competence in competition + health

series

#2017/04

Christopher Cronin, Matthew Forsstrom, and Nicholas Papageorge

Mental Health, Human Capital and Labor Market Outcomes

Imprint

CINCH SERIES

CINCH – Health Economics Research Center
Weststadttürme Berliner Platz 6-8
45127 Essen

Phone +49 (0) 201 183 - 6326
Fax +49 (0) 201 183 - 3716
Email: daniel.avdic@uni-due.de
Web: www.cinch.uni-due.de

competence in competition + health

series

#2017/04

Christopher Cronin, Matthew Forsstrom, and Nicholas Papageorge

Mental Health, Human Capital and Labor Market Outcomes

Christopher Cronin[*], Matthew Forsstrom[†], and Nicholas Papageorge[‡]

Mental Health, Human Capital and Labor Market Outcomes

Abstract

There are two primary treatment alternatives available to those with mild to moderate depression or anxiety: psychotherapy and medication. The medical literature and our analysis suggests that in many cases psychotherapy, or a combination of therapy and medication, is more curative than medication alone. However, few individuals choose to use psychotherapy. We develop and estimate a dynamic model in which individuals make sequential medical treatment and labor supply decisions while jointly managing mental health and human capital. The results shed light on the relative importance of several drawbacks to psychotherapy that explain patients' reluctance to use it: (1) therapy has high time costs, which vary with an individual's opportunity cost of time and flexibility of the work schedule; (2) therapy is less standardized than medication, which results in uncertainty about its productivity for a given individual; and (3) therapy is expensive. The estimated model is used to simulate the impacts of counterfactual policies that alter the costs associated with psychotherapy.

JEL Classifications: I10, I12, J22, J24.

Keywords: Mental Health, Demand for Medical Care, Labor Supply, Structural Models.

---

* *ccronin1@nd.edu.*

† *matthew.forsstrom@wheaton.edu.*

‡ *Corresponding author: papageorge@jhu.edu.*

We gratefully acknowledge helpful comments from: Bill Evans, David Bradford, Michael Dickstein, Donna Gilleskie, Michael Richards, Richard Frank, Robert Moffitt, and Daniel Avdic along with seminar participants at ASHEcon Philadelphia, the Southeastern HESG Richmond, H2D2 Ann Arbor, SOLE Raleigh, and Essen Health and Labour Conference.

1 Introduction

Economists view health as a form of human capital that has the potential to improve an individual's longevity, quality of life, and productivity at work (Grossman, 1972). An implication is that medical treatment can be viewed as a costly investment. The Grossman framework has been widely applied in economics to understand how patients make medical decisions by weighing both current and future costs and benefits of different medical treatments (e.g., Khwaja, 2010). It has also been extended to incorporate other features linked to healthcare decision-making, including learning and uncertainty about drug quality (Crawford and Shum, 2005a), side effects of treatments (Papageorge, 2015), and links to labor market decisions and outcomes (Gilleskie, 1998).

Surprisingly, existing work has rarely applied the Grossman model to mental health.[1] This is troubling since nearly one in five adults in the US experiences mental illness in a given year, the most common being mild to moderate anxiety or depression.[2] Moreover, mental health problems are consistently associated with poor labor market outcomes, including lower productivity, absenteeism, and disability, which seems to suggest that mental health (like physical health) should be analyzed as a form of human capital. One reason for this gap in the literature amounts to measurement problems surrounding both the diagnosis of mental illness and the impact of treatment, the latter posing a formidable empirical challenge due to selection bias, as we document below. Moreover, data limitations make it difficult to relate mental health, treatment, and labor market outcomes. Another unfortunate reason for this gap is that mental health problems—perhaps due to widespread stigma or a general lack of understanding—are often seen as fundamentally different from physical health problems. The implicit suggestion seems to be that rational choice, applied in a wide variety of medical contexts, is somehow inappropriate for an analysis of mental healthcare. This position ignores the fact that the majority of mentally ill individuals manage relatively mild illnesses. According to the National Survey on Drug Use and Health, in 2015 18% of U.S. adults reported experience with a mental illness in the past year, while only 4% report experience with a *serious* mental illness.[3]

This paper examines how mental health and treatment decisions relate to the labor

[1] This observation was made by Currie and Stabile (2006) as well.

[2] According to the National Alliance on Mental Illness. Moreover, as of 2011, antidepressants were the most consumed class of drugs in the United States at roughly 260 million prescriptions per year, generating nearly $20 billion in revenues annually (Mojtabai and Olfson, 2014).

[3] According to the National Institute of Mental Health (NIMH), *serious* mental illnesses are defined as those, "resulting in serious functional impairment, which substantially interferes with or limits one or more major life activities." Even among these individuals, inpatient treatment, much less institutionalization, is rare. In 2008, only 7.5% (NIMH) of individuals reporting a serious mental illness sought inpatient treatment.

market. We focus on individuals with mild to moderate depression or anxiety and on two broad treatment categories: therapy and pharmaceuticals. A key motivation of our analysis is to understand why patients often opt for pharmaceuticals even though medical literature has suggested that therapy may be a more productive treatment. If so, patient choices are not consistent with the objective of solely maximizing their mental health, but instead reflect additional costs associated with therapy. The aim is to explain these costs and to ascertain to what degree they can be mitigated with policy.

Our analysis begins with the estimation of a mental health production function, which links treatment alternatives to improvements in mental health outcomes. We use data from the Medicare Expenditure Panel Survey (MEPS), which follows the treatment decisions and mental health outcomes of more than 200,000 individuals over a two-year period. We identify the effect of endogenously selected medical treatment on mental health outcomes using an instrumental variables strategy. The strategy leverages plausibly exogenous variation in the supply of psychiatrists and prescription drug prices over time within an individual's county of residence. Our first contribution is to use observational data to provide causal estimates of the effect of various medical treatments on mental health. Our analysis suggests that therapy is more effective than pharmaceuticals, which is more effective than no treatment, at improving mental health.

Our findings on the productivity of mental health treatments are in line with recent medical literature.[4] However, they raise new questions since patients are far more likely to forgo therapy in favor of pharmaceuticals. To illustrate this point, Figure 2 shows that the proportion of American adults reporting a mental health issue has nearly doubled in the past 20 years. With the exception of a small increase in attention deficit disorders (ADD), the entirety of this increase is due to increases in depression and anxiety disorders, the illnesses we focus on in this paper. Over the same period of time, the use of psychotherapeutic drugs to treat mental illness has risen substantially (by roughly 50%) while the use of talk therapy has fallen (see Figure 3). These patterns, coupled with estimates on the productivity of therapy, suggest that the returns to therapy do not outweigh the costs of the investment for most patients.

The primary objective of this work is to understand the various costs and informational

[4] In particular, several studies find cognitive therapy to be more efficacious than anti-depressants for patients with both major (Blackburn et al., 1981) and mild/moderate (Gloaguen et al., 1998; Hollon et al., 2005) depression. Additional studies have shown anti-depressants to be no more effective in treating mild/moderate depression than a placebo, while somewhat effective for seriously depressed patients (Kirsch et al., 2008; Fournier et al., 2010). A large number of studies support the efficacy of talk therapy (cognitive behavioral therapy in particular) for patients with varying severities of depression and/or anxiety (Fava et al., 2004; Hofmann and Smits, 2008; Stewart and Chambless, 2009; Hollon et al., 2014).

barriers that prevent patients from seeking out the most effective forms of mental health treatment. To do this, we embed the estimated mental health production function into a dynamic model of mental health treatment, labor supply decisions, and labor market outcomes. Previous literature has demonstrated strong links between mental health and labor market outcomes (Frank and Gertler, 1991; Ettner, Frank, and Kessler, 1997; Stewart et al., 2003). In the model, patients maximize lifetime utility by making treatment and employment decisions, while fully aware of the impact of treatment on future mental health. To rationalize reluctance to use therapy, the model incorporates several downsides, many of which are explicitly linked to employment. First, therapy is time consuming (Howard et al., 1986), which is particularly salient for individuals who work many hours, have a high opportunity cost of time, or have less flexible work schedules. Second, therapy is expensive (Frank, Busch, and Berndt, 1998). Third, therapy poses a mismatch risk in the sense that patients may need to try several therapists before finding one that is effective. Mismatch risk exacerbates already burdensome time and financial costs.[5] Finally, there is social stigma attached to mental illness (Satcher, 2000), which might make repeated personal interaction with a therapist uncomfortable for some patients. Medication, in contrast, is relatively low-cost in terms of time and money; is a fairly standardized product, which reduces uncertainty; and can be used in private, which helps to dispel stigma concerns.[6] Incorporating these factors into a dynamic choice model means we can ascertain to what degree policy-relevant factors make therapy a less attractive option. For example, distaste for therapy would be difficult to overcome via policy, but time constraints or work flexibility could be changed.

To illustrate the dynamic tradeoffs captured by the model, we consider how treatment choices could depend on an individual's level of labor market human capital. A high-earning individual may have a stronger incentive to invest in future health, which would add value to talk therapy. However, a high-earner also has a high opportunity cost of time, which could make medication relatively more attractive. The option that an individual ultimately chooses depends on several factors, including: the severity of the mental health condition, current earnings and employment, and the individual's stage in her life-cycle, which influences the dynamic returns to work experience.

We estimate the dynamic choice model using the 1996-2012 cohorts of the MEPS data, which apart from mental health treatments and conditions also contains rich data on labor supply and earnings. Model estimates suggest that in addition to reducing utility directly, mental illness reduces wage offers and employment, both through lower wages and a higher

[5] We credit Richard Frank with bringing this downside to our attention.

[6] A downside of medication is that it may have undesirable physical side effects (Khawam, Laurencic, and Malone, 2006), a point which we discuss further below. Presumably, side effects would make therapy a relatively more attractive option.

disutility of work. These consequences of mental illness create a strong incentive for patients to use the best mental health treatment available, which is talk therapy. Our estimates reveal that time costs, monetary costs, and mismatch are all important deterrents to therapy use. That said, reducing mismatch costs to zero generates significantly larger increases in therapy use than does reducing both time *and* monetary costs to zero. Beyond these factors, there is a large, negative utility cost associated with therapy, which captures not just preferences but all therapy costs not explicitly modeled, such as social stigma and lost leisure time. The costs of therapy are found to be the most salient for the least educated workers, who display the greatest price sensitivity and contend with the least flexible work schedules. This is true, even though low education workers are likely to be publicly insured, meaning they pay the lowest out-of-pocket price for therapy.

In studying mental health treatment choices, we contribute to a massive medical and public health literature on mental health. This literature includes well-developed scholarship on the determinants and consequences of mental health issues, the effectiveness of mental health treatment and predictors of mental health treatment choices. We do not provide an exhaustive review of this literature, but highlight some key results that we incorporate into our framework. Several papers have discussed that treatments are at least somewhat substitutable (Elkin et al., 1989; Berndt, Frank, and McGuire, 1997) and that consumers are price sensitive (Ellis, 1986; Frank and McGuire, 1986; Keeler, Manning, and Wells, 1988). Moreover, as drugs are less expensive under many insurance plans, another literature explores selection into insurance by the mentally ill (Sturm, Meredith, and Wells, 1996; Deb et al., 1996). Together, these results could help to explain widespread reliance on drugs. Others papers have examined how mental health, treatment and the labor market interact for both adolecents (Currie and Stabile, 2006; Fletcher and Wolfe, 2008; Fletcher, 2008, 2014) and for adults (Frank and Gertler, 1991; Ettner, Frank, and Kessler, 1997; Stewart et al., 2003; Greenberg et al., 2003). These analyses have utilized reduced form methods almost exclusively, limiting the potential for counterfactual welfare analysis. Moreover, few papers have discussed either mismatch or the time-cost of therapy. Our approach therefore incorporates many of the features from the mental health care literature, but adds some new features that could help to explain observed treatment choice decisions.

A second literature to which we contribute is economic research on medical treatments. Several papers have studied consumer choice under uncertainty (Crawford and Shum 2005b; Cronin 2016) in the context of medical treatment. Papageorge (2016) examines how side effects of HIV drugs reduce time in the labor market, which can incentivize patients to avoid effective treatments. Similarly, we study how various drawbacks of effective mental health treatment can lead patients to optimally choose less effective treatments. The difference is not only the context (mental health versus HIV), but also the source of product features that we see as drawbacks. Therapy is costly not because it causes side effects, but because it is time-intensive, fraught with uncertainty, and expensive. We incorporate these features into a decision model with the aim of evaluating policies that could mitigate these downsides and possibly shift treatment choices.

The paper proceeds as follows: Section 2 introduces the data used in this project. Section 3 describes estimation of the mental health production function. Section 4 discusses why patients tend to avoid therapy. Section 5 introduces the dynamic choice model. Section 6 discusses estimation, including identification and parameter estimates. Section 7 presents results from counterfactual policy simulations, which use the estimated dynamic choice model. Section 8 concludes.

2 Data Set and Summary Statistics

In this section we introduce the data set used in this project, which comes from the Medical Expenditure Panel Survey (MEPS). In doing so, we explain some of the terminology surrounding mental health used in the remainder of the paper. We also discuss the construction of the analytic sample and provide summary statistics.

2.1 MEPS Data Set

The MEPS is a nationally representative survey of families and individuals in the United States, collected by the Agency for Healthcare Research and Quality (AHRQ). A new cohort of individuals has been added to the MEPS annually since 1996, drawn randomly from the previous year's National Health Interview Survey (NHIS) sample. For our analysis, we use the 1996-2012 cohorts. Each individual in a cohort is interviewed five times over the course of two years with the time between interviews determined randomly at the individual level.

The MEPS data are well-suited for examining mental health treatment choices and health outcomes. The data set includes both multiple observations on the same individuals over time and observations on individuals across the age distribution. Detailed information is collected on treatment choices, including units consumed, out-of-pocket costs, total costs, and date of treatment. The MEPS data also contain measures of mental health. The most important measure of mental health that we extract from the data is subjective. In each interview, the individual is asked, "In general, would you say that (your) mental health is excellent, very good, good, fair, or poor?" Within our model, the decision to seek medical treatment is derived from the objective to improve this measure of mental health. We are also able to determine whether individuals possess specific mental illnesses (e.g., depression, schizophrenia, dementia, etc.) via three-digit ICD-9 Condition Codes, which are extracted during the interview process. Though these measures are not used in our main analysis, we use them in several robustness checks to show that subjective measures of mental health do indeed measure established symptoms of mental health conditions.

The MEPS data also contain rich information on demographics (e.g., education, age, race, gender, and location) and labor market choices and outcomes, such as hours worked, wages, and occupation. This information allows us to examine how mental health and treatment decisions relate to employment, hours, and wages. Moreover, MEPS restricted-use data allow us to identify an individual's county of residence, which enables us to merge in detailed information on the supply of medical services in an individual's location from the Area Health Resource File. Such information includes, for example, the number of doctors and psychiatrists per capita. This information helps to identify the causal effects of mental health treatment in a manner to be explained below.

2.2 Estimation Sample Construction

The estimation sample begins with individuals from the 1996-2012 cohorts who are interviewed in each of five possible rounds over a two year period. We then restrict the sample to individuals between the ages of 22 and 64, which reflects our focus on individuals for whom education is unlikely to change and who are making decisions with respect to the labor market. As lagged mental health cannot be observed in the first interview, only rounds two through five can be used in most of our analysis.

In each round, MEPS participants answer questions related to behaviors and outcomes occurring since the most recent interview. These interview periods vary in length - on average, they are about 5.4 months long and approximately 85% are between 3 and 8 months long. Figure 1 shows the distribution of period lengths, rounded to the nearest half-month intervals. Period length was randomly allocated as a part of the survey design. The estimation of our structural model requires that each interview period covers an approximately equal amount of time; thus, we eliminate observations where the length of time between interviews is less than 3 months or greater than 8 months. To avoid needing to integrate over missing time periods in the estimation of the structural model, we use the following process to eliminate individuals and observations from the data: (1) drop any observation where length is less than 3 months; (2) drop any observation where length is greater than 8 months; and (3) drop any individual whose 2nd, 3rd, or 4th interview is dropped in (1) or

(2). Finally, we also exclude those who have been diagnosed with a severe mental disorder per the ICD-9 codes (290, 296-299, 307, 309).

Given these restrictions, the final analytic sample consists of 98,056 individuals and 376,234 individual-period dyads. Table 1 details how sample size changes with each restriction imposed on the data.

2.3 Summary Statistics

Tables 2, 3, and 4 provide summary statistics related to mental health and treatment choice. Table 2 shows how mental health and treatment decisions evolve over the lifecycle. On average, as individuals age, subjective mental health worsens and depression and anxiety become more prevalent. This suggests that subjective mental health captures recognized symptoms of anxiety and depression, as indicated by ICD-9 codes. Moreover, use of mental health treatment rises with age. Table 2 also highlights that medication is a much more popular treatment choice than talk therapy at every age, and therapy becomes relatively less popular as individuals age.

Table 3 presents sample means for demographic and labor market variables by treatment choice. The statistics indicate that those individuals who use therapy are younger, more likely to live in a metropolitan statistical area (MSA), and are more highly educated than those who use medication. That those who use therapy are younger and more educated relative to those who use medication supports the possibility that individuals see therapy as an investment in future mental health. Of course, these are unconditional means and those who are younger and more educated differ in many dimensions from those who are older and less educated. Further, those who are more educated may be more likely to take therapy for other reasons. For example, therapy may be more productive for those who are more verbal and those who are more educated may have more flexible work schedules. Finally, subjective mental health is worse for individuals in treatment in comparison to those who are not.

Finally, Table 4 presents sample means by level of subjective mental health, ranging from excellent mental health, $MH = 5$, to poor mental health, $MH = 1$. Those with poor mental health are more likely to be female, older, less educated, and have lower wages relative to those with better mental health. Again, as expected, worse subjective mental health is associated with depression, anxiety, and the use of treatment.

3 The Mental Health Production Function

In this section, we describe the estimation of a production function for mental health. The structural choice model treats individuals as choosing from a menu of medical treatments in an effort to maximize their lifetime utility. One component of lifetime utility is an individual's stock of mental health, which is potentially valuable on its own, but may also generate utility through its impact on other outcomes, such as employment or productivity at work (Grossman, 1972). To capture how treatment choices reflect incentives to improve mental health, an important component of the dynamic choice model is a mental health production function. The function maps treatment choices to mental health outcomes. Identifying the effects of treatment requires overcoming potential bias due to non-random selection into treatment. As we explain below, we use panel data along with an instrumental variables strategy to recover estimates of the causal impact of mental health treatment on mental health outcomes. Estimates reveal that therapy is more productive than medication, which is more productive than no treatment at all. We begin with findings from OLS regressions relating treatment choices to mental health outcomes. We also discuss possible selection problems that preclude assigning a causal interpretation to estimated parameters. Next, we proceed to 2SLS estimates, in which we use instrumental variables (IVs) to overcome endogeneity issues.

Prior to discussing production function results, note one complication involving the use of panel data in this setting is that self-assessed mental health is reported during an interview, which often takes place in the middle of a psychotherapeutic treatment episode (i.e., a sequence of consecutive months in which an individual consumes psycotherapy). Given that our analysis focuses on the effect of *any* therapy, we are compelled to choose whether to code these treatments as one distinct treatment episode, occurring before or after the interview, or two distinct treatment episodes, occurring both before and after the interview. With respect to the structural model, the former may be more intuitive, as only one extensive margin decision is being made about whether to visit a therapist. With respect to the mental health production function, the latter may be more intuitive, as consuming one or two sessions of a longer treatment episode prior to an interview could have a small effect on the evolution of one's mental health. For consistency across our analysis, we have decided on the latter. Formally, if an episode of therapy sessions spans two interview periods, then the individual is coded as having chosen therapy for two consecutive periods.[7] Results from the alternative specification, which are similar to those presented here, are available upon request from the

[7]Because prescription drugs are typically purchased on one day and consumed over the month that follows, it is reasonable to assume that each refill represents a separate decision made by the individual. As such, we code the consumption of any prescription drugs in the sample period observed in the data.

authors.

3.1 Ordinary Least Squares Estimates

Column 1 of Table 5 contains parameter estimates from a linear model where self-reported mental health status is regressed on mental health treatment variables, lagged mental health, demographic characteristics, and county and time fixed effects.[8] The results suggest that both therapy and pills *worsen* mental health - a likely indicator of selection bias. As seen in Table 4, individuals in the worst mental health states are most likely to consume medical care. Controlling for lagged mental health does not correct this problem because while mental health is reported at each interview, treatment is consumed between interviews. Thus, an individual may receive a negative health shock between the two interview periods, which leads them to both (i) consume medical care and (ii) end up in a worse mental health state. Controlling for this type of selection is a key challenge in our paper, as well as many others (Lu, 1999; Blau and Gilleskie, 2008; Cronin, 2016). We solve the selection problem using an instrumental variables approach.

3.2 Two-Stage Least Squares Estimates

The instrumental variables strategy requires a minimum of two instruments that (i) alter mental health treatment decisions (i.e., instruments are not weak) and (ii) have no *direct* effect on mental health (i.e., instruments are exogenous). The first instrument that we consider is the number of psychiatrists per capita in an individual's county of residence. This information can be found in the Area Health Resource File (AHRF), which is collected annually by the US Department of Health and Human Services, for every year between 1995 and 2016, except for 2008. There is substantial variation in the number of psychiatrists per capita across the sample - nearly 10% of individuals live in a county without any psychiatrists, the average individual lives in a county with 1.3 psychiatrists per 10,000 people, and the individual at the 90^{th} percentile lives in a county with 2.5 psychiatrists per 10,000 people.

[8] In our production function analysis, we further restrict the estimation sample discussed in Section 2.2 to include only individuals with private insurance. This restriction strengthens the first stage effect of one of our instruments (i.e., number of psychiatrists per capita) and, thus, the precision of our 2SLS estimate. A separate analysis of publicly insured and uninsured individuals in our estimation sample reveals that their treatment decisions are not responsive to changes in the instrument - these results are available upon request. Many private practice psychiatrists do not accept Medicaid patients (Taube, Goldman, and Salkever, 1990), which comprises nearly all of the publicly insured individuals in our estimation sample. Furthermore, according to our data, the uninsured are simply very unlikely to consume any mental health treatment, making the supply of psychiatrists mostly irrelevant for them. We also drop counties in the bottom 10th percentile of total observations.

Unsurprisingly, this variable is highly persistent over time - regressing the variable on county fixed effects produces an R-squared of 0.97, suggesting that just 3% of the overall variation in psychiatrists per capita is due to within county variation. Because these county fixed effects are included in our 2SLS specification, identification will come from these within-county changes in the number of psychiatrists per capita, which we argue is conditionally random.

A second instrument we consider is an indicator for whether the individual's county of residence has a Walmart with a pharmacy *and* the survey period ends in 2007 or later.[9] On September 21, 2006, Walmart began offering almost 300 generic prescriptions at a price of $4 for a monthly supply at it's stores in Tampa Florida.[10] Initially, Walmart planned to expand the offering to all Florida stores in January of the following year; however, by November 27, 2006, Walmart had expanded the policy to all of its US stores. In a 2006 company newsletter, (then) Executive VP of Professional Services, Bill Simon, explained that, "many customers have greatly benefited from the savings and consumer demand has been a significant factor in the program's expansion." According to the AARP, the average *annual* retail cost of prescription drug therapy for a basket of 280 popular generics in 2006 was $391 (i.e., roughly $33 for a monthly prescription). This suggests that Walmart's offering of $4 monthly prescriptions could represent significant cost savings for individuals and, thus, increase the quantity of drugs demanded. 90% of our sample lives in a county with a Walmart and, therefore, had access to these low cost drugs.

Our first stage results can be found in Table 6. All models control for county and year fixed effects as well as lagged mental health and a robust set of demographic controls. Column 1 displays the relationship between our instruments and whether an individual consumes any therapy. The estimates reveal that the number of psychiatrists per capita significantly increases therapy use, while having access to low cost generic prescriptions via Walmart has no significant effect. Column 2 displays the relationship between our instruments and whether an individual consumes any prescription drugs for mental illness. The estimates reveal that both psychiatrists per capital and low cost generic prescriptions through Walmart significantly increase prescription drug use.

Weak instruments can produce biased, inconsistent 2SLS estimates (Bound, Jaeger, and Baker, 1995). In the standard one-instrument, one-endogenous variable setting, it is generally accepted that the instrument is adequately strong if it's F-statistic is greater than 10, which corresponds to a bias in the 2SLS estimate that is less than (approximately) 10% of the bias

[9]We purchased data from AggData containing information on the 4,618 Walmart stores operating in the US in 2016, including opening dates and whether a store has a pharmacy. These data do not contain information an Walmart closures.

[10]On this list are roughly 28 medications used in the treatment of mental health, including Fluoxetine (Prozac), Citalopram (Celexa), and Paroxetine (Paxil), all popular anti-depressants.

in the OLS estimate. With multiple instruments and endogenous variables, the joint F-test, conducted on the instrument set in each first stage equation, tests the null hypothesis of no correlation between the instruments and the endogenous variables against the alternative of correlation. This information is valuable, however, rejection of the null does not guarantee that there is sufficient variation in the instruments to identify the model. For example, in a two-instrument, two-endogenous variable setting, it is possible that only one instrument explains variation in the two endogenous variables, which can generate large F-statistics, but an underidentified model. Kleibergen and Paap (2006) develop a Lagrange Multiplier (LM) statistic for this scenario, which allows for a test of the null hypothesis that the rank of the instrument set is greater than the number of endogenous variables minus one (i.e., that the model is underidentified). Moreover, using critical values provided by Stock and Yogo (2005), the Kleibergen-Paap rk Wald F-statistic can be used to test the null hypothesis that the instrument set is weak *if* the bias of the 2SLS estimator, relative to OLS, could exceed a certain threshold, such as 10%.[11] Table 6 provides traditional joint F-statistics, as well as Kleibergen-Paap LM and rk Wald F-statistics for each of the models presented.

While the instruments presented in our first specification (Columns 1 and 2) significantly alter treatment decisions, the instruments set is weak. Moreover, with just two instruments and two endogenous variables, we cannot test of the exogeneity of our instruments. In Columns 3 and 4 of Table 6, we present our preferred instrument set, which contains interactions of original instruments with several demographic variables.[12] In Column 3, the presence of psychiatrists significantly increases the use of therapy for previously married and white individuals. Access to low cost drugs through Walmart decreases therapy use, presumably as individuals substitute therapy for prescription drugs. In Column 4, the presence of psychiatrists increases prescription drug use, but for males only, as does the presence of a Walmart after the generic drug price drop. The Kleinberger-Paap LM statistic allows us to reject the null of underidentification at a 7% significance level, while the Kleinberger-Paap rk Wald F-statistic allows us to reject the null hypothesis that our instruments are weak, as defined as a 2SLS to OLS relative bias of greater than 10%.[13]

[11] The Kleibergen-Paap LM and rk Wald F-statistics are cluster-robust alternatives to similar statistics provided in Cragg and Donald (1993), which can only be applied linear models with i.i.d. random errors.

[12] Note that once interactions are added the uninteracted *number of psychiatrists per capita* does not have a significant impact on prescription drug decisions and has only a mildly significant impact on therapy use; thus, to strengthen the instrument set, the variable is included in both stages of the model.

[13] Stock and Yogo (2005) establish critical values for testing whether a set of instruments is weak in a setting with multiple endogenous variables. An F-statistic smaller than the critical value implies that the instruments are sufficiently weak, such that the bias in the 2SLS estimate is greater than a particular percentage of the bias in the OLS estimate. With two endogenous variables and four instruments (relevant for Specifications 2A and 2B), the critical values corresponding to 5, 10, 20 and 30% relative bias are 11.04, 7.56, 5.57, and 4.73, respectively. Thus, for the preferred instrument set, we can reject the null that the

Column 2 of Table 5 contains parameter estimates from our 2SLS specification. The first two rows show that our identification strategy has the desired effect. Both pills and therapy are found to be effective in improving an individual's mental health. Moreover, consistent with the medical literature cited above, therapy is found to have a larger positive effect than prescription drugs.[14] Because our model is over-identified, we are also able to conduct a Hausman J test, which tests the assumption that our instruments are exogenous. This test statistic, which is Chi-squared with 2 degrees of freedom, is 2.924 (p-value 0.233). Thus, we fail to reject the null hypothesis that the instruments are exogenous, which supports our identifying assumptions.

4 Explaining Mental Health Treatment Choices

Estimates from the previous section provide evidence that therapy is more productive than other treatment options. Yet, as alluded to earlier, therapy use is low. In this section, we examine treatment choices more closely. To begin, Section 4.1 provides evidence that mental health is associated with higher rates of employment and higher earnings. This suggests a strong incentive for use of the most effective treatment to improve mental health. However, on average, patients tend to forgo therapy. Widespread reluctance to use therapy suggests that there are important costs. Section 4.2 explores potential costs that help to explain why patients forgo therapy. These costs are incorporated into the dynamic choice model specified in Section 5.

4.1 The Benefits of Mental Health in the Labor Market

Existing studies have established that those suffering from mental illness are less likely to work at all or to miss work and also earn lower wages conditional on working (Bartel and Taubman, 1986; Ettner, Frank, and Kessler, 1997; Druss, Schlesinger, and Allen, 2001; Stewart et al., 2003). These relationships are evident in the MEPS data as well. Table 7 contains regressions of employment, (log) hourly wages, and hours worked on mental health, both without (Columns 1-3) and with (Columns 4-6) individual and time fixed effects. Across specifications, we find that worse mental health is associated with a lower likelihood of em-

instruments are weak at a 10% acceptable relative bias level, which would correspond to a standard joint F-statistic above 10 in a one-instrument, one-endogenous variable setting.

[14] The parameters on any medication and any therapy have p-values of 0.03 and 0.11, respectively. Note that these are not our final, preferred production function estimates. These estimates are presented in Column 3 and are discussed in Section 4.2.1. The corresponding p-values for this specification are 0.03 and 0.09, respectively.

ployment and hours worked, conditional on employment. For hourly wages, effects are no longer significant after including individual and time fixed effects. Of course, significant effects on hours worked imply that total earnings will be impacted even if there are no effects on productivity. By estimating models with individual fixed effects, we confirm that the relationship for employment holds even when the variation in mental health is limited to that within an individual - i.e., negative "shocks" to mental health seem to have immediate consequences for one's employment and hours worked. These findings suggest incentives to invest in mental health that go beyond feeling better, but also extend to other economic outcomes, including employment and income.

4.2 Why Do Patients Forgo Therapy?

Given that therapy produces higher levels of mental health and, moreover, that mental health has benefits in the labor market, a natural question is why patients tend to forgo therapy. We focus on three reasons: therapist mismatch, the time costs of therapy and the monetary costs of therapy.

4.2.1 Therapist Mismatch

A striking feature of the data is a large mass of individuals who attend therapy only once or twice before stopping treatment To show this, we define a therapy *treatment episode* as a consecutive sequence of therapy sessions occurring without a two-month gap in visits. Figure 4 contains a histogram of the number of therapy visits within each treatment episode. Notice that about 40% of these treatment episodes contain only one or two visits, meaning one or two therapy sessions are attended without any sessions attended in the two months preceding or following these visits. Many of those who consume therapy at some point during the two-year survey period (3,990 individuals) are *only* observed to consume these very short treatment episodes (1,315 individuals). Relative to other therapy users, those who only consume short therapy episodes are predominantly from the south, are less educated, earn lower incomes, are less likely to live in an MSA, and have slightly better self-reported mental health.

We assume that treatment episodes containing only one or two visits represent a *mismatch*, which could either mean (i) that the individual is an inexperienced therapy user and, upon their initial visit, learns that they dislike this type of treatment and quits or (ii) that the individual, experienced or inexperienced, visits a new therapist that happens to be a bad match, leading them to quit treatment. Unfortunately, we are not able to see the identity of the therapist that an individual visits and we have limited information on an individuals history of therapy use; thus, distinguishing type (i) and type (ii) individuals from those who consciously visit a therapist every 3-4 months is difficult with our data. That said, we are able to provide evidence that these mismatch episodes do not improve mental health on average using an alternative specification of our mental health production function. In Column 3 of Table 5, we provide 2SLS estimates of the impact of prescription drug and therapy treatment on mental health, where mismatch therapy visits are coded as if therapy was not attended (i.e., *Any Therapy*=0). These estimates lead to an increase in the effect of therapy on mental health, suggesting that mismatch visits are not efficacious.[15] As such, we believe that these sessions generally represent costly, though non-productive medical care consumption that occurs primarily due to a lack of information.

4.2.2 Therapy and Hours Worked

Next, we explore the relationship between hours worked and employment, therapy use, and level of education. Table 8 presents results from a regression of hours worked on therapy sessions attended and interactions of therapy with education with individual and time fixed effects. The models condition on subjective mental health, an indicator of living in an MSA, geographic region, hourly wages, and demographic characteristics. The sample is restricted to employed workers. The results suggest that using therapy is associated with fewer hours worked and that this association is the strongest for those with low levels of education. Note that by controlling for subjective mental health and individual fixed effects, we insure that within-individual changes in therapy use, conditional on changes in mental health, are leading individual to work fewer hours - i.e., the reduction in hours is not a response to illness shocks These results also provide interpretable magnitudes. *Each* therapy session attended in a week is associated with approximately 0.25 fewer hours per week on average for those with 16 years or more of education, 0.65 fewer hours per week for those with between 12 and 15 years of education, and 2.25 fewer hours per week for those with less than 12 years of education.[16]

These patterns suggest the possibility of variation in work-time costs across education groups for attending therapy. This variation could arise for a number of reasons, including relatively limited availability of lower-cost therapists. It might also reflect differences by education group in work-time flexibility, the idea being that workers with lower levels of

[15] Note also that the impact of therapy on mental health is more significant in this specification, 2SLS-B. First stage results are presented in Column 5 of Table 6.

[16] These numbers ignore the quadratic term on sessions, which suggests that the impact of each additional session on hours is diminishing with number of sessions attended.

education tend to have jobs that are less flexible so that seeing a therapist results in fewer hours worked. More flexible workers may be able to make up missed work hours at home or later in the day to avoid lost wages. Heterogeneity in the costs of therapy for different education levels also suggests that lower-earning workers will be of particular concern when considering mental health policies. Here, the reasoning is that work flexibility does not often extend to such workers. Thus, if the goal is to improve the mental health of all workers, and not just those who are highly educated, policy interventions that address differences in work-time flexibility in light of time costs of therapy could play an important role.

4.2.3 Monetary Costs

The third reason that individuals may avoid therapy is that it is relatively expensive. In Table 9 we provide summary statistics of the monetary costs associated with therapy and pharmaceuticals for an interview period across insurance groups. Note that the therapy costs presented in the table exclude mismatch visits. For individuals who are uninsured or have private insurance, which is the majority of the sample, therapy is more expensive on average. For public insurance, medication is more expensive on average. We now turn to the specification of the model.

5 Dynamic Model

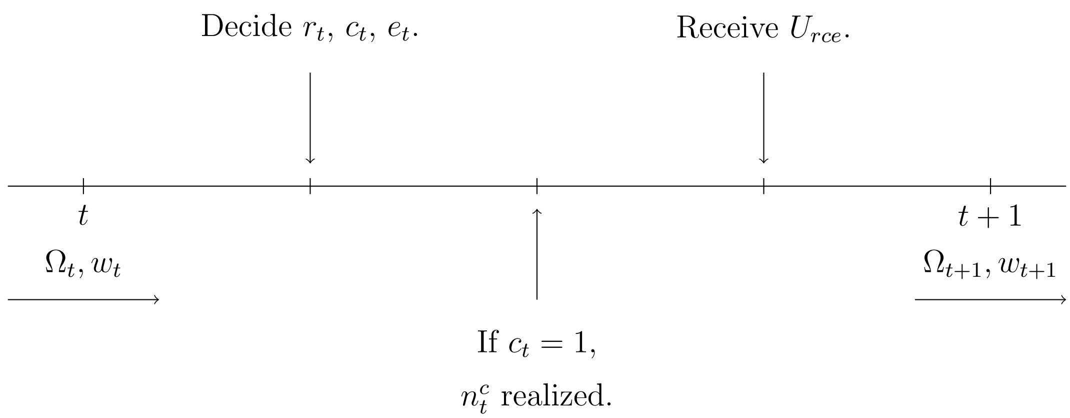
An individual decides among treatment and labor supply alternatives considering both the contemporaneous and the expected future utility associated with each alternative.[17] Current treatment decisions impact the distribution from which mental health is drawn in the next period, while labor supply decisions determine an individual's accumulation of work experience over time. The model has a finite time horizon with discrete time periods and discrete treatment and labor supply alternatives. The average interview period length in the estimation sample is 5.4 months. For simplicity, we assume that all decision periods in the model are six months in length. An individual makes decisions from the ages of 25 to 60, resulting in 72 total decision periods. In the terminal period, T, the individual receives a continuation value that depends on her terminal mental health and human capital.

The timing within a period is summarized in the figure below. Entering period t, the indi-

[17] We acknowledge the important role that physicians play as advisors, and potential gatekeepers, in treatment choices. Unfortunately, unlike Dickstein (2014) our data do not allow us to separately identify the incentives faced and choices made by patients and physicians. Thus, while we describe in this section an optimization problem solved by an individual, the true data generating process is determined by a joint patient-physician optimization problem, and our estimates for treatment preferences will reflect this.

vidual observes her state vector, Ω_t, which consists of current mental health state, education, work experience, demographic characteristics, including age, race, and sex, and characteristics of the county and state of residence. Before making her decision, the individual also observes a draw from her distribution of hourly wage offers, w_t. Given this information, she decides whether or not to use any medication, r_t, and/or any talk therapy, c_t, and also whether to work full time ($e_t = 2$), part time ($e_t = 1$), or not at all ($e_t = 0$). If talk therapy is chosen, she then realizes a draw that determines whether or not she experiences a therapist mismatch, n_t^c. A mismatch requires that the individual incur the monetary and time costs associated with a therapy visit, but yields no benefits in the form of improved mental health.

Figure 1: Timing



The total number of therapy visits that occur within a decision period, in reality, accumulate as the solution to an intra-period problem. A full exposition of this intra-period problem is available upon request; however, the problem can be understood as follows: on day 1, an individual decides whether to attend therapy or not; on day 2, an inexperienced therapy user can decide to visit for the first time or not, while an experienced user must decide whether to continue going to therapy or to stop; this continues to the end of the period. In such a problem, there are several reasons why an individual may start or stop attending therapy. The simplest reason is that they realize a positive/negative utility draw for therapy on any given day. Another reason, particularly for patients new to therapy, is that they may learn about the productivity, monetary costs, or time cost of the treatment. Ideally, we would estimate this intra-period problem in an effort to distinguish how these factors contribute to patient choice; however, the data are not rich enough to estimate such a model. As such, our approach focuses on adequately representing an individual's expectations when making the initial decision of whether or not to attend any therapy. We assume that this initial decision us undertaken with some uncertainty as to how many visits will actually occur, a feature of the intra-period problem. As discussed in Section 4.2.1, we believe that the primary source of uncertainty is due to a "mismatch risk" but recognize that the observed tendency for individuals to quit therapy could also be explained by the above described features of the intra-period problem. As a result of this modeling approach, our simulations speak primarily to effects of policies on the extensive, rather than the intensive, margin for therapy.

5.1 State Vector and State Transitions

Human capital updates deterministically. Part time and full time experience, $K_{1,t}$ and $K_{2,t}$, have an initial value of zero and increase by one each period that the individual decides to be employed in the respective market. Education, E_t, is predetermined entering the first period. Possible education levels are:

$$E_t = \begin{cases} 1 & \text{less than high school} \\ 2 & \text{high school degree} \\ 3 & \text{college degree or more} \end{cases}$$

Mental health takes a discrete, integer value from 1 to 5. M^*_{t+1} is a latent, continuous measure of mental health which depends on an individual's treatment choices in period t, her level of education, and demographic characteristics. The latent variable M^*_{t+1} takes the following form:

$$M^*_{t+1} = \delta_1 \sum_{m=1}^{5} \mathbb{1}[M_t = m] + \delta_2 r_t + \delta_3 \mathbb{1}[c_t = 1 \text{ and } n^c_t = 0] + \epsilon^M_t \tag{1}$$

where ϵ^M_t is drawn from a normal distribution with mean zero and a standard deviation that is estimated within the model. Consumption of medication (through δ_2) and therapy sessions (through δ_3) has direct productive effects on mental health. Mismatch sessions are assumed to have no productive effects on mental health.[18]

The probability of mismatch is allowed to vary by an individual's sex and race and is determined by a linear probability model.

[18] In future iterations of this paper, we will estimate the model and perform counterfactual simulations across the range of possible productivities of mismatch sessions, from mismatch having no productive effect to having the same effect as therapy without mismatch. That mismatching is less productive than not mismatching is supported by the fact that estimating the mental health production function including mismatch in the indicator for any therapy use results in a significantly lower productivity of therapy.

5.2 Preferences

U_{rce} is the flow utility associated with the alternatives $r_t = r$, $c_t = c$, and $e_t = e$. Preference shocks associated with each of the twelve combinations of r_t, c_t, and e_t, denoted ϵ_t^{rce}, are realized at the beginning of each period.

$$
\begin{aligned}
U_{rce} =& \frac{C_t^{1-\theta} - 1}{1-\theta} + \alpha_0 r_t + \alpha_1 c_t + M_t(\alpha_2 + \sum_{e=1}^{2} \alpha_{3,e} \mathbb{1}[e_t = e]) \\
&+ \sum_{e=1}^{2} \alpha_{4,e} \mathbb{1}[e_t = e] + \epsilon_t^{rce}
\end{aligned} \tag{2}
$$

C_t represents consumption of a composite good which is determined by the budget constraint, described below, and θ is a constant relative risk aversion (CRRA) parameter. The CRRA form for the utility function allows for the pecuniary costs of therapy visits and lost income due to time costs to have different effects on utility across the consumption distribution. Preference parameters on prescription drug use, α_0, and talk therapy use, α_1, capture preferences for these treatment alternatives net of budget and time costs. Mental health directly impacts the current period flow utility via α_2. Leisure preferences are captured by $\alpha_{4,1}$ and $\alpha_{4,2}$. The parameters $\alpha_{3,1}$ and $\alpha_{3,2}$ allow for employment to result in more disutility as mental health worsens; these effects of poor mental health on the desirability of being employed are one of the primary costs of poor mental health in the model.

5.3 Budget and Time Constraint

Consumption in a period is constrained as follows:

$$
C_t = \lambda(w_t h_t 26) - \delta_1 r_t - \delta_2 c_t \tag{3}
$$

where λ is the share of family income the individual consumes, w_t is the hourly wage offer, h_t is weekly hours worked, δ_1 is the price of medication, and δ_2 is the price for the median number of sessions when mismatch does not occur.[19] Individuals who mismatch pay the monetary price for one session. Weekly earnings are multiplied by 26 because there are 26 weeks in each 6-month period.

Hourly wage offers are drawn from a log-normal distribution that depends on education, part time and full time work experience, current mental health, and demographic character-

[19]Currently, treatment prices are based on the sample average out-of-pocket costs per round. In future iterations of the model we may estimate the price distribution and allow out-of-pocket costs to vary at the individual level.

istics, including age, race, and sex:

$$\begin{aligned} ln(w_t) = \gamma_0 + \sum_{k=1}^{3} \gamma_{1,k} \mathbb{1}[E_t = k] + \gamma_2 K_{1,t} + \gamma_3 K_{2,t} \\ + \gamma_4 M_t + \gamma_5 D_t + \epsilon_t^w \end{aligned} \tag{4}$$

$$\epsilon_t^w \sim N(0, \sigma_w^2)$$

Weekly hours worked are determined by the following time constraint:

$$\begin{aligned} h_t =& 20 \times \mathbb{1}[e_1 = 1] + 40 \times \mathbb{1}[e_t = 2] - \mathbb{1}[c_t = 1 \text{ and } n_t^c = 0] N_n (\lambda_0 + \sum_{i=2}^{3} \lambda_i \mathbb{1}[E_t = i]) \\ & - \mathbb{1}[c_t = 1 \text{ and } n_t^c = 1](\lambda_0 + \sum_{i=2}^{3} \lambda_i \mathbb{1}[E_t = i]) \end{aligned} \tag{5}$$

Individuals who work part time have a base of 20 hours per week, while those who work full time have a base of 40 hours. The parameters λ_0 through λ_3 represent the time costs of therapy sessions. N_n is the median number of therapy sessions for individuals who attend therapy and do not mismatch; those who mismatch lose the number of work hours associated with attending one session. The time costs of medication are normalized to zero. All employed individuals lose some work time from using therapy. However, interactions with level of education allow for these time costs to vary across education level. These interactions are motivated by the results presented in Subsection 4.2.2.

5.4 The Optimization Problem

The individual's objective is to maximize her expected discounted lifetime utility. The individual makes decisions for T periods and then receives the terminal value V_f, which is a linear function of terminal mental health, terminal part time experience, and terminal full time experience. Let $V_{rce}(\cdot_t)$ denote the expected lifetime utility associated with choosing alternative $r_t = r$, $c_t = c$, and $e_t = e$ at the beginning of time t. $V_{rce}(\cdot_t)$ can be written recursively as the sum of contemporaneous utility and expected future utility associated with that alternative:[20]

$$
\begin{aligned}
V_{rce}(\Omega_t, w_t, \epsilon_t^{rce}) = \mathbb{1}[c{=}1] \Bigg[\sum_{n=0}^{1} P(n_t^c{=}n|\Omega_t, \Phi_t) \Bigg(U_{rce}(r_t, n_t^c, e_t; \Omega_t) \\
+ \beta \sum_{k=1}^{5} P(M_{t+1}{=}k|\Omega_t, r_t{=}r, n_t^c{=}n) \int_{w_{t+1}} EV(\Omega_{t+1}, w_{t+1}) f(w_{t+1}) dw_{t+1} \Bigg) \Bigg] \\
+ \mathbb{1}[c{=}0] \Bigg[U_{rce}(r_t, 0, e_t; \Omega_t) + \beta \sum_{k=1}^{5} P(M_{t+1}{=}k|\cdot) \int_{w_{t+1}} EV(\Omega_{t+1}, w_{t+1}) f(w_{t+1}) dw_{t+1} \Bigg]
\end{aligned}
\tag{6}
$$

When making her decision, the individual must integrate over the distribution of n_t^c, the distribution of M_{t+1}, and the distribution of future wage offers to calculate expected future utility. β is the discount factor. $EV(\Omega_{t+1}, w_{t+1})$ is the expected maximal $V_{rce}(\cdot_{t+1})$, where the expectation is over ϵ_{t+1}^{rce}:

$$
EV(\Omega_{t+1}, w_{t+1}) = E_{\epsilon^{rce}}[\max_{rce} V_{rce}(\Omega_{t+1}, w_{t+1}, \epsilon_{t+1}^{rce})] \tag{7}
$$

The model can be solved using backwards recursion. Starting in the terminal period, the individual can calculate the deterministic value function associated with each combination of r_t, c_t, and e_t for each possible state. Taking expectations over ϵ_t^{rce} allows her to calculate $V(\Omega_T)$ for each Ω_T. The collection of $V(\Omega_T)$ allows her to calculate continuation payoffs for any combination of r, c, and e in period $T-1$ for any state. Hence, she can make the same calculations for $T-1$ and continue working backwards to the first period. The econometrician can solve the model in a similar manner for choice probabilities associated with each time period and state, as discussed below, which can then be matched to the choices observed in the data.

6 Estimation

The structural parameters of the dynamic model are estimated using a nested algorithm. In an inner algorithm, the model is solved using backwards recursion at a given set of parameters. The outer algorithm uses the model solution to calculate the likelihood function (below) and updates the parameter vector using the Berndt, Hall, Hall, and Hausman (1974) (BHHH) algorithm. We estimate 30 parameters using this nested algorithm: 9 preference parameters, the CRRA parameter, 14 parameters in the wage offer function, and 6 parame-

[20] In period T, the expected future utility in this equation is replaced with V_f. Also, whenever mismatch occurs ($\Phi_t = 1$), the number of therapy sessions is zero ($n_t^c = 0$). That is, $P(n_t^c{=}0|\Omega_t, \Phi_t = 1) = 1$.

ters in the terminal value function. As discussed in Sections 3 and 6.3.1, the productivity of mental health treatments is identified and estimated outside of the model using geographic and time variation in psychiatrists per capita and the presence of a Walmart with a pharmacy post-2006. Currently, time costs of therapy are taken from a regression of hours worked on therapy sessions detailed in subsection 4.2.2. The probability of mismatch and treatment intensity are likewise estimated outside of the model using a logit specification.

6.1 Human capital approximation

Recall that the MEPS surveys individuals at different points in the life-cycle, and each individual is followed for two years. MEPS does not include questions about work history, which means that part-time and full-time work experience are not observed. Solving the model requires calculating the expected maximal value function (Equation 7) for each point in the state space. Hence, it is necessary to approximate work experience for each individual in our sample.[21]

To approximate an individual's work experience we first calculate the proportion of the sample at each age, sex, and level of subjective mental health that is working part-time, the proportion that is working full-time, and the proportion that is not working. Using these proportions as the probability of working part-time or full-time given an age, sex, and level of mental health, we simulate an employment status for each age from 22 to 65 for each individual. The simulated work history allows us to calculate part-time and full-time experience entering the first period for each individual in the sample. Figure 5 depicts the simulated part-time and full-time experience profiles for the sample by level of subjective mental health. The figure shows that, using the approximation method, there is a considerable gap in the accumulation of full-time experience across levels of mental health. By age 40, those with poor mental health have approximately five fewer years of full-time experience on average than those with excellent mental health. Differences in part time experience are not as substantial. It is possible that those with poor mental health substitute part-time work for full-time work in some cases. Late in the life-cycle it does appear that those with worse mental health have accumulated slightly less part-time experience.

[21]Ultimately, we plan to use an additional source of data that includes work histories by level of mental health and to have our estimation procedure match the moments in that second source. The National Comorbidity Survey is one data source that includes this information.

6.2 Likelihood function

In each time period, an individual's contribution to the likelihood function includes the probability of her observed choice of treatment alternatives and labor supply given her state. Individuals who are working also contribute through their observed wage. Given that county fixed effects are used in the estimation of the mental health production function, but do not enter the model,[22] it is also necessary to estimate the constant and variance term for the mental health transition function within the model, though all other parameters of the mental health production function are estimated outside the model. Hence, the probability of observing an individual's reported level of mental health given her state is also included in the likelihood function.

The time-varying preference shock ϵ_t^{rce} is assumed to be distributed Type I Extreme Value, which results in the expected maximal value function, $V(\Omega_{t+1}, w_{t+1})$, taking the following closed form:

$$V(\Omega_{t+1}, w_{t+1}) = \gamma + \log\left(\sum_{r=0}^{1}\sum_{c=0}^{1}\sum_{e=0}^{2} exp(\overline{V}_{rce}(\Omega_{t+1}, w_{t+1}))\right) \tag{8}$$

where γ is Euler's constant and $\overline{V}_{rce}(\Omega_{t+1}, w_{t+1})$ is the deterministic portion of the alternative specific value function, $V_{rce}(\Omega_{t+1}, w_{t+1}, \epsilon_{t+1}^{rce})$. The assumption of a Type I Extreme Value and additively separable preference shock also yields the following choice probabilities:

$$P(d_t^{rce} = 1|\cdot_t) = \frac{exp(\overline{V}_{rce}(\Omega_t, w_t))}{\sum\limits_{r=0}^{1}\sum\limits_{c=0}^{1} exp(\overline{V}_{rce}(\Omega_t, w_t)} \tag{9}$$

As it is assumed that individuals observe a wage offer each period, the probability of choosing any combination of r_t, c_t, and e_t must be integrated over the distribution of wage offers for those who are not working. It is also necessary to integrate over the distribution of future wage offers and future amounts of other family income to calculate expected future utility. All of these integrations are simulated using 25 draws from a Halton sequence.

Let the indicator d_t^{rce} take a value of one whenever $r_t = r$, $c_t = c$, and $e_t = e$, and zero otherwise. Also let f_w represent the probability density function for the disitribution of wage offers. Then, for individual i in time period t, the likelihood contribution at a set of

[22]County of residence can only be observed for MEPS participants within a Research Data Center (RDC). Given our limited access to processors on the RDC server, we have chosen to estimate the full structural model, which utilizes OpenMP parallel processing software, outside of the RDC

parameters Θ is:

$$L_{i,t}(\Theta|\Omega_t) = \left(\prod_{r=0}^{1}\prod_{c=0}^{1}\prod_{e=0}^{2} P(d_t^{rce} = 1|\Omega_t, w_t)^{d_t^{rce}}\right) f_w(w_t)^{\mathbb{1}[e_t>0]} \left(\prod_{m=1}^{5} P(M_t = m|\Omega_t, r_{t-1}, c_{t-1})^{\mathbb{1}[M_t=m]}\right) \tag{10}$$

6.3 Identification

The term *identification* is used frequently to describe two different econometric concepts. Traditionally, a model was said to be "identified" if the data and model were such that a unique set of parameters maximized the objective function. The OLS corollary is the rank condition, which allows for the inversion of the $X'X$ matrix. More recently, researchers have begun describing a particular treatment effect as "identified" if the variation in the causal variable used in estimation is uncorrelated with unobserved determinants of the outcome variable. The OLS corollary is the exogeneity condition, or that $E[\epsilon' X] = 0$. In the following two subsections, we separately discuss each of these concepts in relation to the model described in Section 5.

6.3.1 Exogeneity

There are three sets of parameters in the model where bias, due to correlation between observed and unobserved variables, may be of concern. Each of these parameters relates directly to the costs and benefits associated with medical care. First, we posit that the primary benefit of both therapy and pharmaceutical consumption is their positive impact on future mental health. As discussed in Section 3, observed treatment is likely correlated with unobserved determinants of mental health transitions, ϵ_t^M, biasing δ_2 and δ_3, our estimated measures of treatment efficacy. To ensure that we estimate δ_2 and δ_3 without bias, we have chosen to estimate the mental health production function outside of the structural model, using 2SLS, which enables the use of standard econometric strategies to evaluate the strength and exogeneity of our instruments.[23]

Second, estimate lost work hours due to therapy by regressing average weekly hours on the average number of therapy visits an individual consumes in a week - i.e., $(\lambda_0, \ldots, \lambda_3)$ from Equation 5. Correlation between therapy use and unobserved determinants of hours

[23] It is well understood that instrumental variables approaches, such as 2SLS, provide estimates of only local average treatment effects (LATE) (Angrist, Imbens, and Rubin, 1996), which we will apply to our model as if they are average treatment effects (ATE). It is important to consider this in interpreting our results, not only because these effects are specific to "compliers," but also because we estimate the mental health production function on a subsample of privately insured individuals.

worked could exist if, for example, employment characteristics, such as a long commute time, impact both hours on the job as well as therapy use. Currently, we are estimating these parameters outside the model using individual and time fixed effects, which should alleviate concerns about correlation due to all unobserved individual characteristics and most unobserved job characteristics. Eventually, we plan to estimated the parameters of the hours equation inside the structural model, while allowing for unobserved correlation between the models unobservables. Using this joint modeling approach, generating causal estimate of $(\lambda_0, \ldots, \lambda_3)$ requires that some state variables have a significant impact on the decision to consume therapy, but are excluded from the hours equation. In our model, the price of therapy serves as an appropriate exclusion restriction.

Third, an indirect benefit of consuming treatment is increased future wages generated by improved mental health. We estimate the impact of mental health on wages within the model (i.e., γ_4 from Equation 4), which could be biased if, for example, those with a history of mental illness accumulated less unobserved human capital, leading to both lower current wages and a higher likelihood of current mental illness. In the current version of this paper, we do little to address this potential bias; however, in future iterations we intend to allow for correlation between the unobserved determinants of mental health and wages. Given the timing assumptions imposed on the model, identification will again require that some observable affecting mental health is excluded from the wage equation; the treatment variables serve this role.

6.3.2 Uniqueness of Structural Parameters

The likelihood function consists of choice probabilities for combinations of employment (e_t), medication (r_t), and therapy (c_t), as well as contributions from observed wages. Preference parameters found in Equation (2) only impact the likelihood function through impacts on choice probabilities. Hence, the estimation procedure will find the preference parameters that make the choice probabilities generated by the model most closely match the choice probabilities observed in the data. Preferences for treatment (α_0, $\alpha_{1,0}$, $\alpha_{1,1}$) are identified based on the popularity of treatment choices given their productivity levels and given their costs in terms of time and money. The preference for mental health (α_2) is identified by variation in the degree to which individuals are willing to consume costly treatments across the distribution of mental health states.[24] Preferences for employment are identified by the popularity of working part-time and full-time given the amount of income that the

[24]The non-linearity in preferences for mental health (α_3) is identified by the degree to which treatment uptake is increasing at an increasing or decreasing rate as mental health worsens.

model predicts for these types of employment and preferences for consumption of income (determined by the CRRA parameter). Because treatment alternatives have a pecuniary cost, the CRRA parameter is identified by variation in treatment choices across the income distribution.

6.4 Structural Parameter Estimates

Preliminary estimates of the model parameters are presented in Tables 10 and 11. Our preferred estimates of the mental health production function are presented in Column 3 of Table 5 and were discussed in Section 4.2.1.

Preference parameter estimates are found in Table 10. Negative coefficients on therapy consumption suggest that there are unobserved costs to therapy that are not being captured by the model, which accounts for uncertainty, time, and monetary costs. Likewise, the negative coefficient on consumption of medication suggests that there are costs of using medication beyond the monetary costs. Individuals receive positive flow utility from good mental health. Individuals receive disutility from working, which rationalizes the large portion of the sample that chooses not work. Better mental health makes both part time and full time work relatively more attractive. These estimates suggest that the impacts of poor mental health on employment could reduce income contemporaneously by causing an individual to be less likely to work as well as income in the future by reducing the accumulation of human capital.

Table 11 shows wage offer parameters. As expected, wages are increasing in education, accumulated part time and full time work experience, and better mental health.[25]

7 Counterfactual Policy Simulations

Using the parameters from the estimated model, we are able to conduct a number of counterfactual policy experiments to consider how altering the various costs associated with therapy would impact therapy utilization, mental health, and labor market outcomes. Decisions and outcomes are forward simulated over four periods for 10 times the number of individuals that are in the sample used to estimate the structural model.

[25] Coefficients on part time and full time work experience should be interpreted in light of the human capital approximation described in subsection 6.1

7.1 Decomposing the costs of therapy

The model nests four potential deterrents of using therapy: a monetary cost, a time cost, therapist mismatch, and a utility cost. We begin by showing how much each of these costs contributes to treatment decisions. Table 12 shows the percentage of individual-period dyads in which therapy is used across simulations that impose different costs. We begin with a simulation to show what optimal usage would be in a world with no costs.[26] If none of the four costs are included but the benefits of therapy for mental health remain the same, then therapy is consumed in 67.89 percent of observations (whereas we only observe 1.3 percent of observations consuming therapy in the data).[27] Adding in the possibility of mismatch reduces therapy usage by 5.38 percentage points (7.9 percent). Adding in time costs and monetary costs have much smaller effects on usage.

Table 12 also shows how each cost contributes to the distribution of mental health in the population. For example, moving from a world with no costs to a world with mismatch reduces the proportion of the population with the highest level of mental health by 2.39 percentage points (3.6 percent). In line with the effects on therapy utilization, time costs and monetary costs have limited, but non-zero, effects on mental health at the population level.

7.2 Effects of counterfactual policies across policy intensity and education level

To provide a more complete comparison of the effects of policies on therapy use, mental health, wages, and employment, we perform a series of policy experiments adjusting the size of cost reductions. Figures 6, 7, 8, and 9 show effects of reducing monetary costs, time costs, and the probability of mismatch by 25 percent, 50 percent, 75 percent, and 100 percent, respectively. All policy simulations are compared to a baseline of optimal decision-making when facing no utility costs of attending therapy outside of the monetary costs, time costs, and uncertainty.

Figures 6 and 7 indicate that removing the uncertainty that arises from mismatch has the largest impacts on therapy usage and average mental health in the population. The

[26] The zero cost simulation includes the removal of utility costs. We are not able to disentangle all of the potential costs included in this utility cost; it could consist of stigma, lack of access to therapy, or time costs separate from lost work hours.

[27] It is important to note that even if there were no benefits to therapy, a model with no therapy costs would predict usage of 50 percent. This is because the only way in which therapy would relate to an individual's decision would be through a random, mean-zero preference shock and therapy is included in half of all alternatives.

figures also show that effects of decreasing any of the costs (monetary, time, or uncertainty) are decreasing in an individual's level of education. This result is intuitive - according to the model, low education individuals, who also tend to be low income, are relatively more price sensitive due to the estimated decreasing returns to income/consumption, face the highest time costs to begin with, and have the lowest levels of mental health.

Figure 8 shows that reducing any of the costs results in an increase in employment, with the effects being the largest for reductions in monetary costs. However, which types of individuals are incentivized to select into employment differs, as illustrated by the effects on wages shown in Figure 9. A reduction in monetary or time costs incentives those who earn lower wages to move from not working to working, while a reduction in mismatch incentives those who earn higher wages to move from not working to working.

8 Conclusion

This paper uses a dynamic decision-making model and data on consumer behaviors to study relationships between mental health conditions, treatment alternatives, and labor market outcomes. The goal is to identify the key impediments to effective medical treatments for mental illness and to measure the health and labor market returns to policies which promote more effective treatment choices. Our findings suggest that monetary costs, time costs, uncertainty over treatment productivity, and preferences together explain the lack of demand for psychotherapy relative pills.

Counterfactual simulations show that utility costs, over-and-above time costs, monetary costs, and uncertainty, are a much larger impediment to therapy use than any of the other costs nested in the model. Though informative about the magnitude of the full set of costs to using therapy, a high utility cost does not lend itself easily to policy evaluation. It does suggest that lack of therapy use and the problem of poor mental health in the United States would not be solved by parity laws or other reductions in the cost of therapy nor by work-place equity laws that compensate workers for work hours lost due to seeking medical treatment.

One reason the high utility cost of therapy dominates other, more policy-relevant factors that could also discourage use of therapy is our focus on monetary and work-hours costs. Reduced-form evidence suggests that, conditional on working, therapy usage does not change labor supply very much. Wage effects are also fairly small. However, there appear to be large differences in employment among those who use therapy and those who do not. As constructed, the structural model subsumes costs of working while using therapy into the utility cost. In reality, individuals may anticipate that remaining employed while using therapy is difficult, which means that choosing therapy would lead to a job separation. Those who use therapy and remain employed would therefore be a selected sample of individuals who do not anticipate a job separation. In fact, hours costs may be low since individuals simply lack the flexibility to trade an hour of supplied labor for an hour of therapy.

To capture how therapy can affect employment (rather than hours and wages conditional on employment), future iterations of the model will interact the utility cost of therapy with employment. To permit variation in costs by job types, these utility costs will be interacted with education. These interactions would capture how education can affect the degree to which working while using therapy can vary by education and, presumably, job flexibility. If we find that a high utility cost of working while using therapy can help to explain why many people do not use therapy, the next task will be to place additional structure on the joint employment/therapy decision. One possibility would be to incorporate a job separation probability into the model. Job separation risk would be a function of therapy and mental health and occur after the joint therapy/employment decision. If individuals anticipate that therapy increases the risk of job separation, this may be an important reason to avoid it. Counterfactual policies to evaluate would include reductions to the risk of job separation due to therapy use.

Our current analysis has several additional limitations, which will lead to further changes to the model. First, we use average medical care prices that do not vary by individual and that are unrelated to insurance status. In our data, we observed individual-level medical care prices as well as insurance status, so this variation can be integrated into the consumer's optimization problem. Second, our current model is not sufficiently flexible across individuals' ages to allow for simulations over the life-cycle. Third, because individuals enter our data at different ages and we do not observe work history, we must impute their accumulated work experience, which is likely to be correlated with their mental health. Our current imputation is described in Section 6.1. This technique produces variation in entering work history by entering mental health status, but does not account for the evolution of mental health over the life-cycle to affect accumulated experience. Fourth, education gradients can be added to additional processes in the model, including those for wages and hours. Addressing these four limitations (and further explaining the utility cost of therapy by considering employment as outlined above) will be the focus of our work moving forward.

References

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin. 1996. "Identification of causal effects using instrumental variables." *Journal of the American statistical Association* 91 (434):444–455.

Bartel, Ann and Paul Taubman. 1986. "Some economic and demographic consequences of mental illness." *Journal of Labor Economics* 4 (2):243–256.

Berndt, E.R., Richard G. Frank, and Thomas G. McGuire. 1997. "Alternative Insurance Arrangements and the Treatment of Depression: What are the Facts?" *American Journal of Managed Care* 3:243–250.

Berndt, Ernst R, Bronwyn H Hall, Robert E Hall, and Jerry A Hausman. 1974. "Estimation and inference in nonlinear structural models." In *Annals of Economic and Social Measurement, Volume 3, number 4*. NBER, 653–665.

Blackburn, Ivy M, S Bishop, AI Glen, LJ Whalley, and JE Christie. 1981. "The Efficacy of Cognitive Therapy in Depression: A Treatment Trial Using Cognitive Therapy and Pharmacotherapy, Each Alone and in Combination." *The British Journal of Psychiatry* 139 (3):181–189.

Blau, David M and Donna B Gilleskie. 2008. "The role of retiree health insurance in the employment behavior of older men." *International Economic Review* 49 (2):475–514.

Bound, John, David A Jaeger, and Regina M Baker. 1995. "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak." *Journal of the American statistical association* 90 (430):443–450.

Cragg, John G and Stephen G Donald. 1993. "Testing identifiability and specification in instrumental variable models." *Econometric Theory* 9 (2):222–240.

Crawford, B Y Gregory S and Matthew Shum. 2005a. "Uncertainty and Learning in Pharmaceutical Demand." *Econometrica* 73 (4):1137–1173.

Crawford, G.S. and M. Shum. 2005b. "Uncertainty and Learning in Pharmaceutical Demand." *Econometrica* :1137–1173.

Cronin, Christopher J. 2016. "Insurance-Induced Moral Hazard: A Dynamic Model of Within-Year Medical Care Decision Making Under Uncertainty." *SSRN* #2706185:1–73.

Currie, Janet and Mark Stabile. 2006. "Child mental health and human capital accumulation: the case of ADHD." *Journal of health economics* 25 (6):1094–1118.

Deb, P., J. Rubin, V. Wilcox-Gok, and A. Holmes. 1996. "Choice of health insurance by families of the mentally ill." *Health Economics* 5 (1):61–76.

Dickstein, Michael. 2014. "Efficient provision of experience goods: Evidence from antidepressant choice." Working Paper.

Druss, B.G., M. Schlesinger, and H.M. Allen. 2001. "Depressive Symptoms, Satisfaction with Health Care and 2-Year Work Outcomes in an Employed Population." *American Journal of Psychiatry* 158 (5):731–734.

Elkin, Irene, Tracie M. Shea, John T. Watkins, Stanley D. Imber, Stuart M. Sotsky, Joseph F. Collins, David R. Glass, Paul A. Pilkonis, William R. Leber, John P. Docherty, Susan J. Fiester, and Morris B. Parloss. 1989. "National Institute of Mental Health Treatment of Depression Collaborative Research Program." *Archives of General Psychiatry* 46:971–982.

Ellis, RP. 1986. "Rational behavior in the presence of coverage ceilings and deductibles." *RAND Journal of Economics* 17 (2):158–175.

Ettner, Susan L, Richard G Frank, and Ronald C Kessler. 1997. "The Impact of Psychiatric Disorders on Labor Market Outcomes." *Industrial and Labor Relations Review Labor Relations Review* 51 (1):64–81.

Fava, Giovanni, Chiara Ruini, Chiara Rafanelli, Livio Finos, Sandra Conti, and Silvana Grandi. 2004. "Six-Year Outcome of Cognitive Behavior Therapy for Prevention of Recurrent Depression." *American Journal of Psychiatry* 161 (October):1872–1876.

Fletcher, Jason and Barbara Wolfe. 2008. "Child mental health and human capital accumulation: the case of ADHD revisited." *Journal of health economics* 27 (3):794–800.

Fletcher, Jason M. 2008. "Adolescent depression: diagnosis, treatment, and educational attainment." *Health economics* 17 (11):1215–1235.

———. 2014. "The effects of childhood ADHD on adult labor market outcomes." *Health economics* 23 (2):159–181.

Fournier, Jay C, Robert J DeRubeis, Steven D Hollon, Sona Dimidjian, Jay D Amsterdam, Richard C Shelton, and Jan Fawcett. 2010. "Antidepressant drug effects and depression severity: a patient-level meta-analysis." *Jama* 303 (1):47–53.

Frank, R. G., S. H. Busch, and E. R. Berndt. 1998. "Measuring Prices and Quantities of Treatment for Depression." *The American Economic Review, P&P* 88 (2):106–111.

Frank, R. G. and T. G. McGuire. 1986. "A review of studies of the impact of insurance on the demand and utilization of specialty mental health services." *Health services research* 21:241–265.

Frank, R.G. and Paul Gertler. 1991. "An Assessment of Measurement Error Bias for Estimating the Effect of Mental Distress on Income." *The Journal of Human Resources* 26 (1):154–164.

Gilleskie, Donna B. 1998. "A dynamic stochastic model of medical care use and work absence." *Econometrica* :1–45.

Gloaguen, Valérie, Jean Cottraux, Michel Cucherat, and Ivy-Marie Blackburn. 1998. "A meta-analysis of the effects of cognitive therapy in depressed patients." *Journal of affective disorders* 49 (1):59–72.

Greenberg, Paul E, Ronald C Kessler, Howard G Birnbaum, Stephanie A Leong, Sarah W Lowe, Patricia A Berglund, and Patricia K Corey-Lisle. 2003. "The economic burden of depression in the United States: how did it change between 1990 and 2000?" *Journal of clinical psychiatry* 64 (12):1465–1475.

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

Hofmann, Stefan G. and Jasper A.J. Smits. 2008. "Cognitive-Behavioral Therapy for Adult Anxiety Disorders: A Meta-Analysis of Randomized Placebo-Controlled Trials." *Journal of Clinical Psychiatry* 4 (69):621–632.

Hollon, Steven D, Robert J DeRubeis, Jan Fawcett, Jay D Amsterdam, Richard C Shelton, John Zajecka, Paula R Young, and Robert Gallop. 2014. "Effect of Cognitive Therapy with Antidepressant Medications vs Antidepressants Alone on the Rate of Recovery in Major Depressive Disorder: A Randomized Clinical Trial." *JAMA Psychiatry* 71 (10):1157–1164.

Hollon, Steven D, D Ph, Robin B Jarrett, Andrew a Nierenberg, Michael E Thase, Madhukar Trivedi, and a John Rush. 2005. "Psychotherapy and Medication in the Treatment of Adult and Geriatric Depression: Which Monotherapy or Combined Treatment?" *Journal Clinical Psychology* 66 (April):455–468.

Howard, K I, S M Kopta, M S Krause, and D E Orlinsky. 1986. "The dose-effect relationship in psychotherapy." *The American psychologist* 41 (2):159–164.

Keeler, Emmett B., Willard G. Manning, and Kenneth B. Wells. 1988. "The demand for episodes of mental health services." *Journal of Health Economics* 7 (4):369–392.

Khawam, Elias A., Georgia Laurencic, and Donald A. Malone. 2006. "Side effects of antidepressants: An overview." *Cleveland Clinic Journal of Medicine* 73 (4):351–361.

Khwaja, Ahmed. 2010. "Estimating willingness to pay for medicare using a dynamic life-cycle model of demand for health insurance." *Journal of Econometrics* 156 (1):130–147.

Kirsch, Irving, Brett J. Deacon, Tania B. Huedo-Medina, Alan Scoboria, Thomas J. Moore, and Blair T. Johnson. 2008. "Initial severity and antidepressant benefits: A meta-analysis of data submitted to the food and drug administration." *PLoS Medicine* 5 (2):0260–0268.

Kleibergen, Frank and Richard Paap. 2006. "Generalized reduced rank tests using the singular value decomposition." *Journal of econometrics* 133 (1):97–126.

Lu, Mingshan. 1999. "The Productivity of Mental Health Care: An Instrumental Variable Approach." *The Journal of Mental Health Policy and Economics* 2 (2):59–71.

Mojtabai, Ramin and Mark Olfson. 2014. "National Trends in Long-Term Use of Antidepressant Medications." *The Journal of Clinical Psychiatry* 75 (02):169–177.

Papageorge, Nicholas W. 2015. "Why Medical Innovation is Valuable : Health , Human Capital and the Labor Market."

Papageorge, Nicholas W. 2016. "Why Medical Innovation is Valuable: Health, Human Capital and the Labor Market." Forthcoming in *Quantitative Economics*.

Satcher, David. 2000. "Mental health: A report of the Surgeon General–Executive summary." *Professional Psychology: Research and Practice* 31 (1):5–13.

Stewart, Rebecca E and Dianne L Chambless. 2009. "Cognitive–Behavioral Therapy for Adult Anxiety Disorders in Clinical Practice: A Meta-Analysis of Effectiveness Studies." *Journal of Consulting and Clinical Psychology* 77 (4):595.

Stewart, Walter F, Judith A Ricci, Elsbeth Chee, Steven R Hahn, and David Morganstein. 2003. "Cost of Lost Productive Work Time Among US Workers With Depression." *JAMA : the journal of the American Medical Association* 289 (23):3135–3144.

Stock, James H and Motohiro Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* :80.

Sturm, Roland, Lisa S Meredith, and Kenneth B Wells. 1996. "Provider Choice and Continuity for the Treatment of Depression." *Medical Care* 34 (7):723–734.

Taube, Carl A, Howard H Goldman, and David Salkever. 1990. "Medicaid coverage for mental illness: Balancing access and costs." *Health Affairs* 9 (1):5–18.

Tables and Figures

Table 1: Sample Creation

Description	Individuals	Individual-Periods
MEPS Panels 1-16, interviewed each of 5 rounds	211,582	1,057,910
22-64 years old	114,267	571,335
Only rounds 2 to 5	114,267	457,068
After period length restriction	100,481	385,615
Excluding those with severe mental disorders	98,056	376,234

Notes: The panel of individuals are interviewed five times over two calendar years. However, the first period cannot be used in estimation as there is no information on the individual's mental health coming into the period.

Table 2: Mental Health and Treatment Decisions By Age

	Perceived MH	Depression/Anxiety	Medication	Therapy
Ages 22-24	4.207	0.026	0.024	0.007
Ages 25-29	4.153	0.046	0.034	0.010
Ages 30-34	4.081	0.048	0.039	0.011
Ages 35-39	4.029	0.058	0.047	0.011
Ages 40-44	3.941	0.082	0.066	0.013
Ages 45-49	3.881	0.095	0.082	0.018
Ages 50-54	3.844	0.116	0.099	0.019
Ages 55-59	3.831	0.102	0.087	0.013
Ages 60-64	3.798	0.118	0.102	0.017

Notes: An observation is an interview period; thus, sample statistics are calculated across all 376,234 observations in the estimation sample (98,056 individuals). "Perceived MH" is the respondent's subjective assessment of own mental health and ranges from 1 (poor) to 5 (excellent). Depression and anxiety indicators are based on the ICD-9 codes associated with reported diagnoses.

Table 3: Sample Means By Treatment Choice

	Therapy N=1,077	Meds N=18,383	Both N=3,277	Neither N=309,048
Demographics				
Male	0.286	0.292	0.296	0.467
Age	42.735	47.023	46.092	42.54
Live in M.S.A.	0.898	0.785	0.887	0.824
Married	0.483	0.564	0.424	0.640
Northeast	0.184	0.134	0.211	0.151
Midwest	0.238	0.234	0.234	0.196
South	0.211	0.413	0.332	0.387
West	0.367	0.219	0.224	0.266
Black	0.095	0.089	0.123	0.159
Other (non-white)	0.034	0.041	0.048	0.078
Schooling & Employment				
Years of School	12.857	12.694	13.161	12.694
Employed	0.619	0.564	0.499	0.753
Hourly Wage	19.611	18.158	18.889	17.271
Mental Health	2.993	3.160	2.566	4.027

Notes: The mean hourly wage is for those who have a positive hourly wage.

Table 4: Sample Means By Subjective Mental Health

	MH=5 N=120,392	MH=4 N=103,374	MH=3 N=85,713	MH=2 N=18,381	MH=1 N=3,792
Demographics					
Male	0.480	0.453	0.435	0.387	0.394
Age	41.441	42.645	44.186	45.679	47.596
Live in M.S.A.	0.846	0.827	0.797	0.778	0.743
Married	0.668	0.656	0.606	0.459	0.366
Northeast	0.156	0.146	0.148	0.154	0.127
Midwest	0.194	0.210	0.195	0.186	0.198
South	0.387	0.374	0.396	0.420	0.442
West	0.264	0.270	0.261	0.240	0.234
Black	0.151	0.132	0.170	0.218	0.188
Other (non-white)	0.083	0.072	0.070	0.066	0.056
Schooling & Employment					
Years of School	13.343	12.897	11.954	11.230	10.801
Employed	0.800	0.781	0.692	0.436	0.259
Hourly Wage	18.519	17.500	15.582	14.538	13.966
Treatment Decisions					
Therapy	0.002	0.006	0.019	0.078	0.148
Medication	0.022	0.045	0.093	0.279	0.425
Conditions					
Depression/Anxiety	0.023	0.051	0.112	0.357	0.565

Notes: The sample is restricted to those who are at least 22 years old. Mental health categories are: 5— excellent, 4— very good, 3— good, 2— fair, and 1— poor. The mean hourly wage is for those who have a positive hourly wage.

Table 5: Mental Health Production Function, Ordered Logit

	(1) OLS		(2) 2SLS - A		(3) 2SLS - B	
	Coef.	SE	Coef.	SE	Coef.	SE
Any Medication	-0.355	(0.009)	0.741	(0.346)	0.711	(0.337)
Any Therapy	-0.414	(0.016)	1.265	(0.804)	1.524	(0.902)
Lagged MH						
Fair	0.510	(0.035)	0.894	(0.120)	0.883	(0.113)
Good	1.111	(0.037)	1.829	(0.198)	1.806	(0.179)
Very Good	1.565	(0.035)	2.384	(0.222)	2.357	(0.199)
Excellent	2.055	(0.037)	2.920	(0.233)	2.891	(0.209)
Age	-0.005	(0.000)	-0.005	(0.001)	-0.005	(0.000)
Male	0.017	(0.003)	0.079	(0.017)	0.076	(0.0161)
Nonwhite	-0.020	(0.007)	0.058	(0.021)	0.055	(0.019)
Marriage Status						
Never Married	-0.041	(0.007)	-0.044	(0.008)	-0.045	(0.008)
Previously Married	-0.035	(0.007)	-0.052	(0.011)	-0.051	(0.019)
Family Size	0.007	(0.002)	0.017	(0.003)	0.017	(0.003)
Education						
High school Grad.	0.106	(0.008)	0.065	(0.015)	0.067	(0.014)
College Grad.	0.082	(0.005)	0.053	(0.010)	0.053	(0.009)
Income						
Second Quartile	0.030	(0.008)	0.030	(0.007)	0.031	(0.007)
Third Quartile	0.071	(0.007)	0.078	(0.006)	0.078	(0.006)
Fourth Quartile	0.126	(0.006)	0.130	(0.008)	0.129	(0.008)
County & Time FE	X		X		X	
R-Squared	0.333		0.137		0.146	
Hansen J Stat $\rightarrow \chi^2(2)$			2.924		2.957	
(P-value)			(0.233)		(0.228)	

Notes: The sample includes all 22-62 year olds from MEPS cohorts between 1996 and 2012 who are privately insured. Further, we remove counties in the lowest 10th percentile of total observations. There are a total of 179,259 observations. Standard errors are clustered at the state level. In 2SLS - A(B), mismatches are coded as *Any Therapy*=1(0).All models also include the number of psychiatrists per capital as a control variable. The Hansen J Statistic is distributed χ^2 with degrees of freedom equal to the number of instruments minus the number of endogenous variables. The statistic enables a test of the joint null hypothesis that the instruments as uncorrelated with the second stage error term.

Table 6: Mental Health Production Function: First Stage Linear Probability Models

	Specification 1				Specification 2A				Specification 2B‡	
	(1) Any Therapy		(2) Any Rx		(3) Any Therapy		(4) Any Rx		(5) Any Therapy	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Instruments										
Psychiatrists per cap.†	0.041	(0.020)	0.079	(0.035)	0.042	(0.022)	0.046	(0.033)	0.040	(0.020)
Psych. per cap. * Nonwhite	*	*	*	*	-0.025	(0.013)	0.010	(0.012)	-0.023	(0.010)
Psych. per cap. * Prev. Married	*	*	*	*	0.031	(0.008)	0.023	(0.013)	0.028	(0.008)
Psych. per cap. * Male	*	*	*	*	-0.003	(0.008)	0.059	(0.009)	-0.001	(0.007)
Walmart * (Year>2006)	-0.005	(0.003)	0.019	(0.006)	-0.005	(0.003)	0.018	(0.005)	-0.004	(0.002)
Lagged MH										
Fair	-0.100	(0.024)	-0.199	(0.019)	-0.100	(0.024)	-0.199	(0.019)	-0.080	(0.022)
Good	-0.179	(0.024)	-0.381	(0.019)	-0.179	(0.024)	-0.381	(0.019)	-0.141	(0.023)
Very Good	-0.199	(0.025)	-0.442	(0.021)	-0.199	(0.025)	-0.442	(0.021)	-0.156	(0.024)
Excellent	-0.206	(0.025)	-0.474	(0.022)	-0.206	(0.025)	-0.473	(0.022)	-0.162	(0.024)
Age	-0.000	(0.000)	0.001	(0.000)	-0.000	(0.000)	0.001	(0.000)	-0.000	(0.000)
Male	-0.008	(0.001)	-0.043	(0.003)	-0.008	(0.001)	-0.051	(0.003)	-0.006	(0.001)
Nonwhite	-0.011	(0.001)	-0.054	(0.004)	-0.007	(0.002)	-0.055	(0.005)	-0.005	(0.001)
Marriage Status										
Never Married	0.003	(0.001)	-0.002	(0.003)	0.003	(0.001)	-0.001	(0.003)	0.003	(0.001)
Previously Married	0.006	(0.001)	0.006	(0.002)	0.002	(0.002)	0.003	(0.003)	0.001	(0.002)
Family Size	-0.002	(0.000)	-0.007	(0.001)	-0.002	(0.000)	-0.007	(0.001)	-0.002	(0.000)
Education										
High school Grad.	0.008	(0.002)	0.027	(0.003)	0.008	(0.002)	0.027	(0.003)	0.006	(0.001)
College Grad.	0.009	(0.001)	0.012	(0.002)	0.009	(0.001)	0.012	(0.002)	0.001	(0.002)
Income										
Second Quartile	-0.001	(0.001)	0.001	(0.003)	-0.001	(0.001)	0.001	(0.003)	-0.001	(0.001)
Third Quartile	-0.001	(0.002)	-0.004	(0.004)	-0.001	(0.002)	-0.004	(0.004)	-0.001	(0.001)
Fourth Quartile	-0.001	(0.002)	-0.002	(0.005)	-0.000	(0.002)	-0.002	(0.005)	-0.000	(0.001)
County & Time FE	X		X		X		X		X	
Joint F-Stat.	3.12		9.31		17.8		14.6		17.87	
Kleinberger-Paap rk LM Stat.	2.36				6.91				7.72	
(P-value)	(0.13)				(0.07)				(0.05)	
Kleinberger-Paap rk Wald F-Stat.	3.00				8.54				10.08	

Notes: The sample includes all 22-62 year olds from MEPS cohorts between 1996 and 2012 who are privately insured. Further, we remove counties in the lowest 10th percentile of total observations. There are a total of 179,259 observations. Standard errors are clustered at the state level.

† Psychiatrists per capita is part of the instrument set for Specification 1, but is part of the control variables for Specifications 2A and 2B as the instrument set is weakened by it's inclusion.

‡ In Specification 2B, mismatches are coded as *Any Therapy*=0.

Table 7: Mental Health and Labor Market Outcomes

	Specification 1						Specification 2					
	Employment		Ln(Hourly Wage)		Hours		Employment		Ln(Hourly Wage)		Hours	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Mental Health												
Excellent	0.437	(0.007)	0.147	(0.018)	2.848	(0.378)	0.041	(0.005)	0.004	(0.008)	0.350	(0.142)
Very Good	0.435	(0.007)	0.119	(0.018)	2.528	(0.378)	0.042	(0.005)	0.004	(0.008)	0.362	(0.142)
Good	0.376	(0.007)	0.069	(0.018)	2.314	(0.378)	0.039	(0.005)	0.003	(0.008)	0.329	(0.141)
Fair	0.179	(0.007)	0.016	(0.019)	1.389	(0.395)	0.020	(0.005)	0.004	(0.008)	0.322	(0.143)
Individual & Time FE							X		X		X	
Observations	331,652		245,783		245,783		331,652		245,783		245,783	

Notes: All models control for therapy use, prescription drug use, sex, age, race, marital status, MSA, family size, region, and education. Models for hourly wage are estimated on those who are working.

Table 8: Regressing Avg. Weekly Hours Worked on Therapy Sessions per Week

	Coef.	SE
Education		
12 to 15 years	0.002	(0.237)
16 or more years	0.231	(0.340)
Therapy Sessions/week	-2.251	(1.544)
“ $\times$ 12 to 15 years	1.577	(1.595)
“ $\times$ 16 or more years	1.983	(1.595)
Sessions/week Squared	0.727	(0.394)
Medication	0.112	(0.063)
Individual & Time FE	X	

Notes: The sample is restricted to those who are working and for whom period length can be calculated at a weekly level (218,492 observations). The model conditions on subjective mental health, sex, age, race, marital status, MSA, family size, region, and education. The excluded education category is less than 12 years of education. Models with education more finely discretized into less than 12 years, 12 years, 13 to 15 years, and 16 or more years show a similar education gradient.

Table 9: Out-of-Pocket Treatment Costs

	Any Consumption	
	Therapy	Meds
Uninsured ($N = 63,211$)	263.20	220.59
	(735.65)	(363.77)
Managed private insurance ($N = 135,797$)	248.82	95.48
	(469.05)	(161.71)
Other private insurance ($N = 80,431$)	237.60	101.49
	(505.42)	(187.55)
Managed public insurance ($N = 19,494$)	19.69	76.13
	(95.12)	(572.18)
Other public insurance ($N = 26,150$)	61.53	113.28
	(341.90)	(278.90)

Standard deviation in parentheses. The calculations reported are the average dollar amount of out-of-pocket costs for all individuals consuming a given type of treatment within a period.

Table 10: Preference and CRRA Parameter Estimates

Variable	Parameter	Estimate	Std. Error
Any Prescription Drugs	α_0	-3.420	0.031
Any Therapy	α_1	-5.201	0.055
Level of MH	α_2	0.695	0.042
Part Time	$\alpha_{3,1}$	-9.517	0.173
$\times$ MH	$\alpha_{4,1}$	0.223	0.018
Full Time	$\alpha_{3,2}$	-11.625	0.208
$\times$ MH	$\alpha_{4,2}$	0.308	0.017
CRRA	θ	0.865	0.002

Notes: The CRRA parameter is currently fixed at 0.95 in estimation.

Table 11: Wage Offer Parameter Estimates

Variable	Parameter	Estimate	Std. Error
Constant	γ_0	1.550	0.023
High school education	$\gamma_{1,1}$	0.400	0.010
College education	$\gamma_{1,2}$	0.848	0.011
Part Time Experience	γ_2	0.003	0.002
Full Time Experience	γ_3	0.013	0.001
Level of MH	γ_4	0.115	0.004
Female	$\gamma_{5,1}$	-0.285	0.008
Black	$\gamma_{5,2}$	-0.121	0.010
Other	$\gamma_{5,3}$	0.000	0.013
Std. Dev.	σ_w	0.688	0.003

Notes: The excluded category for education are those with 16 or more years.

Table 12: Therapy Cost Decomposition

Policy Regime	Therapy	$MH = 1$	$MH = 2$	$MH = 3$	$MH = 4$	$MH = 5$
No costs	67.89	0.37	2.27	10.88	19.89	66.58
+ add mismatch	62.51	0.43	2.54	11.74	21.11	64.19
+ add time cost	62.22	0.43.	2.55	11.78	21.17	64.07
+ add $ costs	61.38	0.44	2.58	11.88	21.34	63.77
All costs included	1.16	1.51	6.56	21.95	32.83	37.14

Notes: Removing all costs, including utility costs, and allowing therapy to have benefits implies that the model must predict therapy usage greater than 50%.

Figure 1: The Distribution of Period Lengths in MEPS

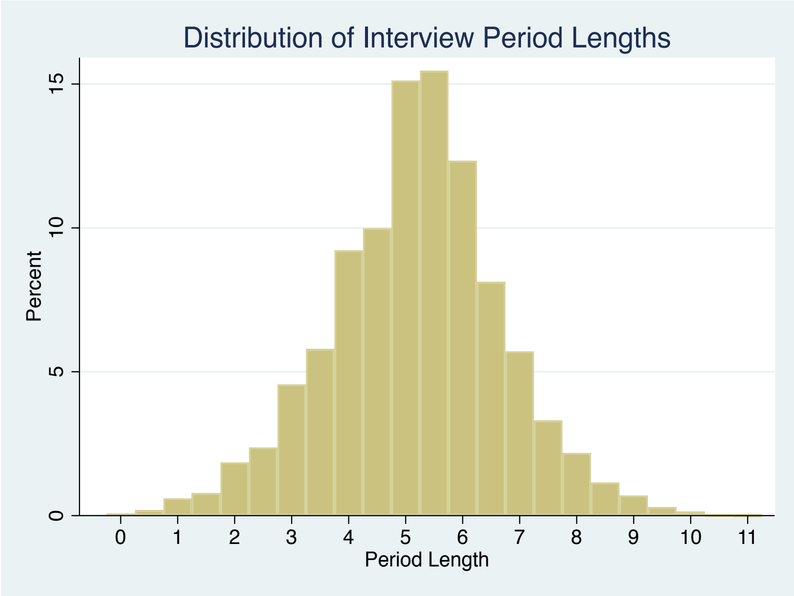


Figure 2: MH Diagnosis Over Time

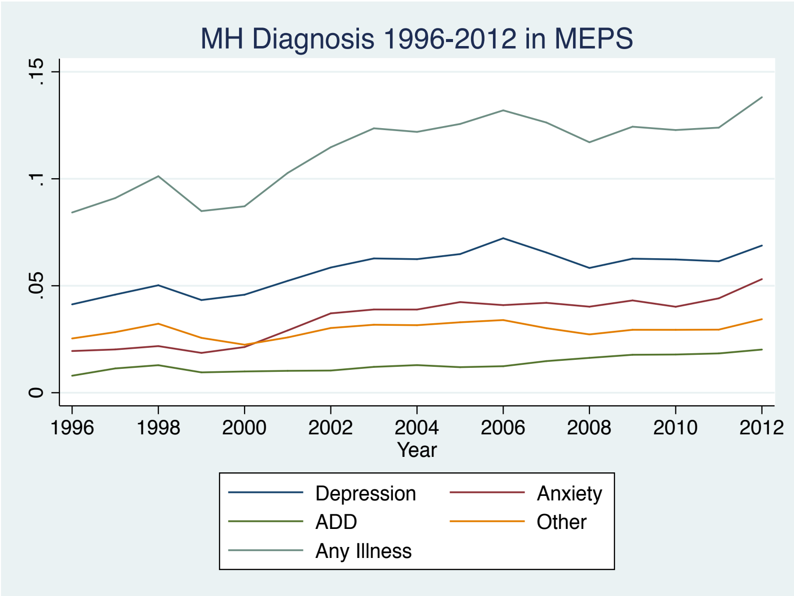


Figure 3: MH Treatment Choices Over Time

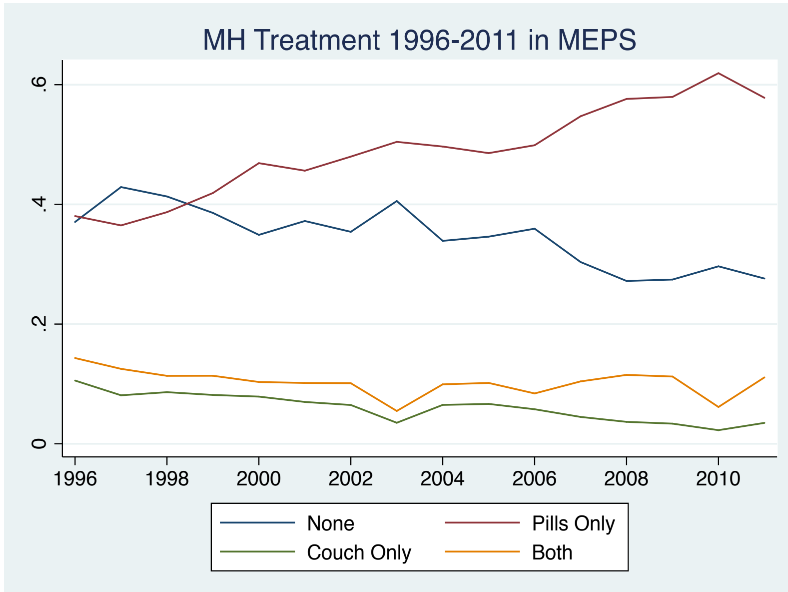


Figure 4: Therapist Mismatch

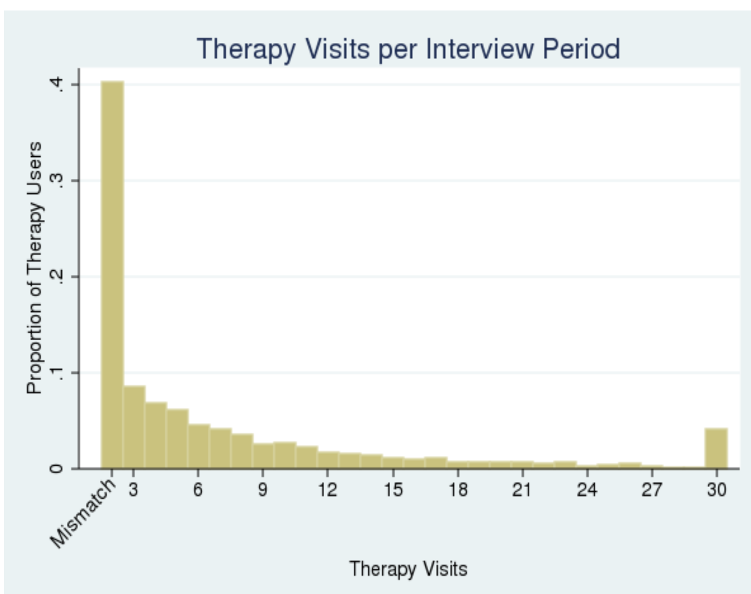


Figure 5: Human Capital Approximation

Figure 6: Effects of counterfactual policies on therapy utilization

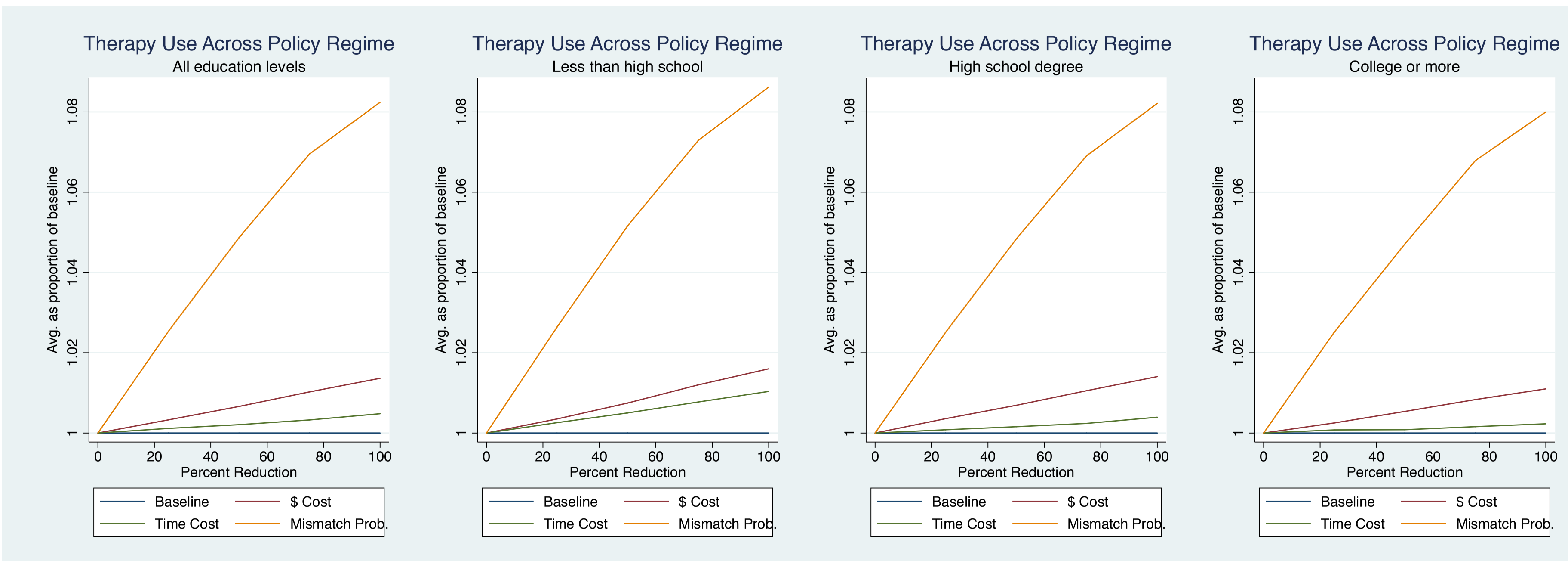


Figure 7: Effects of counterfactual policies on mental health

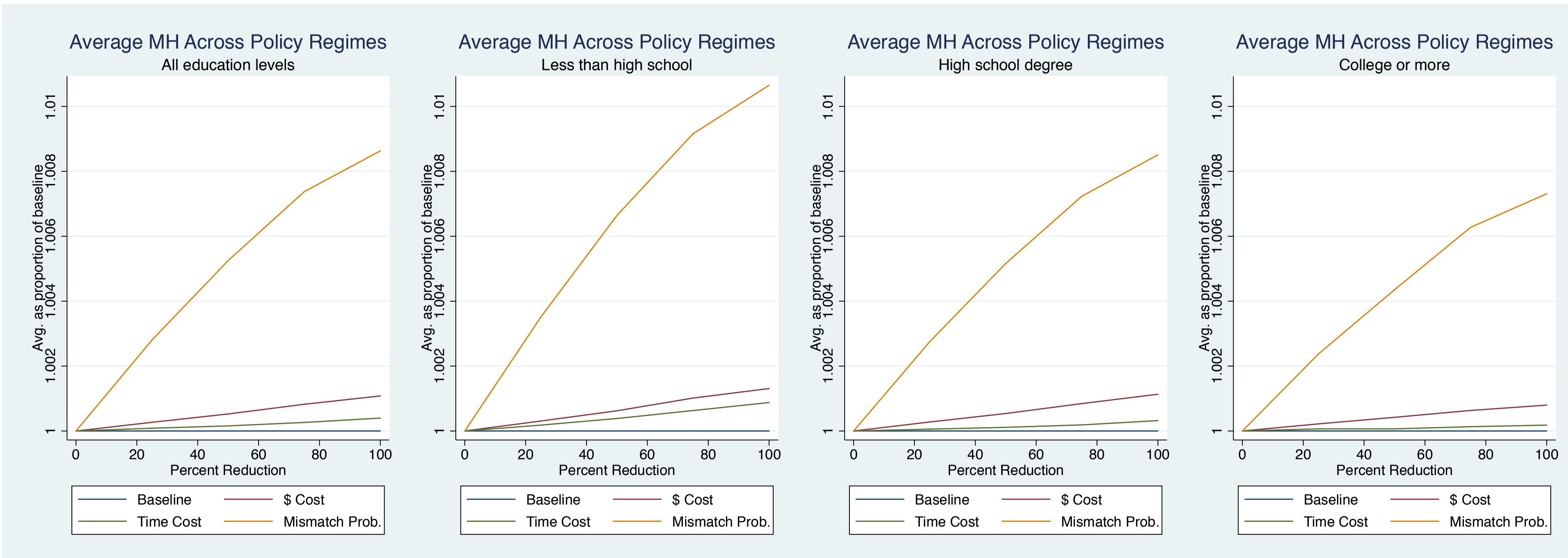


Figure 8: Effects of counterfactual policies on employment

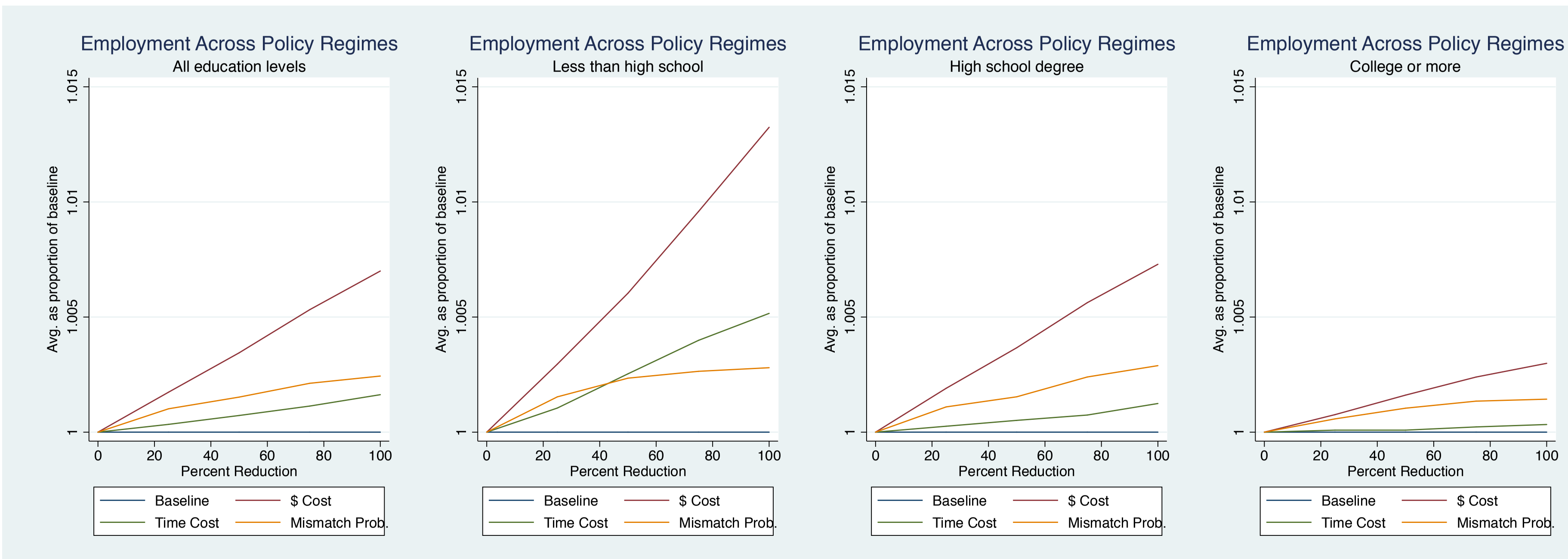
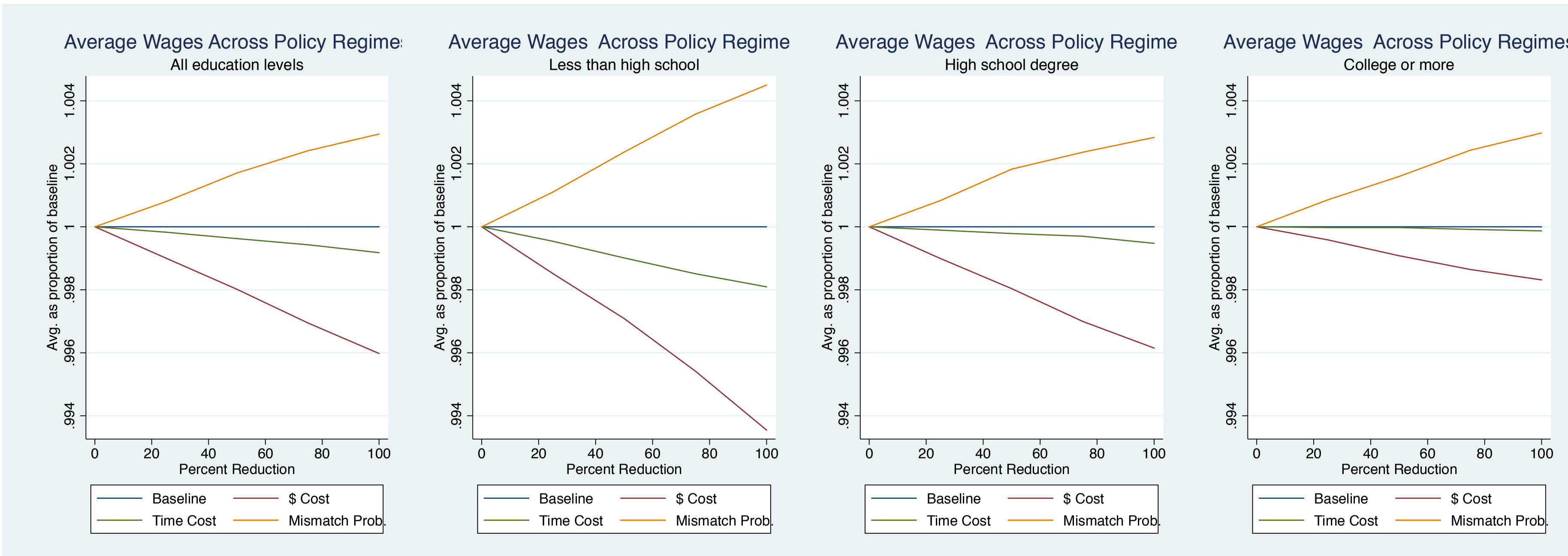


Figure 9: Effects of counterfactual policies on wages



CINCH working paper series

1 Halla, Martin and Martina Zweimüller. **Parental Responses to Early Human Capital Shocks:** Evidence from the Chernobyl Accident. CINCH 2014.

2 Aparicio, Ainhoa and Libertad González. **Newborn Health and the Business Cycle:** Is it Good to be born in Bad Times? CINCH 2014.

3 Robinson, Joshua J. **Sound Body, Sound Mind?:** Asymmetric and Symmetric Fetal Growth Restriction and Human Capital Development. CINCH 2014.

4 Bhalotra, Sonia, Martin Karlsson and Therese Nilsson. **Life Expectancy and Mother-Baby Interventions:** Evidence from A Historical Trial. CINCH 2014.

5 Goebel, Jan, Christian Krekel, Tim Tiefenbach and Nicolas R. Ziebarth. **Natural Disaster, Environmental Concerns, Well-Being and Policy Action:** The Case of Fukushima. CINCH 2014.

6 Avdic, Daniel, **A matter of life and death? Hospital Distance and Quality of Care:** Evidence from Emergency Hospital Closures and Myocardial Infarctions. CINCH 2015.

7 Costa-Font, Joan, Martin Karlsson and Henning Øien. **Informal Care and the Great Recession.** CINCH 2015.

8 Titus J. Galama and Hans van Kippersluis. **A Theory of Education and Health.** CINCH 2015.

9 Dahmann, Sarah. **How Does Education Improve Cognitive Skills?:** Instructional Time versus Timing of Instruction. CINCH 2015.

10 Dahmann, Sarah and Silke Anger. **The Impact of Education on Personality:** Evidence from a German High School Reform. CINCH 2015.

11 Carbone, Jared C. and Snorre Kverndokk. **Individual Investments in Education and Health.** CINCH 2015.

12 Zilic, Ivan. **Effect of forced displacement on health.** CINCH 2015.

13 De la Mata, Dolores and Carlos Felipe Gaviria. **Losing Health Insurance When Young:** Impacts on Usage of Medical Services and Health. CINCH 2015.

14 Tequame, Miron and Nyasha Tirivayi. **Higher education and fertility:** Evidence from a natural experiment in Ethiopia. CINCH 2015.

15 Aoki, Yu and Lualhati Santiago. **Fertility, Health and Education of UK Immigrants:** The Role of English Language Skills. CINCH 2015.

16 Rawlings, Samantha B., **Parental education and child health:** Evidence from an education reform in China. CINCH 2015.

17 Kamhöfer, Daniel A., Hendrik Schmitz and Matthias Westphal. **Heterogeneity in Marginal Non-monetary Returns to Higher Education.** CINCH 2015.

18 Ardila Brenøe, Anne and Ramona Molitor. **Birth Order and Health of Newborns:** What Can We Learn from Danish Registry Data? CINCH 2015.

19 Rossi, Pauline. **Strategic Choices in Polygamous Households:** Theory and Evidence from Senegal. CINCH 2016.

20 Clarke, Damian and Hanna Mühlrad. **The Impact of Abortion Legalization on Fertility and Maternal Mortality:** New Evidence from Mexico. CINCH 2016.

21 Jones, Lauren E. and Nicolas R. Ziebarth. **US Child Safety Seat Laws:** Are they Effective, and Who Complies? CINCH 2016.

22 Koppensteiner, Martin Foureaux and Jesse Matheson**. Access to Education and Teenage Pregnancy.** CINCH 2016.

23 Hofmann, Sarah M. and Andrea M. Mühlenweg**. Gatekeeping in German Primary Health Care –** Impacts on Coordination of Care, Quality Indicators and Ambulatory Costs. CINCH 2016.

24 Sandner, Malte. **Effects of Early Childhood Intervention on Fertility and Maternal Employment:** Evidence from a Randomized Controlled Trial. CINCH 2016.

25 Baird, Matthew, Lindsay Daugherty, and Krishna Kumar. **Improving Estimation of Labor Market Disequilibrium through Inclusion of Shortage Indicators.** CINCH 2017.

26 Bertoni, Marco, Giorgio Brunello and Gianluca Mazzarella. **Does postponing minimum retirement age improve healthy behaviors before retiremnet?** Evidence from middle-aged Italian workers. CINCH 2017.

27 Berniell, Inés and Jan Bietenbeck. **The Effect of Working Hours on Health.** CINCH 2017.

28 Cronin, Christopher, Matthew Forsstrom, and Nicholas Papageorge. **Mental Health, Human Capital and Labor Market Outcomes.** CINCH 2017.