., & Oberfichtner, M. (2017). Bounding the causal effect of unemployment on mental health: Nonparametric evidence from four countries. *Health Economics*, *26*(12), 1844-1861.

De Haan, M. (2011). The effect of parents' schooling on child's schooling: a nonparametric bounds analysis. *Journal of Labor Economics*, *29*(4), 859-892.

Ettner, S. L., Frank, R. G., & Kessler, R. C. (1997). The impact of psychiatric disorders on labor market outcomes. *ILR Review*, *51*(1), 64-81.

Farahati, F., Marcotte, D. E., & Wilcox-Gök, V. (2003). The effects of parents' psychiatric disorders on children's high school dropout. *Economics of Education Review*, *22*(2), 167-178.

Fletcher, J. M. (2008). Adolescent depression: diagnosis, treatment, and educational attainment. *Health Economics*, *17*(11), 1215-1235.

Fletcher, J. (2013a). Adolescent depression and adult labor market outcomes. *Southern Economic Journal*, *80*(1), 26-49.

Fletcher, J. M. (2013b). The effects of personality traits on adult labor market outcomes: Evidence from siblings. *Journal of Economic Behavior & Organization*, *89*, 122-135.

Flores, C. A., & Flores-Lagunes, A. (2013). Partial identification of local average treatment effects with an invalid instrument. *Journal of Business & Economic Statistics*, *31*(4), 534-545.

Frijters, P., Johnston, D. W., & Shields, M. A. (2014). The effect of mental health on employment: evidence from Australian panel data. *Health Economics*, *23*(9), 1058-1071.

Germinario, G., Flores, C.A., & Flores-Lagunes, A. (2021). mpclr: A Stata command for bias-corrected estimation and valid inference for nonparametric bounds on treatment effects. Mimeographed: Department of Economics, Syracuse University.

Gonzalez, L. (2005). Nonparametric bounds on the returns to language skills. *Journal of Applied Econometrics*, *20*(6), 771-795.

Grossman, M. (1972): On the concept of health capital and the demand for health. *Journal of Political Economy,* 80(2), 223–255.

Gruber, J. H. (2005). Religious market structure, religious participation, and outcomes: Is religion good for you?. *The BE Journal of Economic Analysis & Policy*, *5*(1).

Lundborg, P., Nilsson, A., & Rooth, D. O. (2014). Adolescent health and adult labor market outcomes. *Journal of Health Economics*, *37*, 25-40.

Hamilton, V. H., Merrigan, P., & Dufresne, É. (1997). Down and out: estimating the relationship between mental health and unemployment. *Health Economics*, *6*(4), 397-406.

Hill, W. D., Davies, N. M., Ritchie, S. J., Skene, N. G., Bryois, J., Bell, S., ... & Deary, I. J. (2019). Genome-wide analysis identifies molecular systems and 149 genetic loci associated with income. *Nature Communications*, *10*(1), 1-16.

Hirano, K., & Porter, J. R. (2012). Impossibility results for nondifferentiable functionals. *Econometrica*, *80*(4), 1769-1790.

Huang, J., van den Brink, H. M., & Groot, W. (2012). Does education promote social capital? Evidence from IV analysis and nonparametric-bound analysis. *Empirical Economics*, *42*(3), 1011-1034.

Imbens, G., & Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica, 62(2),* 467-475.

Imbens, G. W., & Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, *72*(6), 1845-1857.

Lin, D., Lutter, R., & Ruhm, C. J. (2018). Cognitive performance and labour market outcomes. *Labour Economics*, *51*, 121-135.

McCarthy, I., Millimet, D. L., & Roy, M. (2015). Bounding treatment effects: A command for the partial identification of the average treatment effect with endogenous and misreported treatment assignment. *The Stata Journal*, *15*(2), 411-436.

Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human Resources*, 343-360.

Manski, C. F. (1997). Monotone treatment response. *Econometrica*, 65, 1311-1334.

Manski, C. F., & Pepper, J. V. (2000). Monotone instrumental variables with an application to the returns to schooling. *Econometrica,* 68(July), 997-1010

Marcotte, D. E., & Wilcox-Gök, V. (2001). Estimating the employment and earnings costs of mental illness: recent developments in the United States. *Social Science & Medicine*, *53*(1), 21-27.

Marcotte, D. E., & Wilcox-Gok, V. (2003). Estimating earning losses due to mental illness: a quantile regression approach. *Journal of Mental Health Policy and Economics*, *6*(3), 123-134.

Mousteri, V., Daly, M., Delaney, L., Tynelius, P., & Rasmussen, F. (2019). Adolescent mental health and unemployment over the lifespan: Population evidence from Sweden. *Social Science & Medicine*, *222*, 305-314.

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.

Oswald, Andrew J., Eugenio Proto, and Daniel Sgroi. "Happiness and productivity." *Journal of Labor Economics* 33.4 (2015): 789-822.

Peng, L., Meyerhoefer, C. D., & Zuvekas, S. H. (2016). The Short-Term Effect of Depressive Symptoms on Labor Market Outcomes. *Health Economics*, *25*(10), 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*(3), 385-401.

Richey, J. (2015). Shackled labor markets: Bounding the causal effects of criminal convictions in the US. *International Review of Law and Economics*, *41*, 17-24.

Table 1: Descriptive Statistics

	Full	Men	Women	White	Black	Hispanic
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.51 (0.50)	—	—	0.50 (0.50)	0.52 (0.50)	0.50 (0.50)
White	0.80 (0.40)	0.80 (0.40)	0.80 (0.40)	—	—	—
Black	0.14 (0.35)	0.14 (0.34)	0.14 (0.35)	—	—	—
Hispanic	0.06 (0.24)	0.06 (0.24)	0.06 (0.24)	—	—	—
AFTQ (1981)	47.79 (28.85)	48.29 (30.03)	47.31 (27.65)	53.52 (27.44)	22.41 (20.62)	30.49 (24.67)
Age (1992)	31.06 (2.31)	31.03 (2.33)	31.07 (2.31)	31.07 (2.31)	31.01 (2.33)	30.93 (2.35)
CES-D score (1992)	9.04 (9.00)	8.07 (8.15)	9.99 (9.67)	8.49 (8.76)	11.42 (9.48)	10.82 (9.89)
Depressed	0.21 (0.41)	0.15 (0.35)	0.22 (0.42)	0.16 (0.37)	0.28 (0.45)	0.24 (0.43)
No to mild depressive symptoms	0.82 (0.39)	0.85 (0.35)	0.78 (0.42)	0.84 (0.37)	0.72 (0.45)	0.76 (0.43)
Moderate depressive symptoms	0.10 (0.31)	0.08 (0.28)	0.12 (0.33)	0.09 (0.29)	0.16 (0.37)	0.12 (0.32)
Severe depressive symptoms	0.08 (0.27)	0.06 (0.24)	0.10 (0.30)	0.07 (0.26)	0.11 (0.32)	0.12 (0.33)
Employed (1993)	0.88 (0.33)	0.94 (0.24)	0.82 (0.39)	0.89 (0.31)	0.80 (0.40)	0.84 (0.36)
Earnings (1993)	21,365 (19,862)	27,455 (21,575)	15,405 (15,916)	22,769 (20,594)	14,767 (14,823)	17,944 (16,405)
N	7,665	3,678	3,987	3,907	2,291	1,467

Notes: Summary statistics are weighted using the 1993 panel sampling weights. Standard deviations in parenthesis.

Table 2: Observed Family Background, Ability and Education by Depressive Symptom Severity, Full Sample

Variable (Survey Year)	No-to-Mild	Moderate	Severe	Moderate – No-to-Mild	Severe - Moderate	Severe – No-to-Mild
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Demographics (1979)						
Female	0.48 (0.05)	0.60 (0.06)	0.61 (0.05)	0.12*** (0.03)	0.01 (0.04)	0.13*** (0.03)
White	0.82 (0.04)	0.71 (0.06)	0.71 (0.05)	-0.11*** (0.03)	0.00 (0.03)	-0.11*** (0.03)
Black	0.12 (0.03)	0.22 (0.07)	0.19 (0.05)	0.09** (0.03)	-0.03 (0.03)	0.07*** (0.03)
Hispanic	0.06 (0.01)	0.07 (0.02)	0.10 (0.02)	0.01* (0.01)	0.03** (0.01)	0.04*** (0.01)
Family Background (1979)						
Mother's education	11.73 (0.08)	11.12 (0.14)	10.80 (0.23)	-0.60*** (0.10)	-0.32 (0.24)	-0.93*** (0.19)
Father's education	11.97 (0.13)	11.31 (0.24)	11.16 (0.30)	-0.66*** (0.21)	-0.15 (0.30)	-0.81*** (0.21)
Ability Measures (1981)						
AFQT Score	50.35 (1.32)	38.77 (2.29)	33.59 (1.46)	-11.56*** (1.34)	-5.18***(1.50)	-16.77*** (0.75)
Expect to graduate college	0.41 (0.01)	0.31 (0.02)	0.24 (0.03)	-0.10*** (0.02)	-0.07***(0.02)	-0.17*** (0.02)
Education (1993)						
Years of education	13.41 (0.06)	12.78 (0.11)	12.19 (0.12)	-0.063*** (0.08)	-0.059*** (0.011)	-1.23*** (0.10)
College graduate	0.25 (0.01)	0.18 (0.02)	0.11 (0.02)	-0.08*** (0.02)	-0.07** (0.02)	-0.15*** (0.02)

Notes: Summary statistics are weighted by the 1993 panel sampling weights. The sample size for demographics and AFQT score is 7,665. The sample sizes for the other variables are lower due to missing values: (1) mother's education N=7,190; (2) father's education N=6,554; (3) expect to graduate college N=7,610; (4) Education N=7,654 Standard errors in parentheses. ***statistically significant at 1% **statistically significant at 5% * statistically significant at 10%

Table 3: Labor Market Outcomes by Depressive Symptom Severity, Full Sample

Depressive Symptoms, t	$E[Employed \| T = t]$	$E[Earnings \| T = t]$	$\Pr[T = t]$	N_t
	(1)	**(2)**	**(3)**	**(4)**
No to mild depressive	0.90	22,830	0.82	6,023
Moderate depressive	0.82	16,752	0.10	926
Severe depressive	0.75	12,509	0.08	716

Notes: Summary statistics are weighted by the 1993 panel sampling weights.

Table 4: OLS Estimates and Estimated Bounds on the ATE of being Depressed on Labor Market Outcomes

	OLS	No Assumption	MTS	MTR	MTR+MTS	MIV+MTS+MTR
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Employment						
Depressed	-0.108*** (.019)	[-0.771, 0.229] (-0.788, 0.246)	[-0.108, 0.229] (-0.140, 0.246)	[-0.771, 0.000] (-0.788, 0.000)	[-0.108, 0.000] (-0.143, 0.000)	[-0.091, 0.000] (-0.121, 0.000)
Panel B: Earnings						
Depressed	-7931*** (749)	[-34,516, 66,441] (-35,868, 67,793)	[-7935, 66,441] (-9188, 67,793)	[-34,516, 0.000] (-35,868, 0.000)	[-7935, 0.000] (-9261, 0.000)	[-6082, 0.000] (-7520, 0.000)

Notes: Robust standard errors in (.) in column 1. In columns 2-6 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. Earnings are bounded between $0 and $100,948. AFQT score is used as the MIV with 5 bins. 1993 panel weights are applied to OLS regressions and estimated bounds.

Table 5: OLS Estimates and Estimated Bounds on the ATE of Depressive Symptom Intensity on Labor Market Outcomes

	OLS	No Assumption	MTS	MTR	MTR+MTS	MIV+MTS+MTR
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Employment						
Moderate vs No-to-mild	-0.077*** (0.020)	[-0.832, 0.249] (-0.843, 0.269)	[-0.143, 0.235] (-0.173, 0.253)	[-0.771, 0.000] (-0.788, 0.000)	[-0.083, 0.000] (-0.119, 0.000)	[-0.067, 0.000] (-0.094, 0.000)
Severe vs Moderate	-0.072*** (.018)	[-0.921, 0.894] (-0.929, 0.902)	[-0.218, 0.225] (-0.254, 0.250)	[-0.838, 0.000] (-0.854, 0.000)	[-0.135, 0.000] (-0.172, 0.000)	[-0.122, 0.000] (-0.159, 0.000)
Severe vs No-to-Mild	-0.149*** (.024)	[-0.857, 0.248] (-0.873, 0.265)	[-0.149, 0.248] (-0.189, 0.266)	[-0.857, 0.000] (-0.873, 0.000)	[-0.149, -0.000] (-0.192, -0.000)	[-0.139, -0.017] (-0.182, -0.002)
Panel B: Earnings						
Moderate vs No-to-mild	-6077*** (1087)	[-35,526,73,580] (-36,942,75,159)	[-7435, 66,784] (-9160, 68,218)	[-34,516, 0.000] (-35,868, 0.000)	[-6425, 0.000] (-8292, 0.000)	[-5045, 0.000] (-7009, 0.000)
Severe vs Moderate	-4243*** (871)	[-91,180,92,069] (-92,131,92,913)	[-72,886,78,413] (-74,426,79,722)	[-27,497, 0.000] (-28,621, 0.000)	[-9200, 0.000] (-10,043, 0.000)	[-7390, 0.000] (-8097, 0.000)
Severe vs No-to-Mild	-10,321*** (588)	[-36,259,75,202] (-37,667,76,279)	[-10,324,75,202] (-11,290,76,279)	[-36,259, 0.000] (-37,668, 0.000)	[-10,324, 0.000] (-11,309, 0.000)	[-8400, -1785] (-9328, -97)

Notes: Robust standard errors in (.) in column 1. In columns 2-6 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. Earnings are bounded between $0 and $100,948. AFQT score is used as the MIV with 5 bins. 1993 panel weights are applied to OLS regressions and estimated bounds.

Table 6: OLS Estimates and Estimated Bounds on the ATE of Educational Attainment on Labor Market Outcomes

	OLS	No Assumption	MTS	MTR	MTR+MTS	MIV+MTS+MTR
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Employment						
High school dropout vs High school grad/some college	0.137*** (0.025)	[-0.398, 0.837] (-0.413, 0.851)	[-0.192, 0.166] (-0.207, 0.210)	[0.000, 0.825] (0.000, 0.840)	[0.000, 0.154] (0.000, 0.200)	[0.000, 0.106] (0.000, 0.162)
High school grad/some college vs College grad	0.071*** (.004)	[-0.697, 0.415] (-0.709, 0.430)	[-0.683, 0.170] (-0.694, 0.185)	[0.000, 0.332] (0.000, 0.344)	[0.000, 0.087] (0.000, 0.096)	[0.000, 0.069] (0.000, 0.080)
College grad vs High school dropout	0.208*** (.028)	[-0.749, 0.905] (-0.763, 0.921)	[-0.749, 0.208] (-0.763, 0.254)	[0.000, 0.905] (0.000, 0.921)	[0.000, 0.208] (0.000, 0.258)	[0.004, 0.148] (0.001, 0.196)
Panel B: Earnings						
High school dropout vs High school grad/some college	8355*** (1217)	[-78523, 46114] (-79547, 47951)	[-74105, 27625] (-75259, 30156)	[0.000, 30319] (0.000, 31709)	[0.000, 11829] (0.000, 13955)	[0.000, 9148] (0.000, 10971)
High school grad/some college vs College grad	14785*** (972)	[-39398, 72847] (-40891, 74082)	[-30214, 16896] (-31423, 18495)	[0.000, 71675] (0.000, 72851)	[0.000, 15723] (0.000, 17277)	[0.000, 8578] (0.000, 9392)
College grad vs High school dropout	23139*** (1530)	[-82948, 83984] (-84370, 84938)	[-82948, 23146] (-84370, 25543)	[0.000, 83984] (0.000, 84938)	[0.000, 23146] (0.000, 25606)	[307, 16428] (199, 17919)

Notes: Robust standard errors in (.) in column 1. In columns 2-6 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. Earnings are bounded between $0 and $100,948. AFQT score is used as the MIV with 5 bins. Sample size is 7,659. Average employment and earnings in (1) high school dropouts are 0.74 and $10,480; (2) high school grad/some college are 0.88 and $18,835 and (3) college grads are 0.95 and $33,619. 1993 panel weights are applied to OLS regressions and estimated bounds.

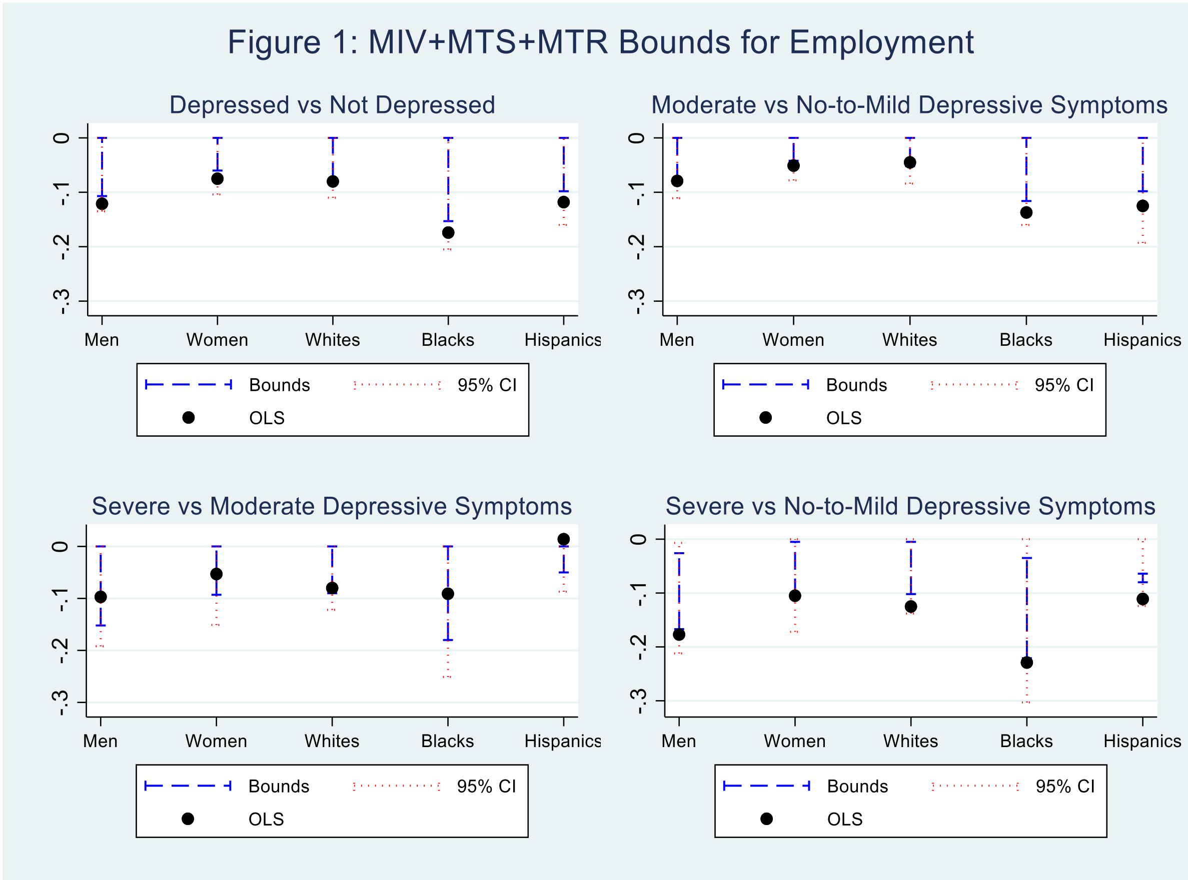
Figure 1: MIV+MTS+MTR Bounds for Employment
Depressed vs Not Depressed
Moderate vs No-to-Mild Depressive Symptoms
Severe vs Moderate Depressive Symptoms
Severe vs No-to-Mild Depressive Symptoms
0
-.1
-.2
-.3
Men
Women
Whites
Blacks
Hispanics
Bounds
95% CI
OLS

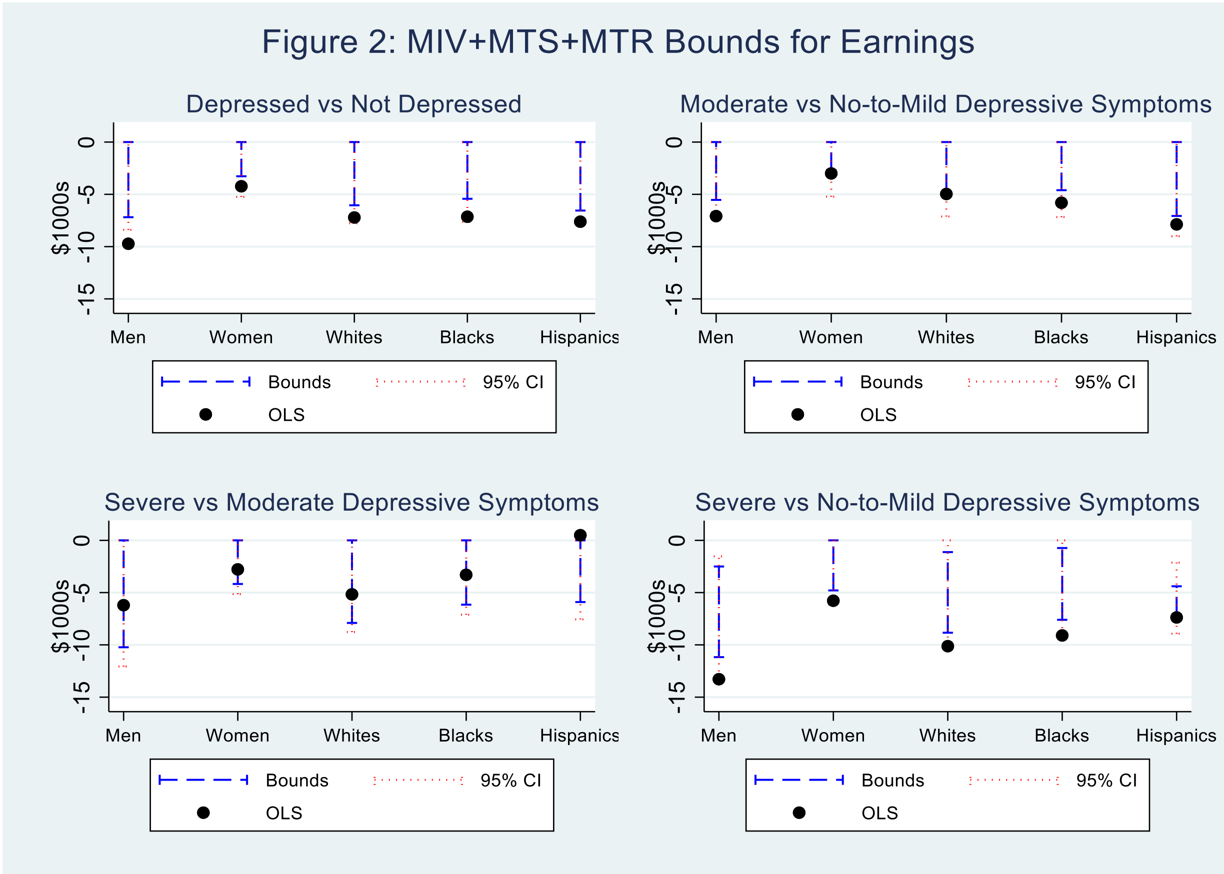
Figure 2: MIV+MTS+MTR Bounds for Earnings
Depressed vs Not Depressed
Moderate vs No-to-Mild Depressive Symptoms
Severe vs Moderate Depressive Symptoms
Severe vs No-to-Mild Depressive Symptoms
$1000s
0
-5
-10
-15
Men
Women
Whites
Blacks
Hispanics
Bounds
95% CI
OLS

ONLINE APPENDIX

Appendix A: Additional Tables

Table A1: Observed Family Background, Ability and Education by Depressive Symptom Severity and Gender

Variable	No-to-Mild	Moderate	Severe	Moderate – No-to-Mild	Severe - Moderate	Severe – No-to-Mild
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Men						
Mother's education	11.74 (0.12)	11.35 (0.25)	10.73 (0.43)	-0.44** (0.18)	-0.62 (0.35)	-1.06*** (0.35)
Father's education	12.06 (0.22)	11.49 (0.45)	11.01 (0.72)	-0.57 (0.32)	-0.47 (0.65)	-1.04* (0.53)
AFQT Score	50.46 (1.91)	39.34 (3.89)	30.95 (2.79)	-11.12*** (2.17)	-8.38** (2.72)	-19.50*** (1.60)
Expect to graduate college	0.42 (0.02)	0.28 (0.02)	0.25 (0.06)	-0.14*** (0.02)	-0.03 (0.04)	-0.17*** (0.05)
Years of education	13.23 (0.08)	12.71 (0.12)	12.01 (0.17)	-0.62*** (0.09)	-0.69*** (0.16)	-1.32*** (0.14)
College graduate	0.25 (0.01)	0.17 (0.02)	0.09 (0.02)	-0.08*** (0.02)	-0.07 (0.03)**	-0.15*** (0.02)
Panel B: Women						
Mother's education	11.66 (0.10)	10.96 (0.16)	10.84 (0.27)	-0.70*** (0.12)	-0.12 (0.31)	-0.82 (0.22)***
Father's education	11.87 (0.15)	11.19 (0.27)	11.24 (0.21)	-0.69* (0.29)	0.06 (0.27)	-0.63*** (0.18)
AFQT Score	50.24 (1.82)	38.38 (2.78)	35.24 (1.57)	-11.86*** (1.81)	-3.14* (1.59)	-15.01*** (0.73)
Expect to graduate college	0.40 (0.02)	0.33 (0.04)	0.23 (0.04)	-0.07** (0.03)	-0.10** (0.02)	-0.16*** (0.02)
Years of education	13.50 (0.07)	12.83 (0.16)	12.30 (0.18)	0.67*** (0.10)	-0.53** (0.16)	-1.20*** (0.13)
College graduate	0.26 (0.01)	0.19 (0.03)	0.12 (0.03)	-0.07*** (0.02)	-0.07* (0.04)	-0.14*** (0.03)

Notes: Summary statistics are weighted by the 1993 panel sampling weights. Standard errors in parentheses. ***statistically significant at 1% **statistically significant at 5% * statistically significant at 10%

Table A2: Observed Family Background, Ability and Education by Depressive Symptom Severity and Race

Variable	No-to-Mild	Moderate	Severe	Moderate – No-to-Mild	Severe - Moderate	Severe – No-to-Mild
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: White Individuals						
Mother's education	12.04 (0.09)	11.61 (0.13)	11.31 (0.34)	-0.43*** (0.11)	-0.30 (0.33)	-0.74** (0.27)
Father's education	12.35 (0.16)	11.98 (0.21)	11.65 (0.39)	-0.38 (0.22)	-0.33 (0.34)	-0.71** (0.26)
AFQT Score	55.39 (1.08)	47.09 (1.44)	40.15 (1.60)	-8.30*** (1.22)	-6.94*** (1.26)	-15.24*** (0.77)
Expect to graduate college	0.41 (0.02)	0.33 (0.03)	0.23 (0.04)	-0.08*** (0.03)	-0.10*** (0.03)	-0.18*** (0.03)
Years of education	13.55 (0.07)	13.04 (0.14)	12.33 (0.18)	-0.49*** (0.10)	-0.71*** (0.15)	-1.21*** (0.15)
College graduate	0.28 (0.01)	0.22 (0.03)	0.13 (0.03)	-0.06* (0.02)	-0.09* (0.04)	-0.15*** (0.02)
Panel B: Black Individuals						
Mother's education	11.12 (0.15)	10.62 (0.20)	10.20 (0.33)	-0.50*** (0.14)	-0.42 (0.34)	-0.92*** (0.29)
Father's education	10.55 (0.22)	9.63 (0.37)	9.97 (0.39)	-0.92*** (0.21)	0.34 (0.33)	-0.58* (0.23)
AFQT Score	24.74 (1.50)	17.26 (1.48)	14.89 (1.35)	-7.48*** (1.61)	-2.36 (1.71)	-9.85*** (0.52)
Expect to graduate college	0.42 (0.02)	0.30 (0.04)	0.28 (0.03)	-0.12*** (0.03)	-0.02 (0.03)	-0.14*** (0.02)
Years of education	12.95 (0.08)	12.32 (0.11)	11.76 (0.13)	-0.64*** (0.09)	-0.56*** (0.13)	-1.19*** (0.15)
College graduate	0.16 (0.01)	0.09 (0.01)	0.03 (0.02)	-0.07*** (0.01)	-0.05** (0.02)	-0.12*** (0.01)
Panel C: Hispanic Individuals						
Mother's education	8.34 (0.18)	7.45 (0.44)	8.17 (0.23)	-0.89* (0.41)	0.72 (0.40)	-0.17 (0.34)
Father's education	8.59 (0.32)	7.86 (0.38)	8.98 (0.55)	-0.73 (0.60)	1.12 (0.61)	0.38 (0.71)
AFQT Score	33.36 (1.61)	20.55 (2.15)	22.20 (1.75)	-12.81*** (1.49)	1.64 (3.26)	-11.17*** (2.33)
Expect to graduate college	0.37 (0.02)	0.21 (0.03)	0.26 (0.03)	-0.16*** (0.02)	0.05 (0.04)	-0.11*** (0.03)
Years of education	12.51 (0.13)	11.50 (0.32)	11.91 (0.11)	-1.02** (0.37)	0.41 (0.36)	-0.60*** (0.08)
College graduate	0.13 (0.01)	0.05 (0.03)	0.07 (0.02)	-0.08* (0.03)	0.02 (0.02)	-0.06** (0.02)

Notes: Summary statistics are weighted by the 1993 panel sampling weights. Standard errors in parentheses. ***statistically significant at 1% **statistically significant at 5% * statistically significant at 10%

Table A3: Labor Market Outcomes by Depressive Symptom Severity, Gender and Race

Depressive Symptoms t	$E[Employed \mid T = t]$	$E[Earnings \mid T = t]$	$\Pr\,[T = t]$	N_t
Panel A: Men				
No-to-mild	0.96	28,891	0.85	3,012
Moderate	0.88	21,807	0.08	394
Severe	0.78	15, 607	0.06	272
Panel B: Women				
No-to-mild	0.83	16,339	0.78	3,011
Moderate	0.78	13,344	0.12	532
Severe	0.73	10,565	0.10	444
Panel C: White Individuals				
No-to-mild	0.91	23,957	0.84	3,263
Moderate	0.86	18,996	0.09	365
Severe	0.78	13,833	0.07	279
Panel D: Black Individuals				
No-to-mild	0.85	16,738	0.72	1,649
Moderate	0.71	10,935	0.16	378
Severe	0.62	7,641	0.11	264
Panel E: Hispanic Individuals				
No-to-mild	0.87	19,772	0.76	1,111
Moderate	0.74	11,903	0.12	183
Severe	0.76	12,396	0.12	173

Notes: Summary statistics are weighted by the 1993 panel sampling weights.

Table A4: OLS Estimates and Estimated Bounds on the ATE of Depressive Symptom Intensity on Labor Market Outcomes, Men

	OLS	No Assumption	MTS	MTR	MTR+MTS	MIV+MTS+MTR
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Employment						
Depressed vs Not Depressed	-0.121*** (.019)	[-0.841, 0.159] (-0.857, 0.175)	[-0.121, 0.159] (-0.154, 0.175)	[-0.841, 0.000] (-0.857, 0.000)	[-0.121, 0.000] (-0.156, 0.000)	[-0.107, 0.000] (-0.135, 0.000)
Moderate vs No-to-mild	-0.079*** (.023)	[-0.890, 0.173] (-0.903, 0.190)	[-0.135, 0.165] (-0.178, 0.180)	[-0.841, 0.000] (-0.857, 0.000)	[-0.085, 0.000] (-0.125, 0.000)	[-0.080, 0.000] (-0.111, 0.000)
Severe vs Moderate	-0.097** (.037)	[-0.941, 0.912] (-0.952, 0.920)	[-0.200, 0.162] (-0.247, 0.206)	[-0.905, 0.000] (-0.920, 0.000)	[-0.165, 0.000] (-0.214, 0.000)	[-0.152, 0.000] (-0.192, 0.000)
Severe vs No-to-Mild	-0.177*** (.028)	[-0.915, 0.169] (-0.927, 0.189)	[-0.177, 0.169] (-0.224, 0.190)	[-0.915, 0.000] (-0.927, 0.000)	[-0.177, 0.000] (-0.226, 0.000)	[-0.167, -0.026] (-0.212, -0.007)
Panel B: Earnings						
Depressed vs Not Depressed	-9726*** (856)	[-36,708,64,253] (-38,861,66,406)	[-9730, 64,253] (-11,169,66,406)	[-36,708, 0.000] (-38,864, 0.000)	[-9730, 0.000] (-11,237, 0.000)	[-7184, 0.000] (-8392, 0.000)
Moderate vs No-to-mild	-7083*** (1151)	[-37691, 69621] (-40085, 71282)	[-8462, 64643] (-10523, 66825)	[-36708, 0.000] (-38862, 0.000)	[-7479, 0.000] (-9496, 0.000)	[-5532, 0.000] (-6734, 0.000)
Severe vs Moderate	-6201*** (1102)	[-93262, 93733] (-94440, 94639)	[-73663, 75148] (-76259, 77057)	[-31850, 0.000] (-34186, 0.000)	[-12241, 0.000] (-13122, 0.000)	[-10229, 0.000] (-12069, 0.000)
Severe vs No-to-Mild	-13,284*** (545)	[-38556, 70960] (-40791, 73274)	[-13287, 70960] (-14208, 73274)	[-38556, 0.000] (-40795, 0.000)	[-13287, 0.000] (-14212, 0.000)	[-11172, -2500] (-13206, -1532)

Notes: Robust standard errors in (.) in column 1. In columns 2-6 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. Earnings are bounded between $0 and $100,948. AFQT score is used as the MIV with 5 bins. 1993 panel weights are applied to OLS regressions and estimated bounds.

Table A5: OLS Estimates and Estimated Bounds on the ATE of Depressive Symptom Intensity on Labor Market Outcomes, Women

	OLS	No Assumption	MTS	MTR	MTR+MTS	MIV+MTS+MTR
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Employment						
Depressed vs Not Depressed	-0.075** (.028)	[-0.703, 0.297] (-0.717, 0.312)	[-0.075, 0.297] (-0.122, 0.312)	[-0.703, 0.000] (-0.717, 0.000)	[-0.075, 0.000] (-0.127, 0.000)	[-0.060, 0.000] (-0.104, 0.000)
Moderate vs No-to-mild	-0.051 (.029)	[-0.774, 0.324] (-0.785, 0.341)	[-0.128, 0.303] (-0.167, 0.320)	[-0.703, 0.000] (-0.717, 0.000)	[-0.057, 0.000] (-0.111, 0.000)	[-0.042, 0.000] (-0.078, 0.000)
Severe vs Moderate	-0.053** (.023)	[-0.902, 0.877] (-0.908, 0.891)	[-0.223, 0.268] (-0.275, 0.301)	[-0.772, 0.000] (-0.783, 0.000)	[-0.093, 0.000] (-0.147, 0.000)	[-0.093, 0.000] (-0.151, 0.000)
Severe vs No-to-Mild	-0.105** (.034)	[-0.799, 0.324] (-0.815, 0.338)	[-0.105, 0.324] (-0.163, 0.339)	[-0.799, 0.000] (-0.815, 0.000)	[-0.105, 0.000] (-0.169, 0.000)	[-0.106, -0.005] (-0.172, 0.000)
Panel B: Earnings						
Depressed vs Not Depressed	-4228*** (981)	[-32,374,68,586] (-34,234,70,446)	[-4234, 68,586] (-5917, 70,446)	[-32,374, 0.000] (-34,237, 0.000)	[-4234, 0.000] (-6119, 0.000)	[-3271, 0.000] (-5239, 0.000)
Moderate vs No-to-mild	-2996* (1348)	[-33410, 77455] (-35238, 79696)	[-4311, 68858] (-6442, 70827)	[-32374, 0.000] (-34234, 0.000)	[-3274, 0.000] (-5750, 0.000)	[-2668, 0.000] (-5223, 0.000)
Severe vs Moderate	-2778** (999)	[-89143, 90441] (-90796, 91379)	[-71018, 80050] (-72510, 81840)	[-23239, 0.000] (-24286, 0.000)	[-5114, 0.000] (-6242, 0.000)	[-4173, 0.000] (-5142, 0.000)
Severe vs No-to-Mild	-5774*** (823)	[-34015, 79355] (-35917, 80353)	[-5777, 79355] (-7126, 80353)	[-34015, 0.000] (-35919, 0.000)	[-5777, 0.000] (-7238, 0.000)	[-4789, 0.000] (-6132, 0.000)

Notes: Robust standard errors in (.) in column 1. In columns 2-6 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. Earnings are bounded between $0 and $100,948. AFQT score is used as the MIV with 5 bins. 1993 panel weights are applied to OLS regressions and estimated bounds.

Table A6: OLS Estimates and Estimated Bounds on the ATE of Depressive Symptom Intensity on Labor Market Outcomes, White Individuals

	OLS	No Assumption	MTS	MTR	MTR+MTS	MIV+MTS+MTR
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Employment						
Depressed vs Not Depressed	-0.080** (.023)	[-0.787, 0.214] (-0.807, 0.234)	[-0.080, 0.214] (-0.119, 0.235)	[-0.787, 0.000] (-0.807, 0.000)	[-0.080, 0.000] (-0.122, 0.000)	[-0.078, 0.000] (-0.110, 0.000)
Moderate vs No-to-mild	-0.045 (.024)	[-0.843, 0.229] (-0.856, 0.253)	[-0.107, 0.219] (-0.140, 0.242)	[-0.787, 0.000] (-0.807, 0.000)	[-0.051, 0.000] (-0.096, 0.000)	[-0.052, 0.000] (-0.084, 0.000)
Severe vs Moderate	-0.080** (.022)	[-0.931, 0.904] (-0.939, 0.911)	[-0.194, 0.184] (-0.243, 0.212)	[-0.854, 0.000] (-0.873, 0.000)	[-0.117, 0.000] (-0.166, 0.000)	[-0.090, 0.000] (-0.122, 0.000)
Severe vs No-to-Mild	-0.125*** (.029)	[-0.867, 0.226] (-0.887, 0.246)	[-0.125, 0.226] (-0.175, 0.247)	[-0.867, 0.000] (-0.887, 0.000)	[-0.125, 0.000] (-0.179, 0.000)	[-0.102, -0.005] (-0.138, 0.000)
Panel B: Earnings						
Depressed vs Not Depressed	-7214*** (912)	[-33,887,67,071] (-35,467,68,651)	[-7219, 67,071] (-8780, 68,651)	[-33,887, 0.000] (-35,467, 0.000)	[-7219, 0.000] (-8901, 0.000)	[-6048, 0.000] (-7764, 0.000)
Moderate vs No-to-mild	-4961** (1379)	[-34881, 73335] (-36538, 75350)	[-6333, 67442] (-8540, 69129)	[-33886, 0.000] (-35466, 0.000)	[-5339, 0.000] (-7786, 0.000)	[-4733, 0.000] (-7122, 0.000)
Severe vs Moderate	-5162*** (1140)	[-92348, 92924] (-93105, 93933)	[-73620, 77059] (-75458, 78744)	[-28041, 0.000] (-29411, 0.000)	[-9309, 0.000] (-10291, 0.000)	[-7902, 0.000] (-8769, 0.000)
Severe vs No-to-Mild	-10,124*** (664)	[-35650, 74677] (-37297, 75958)	[-10127, 74677] (-11254, 75958)	[-35650, 0.000] (-37297, 0.000)	[-10127, 0.000] (-11283, 0.000)	[-8839, -1123] (-9963, 0.000)

Notes: Robust standard errors in (.) in column 1. In columns 2-6 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. Earnings are bounded between $0 and $100,948. AFQT score is used as the MIV with 5 bins. 1993 panel weights are applied to OLS regressions and estimated bounds.

Table A7: OLS Estimates and Estimated Bounds on the ATE of Depressive Symptom Intensity on Labor Market Outcomes, Black Individuals

	OLS	No Assumption	MTS	MTR	MTR+MTS	MIV+MTS+MTR
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Employment						
Depressed vs Not Depressed	-0.174*** (.028)	[-0.704, 0.297] (-0.726, 0.319)	[-0.175, 0.297] (-0.235, 0.319)	[-0.704, 0.000] (-0.726, 0.000)	[-0.175, 0.000] (-0.240, 0.000)	[-0.153, 0.000] (-0.205, 0.000)
Moderate vs No-to-mild	-0.137*** (.024)	[-0.773, 0.339] (-0.795, 0.360)	[-0.217, 0.307] (-0.275, 0.329)	[-0.704, 0.000] (-0.726, 0.000)	[-0.148, 0.000] (-0.207, 0.000)	[-0.116, 0.000] (-0.160, 0.000)
Severe vs Moderate	-0.091* (.043)	[-0.883, 0.841] (-0.897, 0.856)	[-0.302, 0.328] (-0.391, 0.378)	[-0.772, 0.000] (-0.799, 0.000)	[-0.191, 0.000] (-0.274, 0.000)	[-0.180, 0.000] (-0.251, 0.000)
Severe vs No-to-Mild	-0.229*** (.044)	[-0.820, 0.344] (-0.847, 0.366)	[-0.229, 0.344] (-0.312, 0.367)	[-0.820, 0.000] (-0.847, 0.000)	[-0.229, 0.000] (-0.317, 0.000)	[-0.221, -0.035] (-0.303, 0.000)
Panel B: Earnings						
Depressed vs Not Depressed	-7141*** (1123)	[-37,338,63,618] (-38,584,64,863)	[-7148, 63,618] (-9348, 64,863)	[-37,338, 0.000] (-38,584, 0.000)	[-7148, 0.000] (-9568, 0.000)	[-5424, 0.000] (-7632, 0.000)
Moderate vs No-to-mild	-5802*** (1268)	[-38195, 74082] (-39538, 75368)	[-7036, 63987] (-9529, 65273)	[-37338, 0.000] (-38583, 0.000)	[-6179, 0.000] (-8923, 0.000)	[-4600, 0.000] (-7175, 0.000)
Severe vs Moderate	-3294*** (736)	[-85345, 88694] (-87463, 89922)	[-68465, 80779] (-70536, 82422)	[-24380, 0.000] (-26311, 0.000)	[-7498, 0.000] (-8594, 0.000)	[-6153, 0.000] (-7114, 0.000)
Severe vs No-to-Mild	-9096*** (800)	[-39131, 78372] (-40491, 80607)	[-9101, 78372] (-10606, 80608)	[-39131, 0.000] (-40491, 0.000)	[-9101, 0.000] (-10683, 0.000)	[-7602, -733] (-9069, 0.000)

Notes: Robust standard errors in (.) in column 1. In columns 2-6 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. Earnings are bounded between $0 and $100,948. AFQT score is used as the MIV with 5 bins. 1993 panel weights are applied to OLS regressions and estimated bounds.

Table A8: OLS Estimates and Estimated Bounds on the ATE of Depressive Symptom Intensity on Labor Market Outcomes, Hispanic Individuals

	OLS	**No Assumption**	**MTS**	**MTR**	**MTR+MTS**	**MIV+MTS+MTR**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Panel A: Employment						
Depressed vs Not Depressed	-0.118** (.033)	[-0.722, 0.279] (-0.761, 0.318)	[-0.118, 0.279] (-0.181, 0.320)	[-0.722, 0.000] (-0.761, 0.000)	[-0.118, 0.000] (-0.186, 0.000)	[-0.098, 0.000] (-0.160, 0.000)
Moderate vs No-to-mild	-0.125* (.054)	[-0.816, 0.309] (-0.850, 0.357)	[-0.218, 0.277] (-0.300, 0.326)	[-0.722, 0.000] (-0.761, 0.000)	[-0.124, 0.002] (-0.221, 0.000)	[-0.098, 0.000] (-0.193, 0.000)
Severe vs Moderate	0.014 (.058)	[-0.876, 0.884] (-0.890, 0.900)	[-0.179, 0.317] (-0.260, 0.398)	[-0.779, 0.000] (-0.818, 0.000)	[-0.082, 0.000] (-0.147, 0.000)	[-0.050, 0.000] (-0.087, 0.000)
Severe vs No-to-Mild	-0.111** (.032)	[-0.808, 0.308] (-0.844, 0.348)	[-0.112, 0.308] (-0.172, 0.350)	[-0.808, 0.000] (-0.844, 0.000)	[-0.112, 0.000] (-0.178, 0.000)	[-0.080, -0.064] (-0.124, 0.000)
Panel B: Earnings						
Depressed vs Not Depressed	-7614*** (773)	[-36,347,64,607] (-37,335,65,595)	[-7619, 64,607] (-9076, 65,595)	[-36,347, 0.000] (-37,335, 0.000)	[-7619, 0.000] (-9159, 0.000)	[-6549, 0.000] (-7914, 0.000)
Moderate vs No-to-mild	-7869*** (1050)	[-37886, 75605] (-38846, 76750)	[-9353, 64545] (-11177, 65390)	[-36347, 0.000] (-37335, 0.000)	[-7813, -59] (-9735, 0.000)	[-7065, 0.000] (-9012, 0.000)
Severe vs Moderate	493 (1602)	[-89090, 88575] (-90146, 90173)	[-67177, 79532] (-69184, 81894)	[-27407, 0.000] (-28903, 0.000)	[-5494, 0.000] (-8025, 0.000)	[-5901, 0.000] (-7559, 0.000)
Severe vs No-to-Mild	-7376*** (1165)	[-37727, 74929] (-38793, 76378)	[-7383, 74929] (-9657, 76378)	[-37727, 0.000] (-38793, 0.000)	[-7383, 0.000] (-9876, -0.000)	[-7085, -4396] (-8932, -2133)

Notes: Robust standard errors in (.) in column 1. In columns 2-6 estimated bounds are in [.] and corresponding 95% confidence intervals in (.) are from 999 bootstrap replications. Earnings are bounded between $0 and $100,948. AFQT score is used as the MIV with 5 bins. 1993 panel weights are applied to OLS regressions and estimated bounds.

Appendix B: Technical Appendix on the CLR Method

This technical appendix provides additional details about the Chernozhukov et al. (2013; CLR) methodology that yields bias-corrected estimated bounds and valid confidence intervals for intersection bounds. To provide some intuition on the CLR method, we first make explicit the notion of creating the bins of the MIV. We use below 5 MIV bins $\mathcal{B}_m$, $m = 1, \ldots, 5$, each spanning 20 percentiles of the empirical distribution of AFQT scores in the NLSY79 sample. The lower bound on $E[Y(t_1)]$ from Equation (10) can then be rewritten as:

$$(11) \sum_{m=1}^{5} P(Z \in \mathcal{B}_m) \cdot \max_{m_1 \leq m} LB_{m_1}^1$$

where the $LB_{m_1}^1$ are the MTR-MTS lower bounds in bins m_1 up through m.

Instead of expressions like (11) which comprise 5 different maxima, the CLR method requires that these be rewritten as a set of expressions under a single maximum (or minimum, for upper bounds), with each element inside the max operator called a bounding function. Intuitively, each bounding function represents one of the possible outcomes from evaluating (11) in the data. Finally, the full set of bounding functions is defined for the ATE, so we also perform all necessary subtractions. For example, the final bounding functions for the lower bound on $\Delta\,(t_1, t_2)$ are created from all possible subtractions of the $E[y(t_1)]$ upper bound bounding functions from the $E[Y(t_2)]$ lower bound bounding functions. In total, each bound on each ATE implies $(2\text{^}\{5-1\})\text{^}2 = 256$ bounding functions, denoted $\theta^l(v)$ and $\theta^u(v)$, $v = 1, \ldots, 256$, for the respective lower and upper bounds.

The key aspect of the CLR procedure is that the steps for estimation of the bounds and for constructing confidence intervals are completed on the individual bounding functions prior to taking the associated maximum (or minimum). This is referred to as the *precision adjustment* and proceeds as follows.[18] Generally, the adjustment involves taking the product of a critical value $\kappa(p)$ and the pointwise standard error $s(v)$ of the bounding function. For lower bounds, this product is subtracted from the estimator $\widehat{\theta^l}(v)$; for upper bounds, it is added to $\widehat{\theta^u}(v)$. Then—depending on the choice of critical value p—the adjustment yields either the half-median unbiased estimator of the lower an dupper bounds ($p = 0.5$), or the desired lower and upper limits of the confidence interval (see below). In this way, the CLR method offers the convenience that bias correction and inference are carried out within the same procedure. Also, we note that the resulting large number of bounding functions makes it crucial to implement the CLR procedure for estimation of the bounds and the construction of valid confidence intervals, as in our experience, the amount of bias tends to increase with the number of bounding functions.

More specifically, the precision-corrected estimators of the bounding functions for each average treatment effect bound are given by:

$$(12)\ \widehat{\theta^l}(p) = \max_v \{\widehat{\theta^l}(v) - \kappa^l(p) \cdot s^l(v)\}$$

and

$$(13)\ \widehat{\theta^u}(p) = \min_v \{\widehat{\theta^u}(v) + \kappa^u(p) \cdot s^u(v)\}$$

where $\widehat{\theta^l}(v)$ and $\widehat{\theta^u}(v)$ are the unadjusted estimators of the bounding functions, and $s^l(v)$ and $s^u(v)$ are their associated standard errors. The critical values $\kappa^l(p)$ and $\kappa^u(p)$ are computed via simulations as follows.

---

[18] This process requires that the estimators of $\theta^l(v)$ and $\theta^u(v)$ are consistent and asymptotically normal. Since in our case these estimators are made up of sample means and sample proportions, this condition is met.

Let $\widehat{\boldsymbol{\gamma}^l}$ be a 256-dimensional column vector of all the unadjusted bounding functions for the lower bound, with $\widehat{\boldsymbol{\gamma}^u}$ defined likewise for the upper bounds. An initial step obtains from $B = 999$ bootstrap replications a consistent estimate $\widehat{\boldsymbol{\Omega}^l}$ of the asymptotic variance-covariance matrix of $\sqrt{N}\left(\widehat{\boldsymbol{\gamma}^l} - \boldsymbol{\gamma}^l\right)$ (an analogous process is followed for the upper bounds). With $\widehat{\boldsymbol{g}^l}(\boldsymbol{v})'$ the v^{th} row of $\widehat{\Omega^{1/2,l}}$, we can thus define $s^l(v) \equiv \frac{||\widehat{g^l}(v)||}{\sqrt{N}}$. Next, following CLR, we simulate $R = 100{,}000$ draws from a $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ distribution, where $\boldsymbol{I}$ is the 256×256 identity matrix. The draws are labelled $\boldsymbol{Z}_r$, $r = 1, \ldots, 100{,}000$, and are used to compute $Z_r^*(v) \equiv \widehat{\boldsymbol{g}^l}(\boldsymbol{v})'\boldsymbol{Z}_r / \left|\left|\widehat{g^l}(v)\right|\right|$ for each r and v. In each replication, we select the maximum over the set of $Z_r^*(1), \ldots, Z_r^*(256)$. From the resulting R values, we compute $\kappa^l(c)$, which is defined as the c^{th} quantile of the values, where $c \equiv 1 - (0.1/\log N)$. This value is used to construct the following set:

$$\widehat{V^l} = \left\{v \in \mathcal{V}^l : \widehat{\theta^l}(v) \geq \max_{\tilde{v} \in \mathcal{V}^l}\left[\widehat{\theta^l}(\tilde{v}) - \kappa^l(c) \cdot s^l(\tilde{v})\right] - 2\kappa^l(c) \cdot s^l(v)\right\}$$

where $\mathcal{V}^l$ is the indexing set for the lower bound bounding functions $\theta^l(v)$. From the values $Z_r^*(v)$, we next take the maximum from each replication r, this time restricting the search only to $v \in \widehat{V^l}$. The CLR critical value $\kappa^l(p)$ is taken as the p^{th} quantile of the R values, such that $\kappa^l(0.5)$ gives the half-median unbiased estimate of the lower bound $\widehat{\theta^l}(0.5)$ on the average treatment effect.

To obtain the lower bound on a $(1-\alpha)\cdot 100\%$ confidence interval, we must make one final adjustment which accounts for the width of the identified set. Borrowing notation from CLR (2013), define:

$$\hat{\Gamma} \equiv \widehat{\theta^u}(0.5) - \widehat{\theta^l}(0.5)$$
$$\widehat{\Gamma^+} \equiv \max\{0, \hat{\Gamma}\}$$
$$\rho \equiv \max\{\widehat{\theta^u}(0.75) - \widehat{\theta^u}(0.25), \widehat{\theta^l}(0.25) - \widehat{\theta^l}(0.75)\}$$
$$\tau \equiv 1/(\rho \log N)$$
$$\hat{p} \equiv 1 - \Phi\left(\tau\widehat{\Gamma^+}\right) \cdot \alpha$$

where $\Phi(\cdot)$ is the standard normal CDF. We report 95% confidence intervals for the estimates based on, for example, $\widehat{\theta^l}(\hat{p})$, which uses the critical value $\kappa^l(\hat{p})$, with $\alpha = 0.05$ in the expression for $\hat{p}$.