ABSTRACT

AN ANALYSIS OF MENTAL HEALTH AND WORKPLACE ABSENTEEISM

Kristin Marie Vrona, Ph.D.
Department of Economics
Northern Illinois University, 2025
Virginia Wilcox, Director

Untreated mental illness contributes substantially to productivity losses in the workplace. This dissertation investigates how mental health affects health-related absenteeism among working adults in the United States and examines the extent to which fringe benefits and workplace characteristics can moderate this relationship. By focusing on the labor supply consequences of mental illness, the analysis provides evidence to inform employer decisions about investing in healthcare benefits and workplace accommodations.

Using data from the Medical Expenditure Panel Survey (MEPS) from 2010 to 2014, I conduct separate analyses for men and women. A key strength of this study is its focus on absences due exclusively to the worker's own health, excluding those related to others' health conditions. MEPS uniquely provides both exogenous indicators of diagnosed mental illness and continuous measures of point-in-time psychological well-being, enabling a spectral approach to analyzing mental health that extends beyond diagnostic status. The dataset also includes rich information on workplace characteristics, fringe benefits, and physical health, supporting a focused estimation of the causal effects of mental illness on absenteeism.

Results indicate that the effect of mental illness on absenteeism is more robust for women than for men after conditioning on employer-provided fringe benefits. Among men, the cost

of health insurance significantly mitigates the impact of mental illness on absenteeism, a pattern not observed among women. Additionally, workers with employer-sponsored health insurance exhibit lower absenteeism in response to mental illness when they are offered a choice among multiple plans, suggesting beneficial selection for individuals with a known diagnosis.

In examining how diagnostic status interacts with general psychological well-being, a consistent pattern emerges: individuals with high psychological distress but no formal diagnosis have higher absenteeism rates than similarly distressed individuals with a diagnosis. This pattern is also evident in the context of physical health, where diagnosed individuals exhibit better attendance outcomes under poor physical conditions than undiagnosed peers—suggesting that diagnosis may improve coping and health management strategies.

This dissertation demonstrates that mental illness significantly influences labor supply decisions, particularly absenteeism, and that the magnitude of this effect is likely to be underestimated if sample selection bias is not addressed. A classic Heckman selection model, supported by strong exclusion restrictions, produces robust estimates and reveals gender differences: mental illness more strongly predicts absenteeism among women, while selection bias is more pronounced among men. A semi-parametric copula-based selection model, which accommodates the discrete and overdispersed nature of absence data, confirms these findings.

Given the high prevalence of mental illness in the workforce, employer recognition and response are not optional but inevitable. How employers choose to respond—such as through benefit design and early intervention—can improve both productivity and workforce well-being. These findings underscore the importance of accounting for selection effects in labor supply models and highlight the critical role of household dynamics and dependent care responsibilities in shaping absence behavior.

NORTHERN ILLINOIS UNIVERSITY
DE KALB, ILLINOIS

MAY 2025

AN ANALYSIS OF MENTAL HEALTH AND WORKPLACE ABSENTEEISM

BY

KRISTIN MARIE VRONA

A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECONOMICS

Dissertation Director:
Virginia Wilcox

ACKNOWLEDGEMENTS

The expertise and mentorship provided by my dissertation committee have been invaluable, and I would like to take this opportunity to thank them.

First, I would like to express my deepest gratitude to my dissertation director, Dr. Virginia Wilcox. I have long admired her foundational research in the field of health economics, and it was an honor to receive her guidance and advice throughout this process.

I also thank Dr. Anna Klis for her mentorship and academic creativity and Dr. Ponomareva whose expertise in econometrics and insightful feedback were instrumental in producing this piece of research.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

Mental illness is one of the most prevalent and costly health conditions in the United States, affecting one in five adults annually. However, treatment rates for mental illness are some of the lowest of any chronic illness, with 47.2 percent of US adults with mental illness receiving treatment in 2021 (National Alliance on Mental Illness [NAMI], 2023). Additionally, 160 million Americans live in areas with shortages of mental health professionals and US adults report an average delay of 11 years between symptom onset and treatment (NAMI, 2023).

While mental illness is not a communicable disease by definition, it operates similarly, impacting not only the individual suffering but also the other communities. The ripple effect of mental illness describes the mechanism through which a lack of sufficient treatment leads to broader impacts on an individuals functioning, in turn leading to declining outcomes for the household, and ultimately the general community and the world (NAMI, 2023; National Institute of Mental Health [NIMH], 2024).

Untreated mental health induces substantial economic burdens, such as reduced productivity and heightened rates of disability. Depression and anxiety disorders alone contribute to an estimated $1 trillion in lost productivity worldwide each year (World Health Organization [WHO], 2024). The gap in prevalence rates and treatment access in the United States has severe implications, as untreated symptoms lead to higher rates of substance use, bankruptcy, homelessness, and suicide (NIMH, 2024).

Legislative efforts, including the Affordable Care Act's (ACA) mental health parity mandates, have attempted to address some barriers, yet access to mental healthcare remains limited. While parity laws have reduced discriminatory coverage practices, significant barriers persist (NIMH, 2024; WHO, 2024).

Though now required by law to cover mental health treatment services at the same level as physical health services, state enforcement barely exists (The Kennedy-Satcher Center for Mental Health Equity, The Kennedy Forum, The Carter Center, & Well Being Trust, 2018) allowing insurance companies to maintain practices that drive a disparity between the two types of care by driving mental health professionals out of the insured marketplace.

Delays in reimbursement, denial of treatment and psychotropic medication, and late notifications that a service will not receive reimbursement have driven these professionals to private practice care, restricting accessibility to millions of Americans (Bishop et al., 2014; WHO, 2024). Further, mental health professionals willing to participate in the insured market face a professional workforce increasingly strained by retirements outpacing new entrants.

The economic impact of untreated mental illness underscores the need for a shift in focus. Prevalence rates of mental illness make it nearly impossible for employers to never have to acknowledge; how this is acknowledged, however, does give some control back to the employers that recognize and adapt to this fact. Evidence suggests that by treating mental healthcare as an investment rather than a cost of labor, employers benefit from worker wellness through productive gains, improved collaboration, and reduced turnover (Ashwood et al., 2017; Cseh, 2008; Fletcher, 2013). These outcomes can be achieved through employer workplace policies that promote well-being. In doing so, employers can not only mitigate the economic toll to their own operations that may be associated with poor mental health among their own workers, but can also play substantial role in improving the mental health crisis and contribute to a healthier, more resilient workforce.

Much of the focus in recent years has been on the rising cost of mental healthcare and the burden this places on insurers and employers rather than on the importance of improving access to and utilization of mental healthcare. However, given evidence that access to treatment services may improve general productive capacity, employers in particular should not only consider the direct costs associated with providing quality mental healthcare to their workers, but the benefits derived by improved labor market outcomes among employees facing mental illness. My dissertation will analyze the effect of mental health on labor supply outcomes in the form of absenteeism, and empirically estimate how workplace policies and improvements in accessibility to mental healthcare can moderate this effect.

1.1 Historical Context

Though the need for better access to and quality of mental healthcare in the United States has been stressed throughout the last few decades, the issue of its provision has been largely ignored by intermediary organizations that connect individuals to affordable care, and broader legislative initiatives have fallen short in improving the access to and utilization of care. The ill-defined legislative definition of mental health parity – which at the very least acknowledges that mental and physical illness are equitable in the potential destructive impact to one's life, and at the very most requires equity of cost sharing schema of physical and mental health services covered under large group health insurance plans – has been exploited to limit any measurable progress in access to mental healthcare.

In fact, some reactionary policies imposed to minimize the additional costs to firms posed by mental health parity legislation have put additional strains upon the market for mental healthcare, such as mental health carve-outs through managed care plans for mental health services which are completely separated from the employer's broader pool of workers receiving medical insurance through a job.

Managed care plans require that a care manager decide on treatment duration and type and usually have an annual cap on inpatient and outpatient services. The more stringent design of these plans, as well as the headache of having to deal with an additional insurance intermediary, may deter some employees from ever seeking treatment for mental health issues. Additionally, if an employee under a managed care plan does choose to seek treatment, the cost-saving design of the plan might mean they are not receiving the most beneficial type of treatment. This in turn results in substantial productivity loss associated with symptom severity (Goetzel et al., 2002, p. 327), meaning the total cost savings associated with a shift to this type of employee healthcare offering might be overestimated.

In favor of this argument, a study from Rosenheck et al. (1999) finds that the introduction of mental healthcare cost containment methods in a firm decreased the utilization of mental health services by its employees while utilization of medical services was found to increase indicating that untreated mental illness can result in higher medical costs borne unto the employer. On top of this, these researchers found evidence of greater productivity loss in workers with mental illness after the introduction of these policies. Nevertheless, the increasing focus in recent years on containing healthcare costs suggests that employers may not take indirect responses to changes in health coverage like these into account when making investment decisions on employee healthcare.

The Mental Health Parity Act (MHPA) was implemented in 1996 and required insurers of mental healthcare to offer plans with annual and lifetime limits equal to those in place for general medical healthcare. The MHPA mainly impacted employers that offered comprehensive healthcare packages with coverage of both general medical and mental health services. The Act does not require employers offering only general medical health benefits to additionally offer mental health benefits, so employers offering only general health insurance benefits were not impacted by the Act. Additionally, employers offering two separate plans for general medical and mental healthcare (i.e., a mental health carve-out) would not have to meet parity requirements. In theory, mental health carve-out schemes are an attractive option to employers because they mitigate some of the healthcare costs imposed by parity-regulated plans while reducing the risk of turnover or lost worker commitment that could occur if cutting mental health benefits completely (Goldman et al., 1998).

After enactment of the MHPA, mental health carve-outs became increasingly popular among employers. Benefit package designs exhibiting a mental health carve-out have mental health benefits separated out of general medical insurance plans so that individuals wishing to receive employer sponsored mental healthcare must do so through a managed care plan. Employers often instituted such carve-outs in attempts to contain rising costs of healthcare. Under a mental health carve-out, employers did not have to follow MHPA regulations because managed care plans for mental health were provided by a different entity than the medical insurance plans. Even when faced with such cost containment efforts, mental healthcare costs still continued to rise at higher rates while access to such care became even more limited, likely due to the switching of cost burden (rather than actual improvements in costs) to the worker and additional efforts to deter service utilization (Goetzel, 2002).

The Mental Health Parity and Addiction Equity Act (MHPAEA) of 2008 took the idea of equality in coverage a step further, requiring treatment limits, co-pays, and in- and out-of-network coverage to be equivalent for insurers or employers offering medical and mental healthcare coverage. The Act additionally expanded parity regulations to managed care plans, making carve-out initiatives less appealing. However, the Act still does not require that insurers or employers offer any mental health coverage, making it still possible for employers to mitigate costs of healthcare by stripping mental healthcare coverage from benefit offerings. Paired with economic uncertainty of The Great Recession, this MHPAEA loophole likely made the complete removal of mental healthcare from worker benefits appealing to many employers as this would largely cut costs without a large risk of losing worker commitment.

Compared to the MHPA, the enactment of the MHPAEA was followed by stagnant rates of mental healthcare utilization (Alikhan et al., 2019; Azocar et al., 2016), further indicating that the 2008 Act did not do much to address the affordability of mental healthcare, or the incentives that drive the market for mental healthcare.

In recognition of the societal burdens of barriers to mental healthcare access, many individual states have developed stronger parity legislation that require insurers not previously offering comprehensive health plans to do so and requiring plans exhibit the parity of costs and treatment allocations across general medical and mental health. In 2010, the Affordable Care Act (ACA) was introduced, which expanded full-parity policy to the federal level to reduce carve-out practices and inequity of care in states with bare-minimum parity regulations. The ACA again induced higher cost burdens onto employers, many of who met the ACA with opposition.

In an attempt to counteract this rising cost burden, President Donald Trump signed the Executive Order Promoting Healthcare Choice and Competition in 2017 after bills proposing the repeal of certain elements of the ACA failed to pass the senate. This executive order rolled back many regulations imposed by the ACA, including the federally expanded mental health coverage rules, giving power back to the states. Thus, states must also consider burdens associated with relatively weak parity laws to ensure optimal policy decisions on which aspects of the ACA to maintain.

Though managed care plans have been phased out for mental health[1], another restrictive measure has taken its place in the form of self-insurance. Employers offering insurance coverage often define coverage as a reimbursement for premiums and deductibles when a worker is insured through the market place. On its own, self insurance may be a good option for employers trying to mitigate costs of healthcare by only paying for the instances in which an employee seeks care.

[1] Note that managed care plans still exist for other chronic and high-cost diseases.

However, the bundling of self-insurance and stop-loss insurance, a common firm practice, negates the benefits of self-insured policies. This type of bundling brings back the oversight manager system similar to managed care gatekeepers, but leaving the worker uninformed of the communication of the costs that will be covered and what is and is not necessary to cover. As mental illness is a chronic condition, costs are already higher on average per year for those who are lucky to receive consistent treatment, leading to a disparity and possibly loss of cost-sharing if a condition of another kind comes up considering the caps put on employer coverage for the entire pool of employees. That is, chronic condition sufferers in general are at higher risk of loss for the same injury as generally health individuals are more likely to be responsible for costs of the unanticipated injury.

A final note, the COVID-19 pandemic has highlighted the misalignment of employer and laborer incentives. Higher rates of job vacancies may be misrepresented due to the increase in under-employment across the US observed during the business recovery period of the COVID-19. A relatively stable portion of the population is underemployed, a practice that may have been necessary during the pandemic, but has persisted through the recovery period, becoming common practice for employers.[2] Some underemployment offers may allow for the reduction of benefits such as cost sharing for insurance. While my dissertation focuses on pre-COVID data, pooled across years 2010 - 2014, future research directions should control for this pandemic.

In this dissertation, I examine the role of mental health on productive time loss in the form of absenteeism to inform the often ignored benefits that come from promotion of worker wellness.

[2] US Bureau of Labor Statistics. (n.d.). Charts related to the latest "The employment situation" news release: more chart packages. US Bureau of Labor Statistics. Retrieved January 13, 2024, from https://www.bls.gov/charts/employment-situation/alternative-measures-of-labor-underutilization.htm.

1.2 Preview of Findings

Overall, results suggest that men exhibit multiple channels through which job and insurance traits impact absenteeism, while women with mental illness exhibit less sensitive short run labor supply compared to men, regardless of which job factors and insurance characteristics are considered. Results do give some merit to the argument that fringe benefits like paid sick leave and offering health insurance successfully encourage individuals to make informed decisions about when to stay home to nurture health. In addition, results suggest that there may be an additional benefit to allowing employees to choose from multiple plans, though this paper does not consider the relative cost-benefit of such policies on the employer's end.

The main dependent variable represents labor supply decisions made on a daily basis after observing one's current state of health; this is a count of the number of days absent from work due to health, which is weighted by the number of days an individual is employed during the year of interest. The MEPS has separate prompts for working respondents to report days absent from work due to their *own* physical or mental health. A separate prompt asks respondents to report days absent due to *another person's* health. I specify the first variable, recording individual-level health induced absences so that I am able to improve the causality of estimates.

This individual-level measure allows labor supply to be synonymous with productive work time through the channel of health-related absenteeism while an individual is employed. Results suggest that mental illness generates additional absenteeism and that income-oriented fringe benefits, such as the ability to earn bonuses mitigate the effect of mental illness on absenteeism, while those that supplement time away from work like paid sick leave amplify such effect. Findings suggest that using the strictly exogenous indicator of mental illness leads to results consistent with those portrayed in studies that have used more time-dependent and subjective terms that may be posed to attenuation bias for or against.

A diagnosed mental illness at the start of a survey period is on average indicative about a day more in absence before controlling for factors like sick time and insurance that may induce absenteeism in the short run, but may also reduce the aggregate time loss from usual contracted work hours if considered over time and is something thought provoking to look at in the future.

The remainder of this paper is organized as follows: Chapter 2 is a review of the relevant literature intersection mental health and labor force outcomes. Chapter 3 describes the theoretical approach used in this dissertation. Chapter 4 describes the data and variables; I also discuss sample statistics. Chapter 5 describes econometric implementation; Chapter 6 provides results; Chapter 7 concludes.

CHAPTER 2

LITERATURE REVIEW

It is the goal of this paper to utilize modeling techniques of some of the relevant pieces of literature while improving on issues of generalizability and causality. In what follows, I summarize key pieces of literature in health economics related to mental health and labor markets.

Additionally, this dissertation acknowledges the differences observed between men and women and labor supply decisions. A recent study by Jia and Vatto (2021) reinforces this consistent discrepancy in how men and women supply labor. A 2021 study from these researchers utilized an individual-level panel of Norwegian women; this research found that women do not alter labor supply in response to changes in tax policy impacting income even after controlling for spousal features, past work history, and hours and is consistent with the theory that the elasticity of labor supply for women is more likely defined by factors impacting home production.

Though the findings of Jia and Vatto (2021) are consistent with the larger body of literature on labor supply of women in the United States, it is a much more recent piece of literature and it begs the question as to whether the degree to which this pattern is exhibited varies across country and over time, highlighting the need to analyze features of employment contracts as key indicators for individuals in the United States. Norway provides universal healthcare to all citizens as well as mandatory sick leave benefits, giving factors such as time to improve health after an injury or ailment less dependent on income and job-attachment. Employment in the US in specific positions is necessary for many to receive health insurance that is affordable.

Goldman et al. (1998) set out to examine the impact of switching mental healthcare benefits to a managed care plan on a large company's mental healthcare cost and utilization. Data on the company's mental health claims from 1988 to 1996 is used to compare cost and utilization prior to and after the company's switch to managed care. Researchers find that the company's mental healthcare costs fell 43 percent across the years of study while the probability of utilizing any mental healthcare rose at a slower rate over periods after the switch. Average cost per outpatient session fell during the change, then returned to the level observed prior to the switch; meanwhile, the average number of outpatient sessions in a year consistently declined under the new plan, indicating that the fall in aggregate outpatient costs is most likely a result of lower utilization.

In the end, Goldman et al. (1998) conclude that managed care plans are an effective way to cut costs of employee mental healthcare. There are, however, multiple shortcomings of the study. Factors like sex and age were not considered in the analysis, thus failing to account for any changes in workforce composition at the firm over the period studied. Changes in the composition of employee demographics might explain some of the variation in utilization rates, rather than the plan switch. It would be beneficial to include information on employee turnover rates for similar reasons. This study provides insight of the magnitude to which results and implications differ if considering only direct costs when analyzing the success of new policies, versus both direct and indirect costs.

In direct contrast to the 1998 study from Goldman et al. (1998), a study from Rosenheck et al. (1999). The study finds that the introduction of mental healthcare cost containment methods in a large firm (over 1,000 workers, other details unspecified), decreased the utilization of mental health services by employees. Simultaneously, utilization of medical services was found to increase, indicating that untreated mental illness can result in higher medical costs and that cutting costs in the area of mental health does not necessarily mean cutting healthcare costs as a whole. On top of this, these researchers find evidence of greater productivity loss in workers with mental illness after the introduction of these policies. Bir and Eggleston (2009) highlight the importance of considering non-random selection based on specific health conditions in the study of healthcare costs. Ignoring the incentives of insurers to distort the quality of coverage and care offered to individuals facing a given condition likely upward biases the estimated effect of one's condition on medical costs. In particular, they look at those enrolled in managed care programs for more expensive illnesses and find that for at-risk populations, these programs discourage treatment rather than cut costs.

Goetzel et al. (2002) argue that employers providing generous mental healthcare benefits will see benefits in the form of productivity and worker commitment, particularly in the case of workers with depression. The authors discuss multiple barriers to effective depression management in the workplace, including stigma, shortcomings of managed care organizations and insurers, and information gaps between employers and employees regarding sensitive topics. They also discuss barriers to treatment induced by depressive symptoms themselves, as well as lack of confidence in physicians and the shortage of mental health professionals. The authors suggest multiple ways that employers can better manage the mental health of their workers, including implementing programs to inform employees on the persistence and high prevalence rates in the general population, as well as teaching them how to spot mental health issues early. Similar discussions take place in an extensive review of the literature on depression and worker productivity by Simon et al. (2001).

Goetzel et al. (2004) derive important implications on the indirect costs mental illness may pose to firms. The researchers find a large negative relationship between mental illness and productivity. They loosely estimate monetary costs of productivity losses associated with various categories of illness based on responses to multiple self-reported surveys. Goetzel et al. estimate that mental illness is among the top four most costly illness categories in terms of productivity loss. However, these findings should be interpreted with caution, as estimates of these costs are derived using aggregated numbers of illness prevalence and national measures of average dollars per hour earned by employees. Ashwood et al. (2017) also attempt to monetize the cost and benefits associated with mental health, or more specifically, improvements in mental health service utilization. The main focus of their paper is a cost-benefit analysis of a program aimed at reducing stigma associated with mental illness in California. Prior to the cost-benefit portion of the study, researchers empirically test for impacts of program exposure on work absences and the ability to obtain employment and find evidence that adults with mental illness who have been exposed to the program exhibit fewer absences and a greater likelihood of employment compared to adults with mental illness who have not been exposed. This is likely because of the improved treatment utilization numbers observed among those who have been exposed to the program.

In observation of low mental health service utilization numbers worldwide, Kohn et al. (2004) attempt to estimate the treatment gap (percentage of individuals who require care but do not receive treatment) for each type of mental health disorder. These researchers extensively review past empirical studies to find the best estimates of prevalence rates and treatment utilization for each of several mental disorders. These numbers are then used to calculate estimated treatment gaps for each mental disorder. The average treatment gap for the disorders under study is about 56 percent. Due to differences in medical practices and cultural norms across countries, estimated treatment gaps for each disorder vary depending on the availability of treatment or the likelihood of being diagnosed within each of the countries studied.

In regard to the United States, Kohn et al. estimate a treatment gap of 56.9 percent for major depression and find evidence that this gap has decreased as it has become more common for health insurance to cover psychotropic medications, indicating that having health insurance is a significant factor in the decision to seek treatment in the United States. How attitudes toward and characteristics of insurance may influence the effect of mental illness is one point of analysis in this dissertation.

Much of the literature on mental health and the labor force looks to identify a link between mental health and earnings or labor force participation. A study of particular importance among this group of literature comes from Cseh (2008) who analyzes the impact of depression on earnings across years. The data source permits Cseh to control for severity of depression by using a raw depression-scale score. Using this proxy for symptom severity indicates a significant and negative effect of depression severity on wages across years. This finding illustrates the importance of controlling for the severity of emotional symptoms in the analysis. This is a strength of the Cseh study that I attempt to utilize in my own analysis by controlling for a scale scoring variable that measures the severity of emotional distress faced by an individual. Cseh was not able to additionally recognize that the severity of any physical ailments should also be controlled for due to data limitations as well as accounting for exogenous mental illness, I was fortunate to have access to these types of data and so improve the methodology in terms of causality.

Fletcher (2013) attempts to determine the impact of adolescent depression on future labor market outcomes. A strength of the Fletcher study is the elimination of the Cseh study's reverse causality limitation because: while it is plausible that an adolescent's depression may impact their future employment and wage, it is not plausible to theorize that an adolescent's future labor market outcomes are the cause of their depression. Empirical estimates suggest that depression in adolescence has a negative and significant effect on the employment and earnings in adulthood. Fletcher accounts for environmental factors that may impact the magnitude of estimated effects, which is a similar trait to a study by Bubonya et al. (2017) that considers how work environmental factors may impact estimated effects (Bubonya et al. (2017) will be discussed in more detail later). The implications of Fletcher (2013) are particularly important for policy initiative to promote early intervention to manage symptoms of mental illness.

Von Korff et al. (1992) evaluate the impact of untreated depression on work impairment. Participant responses at the start and end of a 12 month period are utilized in analysis. Baseline estimates suggest that depressive symptoms lead to greater work impairment. Findings also indicate no significant changes in these higher levels of work impairment for individuals reporting no change in depressive symptoms as of the follow-up interview. Individuals reporting an improvement in depressive symptoms at the follow-up interview reported a fall in the number of impaired work days by about 36 percent, illustrating the importance of access to early intervention and ongoing treatment.

French and Zarkin (1998) studied the impact of mental health on absenteeism and earnings. They define two models estimating absenteeism: one represents absence within the last 30 days as a dichotomous variable; the other contains a count variable representing total absences within the last 30 days. They employ logit regression to estimate the first absenteeism model and a negative binomial form estimates the second. They control for physical health in their model, however, they do not consider that the interaction between physical and mental health may play a role in predicting absenteeism. Both absenteeism models indicate that poor mental health has a significant positive impact on absenteeism. Additionally, they find evidence that poor mental health has a negative impact on earnings. Unfortunately, the degree to which results generated by French and Zarkin can be generalized to the working population poses major issue due to a sample size of only 408 observations and the fact that all workers in the sample share the same place of employment.

Bubonya et al. (2017) examine how interactions between poor mental health and job characteristics influence presenteeism and absenteeism among workers. This study solves many of the issues of the French and Zarkin study by utilizing a panel data source that is representative of the Australian working population and that yields a large sample size. These researchers find evidence that mental health has a significant impact on productivity, as well as significant differences in the impact of job characteristics on productivity between workers with and without mental illness. Particularly, they find that those who face mental distress see higher rates of both absenteeism and presenteeism. These higher rates of presenteeism among workers in poor mental health are less flexible to changes in job factors than the rates of presenteeism observed for mentally healthy workers, indicating that workers with mental illness may always face a greater level of productivity loss while at work, regardless of environmental factors.

Conversely, absenteeism rates tend to vary more with changes in job factors for the mentally ill workers than mentally healthy workers, indicating that job factors only influence the lower levels of productivity in mentally ill workers through the channel of the decision making process on when to stay home from work. Bubonya et al. are unable to control for things like treatment or severity of the mental illness. This may be problematic because it is likely that individuals who are in intervention for a mental disorder see impacts more in line with mentally healthy workers than mentally ill workers who are not in some form of intervention. Bubonya et al. also do not consider access to health insurance as one of the job characteristics that may impact the effect of mental health on productivity, probably because health insurance is universal in Australia, whereas in the United States, employer-provided health insurance is most prevalent.

Building upon the literature on the topic of mental health and labor force productive outcomes, the following questions will be addressed in this dissertation: How do available fringe benefits moderate the absence behavior of workers with mental illness? How does diagnosed mental illness interact with other measures of health to impact absence behavior?

CHAPTER 3

THEORETICAL APPROACH

This section focuses on the theoretical approach applied to form testable hypotheses. I innovate a common time allocation model of labor supply and health production and resulting in a causal model of health-induced absenteeism. I then examine how fringe benefits can alter behavior after a negative health shock.

3.1 Supply of Labor

Time allocation models of labor supply consider the decision making process associated with market goods as well as time, which is a scarce resource. For intuitive purposes, individuals are modeled as producers that take intermediate inputs such as time and market goods to produce useful commodities to enjoy (Becker, 1965). For example, consider the consumption of a final good: a meal. A decision-maker utilizes inputs such as ingredients purchased from the store and the time spent to cook a meal. If not cooking the meal oneself, the process still involves either the cost of gas or delivery fees, tips, and time. Something as simple as watching television can also be defined by a production process, without loss of generality, as it requires a positive time allotment and market services such as streaming platforms and cable packages, as well as a working television.

For final products c_k for $k = 1, .., K$, an individual faces the following utility function:

$$U = u(c_1, c_2, ..., c_K), \text{ with } \frac{\partial U}{\partial c_k} > 0 \ \forall k. \tag{3.1}$$

Each final good c_k can be defined by its respective production process that turns market goods and time into final units of consumption. For final good c_k, this process is characterized by

$$c_k = f_k(T_k, x_k; e_k), \tag{3.2}$$

where T_k is a vector of time inputs allocated to the production of final good c_k, x_k is a vector of market goods used in production, and e_k represents the efficiency of the process, characterized by exogenous factors such as one's age or education level (referring to the previous example, an individual's cooking skills or experience may define e_k).

In each case, the efficiency factor, e_k, impacts how much time and how many market goods are required to achieve a certain level of a final good, and may or may not be equivalent across production processes. Given each production process, f_k, utility function (3.1) can be rewritten as

$$U = u(f_1, ..., f_K) = u(x_1, ..., x_K; T_1, ..., T_K). \tag{3.3}$$

The separability between time and market inputs exhibited by (3.3) demonstrates that individual production is bounded from above due to the scarcity of income and time resources.

Of particular interest in this paper is the process of producing health which acts as the final good to be "consumed." I define the production of health status, H, in the following manner:

$$H = f_h(T_h, x_h; MH, PH; e_h), \tag{3.4}$$

where T_h is a choice of time allotted to the development of health, such as time spent exercising and utilizing healthcare services, x_h is a vector of market goods utilized in health production, such as vitamins and supplements or health insurance coverage, MH and PH are endowments of mental and physical health, respectively Non-health related factors driving the efficiency of the production process are captured in e_h such as education, age, or quality of care, and (3.4) is a concave function exhibiting diminishing returns to factor inputs. Function (3.4) innovates the health production function proposed by Grossman (1972).

I define C as a conglomerate final good that nests each of the production functions of final goods besides health, $c_k, ..., c_{K-1}$, within it so that one can write

$$C = f_c(T_c, x_c; e_c), \tag{3.5}$$

where $T_c = \sum_{k=1}^{K-1} T_k$ and $x_c = \sum_{k=1}^{K-1} x_k$. Equation (3.1) can therefore be rewritten in the following manner:

$$U = u(C, H) \equiv u(f_c, f_h) \equiv u(x_c, x_h; T_c, T_h), \tag{3.6}$$

so that the level of utility realized depends on choices of market inputs and time allocations. It is assumed that an individual maximizes their utility subject to the time constraint,

$$T = T_c + T_h + N, \tag{3.7}$$

where T is total time available and N is time allocated to occupational work, as well as a budget constraint,

$$I = wN + V = p_c x_c + p_h x_h, \tag{3.8}$$

where I is total income, V is non-earned income such as monetary gifts or inheritances,[1] w is the market wage rate (assumed to be exogenous), and p_i for $i = \{c, h\}$ are vectors of input prices corresponding to the market goods utilized in the production processes (3.4) and (3.5).

[1] Note that money acquired in one's leisure time, such as that from gambling, takes time and market inputs to obtain and will be tied to C, not to V. The same applies to present stock holdings – the person is taking their income and investing it so each dollar invested or gambled makes up part of vector x_c and the price varies with the market.

The time constraint can be substituted into (3.8) to yield a single "full income" constraint,

$$\begin{aligned} w(T - T_c - T_h) + V &= p_c x_c + p_h x_h \\ \implies wT + V &= p_c x_c + p_h x_h + wT_c + wT_h. \end{aligned} \tag{3.9}$$

The left-hand side of the second line in (3.9) represents full income – the income received when an individual allots all available time to occupational labor.

The results of the current utility maximization problem are more intuitive if the production functions (3.4) and (3.5) are redefined. Define $T_c \equiv t_c C$, $x_c \equiv b_c C$, $T_h \equiv t_h H$, $x_h \equiv b_h H$, where t_i and b_i for $i = \{c, h\}$ are vectors of time per unit and market inputs per unit required to produce levels of final goods C and H, respectively.[2] With this, a single resource constraint can be expressed as

$$(p_c b_c + w t_c)C + (p_h b_h + w t_h)H = wT + V, \tag{3.10}$$

where the full price of each unit of the final good, C and H, is the sum of both the direct costs (prices of market goods) and indirect costs (time away from work) associated with each unit produced (Becker, 1965, 6). Now, the individual maximizes their utility by choosing optimal levels of b_i and t_i for $i = \{c, h\}$.

An individual will allot additional units of expenditure and time up to the point for which the marginal utility resulting from an additional unit of the respective input equals zero; this is equivalent to saying that available resources will continue to be allotted to production processes until the marginal product of the input is zero. It should be noted that the choices of market good inputs and time inputs are not independent. A condition of

[2] Note that t_h and/or b_h are decreasing in measures of the efficiency of the health production process, MH, PH, and e_h. t_c and/or b_c are decreasing in efficiency factor e_c.

utility maximization is that the marginal rate of substitution (MRS) between these types of inputs be equal to the ratio of input costs per unit.

At this point in the analysis, it is assumed that an individual has enough information to maximize their utility by choosing the optimal bundle $\{C^*, H^*\}$, which reveals information on the individual's preferences over home production and labor. Manipulation of the model to derive hypotheses to test is more intuitively illustrated by examining the demand for "forgone income."

Call the right-hand side of equation (3.10) as S, which is the full income that would arise if allotting all available time to work. The demand for forgone income $L(C^*, H^*)$, is then

$$L(C^*, H^*) = S - I(C^*, H^*), \tag{3.11}$$

where I is an individual's observed income. The left-hand side, $L(C^*, H^*)$, can be thought of as the indirect cost of utility-seeking; it is the potential earnings lost when an individual allots positive units of time to the production of health and the conglomerate consumption good.[3]

Henceforth focusing on individuals who exhibit some positive level of labor supply, equation (3.11) can be further defined as

$$L^* = w(T_c^* + T_h^*), \tag{3.12}$$

for short run fixed wage, w.

[3]Note that utility maximizing non-workers, without loss of generality, exhibit

$$L(C^*, H^*) = wT + V = V = (p_c b_c + w t_c)C + (p_h b_h + w t_h)H.$$

In such an instance, any general individual could place a price on their own labor (reflected by a reservation wage) after optimizing their potential utility. Determinants of whether an individual enters the labor force or stays out may be based on unobservable features that impact the efficiency of certain inputs in health production. This special case is examined further in Chapter 7.

The expression $I(C^*, H^*)$ from (3.11) can be expressed as

$$I = b_c p_c C + b_h p_h H. \tag{3.13}$$

Substitution of $T_c + T_h = T - N$ into (3.12) and then further substitution of it and (3.13) in Equation (3.11) yields:

$$w(T-N) = wT + V - b_c p_c C - b_h p_h H - V, \tag{3.14}$$

which can be rearranged and simplified as follows:

$$\begin{aligned} -wN &= -b_c p_c C - b_h p_h H \\ \implies N &= \frac{b_c p_c C + b_h p_h H}{w}. \end{aligned} \tag{3.15}$$

Hours of work can now be expressed as a function of health status. Taking the partial derivative with respect to health yields:

$$\frac{\partial N}{\partial H} = \frac{b_h p_h}{w} > 0. \tag{3.16}$$

Equation (3.16) indicates that better health increases labor supply. In (3.16), $b_h p_h$ represents the marginal cost of producing an additional unit of health using market goods and w is the per-unit time cost associated with the level of time allotted to producing health rather than working. The derivative, $\frac{\partial N}{\partial H}$, can additionally be thought of as the marginal product of labor supply with respect to health. Rearranging (3.16) gives

$$\frac{\partial N}{\partial H} w = b_h p_h, \tag{3.17}$$

which states that an individual will continue to produce health up to the point where the marginal benefit of an additional unit of health equals the marginal cost of an additional unit of health.

For the remainder of the analysis, define H^* as an individual's initial utility-maximizing level of health and N^0 as the optimal level of labor supplied when H^* is realized. In the short run, health may fluctuate while income and time constraints do not. In what follows, I describe a simple model of labor demand demonstrating an implicit contract between the demander and supplier such that the labor supplier is rational and conditionally optimizing.

3.2 Demand for Labor

At the commencement of a job, a worker and employer reach a contracted agreement. In the process of reaching this agreement, an employer determines the optimal choice of labor hours per period based on the market wage and the minimum expectations regarding the job. The market wage is assumed to be exogenous and will be equal to the marginal revenue product in equilibrium (Hamermesh, 1993).

A job has minimum expectations for employees, characterized by a predetermined output agenda. Without loss of generality, assume this agenda is characterized by a minimum number of output units produced each period, Y. The minimum number of contracted hours, N_{min}, is increasing in Y. The contract additionally specifies any non-wage compensations available to matched workers, such as health insurance and sick leave.

The derived demand for labor can be represented using a generic profit function. Call π^c an employer's contracted profits for a particular job. Contracted profits persist when the number of hours worked over the period exactly equals N_{min} and all other aspects of the contract are realized (wage rate, non-wage compensation). Mathematically,

$$\pi^c = pY - wN_{min} - B, \tag{3.18}$$

where p is the revenue obtained per unit of output produced for the particular job, Y is the minimum level of output required for a worker to be successful in the role, and B is costs for non-wage characteristics of the job's compensation package, such as the cost of providing health insurance to the employee.

To map out the demand schedule, Y can be replaced by the production function characterizing the particular job, $F(N)$, so that output is allowed to deviate from contracted output. Allowing output to fluctuate implies that the demand for labor, N, may deviate from N_{min} so that wN replaces wN_{min} in (3.18). Taking the first order condition with respect to N gives the following result:

$$p = \frac{w}{\frac{\partial F}{\partial N}}, \tag{3.19}$$

where $\frac{\partial F}{\partial N} > 0$ is the marginal product of labor (MPL) and p is the inverse demand function.

In consideration of the firm's problem thus far, define a general function for the optimal demand for labor,

$$D^* = d(Y, w, N_{min}, B). \tag{3.20}$$

In what follows, I merge the labor supply and demand models to illustrate a reduced form function for absenteeism.

3.3 Absenteeism

Without loss of generality, in terms of a time dimension, define D^* as the total labor hours demanded over the span of one year and N^* as the total labor hours supplied over one year. Prior to the start of the year, an employer and employee must come to an implicit agreement on employment. This condition is met when an employee's evaluation of the optimal labor supply at the point prior to the start of the period, N^0, is equal to N_{min}, the minimum number of optimal hours demanded by a potential employer. The base supply, N^0, is a function of the optimal level of health production at the time of evaluation, H^0, and consumption, C^0.

In this dissertation, absenteeism is defined as an absence from work due to physical or mental illness. Work-related absences occur when contractual constraints induce discrepancies between optimal labor demand and optimal labor supply. This can be illustrated in the following manner:

$$\begin{gathered} A = (D^* - N^*) > 0 \\ \text{for } D^* > N^*, \end{gathered} \tag{3.21}$$

where A represents absenteeism which is characterized by a decision making process that occurs after observing a health shock. The discrepancy between D^* and N^* occur in the short run, as an individual is able to observe their daily health status and reevaluate their optimal labor supply accordingly, while contractual agreements specifying wage, fringe benefits, and minimum time on the job are typically only revisited after a predetermined amount of time.

Substituting functions for D^* and N^* into (3.21) yields an absenteeism function that is decreasing in health,

$$\begin{aligned} A &= [d(w, N_{min}, B) - n(C^*, H^*)] \\ &= [d(w, N_{min}, B) - n(C, f_h(T_h^*, x_h^*; MH, PH, e_h))]. \end{aligned} \tag{3.22}$$

Equation (A.1) results in the hypothesis that absenteeism is inversely related to the endowments of mental and physical health:

$$\frac{\partial A}{\partial MH} < 0, \frac{\partial A}{\partial PH} < 0.$$

In general, illness-related work absenteeism represents the substitutability between time spent earning wages and time spent on improving or maintaining health so that factors which make time inputs in health production more attractive relative to market inputs will exacerbate absenteeism while factors that make market inputs relatively more attractive, such as productivity pay will mitigate absenteeism.

In what follows, I form testable hypotheses derived from this model for certain categories of fringe benefits that workers may receive.

3.3.1 Absenteeism and Fringe Benefits

When in poor mental health, health maintenance requires more inputs per-unit. In what follows, I summarize the impact of fringe benefits as moderators to this phenomena.

I first consider fringe benefits that supplement the costs of time inputs in health production (T_h), such as paid sick leave and paid vacation time. Without loss of generality, I model how paid sick leave alters an individual's labor supply decision when facing a lower endowment of health. Suppose that an individual receives payment s, with $0 < s \leq w$, when they are absent from work due to illness. This reduces the potential earnings lost from $w(T_c + T_h)$ to $wT_c + (w - s)T_h$. Equation (3.15) thus becomes

$$\begin{aligned} N &= \frac{p_c b_c C + p_h b_h H - sT_h - V}{w} \\ &= \frac{p_c b_c C + (p_h b_h - st_h)H - V}{w}. \end{aligned} \tag{3.23}$$

The relationship between health and labor supply is now represented as

$$\frac{\partial N}{\partial H} = \frac{p_h b_h - st_h}{w}. \tag{3.24}$$

The sign of (3.24) illustrates that the direction of the relationship between labor supply and health depends on the choice of b_h relative to t_h. For positive s, there is some degree of reduction in the fall in consumption that would occur if an individual decreased their labor supply and increased t_h without receiving compensation s. An individual in poorer health will choose to increase the share of time-based inputs used in health production relative to market inputs when receiving paid sick leave because this type of fringe benefit supplements the cost of time allocated to health production (reduces the opportunity cost of missing work).

I next consider how fringe benefits that directly impact the ability to purchase market goods used in health production, such as monetary bonuses, commissions, or tips, influence absence behavior.

Consider an individual with mental health endowment, MH_H, who has the opportunity to receive a monetary bonus at his or her job and that this bonus is merit-based. This payment is completely separate from the typical wage they received by the individual. Assume the amount of the monetary bonus received is a function, m, which maps the discrepancy between actual output that the worker produces to some output level, Y determined by the firm.[4] Potential lost income becomes:

$$w(T_h + T_c) + m(Y - F(N)) = w(T - N), \tag{3.25}$$

where $F(N)$ represents the output generated by a worker supplying N hours of labor, with $F' > 0$ and $F'' < 0$.

After a negative shock to mental health that reduces MH_H to MH_L (with $MH_L < MH_H$), the optimal amount of labor supply falls from level, N^H, to a lower level, N^L. However, if the individual chooses to supply this lower level of labor, N^L, then potential lost income rises proportionally to the difference between Y and $F(N^H)$ and Y and $F(N^L)$. To avoid the rise in potential lost earnings, the worker may prefer to increase market inputs (x_h) in health production to mitigate the impact of the health shock rather than choosing time-intensive inputs (T_h). If these are accurate assumptions, potential to earn bonus pay will mitigate some of the change in labor supply induced by poor mental health.

Next, consider how a fringe benefit such as health insurance might impact the behavior of individuals in poor mental health. Healthcare service utilization takes both market and time inputs, so that T_h and x_h are complements in this setting. To simplify the discussion, assume that T_h and x_h are perfect complements so that the utilization of one additional unit of the healthcare market good (the care received) is paired with an additional unit of time spent on producing health. The marginal cost of utilizing one additional unit of healthcare is $(1-c)p_h b_h + wt_h$ when covered by insurance at rate $0 < c \leq 1$, and $p_h b_h + wt_h$ when the individual is uninsured ($c = 0$). It directly follows that health insurance coverage incentivizes individuals to utilize healthcare services at higher rates; yet the overall impact that this access to care has on productive outcomes is ambiguous.

[4] This setup is without loss of generality and imposes the particular framework due to streamline understanding of the model narrative.

Health insurance premiums as well as cost-sharing between individuals and insurers both work to incentivize individuals to utilize healthcare services when ill. In the particular case of poor mental health, health insurance acts as a mechanism toward professional intervention and possibly treatment. Since this type of service utilization requires both time and market inputs and health insurance supplements the cost of the latter, but does not directly supplement the former, I hypothesize that poor mental health will be associated with higher degrees of absenteeism among the employed with health insurance coverage relative to workers that are uninsured.

This hypothesis is applicable to the current theoretical framework in which labor supply decisions are made on a daily basis and are short-run decisions and in which one observes only healthcare utilization, not specific treatment. From a longer term perspective, or from a treatment perspective in which insured care induces one toward the treatment of mental health symptoms, health improves over time and it may be that one's efficiency of health production in general increases. That is, compounding treatment effects may close the gap between optimal labor supply and contracted hours even if requiring additional time allotments to utilize treatment.

The above discussion incorporates two simplifying assumptions, both of which are related to the phenomenon of moral hazard. First, I assume that an individual utilizes healthcare services when experiencing poor (mental) health to address the adverse effects of poor mental health on their overall health stock. In reality, the individual may choose to utilize other services due to the availability of health insurance. Additionally, I assume that the individual's potential income (or full income) remains the same regardless of their insurance enrollment status, and thus premium payments are not considered in the current framework. In practice, however, premium payouts may induce moral hazard, further inflating the utilization of medical services covered by insurance (Riphahn et al., 2003).

The following chapter describes the main data source utilized for the empirical analyses and defines the criteria met by individuals in each sample. I also define the variables used in the empirical analyses and categorize them into named matrices based on similarities in the information they represent.

CHAPTER 4

DATA

The main source of data in this dissertation is provided by public use files of the Medical Expenditure Panel Survey (MEPS) which provides information on demographic and employment characteristics, healthcare utilization, and measures of health and well-being at the individual level. The MEPS consists of several data files. I utilize the Full-Year Consolidated Datafile (FYCD), the Medical Conditions File (MCF), and the Jobs File (JF). I also collect data from the Bureau of Labor Statistics (BLS) on historical unemployment rates per month and across regions. In what follows I describe the data generating process (DGP).

The MEPS follows a one-stage cluster random sampling design. The DGP starts with the random selection of households from the participating household of the most recent National Health Interview Survey (NHIS). The MEPS then collects information on each individual within the selected household units. Each panel of the MEPS consists of five rounds spanning across two consecutive years.

Upon consolidating the data sources, some persons are observed once for a single calendar year while other individuals are observed twice – once for each calendar year they participated in the survey. It should be noted that the reference period for the third interview round of the survey spans across the two consecutive calendar years; fortunately, the data sources used in the analysis ascribe round three data to the appropriate calendar year so that this does not cause a problem. Responses to most of the MEPS interview questions are reported separately for each of the three rounds per year (with the first year's round three variables pertaining to the start of round three up until the end of the calendar year and the next year's round three variables reference the time spent in round three after the start of the new year) so that FYCD annual data files contain three variables per interview prompt.

I use several procedures to annualize variables appearing in groups of three. The MEPS collects information on each individual in a surveyed household. While late entry of *house-holds* into the survey is not permitted, *individual* participants may enter the survey late if they enter a participating household during the survey period. For example, late entry may be observed for a newly married individual that moves into the residence of his or her spouse, who is a current MEPS participant. A binary variable indicating whether an individual moved into a participating household during the survey period is utilized as a control variable in the analysis.

Aside from the possibility of the late entry of individual participants, the length of each reference period round may vary across individuals due to extenuating circumstances that interfere with an individual's availability on the originally scheduled interview date for a particular round. In such a case, the interview may be rescheduled to occur at an earlier or later date; in the former instance, an individual may exhibit fewer reference period days than the average participant and in the latter, may exhibit a greater than average number of days included in the particular survey round. Information on individual-specific reference period start and end dates for each of the three rounds occurring in a given year is utilized to generate control variables that account for heterogeneous exposure to survey prompts.

Only observations reporting employment at one or more of the three interview dates are retained. Some observations indicate employment as of a particular interview date but list a job start date equal to that interview date. This situation may occur when an individual has been hired but has not yet started working at the time of the interview, meaning they did not work during the corresponding reference period.

These observations are omitted from the sample of actively employed individuals if unemployment is reported in both of the other two interview rounds. This is because any survey prompts related to a period when the individual is not actively working cannot capture absenteeism in the context of this dissertation. The FYCD documentation files support this approach, explaining that the portion of the survey collecting work-loss information is independent of the section collecting detailed job-related measures. As a result, individuals who would logically have unobserved values for the *sickdays* variables in this paper may be recorded as reporting zero absences.[1,2]

Variables indicating job changes that occur between interview rounds are utilized to match the proper job characteristics to absenteeism reports relevant to the correct time period. For observations indicating a job change at some point between interview rounds, only the information for the first reported job is considered to mitigate the risk of matching reported work absences to characteristics of the wrong job. Individuals that have ever retired, have a disability, and military personnel are not included. For this research, I keep individuals observed only once in the sample and use pooled data and conduct cluster-robust inference at the individual level.[3]

After these alterations discussed above have been made, the final sample represents adults who were employed for a positive number of days during the respective calendar year and consists of 31,929 observations. The maximum number of absence days reported over an annual period is 160 days.

[1]Codebook Source: MEPS-HC Panel Design and Collection Process, Agency for Healthcare Research and Quality, Rockville, Md. https://meps.ahrq.gov/mepsweb/survey_comp/hc_data_collection.jsp.

[2]Future research could leverage such indicators to explore differences between individuals reporting long periods of layoffs or leave—due to disability, workers' compensation, or maternity/paternity leave—and the actively working population. These sub-populations are excluded in the current paper.

[3]See the following for more detailed descriptions of models estimating discrete outcomes for panel data: (Cameron and Trivedi, 2015; Cameron and Trivedi, 1986; Mundlak, 1978; Greene, 2002).

4.1 Variables

Table A.1 in Appendix A reports variable definitions and descriptive statistics for the study sample. In this section, I discuss the variables used in my analyses and in the following section I describe the descriptive statistics.

Dependent Variable (A_i): The dependent variable of interest is a count variable ($sickdays$) representing absences from work due to an injury or physical or mental illness or ailment. The MEPS includes a separate variable counting the days absent from work due to someone else's illness. I include only the variable prompting for a count of own-health related absence.

Mental Health (MH_i): The main explanatory variable of interest in this study is a binary variable indicating diagnosed mental illness ($keyMHdis$). This variable is based on responses reported in the Medical Conditions File (MCF) which provide condition-specific codes for various forms of mental illness. This variable is equal to one for individual's reporting diagnosis(es) of mood, anxiety, personality, or psychotic disorders. These categories of mental illness are chosen due to population prevalence, standard treatment protocols, and comorbidity hazards among them. I henceforth refer to these categories of mental illness as "key disorders." If an individual indicates a diagnosis of one or more of these key disorders, $keyMHdis$ equals one and is zero otherwise.

It should be noted that other classes of mental illness are also reported in the MCF, such as sexual disorders, conduct disorders, and developmental disorders. I choose not to include these diagnostic categories in variable $keyMHdis$ due to the differing nature of the diagnostic criteria associated with these groups of disorders according to the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM–5; American Psychiatric Association, 2013). Sexual and conduct disorders are often times underreported and symptoms associated with these disorders are often external in nature. Developmental disorders are associated with highly heterogenous symptoms that may be internalized or externalized and that have a broad range of the degree of limitations associated with symptoms. The MEPS separately categorizes one form of developmental disorder, Attention Deficit/Hyperactivity Disorder (ADHD), "because of [its] relatively high prevalence, and because generally accepted standards for appropriate clinical care have been developed." As ADHD is technically a developmental disorder, the nature of its symptoms are less clear than the symptoms associated with the diagnostic categories included in the formation of $keyMHdis$; however, ADHD is highly comorbid with the diagnoses considered as "key" disorders in this paper so that I include a binary variable equal to one for individuals with ADHD as a control in the analysis.

It is important to further stress that the $keyMHdis$ variable represents only *diagnosed* mental illness across the sample so that mental disorder prevalence rates illustrated by the sample may not be representative of actual population prevalence rates in the US. That is, individuals may exhibit symptoms of mental illness even without a formal diagnosis. There is also likely variation in the degree of symptom severity associated with a specific disorder across individuals and failing to account for this possibility may bias the estimated effect of diagnosed mental illness.

The consideration of potential homogeneous symptoms between individuals with and without a given diagnosis, as well as heterogeneous symptoms across individuals within a particular diagnostic category are highlighted in the DSM-5, which separates itself from earlier DSM editions by its focus on a spectral approach to mental illness. DSM-5 states: "Earlier editions of DSM focused on excluding false-positive results from diagnoses; thus its categories were overly narrow, as is apparent from the widespread need to use NOS [not otherwise specified] diagnoses. Indeed the once plausible goal of identifying homogeneous populations for treatment and research resulted in narrow diagnostic categories and did not capture clinical reality, symptom heterogeneity within disorders, and significant sharing of symptoms across multiple disorders".

In acknowledgment of within-disorder heterogeneity and that mental health may be, to some degree, independent of a particular diagnosis, I utilize an index variable from the FYCD that measures one's general level of emotional distress over the last 30 days on a Kessler-6 scale with scores ranging from 0-24 and higher values indicating more psychological distress (Kessler et al., 2003; Ashwood et al., 2017). This information is collected as part of the MEPS Self-Administered Questionnaire (SAQ) which is collected at rounds 2 and 4 of the MEPS.

General (Physical) Health (PH_i): Physical and mental health are inherently intertwined. The MEPS has numerous physical health markers. The FYCD files of the MEPS provide self-rated general health scores for each of the three interview rounds of the respective calendar year. Responses in each round range from 1-5 with higher values indicating *worse* perceived general health. These three variables are summed to create an annual index of general health originally with range 3-15; the index is rescaled by subtracting two from each observation, yielding a poor-health index (*physhlth*) ranging from 1-13 that acts as the main measure of general health.

As the index just described may be prone to attenuation bias due to self-perceptions of health, various other measures of health are utilized in the analysis. The second analytical variable in this category, named *prtycnds*, represents the number of diagnosed priority conditions as defined by the MEPS. The priority conditions specified by the FYCD are cancer, heart conditions, asthma, stroke, chronic bronchitis, emphysema, high cholesterol, high blood pressure, diabetes, arthritis, joint pain and ADHD. ADHD is not counted in *prtycnds* because it is psychological in nature; instead a dummy variable indicating diagnosed ADHD is utilized as a separate control. The MEPS singles out these priority condition categories "because of their relatively high prevalence, and because generally accepted standards for appropriate clinical care have been developed."

A binary variable equal to one if an individual has suffered an injury or illness (*injury*) in the past year that required immediate medical help (e.g. an unanticipated hospitalization) as well as a variable indicating that they have received a routine checkup within the past year (*routine*) are especially important control variables given the impossibility of these events co-occuring with absenteeism in the current year, eliminating the risk of simultaneity. There is no variable available to identify mental-health related visits over the past year due to confidentiality, but it should be noted that severe mental illness requiring hospitalization in the past year may be represented in variable *injury*.

Behavioral variables equal to one for individuals that exercise at least three days a week (*exercise*) and a binary variable identifying smokers (*smoke*) are other controls. For the subsample of women, a variable indicating pregnancy at the beginning of the year is also used as a control variable in analysis.

Job Traits (J_i): Variables indicating the fringe benefits offered to an individual by their employer such as paid sick leave (*sickpay*) and bonuses (*bonus*) are utilized as analytical variables in analysis as these factors likely influence an individual's decisions on short-run labor supply. It should be noted that binary variables indicating whether one receives paid vacation time and whether one receives paid leave to visit the doctor are also analyzed, but are determined to induce issues of collinearity with the *sickpay* variable. In addition, the inclusion of either of these binary variables does not significantly improve model fit so that they are excluded from analysis and I instead only focus on the analytical variable *sickpay*.

The decision making process on whether to go to work when facing illness is very likely to depend on the length of time for which an individual has been at their job; for example, an employee who has only worked at a firm for a year might be more averse to absence when ill than an equivalent worker who has worked at a firm for 15 years because less job-specific experience might make the newer employee relatively more expendable; there may also be varying degrees of rapport between the the firm and the employee based on the length of the worker's tenure, another channel through which tenure may impact labor supply decisions. A variable measuring an individuals job tenure in years (*tenure*) is created using information on job start and end dates provided by the MEPS.

Additional job-related controls include indicators for labor union membership (*union*) and dummy variables identifying seasonal workers (*ssnl*), temporary contracts (*temp*), and part-time employees (*parttime*), with individuals that typically work less than 35 hours per week considered part-time. Other job controls include indicator variables for public sector positions (*pubsect*), industry and occupational categories (see Appendix A: Table A.1), and firm size (represented by variables $1to19$, $20to99$, $100to499$, and $500plus$ based on sample quartiles).

Health Insurance (I_i): The FYCD provides a copious amount of information on individuals' health insurance characteristics. An imputed variable that indicates the category of insurance the individual reported for the majority of the calendar year (categories are private, public, and uninsured) as well as variables reporting whether one held insurance through their employer from the raw FYCD files are used to derive a group of binary variables representing one's source of health insurance. A binary variable equal to one for individuals who have insurance through their job (*jobins*) is included in the estimation of Baseline Model 1 ($BM1$) described later.

Variables indicating uninsured individuals (*unins*), individuals insured privately through a source other than their employer (*otherins*), and individuals with a public source of health insurance (*pubins*) are created and the group of uninsured individuals acts as the reference category. A categorical variable from the FYCD is utilized to create a binary variable (*inscostly*) equal to one for individuals who either somewhat agree or strongly agree to the survey prompt, "health insurance is not worth its cost". Variable *inscostly* acts as a proxy of health insurance generosity.

($\tilde{I}_i$): A model specification henceforth referred to as Baseline Model 2 ($BM2$), to be defined in the next chapter, replaces $jobins$ with a group of three binary variables, *plnchoic*, *nochoic*, and NR*.choic*. Two of these three variables will be analyzed in depth in later sections; these are *plnchoic* and *nochoic*. *plnchoic* identifies individuals with a choice between multiple insurance plans offered by an employer and *nochoic* identifies the individuals that receive insurance through their employer but are only offered one plan option. Variable NR*.choic* indicates observations that report having insurance through an employer but have missing values for prompts on whether or not the employer offers an array of plans to choose from.

Demographics (X_i): The highest level of educational attainment is controlled for using a group of binary variables (*belowhs*, *hsdeg*, *somecoll*, *bachdeg*, *bachplus*). Other variables control for race (*black*, *asian*), Hispanic ethnicity (*hispanic*), native-status (*bornus*), age (*age*), family size ($famsz$), the number of young children in the household (*yngchldrn*), marital status (*married*), and socioeconomic status (*poor*, *lowinc*, *midinc*, *highinc*).

Other Control Variables (C_i): Regional indicator variables (NE, MW, W, S) are generated using FYCD variables indicating one's region of residence for the majority of the calendar year. Monthly data from the U.S. Bureau of Labor Statistics on regional unemployment rates is utilized to estimate the average unemployment rate faced by an individual (*unemp.rt*). Reference period start and end dates are used to generate these estimates for individuals that move to a different region during the year of interest; the unemployment rates are averaged across the months for which an individual reported a certain region of residence, then the average is taken across each of the regions that one resided in during the year. This measure of the average unemployment rate faced by an individual in a particular year provides the benefit of controlling for heterogeneity induced by macroeconomic features of the regional economy.

Following the theoretical framework, it is assumed that the short run decision-making process on work absence is only faced on the days that an individual is employed, as the model inherently assumes that an optimal employment contract has already been reached. Thus, differences in exposure to this decision-making process across individuals should be accounted for. I use information on interview start and end dates as well as job start and end dates and employment status to generate an estimate of the number of days employed during the respective calendar year ($empUB$). The log of this estimate is used as a control variable in analyses and roughly accounts for differences in exposure to work days.

Further considering exposure, as reference period rounds can be longer or shorter for some individuals, and because I keep individuals that may have been unemployed at some point during the calendar year in the sample, I include indicator variables that specify whether the individual was unemployed for one or two survey rounds ($unemp.1, unemp.2$). If an individual became unemployed during the reference period, a binary variable, *partialemp* equals one.

Persons observed for a single year are split into two categories to indicate which of the two years of participation in the panel that they are observed for. Two dummy variables indicate these two sub-groups of individuals, one indicates that the individual is observed in his or her first year of participation ($yearone$). The second identifies individuals observed in the second ($yeartwo$) year of participation. Sample statistics for a third variable, $bothyears$, for individuals observed in the sample for both years in which their household participated, are included in Appendix A: Table A.1, but it should be noted that $bothyears$ acts as a reference variable in the empirical analyses.

I control for attrition induced by the data generating process. I control for other outside factors that might impact absence decisions such as a move from one region of the US to another at some point in the calendar year (*moved_US*) and a move from one participating household to another (*moved_RU*). Finally, binary variables indicating each observation's respective calendar year (2010 - 2014) are used to account for year fixed effects.

4.2 Sample Characteristics

Throughout this dissertation, I limit the number of variables included in the embedded tables which I refer to as "analytical variables". Table A.1 in Appendix A lists *all* variable names, definitions, and summary statistics for the entire sample.

Table 4.1 reports sample means for a subset of explanatory variables by sex and reports test statistics comparing the sample means of men and women. The prevalence rate of diagnosed mental illness is twice as high for the sample of women than for men. The difference in the percentage of observations with $keyMHdis = 1$ between men and women in the sample is about 7 percent. This is similar to the population discrepancy in the percentage of diagnosed mental illness by sex reported for the US in 2017 (7.2 percent).[4]

Women in the sample report significantly more emotional distress and significantly poorer health compared to men in the sample. The sample of women additionally reports a higher average number of priority conditions ("priority" as defined by the MEPS and represented by variable *prtycnds*). Regarding insurance characteristics, Table 4.1 reports that fewer women in the sample believe that health insurance is not worth its cost. As the average woman in the US exhibits higher rates of healthcare utilization than the average man in the US, suggesting that women and men vary significantly in their preferences for being insured.

Table 4.2 reports the sample means for analytical variables by sex and by the levels of $keyMHdis$.

[4]Mental Health by the Numbers. (n.d.). Retrieved from https://www.nami.org/learn-more/mental-health-by-the-numbers From the National Alliance on Mental Illness.

Table 4.1: Means of Analytical Variables by Sex

	Variable Name	Men	Women	T Stat
Dependent Variable (A)	*sickdays*	2.50	3.85	-12.11***
Mental Health (MH)	*distress*	2.30	2.90	-15.02***
	physhlth	4.44	4.87	-16.01***
	prtycnds	1.26	1.35	-5.57***
Job Characteristics (J)	*sickpay*	0.61	0.64	-5.04***
	bonus	0.22	0.18	9.92***
Health Insurance Characteristics (I, $\tilde{I}$)	*inscostly*	0.30	0.26	7.69***
	jobins	0.66	0.61	9.38***
	plnchoic	0.34	0.34	-0.30
	nochoic	0.26	0.22	8.76***
	otherins	0.11	0.15	-12.01***
	pubins	0.05	0.09	-14.71***
	unins	0.19	0.15	8.57***
Observations		15,713	16,216	

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4.2: Means of Analytical Variables by Sex and Presence of Mental Illness

	Men			Women		
Variable Name	$keyMHdis$ =1	$keyMHdis$ =0	T Stat	$keyMHdis$ =1	$keyMHdis$ =0	T Stat
sickdays	4.64	2.35	6.35***	6.14	2.48	9.02***
distress	5.33	2.09	20.53***	5.57	2.46	28.44***
physhlth	5.59	4.36	15.34***	5.78	4.72	18.50***
prtycnds	2.04	1.21	16.47***	1.96	1.25	19.24***
sickpay	0.64	0.61	2.45**	0.66	0.63	2.10**
bonus	0.26	0.22	2.40**	0.19	0.18	1.90*
inscostly	0.27	0.30	-1.61	0.22	0.26	-4.96***
jobins	0.71	0.65	3.87***	0.63	0.60	2.70***
plnchoic	0.38	0.33	3.29***	0.36	0.33	2.87***
nochoic	0.27	0.26	0.47	0.24	0.22	1.63
otherins	0.10	0.11	-0.64	0.16	0.15	1.43
pubins	0.05	0.05	0.40	0.10	0.09	1.42
unins	0.14	0.19	-4.75***	0.11	0.16	-7.05***
Observations	15,713			16,216		

Note: *p<0.1; **p<0.05; ***p<0.01

Men with a diagnosed mental disorder report significantly higher absences, on average; the discrepancy in mean absences for women with and without a disorder diagnosis is much larger. This may suggest that there are additional (or different) factors that influence the absence decisions of women with diagnosed mental illness compared to factors that drive the mentally ill absence decision for men; for example, symptoms may exhibit themselves in different manners between sexes or women may put a higher degree of importance on nurturing their mental health so that they are more prone to illness-related absence. As anticipated, average levels of general psychological distress are significantly higher for individuals with a diagnosed key mental disorder, regardless of sex.

Discrepancies in mean values reported for $physhlth$ by $keyMHdis$ are interestingly very similar for both men and women. This is also the case for the discrepancy in mean values for $prtycnds$ by levels of $keyMHdis$. The magnitude of the within-sex discrepancy in means of these variables are highly statistically significant for both men and women, further suggesting significant correlations between mental illness and other measures of (physical) health.

Means for job characteristics suggest that for individuals with a diagnosed mental illness, one of the following phenomenons may occur: individuals with mental illness select into jobs that provide fringe benefits to supplement the extra time and income constraints necessary to achieve a certain level of health; or, for individuals diagnosed while on the job, fringe benefits that supplement time lost from work (*sickpay*) increase the likelihood of seeking care when experiencing symptoms.

Regarding health insurance characteristics, Table 4.2 reports that 22 percent of women in the sample with a key mental disorder diagnosis believe that health insurance is not worth the cost, while 26 percent of women without any such diagnosis share this belief. This discrepancy in means is significant at the one percent level. Finally of note, women and men in the sample who have a key diagnosis ($keyMHdis = 1$) have fewer uninsured individuals in this category.

The following chapter defines econometric models and methods utilized in empirical estimation.

CHAPTER 5

EMPIRICAL METHODOLOGY

Unless explicitly mentioned *all* of the variables grouped by category labels in Appendix A are included in empirical modeling *with the exception* of the final two variables in the category titled "Exclusion Restrictions (ER)". These last two variables in category ER are not utilized until the very end of this paper and should be ignored for now.

5.1 Econometric Modeling

In what follows, I define two baseline conditional mean functions of absenteeism, define extended model specifications, and form hypotheses about how various measures of health interact with one another to influence absenteeism, and finally, I discuss estimation strategies and how the models are applied to test hypotheses.

5.1.1 Baseline Model Specifications

The dependent variable used in each of the analyses is a count of the number of days absent from work due to *one's own* physical or mental illness over the span of one calendar year, A_i for observation i. Due to the count nature of dependent variable A_i, I specify an exponential conditional mean function.

I specify two separate baseline specifications of the conditional mean of absence days. The first baseline conditional mean specification is defined as follows:

$$\begin{aligned} E[A_i|MH_i, PH_i, J_i, I_i, X_i, C_i] = \exp(&\beta^0_{BM1} + MH_i\beta^{MH}_{BM1} + PH_i\beta^{PH}_{BM1} \\ &+ J_i\beta^J_{BM1} + I_i\beta^I + X_i\beta^X_{BM1} + C_i\beta^C_{BM1}), \end{aligned} \tag{5.1}$$

for every observation, $i = 1,...,n$. Equation (5.1) will henceforth be referred to as $BM1$ (baseline model one). Matrices MH, PH, J, I, X, and C hold observed values of the corresponding explanatory variables for all i = 1, ..., n observations. The previous chapter discussed the explanatory variables included in each of these matrices and Appendix A also groups explanatory variables in terms of these matrices and provides definitions. Each matrix of explanatory variables is of dimension $n \times k_l$ where n is the total number of observations in the sample and k_l is an integer equal to the number of explanatory variables in corresponding matrix l for $l = \{MH, PH, J, I, X, C\}$. In a similar fashion, each β^l for $l = \{MH, PH, J, I, X, C\}$ is a $k_l \times 1$ vector of parameters.[1]

A second specification of the conditional mean of A_i that I reference as $BM2$ (baseline model two) throughout the rest of the paper is defined in a similar manner to (5.1). $BM2$ is identical to specification $BM1$ *except* that matrix $\tilde{I}$ replaces matrix I for the $BM2$ specification. As noted in Chapter 4, $\tilde{I}$ in $BM2$ holds variables *inscostly* and the set of dummy variables *nochoic*, *plnchoic*, *NR.choic*, *pubins*, and *otherins*, with *unins* acting as the reference category variable. Variables *plnchoic*, *nochoic*, and *NR.choic* further break down the variable *jobins* which is included in the $BM1$ specification in place of those three variables. The purpose of this break down is to analyze and compare how the implemen-

[1] For example, β^{MH}_{BM1} is a 3×1 vector holding parameter values for the three explanatory variables represented in matrix MH, which are *keyMHdis*, *distress*, and *adhd*.

tation of different types of benefit package designs influence worker absence behavior. The $BM2$ specification takes the form

$$\begin{aligned} E[A_i|MH_i,PH_i,J_i,\tilde{I}_i,X_i,C_i] = \exp(&\beta^0_{BM2} + MH_i\beta^{MH}_{BM2} + PH_i\beta^{PH}_{BM2} \\ &+ J_i\beta^J_{BM2} + \tilde{I}_i\beta^{\tilde{I}} + X_i\beta^X_{BM2} + C_i\beta^C_{BM2}), \end{aligned} \tag{5.2}$$

where each β^j for $j = \{MH, PH, J, \tilde{I}, X, C\}$ is a $k_j \times 1$ vector of parameters, where k_j is an integer equal to the number of explanatory variables held in corresponding matrix j.

Thus far, I have used notation that assigns parameter subscripts that indicate the specific baseline specification in order to highlight that coefficient estimates may differ between $BM1$ and $BM2$. For notational efficiency I henceforth drop these subscripts and work within a more general framework.

Coefficients are not directly interpretable in nonlinear models, especially if the goal of research is to form policy implications (Braumoeller, 2004; Buis, 2010; Chunrong & Norton, 2003; Long & Freese, 2006; Williams, 2009). Coefficients represent the marginal change in the *log* of the conditional mean outcome and thus are not very valuable in addressing the main research questions presented in this paper. Instead, I prefer to test hypotheses on the difference in the conditional mean expectation of absences resulting from a discrete change in an explanatory variable; this will allow for more intuitive inference (Wooldridge, 2010, *Econometric Analysis of Cross Section and Panel Data*, 737).

Define ME_i as the effect of a diagnosed mental illness on the expected value of annual illness-related work absences, which is equal to the change in expected absences when the binary variable $keyMHdis$ changes between zero and one. Following from the theoretical framework laid out in Chapter 3, consider an individual endowed with a low level of mental health who faces the same time constraints, budget constraints, and preferences over health and conglomerate consumption as a second individual endowed with a higher level of mental health.

The former individual will, in theory, exhibit the same optimal level of health production, H^*, as the latter individual; however, the former individual will require more time and market inputs to reach or maintain level H^* compared to the latter individual due to the less-efficient health production process imposed by the low health endowment. It follows that the individual with the lower endowment of mental health will exhibit a lower optimal level of labor supply relative to the individual with high mental health endowment. This result can easily be used to form a testable empirical hypothesis defined as follows:

$$\begin{aligned} E[ME_i] &= E[A_i|keyMHdis_i = 1] - E[A_i|keyMHdis_i = 0] > 0 \\ &= \exp(\beta^{MH_1} + \sum_{l \neq MH_1} x_{i,l}\beta^l) - \exp(\sum_{l \neq MH_1} x_{i,l}\beta^l) > 0, \end{aligned} \tag{5.3}$$

where β^{MH_1} is the parameter corresponding to the first variable in vector MH, which is assumed to be binary variable $keyMHdis$. Matrix x_l holds observed values for every other explanatory variable $l \neq MH_1$,[2] and β^l is the coefficient corresponding to explanatory variable l. In words, I hypothesize that diagnosed mental illness is associated with higher expected annual absences.

In what follows, I consider how the magnitude of (5.3) may change across levels of various job, health insurance, and health variables. While this would require the use of the appropriate interaction terms if utilizing a linear model, in nonlinear models of exponential form, the effect of one variable ($keyMHdis$, in this dissertation) may vary with the value of a second variable without the additional consideration of an interaction term between the two variables. Aspects of employment contracts and health insurance coverage are theoretically

[2] Note that $x_{i,l}$ is a generalization for notational efficiency. In practice, $E[ME_i]$ will be estimated for both $BM1$ and $BM2$ specifications so that the explanatory variables in matrix x_l for $l \neq MH_1$ will be different based on the model specification used.

different in nature than variables characterizing one's mental health endowment because of the different channels through which each of these types of factors influence absence behavior (e.g., sick leave and health insurance generosity influence choices of time and market inputs in health production, while mental illness is tied to an efficiency factor. See Chapter 3 for more detailed discussion).

Conditional marginal effects, where job and insurance characteristics act as the conditioning variables, are more useful in forming testable hypotheses compared to the use of explicit interaction terms (Karaca-Mandic et al., 2012). I later define extended model specifications that include interaction terms between $keyMHdis$ and other variables measuring health and form hypotheses on the sign of these terms.

In the following section, I present hypotheses on the conditional marginal effect of mental illness at different levels of employment contract and health insurance variables.

5.1.2 Conditional Effects Models of Fringe Benefits

Mental illness is anticipated to impose higher discrepancies between contracted hours and short-term labor supply decisions (i.e., higher absenteeism). It is of interest to consider how this discrepancy may be exacerbated or mitigated in different "states" of employment contracts and access to care. Consider a binary moderating variable with each level of the variable representing a different state. In particular, for a binary fringe benefit variable, a value of zero represents state "not offered" and a value of one represents state "offered". For health insurance source variables, there are states "enrolled" and "uninsured" for levels one and zero, respectively. I form testable hypotheses on the relative magnitude of the change in expected absences occurring when $keyMHdis$ changes from zero to one at each level (state),

s, of a binary fringe benefit or health insurance variable. For arbitrary moderating variable, mod, the conditional marginal effect of mental illness on expected absence days is

$$\begin{aligned} E[ME_i|mod_i = s] = &\exp(\beta^{MH_1} + s\beta^{mod} + \sum_{l \neq mod, MH_1} x_{i,l}\beta^l) \\ &- \exp(s\beta^{mod} + \sum_{l \neq mod, MH_1} x_{i,l}\beta^l), \end{aligned} \tag{5.4}$$

for each value of $s = \{0, 1\}$.The moderating effect of mod on the size of the effect of mental illness on absenteeism is defined as

$$\begin{aligned} &E[ME_i|mod_i = 1] - E[ME_i|mod_i = 0] = \\ &\quad \{E[A_i|keyMHdis_i = 1, mod_i = 1] - E[A_i|keyMHdis_i = 0, mod_i = 1]\} \\ &\quad - \{E[A_i|keyMHdis_i = 1, mod_i = 0] - E[A_i|keyMHdis_i = 0, mod_i = 0]\}. \end{aligned} \tag{5.5}$$

I first consider moderating effects of fringe benefits that supplement time off from work. The MEPS provides information on whether or not individuals receive paid sick leave and paid vacation time at their job. Unfortunately, including both of these factors in a model induces issues of collinearity because only a small portion of the sample has one of these benefits but not the other. As paid sick leave seems more relevant to the context of this paper, I choose to omit information on whether workers receive paid vacation time and instead focus on hypotheses regarding paid sick leave only. Paid sick leave supplements some of the lost wages associated with illness-related work absence and thus acts as a mechanism to reduce the opportunity cost of missing work when ill. This leads to the hypothesis that $E[ME_i|sickpay = 1] > E[ME_i|sickpay = 0]$, i.e., I anticipate that the moderating impact of paid sick leave on the level of additional expected absence days associated with mental illness will be positive.

In contrast to a fringe benefit like paid sick leave, the ability to receive bonus pay at a job increases the opportunity cost of missing work (by increasing potential earnings lost).Thus, even though individuals with mental illness are anticipated to value time spent on nurturing health more than counterparts without mental illness, when given the opportunity to receive bonus pay, relatively higher levels of absenteeism come at a greater cost because, in addition to the opportunity cost of lost wags, absenteeism also hinders one's ability to receive bonus pay. Thus, I hypothesize that mental illness is associated with fewer additional expected absences when bonus pay can be earned relative to when bonus pay is not offered at a job ($E[ME_i|bonus = 1] < E[ME_i|bonus = 0]$). This translates to the hypothesis that the variable *bonus* will mitigate the degree of additional expected absences associated with the variable $keyMHdis$ switching between zero and one.

I now form hypotheses on the group of dummy variables representing one's source of health insurance. Health insurance covers some of the cost of utilizing healthcare services so that it is likely that insured individuals with mental illness will be more likely to receive treatment or intervention services than uninsured mentally ill individuals.The utilization of such services imposes a time cost. I hypothesize that mental illness is associated with a greater degree of additional expected absences over the baseline category of no mental illness when insured, regardless of the source, relative to the degree of this association when uninsured. That is, I anticipate the moderating effect of health insurance to be positive. Finally, I theorize that a more generous insurance plan will also have an amplifying impact on the positive change in expected absences associated with a change in $keyMHdis$ from zero to one.

The variable *inscostly* represents either somewhat or totally agreeing with the survey prompt "health insurance is not worth its cost" and acts as a proxy for insurance plan generosity. The argument for this is that individuals who agree with this prompt exhibit a marginal cost of receiving care under a plan that outweighs the marginal benefit of care.

Such individuals may face financial and/or accessibility barriers under their health insurance plan (for the group of insured individuals), or may believe that the plans available to them would not benefit them anymore than if they remained uninsured so that the *inscostly* variable proxies exposure to "skimpy" health insurance plans. Thus, I hypothesize that $E[ME_i|inscostly = 1] < E[ME_i|inscostly = 0]$.

I next discuss an extension of these baseline models to consider the how mental illness and other measures of health interact to influence absenteeism.

5.1.3 Mental Health, General Health, and Absenteeism

As a final analysis, I consider three extensions of the $BM1$ specification. It is clear that general measures of health are highly related to health condition diagnoses. Define $I1$ as the first specification of this sort. $I1$ is defined as follows:

$$\begin{aligned} E[A_i|MH_i, PH_i, J_i, I_i, X_i, C_i] = \exp(&\beta^0 + MH_i\beta^{MH} + PH_i\beta^{PH} \\ &+ J_i\beta^J + I_i\beta^I + X_i\beta^X + C_i\beta^C + MH_{1i} \times MH_{2i}\beta^{MH_1 \times MH_2}), \quad (5.6) \end{aligned}$$

where MH_1 is the first column of matrix MH and it holds values of $keyMHdis$ and MH_2 holds values of $distress$, the second variable present in matrix MH.

Next, define $I2$, a model that includes an interaction term between $keyMHdis$ and the variable representing general health status, *physhlth*. The first variable in matrix PH, PH_1 is the *physhlth* variable. Then for the $I2$ specification yields:

$$E[A_i|MH_i,PH_i,J_i,I_i,X_i,C_i] = \exp(\beta^0 + MH_i\beta^{MH} + PH_i\beta^{PH} \\ + J_i\beta^J + I_i\beta^I + X_i\beta^X + C_i\beta^C + MH_{1i} \times PH_{1i}\beta^{MH_1\times PH_1}). \quad (5.7)$$

Finally, a third specification utilizes an interaction between the mental illness indicator variable and the variable representing the number of chronic physical condition diagnoses, *prtycnds*. Define the $I3$ specification as follows:

$$E[A_i|MH_i,PH_i,J_i,I_i,X_i,C_i] = \exp(\beta^0 + MH_i\beta^{MH} + PH_i\beta^{PH} \\ + J_i\beta^J + I_i\beta^I + X_i\beta^X + C_i\beta^C + MH_{1i} \times PH_{2i}\beta^{MH_1\times PH_2}), \quad (5.8)$$

where PH_2 from matrix PH holds values of the *prtycnds* variable.

I hypothesize that there may be an interaction between mental illness and *distress*, *physhlth*, and *prtycnds* so that the total impact of diagnosed mental illness on absenteeism varies with changes in these variables, and vice versa. Each of these three measures of health are anticipated to have a positive effect on absenteeism *in isolation* (recall that higher values of *physhlth* indicate reports of *poorer* levels of general health). I anticipate that having a diagnosed mental illness will exacerbate the effect of these variables on the number of absence days reported by an individual and that higher levels of *distress*, poorer physical health (indicated by higher levels of *physhlth*), and a greater number of priority conditions (higher levels of *prtycnds*) will increase the total effect of diagnosed mental illness on absenteeism. In sum, I hypothesize

$$\beta^{MH_1\times MH_2} > 0,$$

$$\beta^{MH_1\times PH_1} > 0, \text{ and}$$

$$\beta^{MH_1\times PH_2} > 0.$$

As these three additional measures of health are measured as either a discrete index or continuous variable, I plot the interaction effects, as it is likely that the relationship is nonlinear and may change in magnitude at extreme values of index or continuous variables. After observing plots of interactions, I choose a few interesting values of each of these three variables and compute conditional average marginal effects (CAME) for variable $keyMHdis$, conditioning on these interesting values for further inference.

5.2 Estimation Strategy

I finally discuss the applied econometric strategy used to generate estimates and test hypotheses.

The full sample resulting from the data-generating process discussed in Chapter 4 is unbalanced; some individuals are observed once while others are observed twice. This poses the issue of potential attrition bias. I include multiple control variables that identify possible sources of attrition for individuals observed once, including variables indicating which year of participation (first year or second) in the MEPS these individual's are observed for and whether these individuals participated for both years of MEPS and thus are observed once as a result of my own data-generating process. The inclusion of these variables is inspired by strategies proposed by Nijman & Verbeek (1992) to roughly control for some of the attrition bias induced by the unbalanced sample design.

Given the significant differences exhibited by the sample between men and women, I conduct analyses for men and women separately. I consider two conditional mean distributions: Poisson and negative binomial, which inherently account for the fact that A_i is restricted to non-negative values. The negative binomial distribution innately allows for overdispersion and includes the Poisson distribution as a special case, allowing for a more flexible specification. I thus choose to utilize a negative binomial model to estimate $BM1$, $BM2$, $I1$, $I2$, and $I3$ conditional mean specifications.

Though estimates are only reported for a subset of explanatory variables for organizational purposes, it should be noted that no variables are dropped from any of these model specifications at any point in the estimation process. The MEPS design is such that information on every individual living within a household is collected. Thus, some individuals in my sample are from the same family unit. Due to possible correlation of unobservables within the same family unit, I cluster standard errors at the family level. Further, there is likely an individual-specific error component that will be correlated within individuals that have two observations in the sample leading to bias in variance estimates for these individuals. Thus, standard errors are additionally clustered at the individual level.

Unconditional average marginal effects (AME) of a subset of explanatory variables are estimated for both of the baseline model specifications. Conditional AME (CAME) estimates for $keyMHdis$ are estimated using the $BM1$ specification for conditioning variables *sickpay*, *bonus*, $jobins$, and $inscostly$. Estimated CAME for $keyMHdis$ conditional on variables *plnchoic* and *nochoic* are generated using the $BM2$ specification. CAME estimates of $keyMHdis$ given conditioning variables *pubins* and *otherins* are estimated for each of the $BM1$ and $BM2$ specifications. Estimates of the moderating effect of each of the conditioning variables on the AME of $keyMHdis$ are calculated as the mean difference in conditional fitted value estimates across levels of the moderating variable and are based on manipulating the sample to assign a particular level of a variable to the entire sample; thus, the moderating effect estimates are based on counterfactual information. Standard errors for all AME, CAME, and moderating effect estimates are computed using the delta method.

CHAPTER 6

RESULTS

Prior to estimation, I consider the possibility of harmful collinearity that may persist when including both $keyMHdis$ and $distress$ together in each model. I conduct a VIF test for variables $keyMHdis$ and $distress$ for each sex separately and for both $BM1$ and $BM2$ specifications. The results of these tests suggest no statistical evidence of significant multicollinearity by including both of these measures of mental health in a model in tandem. I perform a likelihood ratio test of the null hypothesis that a model specification excluding $distress$ provides a superior fit to a model that includes only the diagnosis binary variable. The test indicates rejection of the null hypothesis of the restricted model in favor of the model including both measures of mental health at the 1 percent level.

6.1 Baseline Model ($BM1$, $BM2$)

The estimated dispersion parameter is approximately 0.21 for the baseline negative binomial model applied to the sub-sample of men with an estimated standard error of 0.004; the equivalent values for the baseline estimator for women in the sample are 0.28 wand 0.01, respectively.

Table 6.1 reports AME for analytical variables. Pseudo R^2 are also reported. AME estimates represent the expected change in the count of absence days per year in response to a unit change in the respective explanatory variable, on average. Estimates are robust across both $BM1$ and $BM2$ specifications.

Table 6.1: Baseline Models 1 and 2: AME Estimates

Dependent Variable:	Men		Women	
sickdays	$BM1$	$BM2$	$BM1$	$BM2$
keyMHdis	1.11 (0.35)***	1.10 (0.34)***	0.92 (0.31)***	0.92 (0.31)***
distress	0.11 (0.06)**	0.11 (0.03)***	0.15 (0.03)***	0.15 (0.03)***
physhlth	0.37 (0.05)***	0.37 (0.05)***	0.55 (0.06)***	0.55 (0.06)***
prtycnds	0.75 (0.09)***	0.67 (0.09)***	0.69 (0.09)***	0.69 (0.09)***
sickpay	0.49 (0.23)**	0.45 (0.23)**	0.63 (0.30)**	0.62 (0.30)**
bonus	-0.19 (0.22)	-0.21 (0.22)	0.08 (0.28)	0.07 (0.28)
jobins	0.82 (0.26)***		1.23 (0.34)***	
plnchoic		1.04 (0.30)***		1.41 (0.38)***
nochoic		0.80 (0.28)***		1.09 (0.37)***
pubins	0.90 (0.45)**	0.87 (0.44)**	1.15 (0.44)***	1.15 (0.43)***
otherins	0.65 (0.33)**	0.68 (0.32)**	0.39 (0.36)	0.41 (0.36)
inscostly	-0.42 (0.19)**	-0.43 (0.19)**	-0.34 (0.24)	-0.32 (0.24)
Observations	15,713		16,216	
Pseudo R^2:	0.164		0.198	

Note: Values in parentheses are standard errors computed using the delta method. *p<0.1; **p<0.05; ***p<0.01

Both men and women with a diagnosed mental illness are expected to exhibit a higher number of absence days than counterparts without a diagnosed mental illness, on average. These estimates support the empirical hypothesis that the unconditional AME of $keyMHdis$ is positive and are consistent with the literature on mental health and worker productivity. AME estimates for $keyMHdis$ reported in columns one through four are all statistically significant at the one-percent level. Men with a diagnosed mental disorder are estimated to report 1.11 days of additional absence, on average, compared to identical counterparts without a diagnosed mental disorder. The average woman in the sample is expected to report about 0.92 additional annual absences when diagnosed with a mental illness.

Estimates for the other three variables measuring health that are reported in Table 6.1 are highly statistically significant at the 0.1 percent level for both sexes. A marginal increase in general distress is estimated to increase absence by an average of 0.11 days for men and 0.15 days for women. The generalizability of these estimates should be interpreted with caution; this measure of general distress is based on responses to survey prompts given in only the second of the three interview rounds that occurs per year and so it is unclear how issues of timing may influence these estimates. For example, some individuals in the sample may have been unemployed at the time of the second interview so that the estimated effect of distress on absenteeism is not representative of the true effect of distress on absences from work for the employed.

In other words, the estimates for variable *distress* are only generalizable if most individuals in the sample who are unemployed at the second interview round report levels of distress that are homogeneous to (unobserved) levels of distress exhibited during periods of employment. AME estimates regarding self-reported measures of general physical health suggest that poorer degrees of physical health are associated with higher absences from work, on average, as anticipated.

A one-point increase in the rating of one's own degree of poor physical health is estimated to increase absences by under half a day for men (by a factor of about 0.37) on average, and over half a day (a factor of 0.55) for women in the sample, on average. This discrepancy across the sexes could be driven by differences in preferences over health across sex, but it is unclear to what degree this is the case. AME estimates suggest that men and women respond similarly to an additional priority condition diagnosis on average.

Table 6.1 reports that AME estimates for variable *sickpay* are significant at the five-percent level and are robust across models for both men and women. On average, receiving compensation for health-related absence from work is anticipated to increase expected days absent by about half a day for the sample of men. Estimates reported in columns three and four suggest that women may be slightly more responsive to a fringe benefit that supplements time lost from work due to illness, on average.

Having health insurance through a job is estimated to have a highly significant, positive association with absenteeism. Men enrolled in health insurance through an employer reporting an average of 0.82 additional days of absence compared to the average of 1.23 additional expected absences estimated for women. Women exhibit larger magnitudes of AME estimates compared to men when having a public source of insurance as a well as some other form of private insurance. This is not unusual considering that the empirical literature suggests that women have a greater propensity to utilize healthcare.

Results for variable *pubins* suggest that having a public source of health insurance slightly amplifies expected absences, on average, relative to having insurance through an employer. This phenomena might be partially driven by the association between poverty and health – low-income individuals who are eligible for public health insurance may face more health issues and thus require more time away from work; public health insurance may also be offered to individuals with debilitating health problems who require more time outside of work to receive care.

Having a source of private insurance provided by an entity other than an employer (represented by variable *otherins*) is significant at the 5 percent level for men; this is not significant for women in the sample. Men reporting agreement to the prompt "my health insurance is not worth its cost" exhibit a decline in absence days of 0.42 to 0.43 days with estimates robust at the 5 percent level across models. One the other hand, women exhibit a similar magnitude and sign of the AME estimate for *inscostly*, however, these are not statistically significant. This result is consistent with literature on the difference between men and women in seeking healthcare when ill.

The $BM2$ specification yields AME estimates for variables *plnchoic*, *nochoic*, and $NR.choic$ that further break down the $jobins$ variable into three additional categories of insurance. Individuals enrolled through an employer-sponsored plan are grouped based on whether their employer provides a choice between a portfolio of health plans (*plnchoice* $= 1$) or offers one health plan (*nochoic* $= 1$). Employed individuals that report being enrolled in an employer-sponsored plan but fail to answer survey prompts on whether or not their employer offers multiple plans to choose from ($NR.choic$). Variable $NR.choic$ is utilized as a necessary control so that the reference insurance source variable is still $unins$. Variable $NR.choic$ is not considered an analytical variable. Therefore, estimates are not reported for this variable in what follows.

Men with employer-provided insurance that chose from a catalog of plans exhibit an average of 1.04 additional days absent relative to uninsured men, and this AME estimate is significant at the 1 percent level. AME estimates of the *plnchoic* variable for sample women are also significant at the 1 percent level, with results suggesting the ability to select into an insurance plan is associated with an average increase in absenteeism of about 1.41 days for employed women. AME estimates of the additional absence days induced by a single-plan employer-sponsored package (when *nochoic* $= 1$) are comparatively smaller to the results for variable *plnchoic*.

In the next section, I examine results for the CAME of diagnosed mental illness, comparing groups that have access to a certain fringe benefit of interest relative to the identical counterpart that does not receive this benefit. Clarification on the interpretation of results is presented in what follows.

6.2 CAME Results: Job and Insurance Moderators

Table 6.2 reports CAME estimates separately for men and women. CAME estimates represent the average change in expected absence days associated with having a diagnosed mental illness in each "state" (each level of the binary moderating variable). The first column lists the name of the conditioning explanatory variable. CAME estimates of variable $keyMHdis$ at each level of a moderating variable are presented in columns three through eight along with delta method standard errors. The estimates in columns three and six are derived using counterfactual datasets and may be slightly different in magnitude than the discrete difference in the CAME estimates presented in columns one and two for men and four and five for women.

Table 6.2 reports that, conditional on having paid sick leave at a job, diagnosed mental illness is anticipated to increase expected annual absence by about 1.18 days for men and 0.98 days for women, on average. These estimates are larger by factors of about 0.7 and 0.6 days, respectively, than the unconditional AME estimates of diagnosed mental illness. Columns 3 and 6 suggest that, on average, the discrete difference in expected absences when sample men exhibit $keyMHdis = 1$ versus when sample men exhibit $keyMHdis = 0$ is larger by a factor of about 0.20 when conditioning on level $sickpay = 1$ compared to conditioning this discrete difference on $sickpay = 0$. This value is 0.14, on average, for sample women.

Table 6.2: Conditional Average Marginal Effects

Dep. Var.: *sickdays*		Men			Women		
Moderator	Model	$mod = 1$	$mod = 0$	Mean Change	$mod = 1$	$mod = 0$	Mean Change
sickpay	$BM1$	1.18 (0.36)***	1.00 (0.31)***	0.20 (0.08)***	0.98 (0.33)***	0.85 (0.28)***	0.14 (0.06)**
bonus	$BM1$	1.06 (0.33)***	1.13 (0.35)***	-0.08 (0.06)	0.95 (0.32)***	0.93 (0.31)***	0.02 (0.05)
inscostly	$BM1$	1.00 (0.31)***	1.15 (0.35)***	-0.17 (0.06)***	0.88 (0.30)***	0.95 (0.31)***	-0.07 (0.04)*
jobins	$BM1$	1.15 (0.35)***	0.85 (0.37)***	0.34 (0.11)***	0.98 (0.33)***	0.73 (0.25)***	0.27 (0.09)***
pubins	$BM1$	1.18 (0.39)***	0.85 (0.37)***	0.38 (0.16)**	0.97 (0.33)***	0.73 (0.25)***	0.26 (0.09)***
otherins	$BM1$	1.09 (0.35)***	0.85 (0.37)***	0.27 (0.11)**	0.81 (0.27)***	0.73 (0.25)***	0.09 (0.06)
plnchoic	$BM2$	1.20 (0.37)***	0.82 (0.26)***	0.43 (0.14)***	1.01 (0.34)***	0.72 (0.25)***	0.31 (0.10)***
nochoic	$BM2$	1.11 (0.35)***	0.82 (0.26)***	0.33 (0.11)***	0.94 (0.32)***	0.72 (0.25)***	0.24 (0.08)***
pubins	$BM2$	1.14 (0.38)***	0.82 (0.26)***	0.36 (0.15)**	0.95 (0.33)***	0.72 (0.25)***	0.25 (0.09)**
otherins	$BM2$	1.07 (0.34)***	0.82 (0.26)***	0.28 (0.11)**	0.80 (0.27)***	0.72 (0.25)***	0.09 (0.06)
Observations			15,713			16,216	

Note: Values in parentheses are clustered standard errors computed using the delta method. *p<0.1; **p<0.05; ***p<0.01

This is a key finding considering that the unconditional AME estimate of *sickpay* on expected absences reported in Table 6.1 is larger for women than for men, suggesting that the average additional absences induced by diagnosed mental illness may be less sensitive for women in the sample relative to men. The economic significance of the estimates reported in columns three and six for the moderator of paid sick leave is negligible, suggesting that paid sick leave is a means for increasing absenteeism regardless of whether one has a mental illness or not. A similar issue of economic significance is estimated for variable *inscostly* as well.

When insured, men exhibit CAME estimates of mental illness between 1.18 days and 1.09 days for the $BM1$ specification, and public sources of insurance are still anticipated to have the largest impact on absenteeism relative to the uninsured. In contrast, women report CAME estimates that are largest in magnitude for the *jobins* variable, though almost identical in magnitude to CAME estimates conditional on levels of *pubins* for the $BM1$ specification, there may be greater accessibility to mental healthcare, increasing healthcare service utilization even further over the group of the uninsured.

Another consideration is that public insurance indicates a low income household, a single parent, or those that have expensive health conditions. Individuals with low income at less lucrative jobs may not be offered certain benefits, especially more expensive offerings of fringe benefits such as health insurance. Though I control for age, this might also represent heterogeneity induced by the health of elderly persons who elect into Medicare Part B.

CAME estimates for men suggest that when uninsured, mental illness is anticipated to increase predicted absences between 0.82 and 0.85 days, on average. CAME estimates for women reported in column five report an average of 0.72 to 0.73 additional absences induced by diagnosed mental illness when uninsured. Women consistently report lower CAME estimates compared to men for the moderating variables listed in Table 6.2, as well as smaller estimated mean changes in predicted absences reported in column six, suggesting that the additional absences associated with diagnosed mental illness are less sensitive to values of the moderating variables for women.

CAME estimates of mental illness for the $BM2$ specification suggest that being enrolled in employer-provided health insurance and having the choice between multiple plans to enroll in is associated with about 1.20 additional absences when diagnosed with mental illness relative to this conditional effect when there is no diagnosis, on average for men, and about 1.01 additional absences over the baseline category of no diagnosed mental illness, on average, for women. These are the largest CAME estimates of diagnosed mental illness on expected absences reported for both genders, suggesting that the choice among plans may induce selection into plans that improve accessibility to care and time allotted to receiving care.

CAME estimates for men additionally suggest that, conditional on being enrolled in an employer-sponsored health insurance plan that was assigned without selection on behalf of the employee (i.e., conditional on $nochoic = 1$), expected absences are greater by a factor of about 1.11 days, on average, which is identical to the estimated average unconditional effect of mental illness on sampled men, suggesting that diagnosed mental illness on predicted absences suggest that Compared to AME estimates of $keyMHdis$ on expected absences conditional on variable *nochoic*, AME estimates conditional on *plnchoic* are higher by a factor of 0.9 for men and 0.7 for women. Results for conditioning variable *nochoic* also suggest that employer-sponsored plans that are not self-selected (i.e., only one plan offered) may be more restrictive and induce less health service utilization among the mentally ill than sources of public health insurance.

6.3 Interactions Between Diagnosed Mental Illness and Other Measures of Health

This section serves to provide more insight into the relationship between the binary measure of mental illness, $keyMHdis$, and variables $distress$, $physhlth$, and $prtrycnds$, which act as measures of general health that are likely highly correlated with a diagnosed mental disorder. I choose to graphically illustrate these relationships; then, I choose interesting values to demonstrate the empirical estimates of model specifications $I1$, $I2$, and $I3$ (refer to Section 5.1.3 for definitions of these models).

Each of the figures discussed hereafter plot predicted absences by index variables. Curves are plotted for each of two groups: those reporting a diagnosed mental illness and their counterparts reporting no such diagnosis. Dashed curves (blue dashed curves, if viewed in color) illustrate the relationship between a specified measure of health and predicted absences when $keyMHdis = 1$; dotted curves (red dotted curves, if viewed in color) illustrate this relationship when $keyMHdis = 0$. This applies to both sexes.

Figures 6.1 and 6.2 depict the relationship between general psychological distress (index variable $distress$) and predicted absences for men and women, respectively, by levels of $keyMHdis$. In general, both Figures 6.1 and 6.2 illustrate that at lower levels of $distress$, individuals that have a mental illness are anticipated to see higher levels of absences than individuals without a mental illness. However, at levels of the $distress$ index variable consistent with symptoms of severe mental illness (Kessler et. al, 2003), the inverse of this relationship is observed.

An intersection between the groups of men with and without a key mental disorder diagnosis is observed around a scale score of 11. At this point of intersection, there is

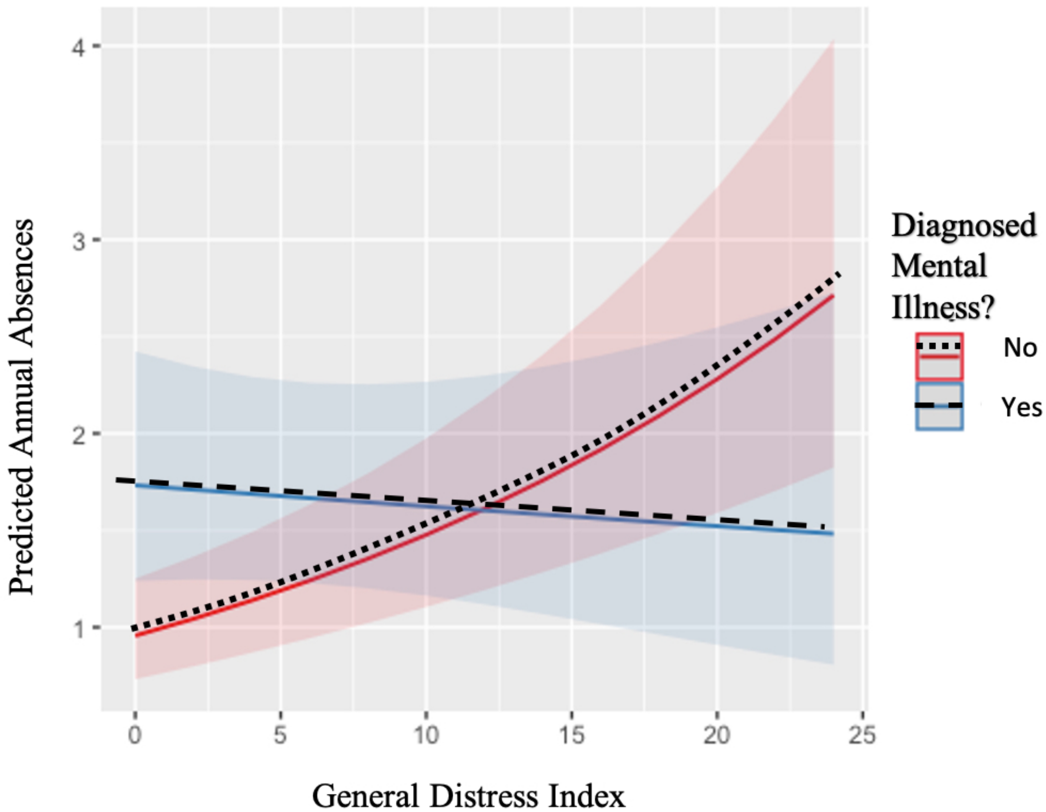

Figure 6.1: Between-Group Predicted Absences by General Distress: Men

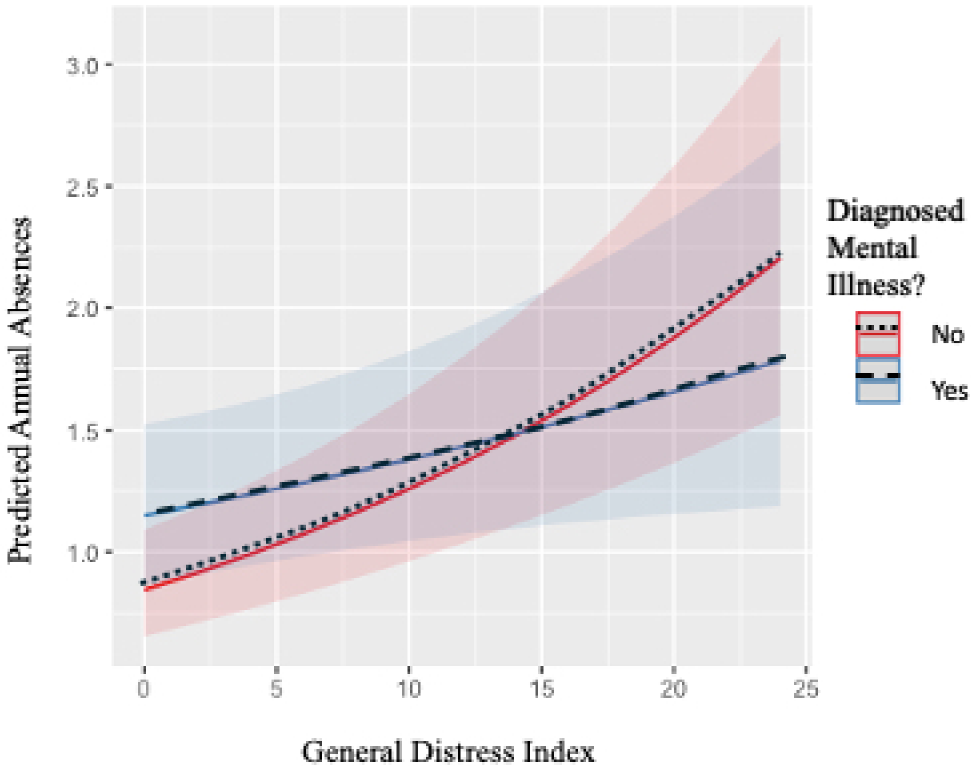

Figure 6.2: Between-Group Predicted Absences by General Distress: Women

anticipated to be no statistically significant difference in expected absence days between the two groups. From this intersection onward, it appears that at higher levels of distress, men without a diagnosed mental illness are anticipated to exhibit higher levels of absence than men with a diagnosis.

After observing this pattern in Figure 6.1, I estimate the AME of *diagnosed* mental illness ($keyMHdis$) on expected absences at three distinct levels of the variable $distress$ for men in the sample: 5, 11, and 15. These estimates are reported in Table 6.3 along with coefficient estimates for the interaction term $keyMHdis \times distress$.

CAME results are consistent with the relationship suggested by Figure 6.1 − at a score of 5 on the $distress$ index, the CAME of $keyMHdis$ on expected absences is estimated to be positive and statistically significant; on the other hand, at a $distress$ level of 11 which is approximately the point of intersection exhibited in Figure 6.1, this estimate is not statistically significant, nor is it at the higher index score of 15. Though the sign and magnitude of the coefficient on the interaction term is not directly interpretable here, the estimated degree of statistical significance of this coefficient is of value and illustrates that there is a distinct association between the interaction of diagnosed mental illness and general psychological distress and absenteeism for men in the sample.

A phenomenon similar to that exhibited for men can be observed in Figure 6.2 for women after adding an interaction between $distress$ and $keyMHdis$ to the $BM1$ specification.

The point of intersection between the lines representing the marginal effect of $distress$ on predicted absences at levels $keyMHdis = 0$ and $keyMHdis = 1$ occurs at slightly higher levels of $distress$ for sample women compared to sample men (approximately between levels 14 and 15), but the pattern exhibited by the relationship between $distress$ and $keyMHdis$ and its impact on predicted absences more or less mimics that observed for sample men. At lower levels of general distress, women display a positive and significant difference in

Table 6.3: CAME and Interaction Effects for Model $I1$

Dependent Variable sickdays				
Sample	Variable	Conditioning Level	CAME	Interaction Coefficients
Men	*distress*	5	1.08 (0.35)***	-0.05 (0.02)***
		11	0.18 (0.68)	
		15	-0.70 (1.13)	
	Pseudo R^2: 0.177			
	LR Test (H0: $BM1$, HA: $I1$): ***			
Women	*distress*	5	0.92 (0.32)***	-0.03 (0.01)**
		15	-0.11 (1.00)	
		20	-0.98 (1.72)	
	Pseudo R^2: 0.208			
	LR Test (H0: $BM1$, HA: $I1$): ***			

Note: *p<0.1; **p<0.05; ***p<0.01

predicted absences between the group reporting diagnosed mental illness and the baseline group of women with no such diagnosis, while higher levels of $distress$ suggest no statistically relevant difference in the absence behavior between the group of women with a diagnosed mental illness and no diagnosed mental illness. The coefficient estimate on the interaction term for the $I1$ specification for women is statistically significant at the five-percent level and is reported in Table 6.3.

Taking together the findings discussed so far for men and women, there is evidence of harmful impacts on labor productivity associated with exhibiting symptoms of moderate-to-severe mental illness, but failing to be diagnosed. We might indeed see lower levels of absences attributed to extreme levels of distress across individuals with diagnosed mental illness compared to individuals without a diagnosis that may not have the same knowledge or resources to be proactive. Further study of the structural mental health formula will provide valuable insight of the validity of this argument.

Figure 6.3 and Figure 6.4 plot predicted absences by a self-reported score on the $phsyhlth$ index, with higher scores indicating poorer physical health, for men and women, respectively. Figure 6.3 illustrates a similar trend among the group of men with a mental illness in the sample (linear and decreasing) as that observed in Figure 6.1, though men without a diagnosed mental illness clearly exhibit predicted absences that compile at a faster rate with increasing levels of $physhlth$ compared to the rate of change exhibited in Figure 6.1.

Conditional average marginal effects and coefficient estimates are reported for men and women in Table 6.4. CAME estimates suggest that the AME of diagnosed mental illness on absenteeism changes sign as self-rated scores of poor mental health increase from lower levels to very high levels. The coefficient estimate for term $keyMHdis \times physhlth$ is significant at the 0.1 percent level for men after clustering on individual and family units.

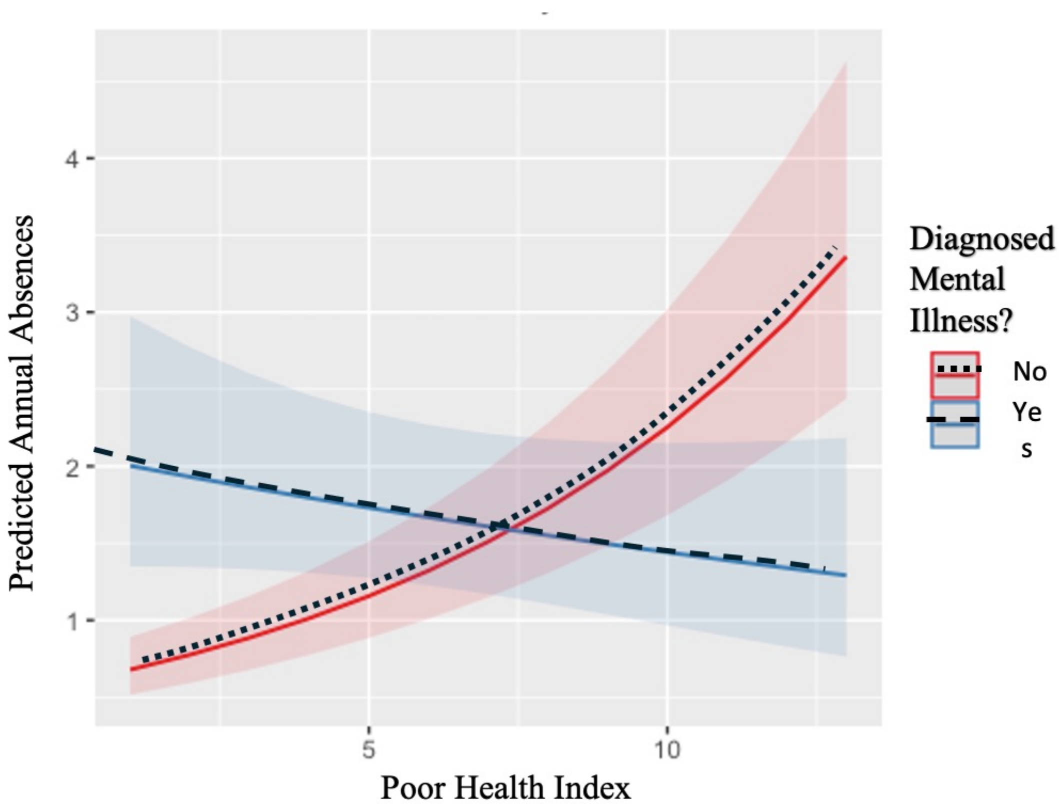

Figure 6.3: Between-Group Predicted Absences for Physical Health: Men

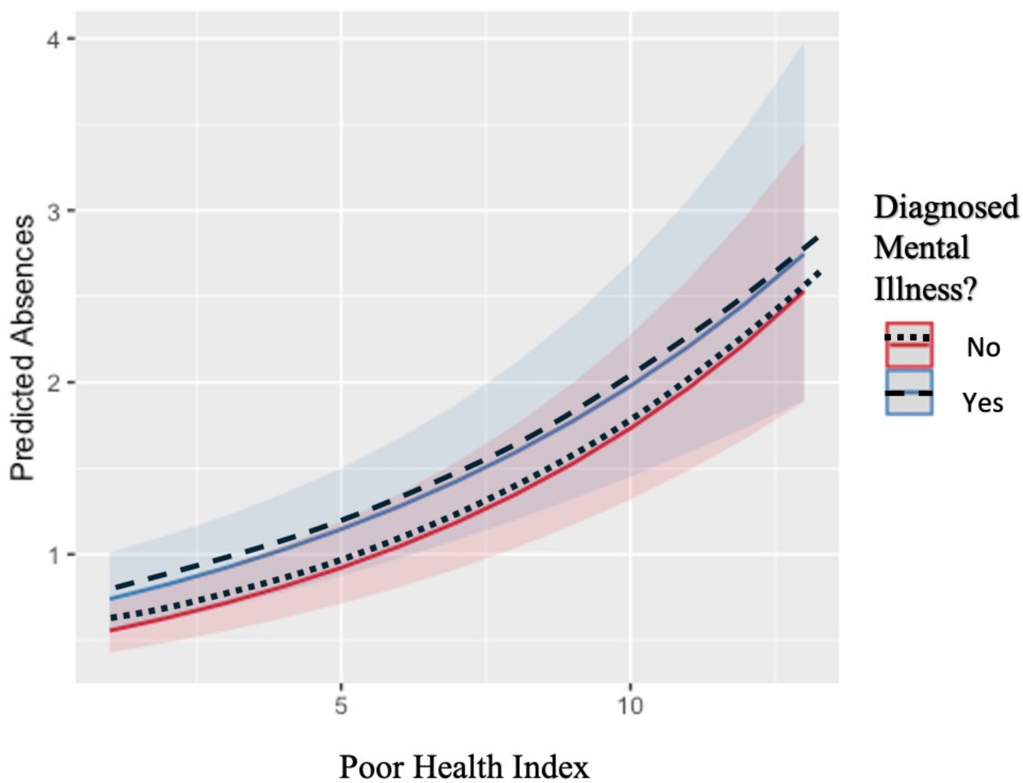

Figure 6.4: Between-Group Predicted Absences for Physical Health: Women

Table 6.4: CAME and Interaction Effects for Model $I2$

Dependent Variable sickdays				
Sample	Variable	Conditioning Level	CAME	Interaction Coefficients
Men	*physhlth*	3	1.64 (0.38)***	-0.19 (0.04)***
		7	0.21 (0.42)	
		10	-2.11 (1.04)**	
	Pseudo R^2: 0.178			
	LR Test (H0: $BM1$, HA: $I2$): ***			
Women	*physhlth*	3	0.79 (0.31)**	-0.02 (0.03)
		7	0.94 (0.39)***	
		10	0.98 (0.95)	
	Pseudo R^2: 0.207			
	LR Test (H0: $BM1$, HA: $I2$): Fail to Reject			

Note: *p<0.1; **p<0.05; ***p<0.01

On the other hand, there is no clear difference in the rate of change of predicted absences across levels of physical health for either group of women.

Figure 6.5 plots predicted absences by the count of the number of priority condition diagnoses as defined by the MEPS for sample men. These conditions include chronic heart conditions, high blood pressure, emphysema, asthma, cancer, and other chronic issues. In contrast to Figure 6.1 and Figure 6.3, Figure 6.5 does not exhibit a downward sloping curve for the group of men with a diagnosed mental illness; instead, it appears that this curve may exhibit a slightly positive or near-zero slope on average across all levels of the *prtycnds* variable exhibited by men in the sample.

In addition, some unobservable features of one's personality that may characterize the efficiency of the health production process likely come into play. For example, are individuals with a higher number primary condition diagnoses who also have a diagnosed mental illness simply more proactive when it comes to intervention and treatment and this is what is driving the distinction in the shape of the relationship in Figure 6.5. Similar questions come into play after observing the positive slope of the blue curve in Figure 6.6 for sampled women. Taking all of these considerations into account, inference on the implication of the relationship plotted in Figure 6.5 for men, as well as Figure 6.6 for women, would be pure conjecture; however, future research may address these possibilities by modeling structural equations of health rather than the reduced form focused on in this paper.

Noting the steeply increasing rate of change in predicted absences for the group of men without a diagnosed mental illness depicted in Figure 6.5 as well as values reported in Table 6.5, it can be deduced that men are more sensitive to shocks.

Figure 6.6 plots the predicted absences by condition count for women and depicts a higher positive rate of change per additional diagnosis for women without a diagnosed mental illness, and an intersection in prediction only at a very high number of diagnoses.

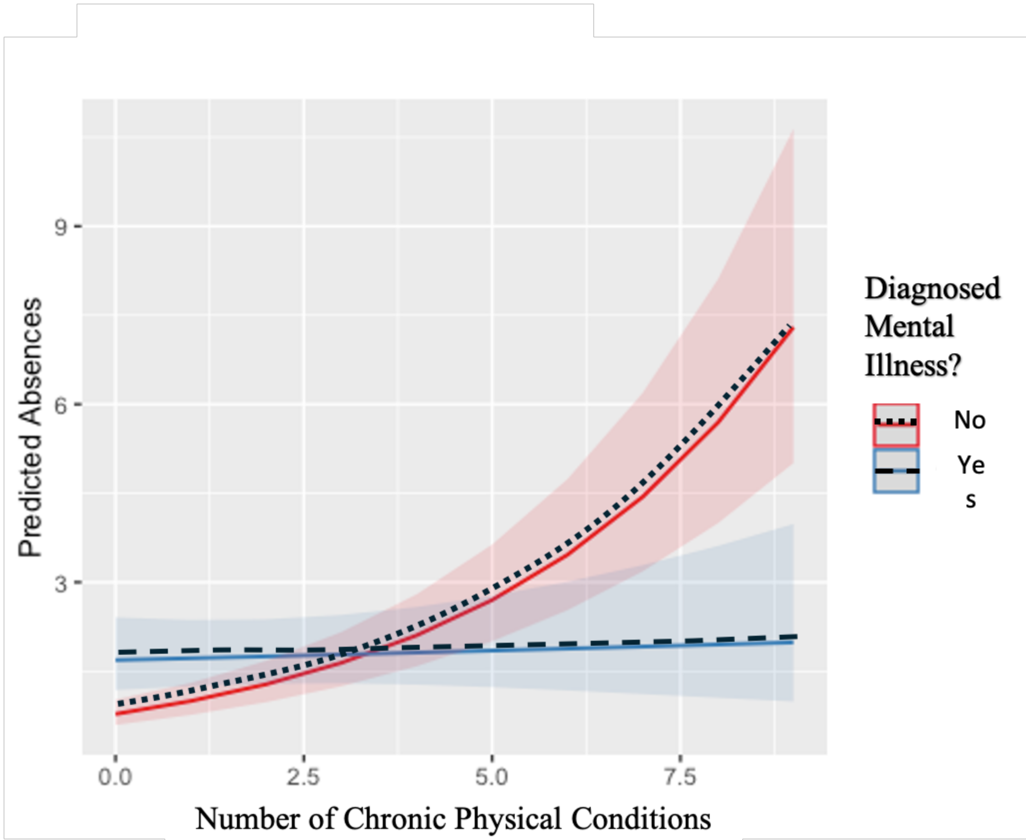

Figure 6.5: Between-Group Predicted Absences by Number of Chronic Physical Conditions: Men

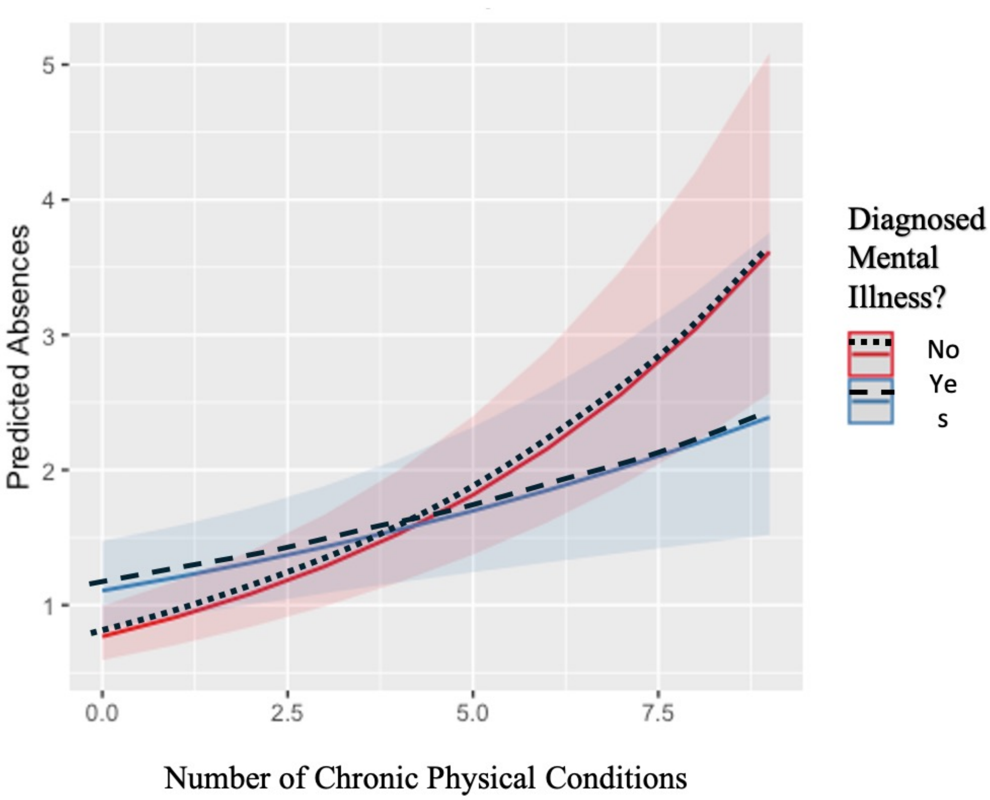

Figure 6.6: Between-Group Predicted Absences by Number of Chronic Physical Conditions: Women

Table 6.5: CAME and Interaction Effects for Model $I3$

Dependent variable: sickdays				
Sample	Variable	Conditioning Level	CAME	Interaction Coefficients
Men	*prtycnds*	1	1.25 (0.32)****	-0.24 (0.07)****
		3	0.29 (0.44)	
		6	-4.22 (2.06)**	
	Pseudo R^2: 0.178			
	LR Test (H0: $BM1$, HA: $I1$): ****			
Women	*prtycnds*	1	1.10 (0.31)****	-0.08 (0.04)**
		4	0.12 (0.66)	
		7	-2.39 (2.11)	
	Pseudo R^2: 0.208			
	LR Test (H0: $BM1$, HA: $I1$): ***			

This may suggest that the higher absence rates of women with mental illness are less responsive to additional condition diagnoses; said differently, a woman who is already dealing with symptoms of mental illness may have a higher tolerance level in terms of her threshold in which she decides to be absent than an identical woman who receives a primary condition diagnosis who did not face mental illness, all else equal. Of final note is the consideration of whether individuals facing multiple chronic physical conditions on top of mental illness may supply less labor in general or may be on disability leave that technically may not count as short-run absence in the context of this current paper.

In what follows I address the potential bias induced by non-random selection into the employed sample that is likely present for a number of reasons, such as sample error, attrition in sample editing, and unobservable data for all except a sub-population of unemployed individuals.

6.4 Additional Model Specifications and Estimation

6.4.1 Correlated Random Effects

Throughout this paper, I have argued and found evidence that symptomatology plays a significant role in labor supply that varies between and within diagnostic status. An equally critical consideration is how people respond to mental health shocks. Ignoring individual-specific patterns within the sample assumes that all individuals respond uniformly to a given shock, holding all else constant. This assumption overlooks the nuanced variability in responses across individuals in the panel data framework.

Sources of heterogeneity may capture worker emotional "grit" and how this may generate omitted variable bias. While a worker's grit is not directly observable until after a contract is undertaken, it fundamentally shapes their production process and must therefore be accounted for. In consideration of these factors, I employ the correlated random effects (CRE) method proposed by Wooldridge (2019).[1] This approach leverages the panel design by incorporating individual-specific means of time-varying explanatory variables. The CRE method accounts for the for individual-level heterogeneity by using these means as control variables (Benson et al., 2022; Heckman, Urzua, and Vytlacil, 2006). The following provides the generalizations of the model and definitions of necessary assumptions.

I utilize the Wooldridge method due to the unbalanced panel I work with, in which some individuals are observed for one year, while others are observed twice – once in each year of participation in the MEPS. For individual i at time t:

$$y_{it}|x_{it},\alpha_i \sim \text{NegBin}(\mu_{it},\theta) \quad \text{with} \quad \mu_{it} = \exp(x_{it}'\beta + \alpha_i).$$

For CRE to be feasible, the following assumptions should hold:

$$\alpha_i = \bar{x}_i'\xi + a_i, \quad E[a_i|x_i] = 0.$$

Finally, marginalize over $a_i \sim \text{Gamma} \Rightarrow$ Negative Binomial likelihood:

$$E[y_{it}|x_{it},\bar{x}_i] = \int \exp(x_{it}'\beta + \bar{x}_i'\xi + a)h(a)da.$$

A final estimate of the average marginal effects (AME) can be estimated simply after accounting for the estimated within-individual effects. Table 6.6 and Table 6.7 show estimates for men and women, respectively. The columns labeled $BM1$ in each table report the AME estimates initially reported in Table 6.1 as well as the coefficient estimates from the $BM1$

[1]Wooldridge 2019 is a method for nonlinear estimators and unbalanced samples as an extension of the Mundlak approach (1978).

Table 6.6: Estimates for Pooled Negative Binomial and Correlated Random Effects Negative Binomial: Men

Dependent variable:	*AME*		Coefficients	
sickdays	$BM1$	Correlated RE	$BM1$	Correlated RE
keyMHdis	1.11	0.95	0.37***	0.33***
	(0.43, 1.77)	(0.39, 1.51)	(0.10)	(0.11)
distress	0.11***	0.10***	0.04***	0.03***
	(0.06, 0.16)	(0.05, 0.14)	(0.01)	(0.01)
physhlth	0.37	0.35	0.12***	0.12***
	(0.27, 0.46)	(0.27, 0.43)	(0.01)	(0.01)
prtycnds	0.67***	0.62***	0.22	0.21***
	(0.50, 0.84)	(0.48, 0.75)	(0.02)	(0.02)
sickpay	0.49**	0.44	0.16**	0.15**
	(0.04, 0.93)	(0.02, 0.85)	(0.08)	(0.07)
bonus	-0.19	-0.16	-0.06	-0.06
	(-0.61, 0.24)	(-0.56, 0.23)	(0.07)	(0.07)
jobins	0.82***	0.81***	0.30***	0.36***
	(0.31, 1.32)	(0.32, 1.30)	(0.10)	(0.10)
pubins	0.90**	0.76*	0.33**	0.30**
	(0.003, 1.78)	(-0.05, 1.56)	(0.15)	(0.15)
otherins	0.65**	0.61**	0.25**	0.25**
	(0.01, 1.29)	($-1e^{-3}$, 1.22)	(0.12)	(0.12)
inscostly	-0.42**	-0.38**	-0.14**	-0.13**
	(-0.78, -0.04)	(-0.73, -0.04)	(0.06)	(0.06)
Observations		15,713		

*p<0.1; **p<0.05; ***p<0.01

Table 6.7: Estimates for Pooled Negative Binomial and Correlated Random Effects Negative Binomial: Women

Dependent variable:	*AME*		Coefficients	
sickdays	$BM1$	CRE	$BM1$	CRE
keyMHdis	0.92**	0.91***	0.21***	0.21***
	(0.33,1.54)	(0.35, 1.46)	(0.07)	(0.06)
distress	0.15***	0.14***	0.03***	0.03***
	(0.09, 0.21)	(0.09, 0.20)	(0.01)	(0.01)
physhlth	0.55**	0.53***	0.12***	0.12***
	(0.44, 0.66)	(0.43, 0.64)	(0.01)	(0.01)
prtycnds	0.69***	0.66***	0.16***	0.15***
	(0.52, 0.88)	(0.49, 0.82)	(0.02)	(0.02)
sickpay	0.63**	0.55*	0.14**	0.13*
	(0.04, 1.22)	(-0.01, 1.11)	(0.07)	(0.07)
bonus	0.08	0.06	0.02	0.01
	(-0.46, 0.64)	(-0.46, 0.57)	(0.06)	(0.06)
jobins	1.23***	1.24***	0.31***	0.31***
	(0.56, 1.90)	(0.59, 1.88)	(0.09)	(0.09)
pubins	1.15***	1.17***	0.29***	0.30***
	(0.30, 2.01)	(0.35, 1.99)	(0.10)	(0.10)
otherins	0.39	0.39	0.11	0.11
	(-0.31, 1.10)	(-0.30, 1.07)	(0.10)	(0.10)
inscostly	-0.34	-0.31	-0.07	-0.07
	(-0.80, 0.15)	(-0.76, 0.14)	(0.05)	(0.05)
Observations		16,216		

*p<0.1; **p<0.05; ***p<0.01

model specification. Columns titled "Correlated RE" in Tables 6.6 and 6.7 report estimates after including the individual-specific means of explanatory variables as control variables in the analysis.

The third and sixth columns in both Table 6.6 and Table 6.7 provide estimates of the AME and coefficients when additionally conditioning on the individual-specific intercept parameters. The fringe benefit variables are robust, with *sickpay* coefficients remaining statistically significant and *bonus* consistently not exhibiting statistical nor economic significance.

Access to insurance (job, public, or other) consistently increases sick days for both genders, likely due to reduced financial barriers to taking time off.

The limitations of the CRE method is the main assumption that x_{it} is strictly exogenous after conditioning on the individual level heterogeneity, α_i. Said differently, an individual is observed randomly conditional on their individual-specific intercept and there is no mechanism that induces an individual to be observed after a shock occurs. In what follows, I test for a non-random mechanism of observation.

6.4.2 Addressing Sample Selection on Unobservable Components

Individuals who are not presently employed do not have observable information for absenteeism. The decision not to work may relate to the factors that affect absenteeism. Thus, I conduct a final analysis using methods to account for unobserved sample selection. First, I use the classic Heckman sample selection model even though the outcome of absenteeism is a count variable. This approach ignores the discrete nature of the outcome variable and the overdispersion in the data. Later on, I invoke an estimator that uses copulas and penalized maximum likelihood to estimate the dependence between the selection and outcome equations, allowing me to use a negative binomial distribution in the second stage.

In what follows, I only am interested in analyzing health-related variables. However, some job factors are utilized as controls in the outcome equations. Because the first stage model estimates the probability of employment, I also include a sample of unemployed persons in this analysis these individuals make up a sample representing adults in the labor force. For consistency, the same data-generating processes are implemented to obtain the sample of unemployed persons so that it excludes individuals who have ever retired, military personnel, disabled individuals reporting significant difficulty completing activities of daily life, and students.

Additionally, I impose the condition of only excluding persons who have never worked and those currently out of the labor force at the time of the MEPS interviews so that I am focusing on behavior of people who *could* work (Jones et al., 2008; Certo et al., 2016). The inclusion of unemployed individuals yields a full sample size of 23,093 for women and 19,823 for men.

As a final comparison, I consider working-age individuals who persist after the data-generating process previously described and who are outside the labor force (and thus are not unemployed). The extended sample for men consists of 24,562 observations; for women, there are 29,947 observations.

Though estimates of sample selection models can be estimated when the same explanatory variables are included in the first stage equation and second stage equation, the estimation can be improved by including some variables in the selection (first stage) equation that are not included in the outcome equation. These variables are called exclusion restrictions, and strong exclusion restrictions are variables that directly impact the likelihood of employment, but do not directly impact the count of health-related workplace absenteeism.

I propose that two variables may act as exclusion restrictions. The names of these variables are *depout* and *spou.emp*. I will start with the definition and my argument for the variable *depout*. This variable represents the presence of dependents living outside the household, such as college students, non-custodial children, or elderly parents in assisted care. A binary variable equal to one if a person's spouse in employed, *spou.emp*, is also considered as an exclusion restriction.

Dependents inside the household may directly affect daily labor supply decisions, as their health or needs could influence a worker's ability to work. In contrast, dependents outside the household are unlikely to affect daily health-related labor supply decisions but may increase the need for income, thereby influencing employment decisions. Therefore, I argue that the variable *depout* will increase the probability of employment due to higher income needs but is not directly related to health-specific absenteeism.

A similar argument applies to the variable *spou.emp*: while the employment status of one's spouse is likely to strongly impact the individual's employment status, it does not directly influence an individual's decision to be absent from work after experiencing a negative health shock.[2]

The first stage probit includes all categories of explanatory variables as defined in Section 4.1 and also defined in Appendix A, *except for* the job characteristics variables (those grouped into variable category J). Variables *jobins* and *otherins* in category I of explanatory variables are additionally merged into a single variable representing whether one has a private source of insurance information. *Additionally* the two exclusion restrictions discussed previously are included in this probit equation. The binary variable EMP, equal to one if an individual is employed, zero otherwise, is the dependent variable in the probit equation. The second stage outcome equation is estimated using OLS and the equation includes both category J, and the inverse mills ratio, which controls for the dependence between selection into employment and the associated outcome.

[2]It may be argued that if a spouse is unemployed, a worker will be less likely to be absent when ill to ensure job security. However, I contend that such decisions depend on the severity of the health shock and the individual's ability to report to work—factors that are highly individual-specific.

The probit model estimating probability of employment in the first stage is:

$$
\begin{aligned}
P(Y=(EMP=1)|X=\{MH,PH,I,X,C,ER\}) =& F(\beta_0+MH_i\beta^{MH}+PH_i\beta^{PH}\\
&+I_i\beta^{I}+X_i\beta^{X}+C_i\beta^{C}+ER_i\beta^{ER}),
\end{aligned}
$$

for every observation, $i=1,...,n$. Matrices MH, PH, I, X, C, and ER hold observed values of the corresponding explanatory variables for all $i=1$, ..., n observations. Matrix ER holds the exclusion restrictions *depout* and *spou.emp*. The function $P(.)$ represents the probability of being employed, and F represents the Normal distribution.

Each matrix of explanatory variables is of dimension $n\times k_l$ where n is the total number of observations in the sample and k_l is an integer equal to the number of explanatory variables in corresponding matrix l for $l=\{MH,PH,I,X,C,ER\}$. In a similar fashion, each β^l for $l=\{MH,PH,I,X,C,ER\}$ is a $k_l\times 1$ vector of parameter to be estimated.
The outcome equation is:

$$
\begin{aligned}
E[A_i|EMP_i>0;\{MH,PH,J,I,X,C\};\lambda] =& \alpha_0+MH_i\alpha^{MH}+PH_i\alpha^{PH}\\
&+I_i\alpha^{I}+X_i\alpha^{X}+C_i\alpha^{C}+\lambda_i\alpha^{bias},
\end{aligned}
$$

for every observation, $i=1,...,n$. Matrices MH, PH, J, I, X, and C hold observed values of the corresponding explanatory variables for all $i=1$, ..., n observations.

The additional conditioning variable, λ is equal to the inverse Mills ratio, and $\lambda_i\alpha^{bias}$ is the estimated bias imposed by endogenous sample selection. Each matrix of explanatory variables is of dimension $n\times k_m$ where n is the total number of observations in the sample and k_m is an integer equal to the number of explanatory variables in corresponding matrix m. Each α^m for $m=\{MH,PH,J,I,X,C,\lambda\}$ is a $k_m\times 1$ vector of parameter to be estimated.
The results of the Heckman process are presented and discussed below.

Table 6.8 reports the estimates of the equation of the first stage of the probit model and the estimates of the marginal effect of the outcome equation for a select group of analytical variables after employing the Heckman procedure for men.

Table 6.8: Employment and Outcome Equation Estimates, Men

	In Labor Force		Extended Sample	
	Probit	Outcome	Probit	Outcome
keyMHdis	-0.27*** (0.04)	0.91*** (0.15)	-0.40*** (0.05)	0.22 (0.29)
physhlth	-0.92** (0.04)	0.25** (0.20)	-0.03** (0.01)	0.35** (0.03)
distress	-0.56*** (0.05)	0.33** (0.12)	-0.01** (0.004)	0.09*** (0.02)
prmrycnds	-0.05*** (0.01)	0.51*** (0.01)	-0.05*** (0.01)	0.62*** (0.06)
depout	0.45*** (0.06)		0.85** (0.10)	
spou.emp	0.19*** (0.03)		0.21*** (0.04)	
Inverse Mills Ratio	-1.32*** (0.37)		0.81*** (0.16)	
Observations	19,823	15,713	24,562	15,713

*p<0.1; **p<0.05; ***p<0.01

Table 6.8 reports that the coefficient for *keyMHdis* is approximately 0.10 lower than the average marginal effect (see Table 6.1 for these estimates), while the coefficient for *physhlth* is lower by a similar magnitude. In contrast, the coefficient for *distress* is about two times higher, and the coefficient for *prtycnds* is also higher. Both *depout* and *spou.emp* are statistically significant at the 5 percent level and have positive coefficients, indicating that external dependents and the status of spousal employment significantly influence employment decisions. All estimates for men are robust, suggesting reliable results.

Interestingly, for the Heckman model of the extended sample, the IMR is still significant but the sign is now positive. Physical health index variables and distress score variables are significantly smaller in magnitude for the selection equation for men. However, the exogenous diagnostic variables are more robust to the change in samples for the employment equation (*keyMHdis* and *prmrycnds*). In the outcome equation in Table 6.8, *keyMHdis* becomes insignificant.

Table 6.9 reports the first-stage equation estimates of the probit model and the outcome equation marginal effect estimates for a select group of analytical variables after employing the Heckman procedure for women.

Table 6.9: Employment and Outcome Equation Estimates, Women

	In Labor Force		Extended Sample	
	Probit	Outcome	Probit	Outcome
keyMHdis	-0.02 (0.03)	0.95*** (0.15)	-0.34 (0.03)	0.68** (0.26)
physhlth	-0.79*** (0.03)	0.91*** (0.14)	-0.01 (0.01)	0.49*** (0.04)
distress	-0.12*** (0.04)	0.25** (0.20)	-0.01** (0.003)	0.14*** (0.02)
prtycnds	-0.03*** (0.01)	0.52*** (0.03)	-0.06*** (0.01)	0.52*** (0.07)
depout	0.44*** (0.05)		1.12*** (0.15)	
spou.emp	-0.46*** (0.02)		0.21*** (0.03)	
Inverse Mills Ratio	-1.85** (0.80)		0.54 (0.40)	
Observations	23,093	16,216	29,947	16,216

*p<0.1; **p<0.05; ***p<0.01

The results for women, illustrated in Table 6.9, report that the coefficient estimate for *keyMHdis* is slightly higher than AME estimates (see Table 6.1) but remains robust, while the coefficient for *distress* is higher by about 0.10 and also robust. The coefficient for is similarly higher and robust. Notably, the coefficient for *depout* is positive and statistically significant, highlighting the role of external dependents in increasing the likelihood of employment. Interestingly, the coefficient for *spou.emp* is negative and statistically significant, suggesting that spousal employment may reduce the likelihood of employment for women, possibly due to household labor dynamics. The Inverse Mills Ratio (IMR) is highly significant and negative, similar to the results for men but larger in magnitude, indicating a stronger correction for selection bias among women. The extended sample for the Heckman model indicates an IMR that is not significant when including women outside of the labor force.

These results reveal important gender differences in the factors influencing employment decisions. While men and women share some similarities, such as the significance of *depout* and the robustness of estimates, key differences emerge in the magnitude and direction of coefficients for variables like *distress* and *spou.emp*. The significant IMR for both genders confirms the presence of selection bias, with a stronger effect among women. These findings underscore the importance of considering sex-specific factors in labor market analyses and policy design, particularly regarding the role of external dependents and spousal employment status.

Due to the discrete nature of absenteeism outcome variable and the characteristic of overdispersion exhibited by this variable, I consider a sample selection method that allows for a probit specification in the first stage and a negative binomial specification in the second stage. The dependence parameters between the two functions, as well as the coefficient estimates, are estimated using penalized maximum likelihood.[3]

In the semi-parametric model, the dependence between the selection and outcome equations is captured by Kendall's τ and the copula parameter, θ. These values as well as coefficient estimates for analytical variables are reported in Table 6.10 for men. I emphasize that the results in Table 6.10 are based on estimates of men in the labor force, including the unemployed. Individuals outside of the labor force are not considered at this point.

[3]I use the R software repository to locally rebuild the unsupported package, `SemiParSampleSel`. While some functions were not salvageable, the estimation of model coefficients for the first and second stages of an endogenous sample selection model is attainable. The Open AI source, GitHub Copilot, was utilized to edit the manual code necessary to implement these estimates.

Table 6.10: Estimates for Each Stage of Semi-parametric Selection Model: Men

	Probit		Outcome (Negative Binomial)	
Variables	Coefficients	Standard Errors	Coefficients	Standard Errors
keyMHdis	-0.15***	(0.04)	0.41***	(0.11)
physhlth	-0.70***	(0.04)	0.14***	(0.01)
distress	-0.81***	(0.10)	0.03***	(0.01)
prtycnds	-0.09***	(0.01)	0.20***	(0.02)
depout	0.57**	(0.06)		
spou.emp	0.28**	(0.03)		
Sample Size	19,823[1]		15,713	
Kendall's τ	0.124 (0.05,0.20)			
θ	0.194 (0.08,0.31)			

[1] Sample size includes a the sample of unemployed persons in the labor force.

For men, Kendall's τ is estimated at 0.124 (95% confidence interval (CI): 0.05, 0.20), and θ is 0.194 (95% CI: 0.08, 0.31). For women, these estimates are higher, with Kendall's τ at 0.15 (95% CI: 0.07, 0.24) and θ estimates of 0.24 (95% CI: 0.104, 0.37). These results indicate a moderate degree of dependence between the selection and outcome processes. The estimates reported in Table 6.11 for the extended sample of both unemployed men in the labor force and men not in the labor force are robust when compared to results reported in Table 6.10.

Tables 6.12 and 6.13 indicate that the semi-parametric sample selection model estimates for women are robust to broadening the scope of the sample to include women outside of the labor force. However, the point-estimates are not as consistent as those estimated for men.

Table 6.11: Estimates for Each Stage of Semi-parametric Selection Model: Men – *Including men out of the labor force*

	Probit		Outcome (Negative Binomial)	
Variables	Coefficients	Standard Errors	Coefficients	Standard Errors
$keyMHdis$	-0.15***	(0.04)	0.35***	(0.05)
physhlth	-0.73***	(0.04)	0.14***	(0.01)
$distress$	-0.52***	(0.06)	0.03***	(0.01)
prtycnds	-0.09***	(0.01)	0.20***	(0.02)
depout	0.55***	(0.06)		
spou.emp	0.25***	(0.03)		
Sample Size	24,562[1]		15,713	
Kendall's τ	0.126 (0.04,0.20)			
θ	0.197 (0.07,0.31)			

[1]Sample size includes unemployed persons and those outside of the labor force.

Table 6.12: Estimates for Each Stage of Semi-parametric Selection Model: Women

	Probit		Outcome (Negative Binomial)	
Variables	Coefficients	Standard Errors	Coefficients	Standard Errors
$keyMHdis$	-0.10***	(0.02)	0.28***	(0.11)
$physhlth$	-0.64***	(0.06)	0.09***	(0.01)
$distress$	-0.70***	(0.10)	0.06***	(0.01)
$prtycnds$	-0.03**	(0.01)	0.18***	(0.02)
$depout$	0.45**	(0.06)		
$spou.emp$	-0.12**	(0.03)		
Sample Size	23,093[1]		16,216	
Kendall's τ		0.154 (0.07,0.24)		
θ		0.24 (0.104,0.37)		

[1] Sample size includes a the sample of unemployed persons in the labor force.

Table 6.13: Estimates for Each Stage of Semi-parametric Selection Model: Women – *Including women out of the labor force*

	Probit		Outcome (Negative Binomial)	
Variables	Coefficients	Standard Errors	Coefficients	Standard Errors
keyMHdis	-0.28***	(0.04)	0.33***	(0.05)
physhlth	-0.94***	(0.04)	0.09***	(0.01)
distress	-0.76***	(0.06)	0.06***	(0.01)
prtycnds	-0.03**	(0.01)	0.18***	(0.02)
depout	0.47***	(0.06)		
spou.emp	-0.15***	(0.03)		
Sample Size	29,947[1]		16,216	
Kendall's τ	0.139 (0.05,0.21)			
θ	0.217 (0.08,0.33)			

[1]Sample size includes unemployed persons and those outside of the labor force.

The discrepancy between the IMR and the copula parameters highlights the strengths and limitations of each method. The Heckman model provides a clear and interpretable measure of selection bias through the IMR, which is robust and statistically significant. In contrast, the semi-parametric model offers flexibility in modeling the dependence structure but produces less precise estimates, as evidenced by the wide confidence intervals for Kendall's τ and θ. This implies that while the semi-parametric approach is useful for exploring the dependence structure, the Heckman model remains a more reliable tool for correcting selection bias in this context.

The significant IMR in the Heckman model, coupled with the moderate dependence indicated by Kendall's τ and θ in the semi-parametric model, underscores the importance of addressing selection bias in labor market analyses. Both methods confirm that failing to account for non-random selection can lead to biased estimates, particularly for variables that exogenously define one's health and therefore, may not be anticipated to be associated with selection bias at first. As we have seen, $keyMHdis$ and *prtrycnds* show significant changes in magnitude and significance after correction.

While still up for debate, I end the empirical portion of my dissertation by acknowledging that each methods of sample selection has its benefits – the Heckman model is better suited for providing precise and actionable insights, while the semi-parametric approach offers complementary information about the underlying dependence structure.

CHAPTER 7

CONCLUSIONS

Employed persons separate themselves from the remainder of the labor force in that they are subject to implicit contractual constraints defined by one's own utility, employment benefits and earnings defined prior to acceptance of an offer. Features of the contract such as fringe benefits and positional expectations (such as quotas, meeting mandatory deadlines for projects, etc.) become additional factors that indirectly impact the health stock of the individual, and thus, their utility.

Results regarding each of the binary variables representing sources of health insurance imply that any source of health insurance is anticipated to increase absenteeism relative to the category of uninsured individuals. The implications of this finding are intuitive – the higher the price of inputs in health production, the less time allotted to health production and the more time available to allot to occupational labor. Results suggest that health insurance impacts absence rates through two different channels: the substitution of labor time for time spent utilizing healthcare services (leading to an increase in absenteeism), and through the channel of mitigating the impact that mental illness has on absenteeism. The latter of the two cases is suggestive that being insured increases the likelihood of seeking treatment for mental illness so that the additional absences exhibited by individuals with mental illness are further amplified over counterparts without mental illness for both genders. This is further suggested by the estimates for variable *inscostly*.

For the sample of men, results for the analysis of the interaction between mental illness and other measures of general health provide evidence in favor of the hypothesis that the interaction between mental illness and other measures of health are highly correlated and interact significantly to influence absenteeism. I showed in the theoretical framework of this paper that better health should result in fewer absences. This is the case for men in the sample for all measures of general health considered. The interaction between diagnosed mental illness and a self-rated index of one's health is not found to be significant in determining predicted absences for women in the sample, though this may be due to other family members responding on the behalf of a woman family member. This brings up sources of bias due to the use of observational data that should be addressed.

The results suggest that mental and physical health, workplace conditions, and access to insurance are significant predictors of sick days for both men and women. The CRE model provides robust estimates by accounting for unobserved heterogeneity, and the findings highlight the importance of addressing mental and physical health issues, improving work conditions, and understanding the role of insurance in shaping health-related behaviors. The findings additionally suggest that there may be an additional benefit to allowing employees to choose from multiple plans in terms of absenteeism.

Taken together, the findings of this study highlight the significant, often hidden, economic impact of undiagnosed moderate-to-severe mental illness on labor productivity. Individuals without a formal diagnosis may lack the resources or awareness to proactively address worsening symptoms, resulting in higher absenteeism and reduced productivity. Conversely, diagnosed individuals, while potentially facing more frequent *reported* health shocks, may have greater access to interventions that help mitigate extreme levels of distress and promote workplace attendance. This dynamic underscores the importance of more advanced analysis encapsulating mental health diagnoses alongside symptom severity to engage in dialogue of real effects of acute episodes of mental illess.

Findings suggest that workers with a mental health diagnosis in severe distress and fair-to-low physical health exhibit better productive outcomes compared to similar but undiagnosed peers. This supports the notion that employers can mitigate productivity losses by promoting working wellness and by adopting health policies that encourage early treatment and support for mental health. Notably, AME estimates suggest that absenteeism for women in the sample may be more sensitive to presently observed health shocks than men. This may be due to a multitude of reasons, including differences over preferences for the sexes. Also of note is that when including the interaction of diagnosed mental illness and self-reported physical health status, women do not exhibit clear evidence that a diagnosed mental illness eventually makes them better off at poor levels of self-reported physical health.

The number of chronic physical health conditions is found to significantly influence absenteeism schedules for both men and women when comparing across groups of diagnosed mental illness and no such diagnosis. This finding is consistent with the logic surrounding the argument that adverse health effects may pose less harm to the labor market outcomes of individuals who have knowledge of their diagnoses and risks; this is especially true if an individual has a typical healthcare provider, which reduces the search costs associated with improving health through the route of utilizing healthcare. This is an area of interest for future research that addresses these relationship more clearly by modeling structural equations of health rather than the reduced form focused on in this paper. Of final note is the consideration of whether individuals facing multiple chronic physical conditions on top of mental illness may supply less labor in general or may be on disability leave that technically may not count as short-run absence in the current context, which can be studied in the future.

The analysis demonstrates that mental illness ($keyMHdis$) significantly impacts labor supply decisions, particularly absenteeism. However, the magnitude of this effect may be underestimated if sample selection bias is not properly addressed. The Heckman procedure, combined with strong exclusion restrictions (*depout* and *spou.emp*), provides robust estimates that correct for potential selection bias. Gender differences are evident, with women showing stronger associations between mental illness and absenteeism, though selection bias appears less pronounced for women compared to men. The findings underscore the importance of addressing sample selection in labor supply models and highlight the role of dependents and household dynamics in shaping employment and absenteeism decisions.

REFERENCES

Agency for Healthcare Research and Quality (AHRQ). *Medical Expenditure Panel Survey (MEPS) Household Component (MEPS-HC) Panel Design and Collection Process.* https://meps.ahrq.gov/survey_comp/hc_data_collection.jsp

American Psychiatric Association. (2024, March). *Stigma, Prejudice and Discrimination against People with Mental Illness.* https://www.psychiatry.org/patients-families/stigma-and-discrimination

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Ashwood, J. S., Briscombe, B., Collins, R. L., Wong, E. C., Eberhart, N. K., Cerully, J., May, L., Roth, B.,& Burnam, M. A. (2017). *Investment in Social Marketing Campaign to Reduce Stigma and Discrimination Associated with Mental Illness Yields Positive Economic Benefits to California. Rand Health Quarterly, 6*(2).

Becker, G. S. (1965). A Theory of the Allocation of Time. *The Economic Journal, 75*(299), 493 − 517. https://doi.org/10.2307/2228949

Braumoeller, B. F. (2004). Hypothesis Testing and Multiplicative Interaction Terms. *International Organization, 58*(4), 807 − 820. https://doi.org/10.1017/S0020818304040251

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. https://doi.org/10.1016/j.labeco.2017.05.001

Buis, M. L. (2010). Direct and Indirect Effects in a Logit Model *The Stata Journal,10*(2), 305 − 308. https://doi.org/10.1177/1536867X1001000104

Cameron, A. C. & Trivedi, P. (2004). *Microeconometrics: Methods and Applications.* Cambridge University Press.

Cameron, A. C. & Trivedi, P. K. (1986). Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests. *Journal of Applied Econometrics, 1*(1), 29 − 53. https://doi.org/10.1002/jae.3950010104

Center for Disease Control and Prevention, National Center for Health Statistics. (2022). *2021 NHIS Questionnaires, Datasets, and Documentation.* https://www.cdc.gov/nchs/nhis/documentation/2021-nhis.html

Chunrong, A. & Norton, E. C. (2003). Interaction Terms in Logit and Probit Models. *Economics Letters, 80*(1), 123 – 129. https://doi.org/10.1016/S0165-1765(03)00032-6

Cseh, A. (2008). The Effects of Depressive Symptoms on Earnings. *Southern Economic Journal, 75*(2), 383 – 409. https://doi.org/10.1002/j.2325-8012.2008.tb00910.x

Eggleston, K., & Bir, A. (2009). Measuring Selection Incentives in Managed Care: Evidence from the Massachusetts State Employee Insurance Program. *Journal of Risk and Insurance, 76*(1), 159 – 175.

Fletcher, J. (2013). Adolescent Depression and Adult Labor Market Outcomes. *Southern Economic Journal, 80*(1), 26 – 49. https://doi.org/10.4284/0038-4038-2011.193

French, M. T. & Zarkin, G. A. (199). Mental Health, Absenteeism and Earnings at a Large Manufacturing Worksite. *The Journal of Mental Health Policy and Economics, 1*(4), 161 – 172. https://doi.org/10.1002/(SICI)1099-176X(199812)1:4%3C161::AID-MHP21%3E3.0.CO;2-I

Goetzel, R. Z., Long, S. R., Ozminkowski, R. J., Hawkins, K., Wang, S., & Lynch, W. (2004). Health, Absence, Disability, and Presenteeism Cost Estimates of Certain Physical and Mental Health Conditions Affecting U.S. Employers. *Journal of Occupational and Environmental Medicine , 46*(4), 398 – 412. https://doi.org/10.1097/01.jom.0000121151.40413.bd

Goetzel, R. Z., Ozminkowski, R. J., Sederer, L. I., & Mark, T. L. (2002). The Business Case for Quality Mental Health Services: Why Employers Should Care About the Mental Health and Well-Being of Their Employees. *Journal of Occupational and Environmental Medicine, 44*(4), 320 – 330. https://doi.org/10.1097/00043764-200204000-00012

Goldman, W., McCulloch, J., & Sturm, R. (1998). Costs And Use Of Mental Health Services Before And After Managed Care. *Health Affairs, 17*(2), 40 – 52. https://doi.org/10.1377/hlthaff.17.2.40

Greene, W. (2002). *The Behavior of the Fixed Effects Estimator in Nonlinear Models.* [Unpublished Manuscript]. Stern School of Business, NYU.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy, 80*(2), 223 – 255. https://doi.org/10.7312/gros17812-004

Hamermesh, D. S. (1993). *Labor Demand*. Princeton Press.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica, 47*(1), 153–161. https://doi.org/10.2307/1912352

Jia, V. & Vatto, T. E. (2021). Predicting the Path of Labor Supply Responses when State Dependence Matters. https://doi.org/10.1016/j.labeco.2021.102004

Jones, A., García-Gómez, P., Rice, N. (2010). Health Effects on Labour Market Exits and Entries. *Labour Economics, 17*(1), 62 − 76. https://doi.org/10.1016/j.labeco.2009.04.004

Karaca-Mandic, P., Norton, E. C., Dowd, B. (2012). Interaction Terms in Nonlinear Models. *Health Services Research, 47*(1), 255 − 274. https://doi.org/10.1111/j.1475-6773.2011.01314.x

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S. L., Manderscheid, R. W., Walters, E. E., & Zaslavsky, A. M. (2003). Screening for Serious Mental Illness in the General Population. *Archives of General Psychiatry, 60*(2), 84 − 189. https://doi.org/10.1001/archpsyc.60.2.184

Kessler, R.C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D.K., Normand, S. L., Walters, E. E., and Zaslavsky, A. M. (2002). Short Screening Scales to Monitor Population Prevalence and Trends in Non-Specific Psychological Distress. *Psychological Medicine, 32*, 959 − 976.

Kohn, R., Saxena, S., Levav, I., & Saraceno, B. (2004). The Treatment Gap in Mental Health Care. *Bulletin of the World health Organization, 82*(11), 858 − 866.

Kroenke, K., Spitzer, R. L., and Williams, J. B. (2003). The Patient Health Questionnaire-2: Validity of a Two-Item Depressive Screener. *Medical Care, 41*, 1284 − 1292.

Lee, S. (2013). *Interaction and Marginal Effects in Nonlinear Models: Case of Ordered Logit and Probit Models.* [Unpublished Manuscript]. University of Texas at Austin.

Long, J. S. & Freese, J. (2006). *Regression models for categorical dependent variables using Stata* (Vol. 7). Stata Press.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica, 46*(1), 69 − 85. https://doi.org/10.2307/1913646

National Alliance on Mental Illness. (2023, April). *Mental Health by the Numbers.* https://www.nami.org/learn-more/mental-health-by-the-numbers

National Institute of Mental Health. (2024). *Mental Health Information: Statistics.*

Nijman, T., & Verbeek, M. (1992). Nonresponse in Panel Data: The Impact on Estimates of a Life Cycle Consumption Function. *Journal of Applied Econometrics, 7*(3), 243 − 257.

Riphahn, R. T., Wambach, A., & Million, A. (2003). Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation. *Journal of Applied Econometrics, 18*(4), 387 − 405. https://doi.org/10.1002/jae.680

Rosenheck, R. A., Druss, B., Stolar, M., Leslie, D., & Sledge, W. (1999). Effect of declining mental health service use on employees of a large corporation. *Health Affairs, 18*(5), 193 − 203. https://doi.org/10.1377/hlthaff.18.5.193

Shang, S., Nesson, E., & Fan, M. (2017). Interaction Terms in Poisson and Log Linear Regression Models. *Bulletin of Economic Research, 70*(1), 89 − 96.

Simon, G. E., Barber C., Birnbaum, H. G., Frank, R. G., Greenberg, P. E., Rose, R. M., Wang, P. S., & Kessler, R. C. (2001). Depression and Work Productivity: The Comparative Costs of Treatment Versus Nontreatment. *Journal of Occupational and Environmental Medicine, 43*(1), 2 − 9. https://doi.org/10.1097/00043764-200101000-00002.

Stratmann, T. (1999). What Do Medical Services Buy? Effects of Doctor Visits on Work Day Loss. *Eastern Economic Journal, 25*(1), 1–16.

The Kennedy-Satcher Center for Mental Health Equity, The Kennedy Forum, The Carter Center, & Well Being Trust. (2018, October 3). *32 States Receive Failing Grades on New Report Cards Scoring Statutes for Mental Health and Addiction Treatment 10 Years After Federal Parity Law.* https://www.thekennedyforum.org/press-releases/32-states-get-a-failing-grade-on-new-report-cards-scoring-statutes-for-mental-health-and-addiction-treatment-10-years-after-federal-parity-law/

United States Bureau of Labor Statistics. (2024, January). *The employment situation news release: more chart packages.* https://www.bls.gov/charts/employment-situation/alternative-measures-of-labor-underutilization.htm.

Van Eeuwijk, F. A. (1995). Multiplicative Interaction in Generalized Linear Models. *Biometrics, 51*(3), 1017–1032. https://doi.org/10.2307/2533001

Von Korff M., Ormel J., Katon, W., & Lin, E. H. (1992). Disability and Depression Among High Utilizers of Health Care. *Archives of General Psychiatry, 49*(2), 91 – 100.

Wang, P., Puterman, M. L., Cockburn, I., & Le, N. (1996). Mixed Poisson Regression Models with Covariate Dependent Rates. *Biometrics, 52*, 381–400.

Williams, R. (2009). Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients Across Groups. *Sociological Methods & Research, 37*(4), 531 – 559.

Wooldridge, J.M. (2019). Correlated Random Effects Models with Unbalanced Panels. *Journal of Econometrics, 211*(1), 137 – 150. https://doi.org/10.1016/j.jeconom.2018.12.010

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). The MIT Press.

World Health Organization. (2024, September 2). *Mental Health: Mental Health at Work*. https://www.who.int/news-room/fact-sheets/detail/mental-health-at-work

Wyszynski, K., Marra, G. (2017). Sample Selection Models for Count Data in R. *Computational Statistics 33*, 1385 – 1412. https://doi.org/10.1007/s00180-017-0762-y

APPENDIX A

TOTAL SAMPLE SUMMARY STATISTICS AND VARIABLE DEFINITIONS

Table A.1: Total Sample Summary Statistics

Variable	Description	Mean	Min	Max	SD
Dependent Variable (A):					
sickdays	Dependent variable. Count of the total days an individual has been absent from work due to illness or injury in the past year.	3.18	0	160	9.97
Mental Health (MH):					
keyMHdis	=1 if individual has a diagnosed mental disorder of interest, =0 otherwise.	0.10	0	1	0.30
distress	Discrete scale from 0 to 24 with higher scores indicating greater emotional distress.	2.60	0	24	3.57
adhd	=1 if individual reports an ADHD diagnosis, =0 otherwise.	0.01	0	1	0.08
Physical Health (PH):					
physhlth	A discrete scale score with range 1 to 13 with higher scores indicating poorer general health.	4.66	1	13	2.40
prtycnds	The number of priority condition diagnoses.	1.31	0	9	1.45
routine	=1 if the individual received a routine medical evaluation in the past year, =0 otherwise.	0.03	0	1	0.16
injury	=1 if individual suffered an injury or illness requiring immediate medical care in the past year.	0.24	0	1	0.43

(Continued on following page)

Table A.1 Continued					
Variable	Description	Mean	Max	Min	SD
smoke	=1 if individual smokes cigarettes, =0 otherwise.	0.17	0	1	0.38
exercise	=1 if individual exercises at least 3 times per week, =0 otherwise.	0.53	0	1	0.50
pregnt	Sub-sample of women only. =1 if female was pregnant at any point during the year.	0.03	0	1	0.18
Job Traits (J):					
$selfemp$	= 1 for self employed individuals, =0 otherwise.	0.13	0	1	0.02
$one4$	Benchmark group. =1 for individuals with tenure between 1 and 4 years, =0 otherwise.	0.37	0	1	0.48
$five14$	=1 for individuals with tenure between 5 and 14 years, =0 otherwise.	0.35	0	1	0.48
$fifteen24$	=1 for individuals with tenure between 15 and 24 years, =0 otherwise.	0.11	0	1	0.32
$25plus$	=1 for individuals with tenure of 25 years or more, =0 otherwise.	0.06	0	1	0.24
temp	=1 if individual has a temporary employment contract, =0 otherwise.	0.05	0	1	0.22
parttime	=1 if individual reports working 35 hours per week or more, =0 otherwise.	0.17	0	1	0.37

(Continued on following page)

Table A.1 Continued					
Variable	Description	Mean	Max	Min	SD
union	=1 if individual is part of a labor union, =0 otherwise.	0.13	0	1	0.33
ssnl	=1 if individual's main job is a seasonal positions, =0 otherwise.	0.05	0	1	0.21
pubsect	=1 if individual works in the public sector, =0 otherwise.	0.18	0	1	0.38
*occ*1	Management, business, and financial operations.	0.13	0	1	0.33
*occ*2	Professional and related occupations.	0.22	0	1	0.41
*occ*3	Service occupations.	0.19	0	1	0.39
*occ*4	Sales and related occupations.	0.08	0	1	0.27
*occ*5	Office and administrative support.	0.14	0	1	0.35
*occ*6	Farming, fishing, and forestry	0.01	0	1	0.10
*occ*7	Construction, extraction, and maintenance.	0.08	0	1	0.27
*occ*8	Benchmark group. Production, transportation, material moving.	0.15	0	1	0.36
*occ*9	Unclassifiable occupation.	0.004	0	1	0.06

(Continued on following page)

Table A.1 Continued					
Variable	Description	Mean	Max	Min	SD
$1to19$	=1 if firm has between 1 and 19 workers.	0.30	0	1	0.46
$20to99$	=1 if firm has between 20 and 99 workers.	0.33	0	1	0.47
$100to499$	=1 if firm has between 100 and 499 workers.	0.21	0	1	0.41
$500plus$	Benchmark group. =1 if firm has 500 or more workers.	0.16	0	1	0.37
$NRunemp$	=1 if individual did not respond to questions pertaining to firm size.	0.05	0	1	0.22
Health Insurance (I, $\tilde{I}$):					
$inscostly$	=1 if individual either somewhat agrees or strongly agrees with the statement,"health insurance is not worth the cost", =0 otherwise.	0.28	0	1	0.45
$jobins$	=1 if individual is insured through their job, =0 otherwise.	0.63	0	1	0.48
$plnchoic$	=1 if insured through job AND selected from a catalog of plans, =0 otherwise	0.34	0	1	0.47
$nochoic$	=1 if insured through job AND was only eligible for one plan, =0 otherwise.	0.24	0	1	0.43
$NR.choic$	=1 for observations reporting coverage through job but no response about whether they had a choice of plans, =0 otherwise	0.05	0	1	0.22

(Continued on following page)

Table A.1 Continued					
Variable	Description	Mean	Max	Min	SD
otherins	=1 if individual has private health insurance coverage through a non-job source, =0 otherwise.	0.13	0	1	0.34
pubins	=1 if individual has health insurance coverage through a public source, =0 otherwise.	0.07	0	1	0.25
unins	Benchmark group. =1 if individual is uninsured, =0 otherwise	0.17	0	1	0.38
Demographics (X):					
poor	=1 if household income as a % of poverty line puts them into "poor" or "near poor" groups.	0.14	0	1	0.34
lowinc	=1 if household income as a % of poverty line puts them into "low income" group.	0.16	0	1	0.36
midinc	=1 if household income as a % of poverty line puts them into "middle income" group.	0.34	0	1	0.47
highinc	Benchmark group. =1 if household income as a % of poverty line puts them into "high income" group.	0.37	0	1	0.48
married	=1 if individual is married, =0 otherwise.	0.55	0	1	0.50
famsz	Number of individuals within the surveyed household.	3.02	0	14	1.66

(Continued on following page)

Table A.1 Continued					
Variable	Description	Mean	Max	Min	SD
yngchldrn	Number of children aged 6 and under.	0.38	0	6	0.71
age	Age in years.	41.86	18	84	11.82
belowhs	Individuals with less than high school education.	0.13	0	1	0.34
hsdeg	Benchmark group. Individuals with a high school degree or GED.	0.32	0	1	0.47
somecoll	Individuals with some college or an associate's degree, but no 4-year degree.	0.25	0	1	0.43
bachdeg	Individuals with a bachelor's degree.	0.20	0	1	0.40
bachplus	Individuals with schooling beyond a bachelor's degree.	0.10	0	1	0.31
hispanic	=1 if individual is Hispanic, =0 otherwise	0.28	0	1	0.45
black	=1 if individual is Black, =0 otherwise	0.18	0	1	0.38
asian	=1 if individual is Asian, =0 otherwise	0.08	0	1	0.28

(Continued on following page)

Table A.1 Continued					
Variable	Description	Mean	Max	Min	SD
bornus	=1 if individual was born in the US.	0.65	0	1	0.48
Other Controls (C):					
unemp.1	=1 if individual reports unemployment at exactly one of the three interviews.	0.07	0	1	0.26
unemp.2	=1 if individual reports unemployment at exactly two of the three interviews.	0.06	0	1	0.23
partialemp	= 1 if an individual is unemployed at one of the three interviews, but worked for at least part of the reference period prior to employment termination.	0.04	0	1	0.20
$empUB$	The total number of days for which an individual was employed during the year.	338	365	28	0.19
moved_US	=1 if individual moved within the US during the calendar year.	0.03	0	1	0.18
moved_RU	=1 if individual joined a new reference unit (household) at any point during the calendar year.	0.01	0	1	0.10
NE	=1 if individual resides in the Northeastern region for most of the year.	0.16	0	1	0.36
MW	=1 if individual resides in the Midwestern region for most of the year.	0.20	0	1	0.40

(Continued on following page)

Table A.1 Continued					
Variable	Description	Mean	Max	Min	SD
W	=1 if individual resides in the Western region for most of the year.	0.28	0	1	0.45
S	Benchmark group. =1 if individual resides in the Southern region for most of the year.	0.37	0	1	0.48
unemp.rt	The average annual unemployment rate faced by the individual estimated using monthly regional unemployment reports from the Bureau of Labor Statistics and region of residence reported for each round.	8.05	5.5	10.9	1.29
yearone	=1 if individual is only observed for the first year of their designated panel.	0.26	0	1	0.44
yeartwo	=1 if individual is only observed for the second year of their designated panel.	0.20	0	1	0.40
bothyrs	Benchmark. =1 if individual is observed for both years of their panel.	0.54	0	1	0.50
Exclusion Restrictions (ER:)					
depout	=1 if dependent(s) reside outside of household, =0 otherwise.	0.04	0	1	0.01
spou.emp	=1 if individual claims dependents residing outside of the individual's lace of residence.	0.84	0	1	0.12

APPENDIX B

SAMPLE SELECTION DETAILS

Generic Models of Endogenous Sample Selection

Define variable E_i^* for individual i as one's propensity toward work; it can be thought of as the latent utility of working for individual i. This function of latent utility is defined as

$$E_i^* = \delta^0 + MH_i'\delta^{MH} + PH_i'\delta^{PH} + X_i'\delta^X + C_i'\delta^C + ER_i'\delta^{ER} + u_i, \tag{B.1}$$

for every observation, $i = 1,...,n$, δ^m for $m = \{MH, PH, X, C, ER\}$, are parameter estimates corresponding to the previously defined data matrices, and new matrix ER, which contains variables utilized in equation (B.1) that are not in functions $BM1$ nor $BM2$ as previously defined, and u_i is an error term for observation i. In practice, E_i^* is not directly observable and instead, an individual's employment status is observed. E_i^* – the individual's theoretically preferred level of work is ignored when considering only the sample of employed adults. The following translates the latent utility process into the observable processes. Conditionally define E_i^* is defined as follows:

$$E_i = \begin{cases} 1 & \text{if } E_i^* \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Thus, when the utility of supplying labor outweighs the disutility of supplying labor, $E_i = 1$ is observed. If E_i is observed, then absenteeism (A_i) is observed.

If error terms ϵ_i and u_i are independent, then there is exogenous sample selection, and a function of likelihood of employment can be estimated by OLS or other methods that ignore the selection issue. This process makes selection adjustments based on observable characteristics that determine the likelihood of employment. On the other hand, if ϵ_i and u_i are correlated, endogenous sample selection occurs and the conditional expectation of outcome A_i for the sample with observed levels of A_i (i.e., when $E_i = 1$), the endogenous selection can be treated similarly to an omitted variable issue (Heckman, 1979).

To illustrate this, define vectors $w_i = \{MH_i, PH_i, X_i, C_i, ER_i\}$ and $z_i = \{MH_i, PH_i, X_i, C_i\}$ for observation i, and matrix $\beta = \{\beta^0, \beta^{MH}, \beta^{PH}, \beta^X, \beta^C\}$ and consider the following derivation of the expected value of absenteeism for the selected sample given exogenous variables, $E(A_i|w_i, E_i = 1)$:

$$\begin{aligned} E(A_i|w_i, u_i) &= z_i'\beta + E(\epsilon_i|w_i, u_i), \\ &= z_i'\beta + E(\epsilon_i|u_i), \\ &= z_i'\beta + u_i'\gamma. \end{aligned} \tag{B.2}$$

As u_i is not observed and thus cannot be directly utilized in practice, the law of iterated expectations can be used to obtain a function that can be estimated,

$$E(A_i|w_i, E_i) = z_i'\beta + E(u_i|w_i, E_i)'\gamma, \tag{B.3}$$

where $E(u_i|w_i, E_i)$ can be generalized to some function $h(w_i, E_i)$.

Given the modeling definitions thus far and knowing that A_i is only observable for the observations of $E_i = 1$, we can define $h(w_i, E_i = 1)$ as follows:

$$E(u_i|w_i, E_i = 1) = E(u_i|u_i > -w_i'\delta).$$

Assume u_i follows a standard normal distribution so that the above can be further defined as follows:

$$\begin{aligned} E(u_i|u_i > -w_i'\delta) &= \frac{\phi(-w_i'\delta)}{1 - \Phi(w_i'\delta)} \\ &= \frac{\phi(w_i'\delta)}{\Phi(w_i'\delta)} \equiv \Lambda(w_i'\delta), \end{aligned} \tag{B.4}$$

where ϕ denotes the standard normal probability density function and Φ is the standard normal cumulative density function. $\Lambda(.)$ is the inverse Mills ratio (IMR) which can be estimated using a two stage procedure which treats endogenous sample selection similarly to an issue of ommitted variable bias (Heckman, 1979). When using Heckman's method, the function of interest is

$$E(A_i|w_i, E_i) = z_i'\beta + \gamma\Lambda(w_i'\delta), \tag{B.5}$$

where $\gamma\Lambda(w_i'\delta)$ will be part of the residual if OLS estimation is employed. Thus, a simple test of the null hypothesis that $\gamma = 0$ will be sufficient to test if there is any endogenous selection.

Example: Logic Behind Hypothesis of Endogenous Selection

A concrete example of a factor that is unobservable to researchers, but that is likely to induce some of the distinction between mentally ill individuals that are employed and not is ability. Whether or not mental illness is correlated with the unmeasured ability in the overall population, these two variables will be correlated in the selected sample. If ability does have an impact on absenteeism, then the effect of mental illness on absenteeism will be underestimated because in the selected sample, people with mental illness that choose to work have above average ability. Besides ability, the number of years between symptom onset and treatment, one's ability to cope with the illness as well as one's susceptibility to negative side effects associated with medications, severity of overall symptoms (short or long term), and other individual-level features of illness presentation and ability to cope may determine one's level of disutility per labor hour, and thus, employment and absenteeism.