Mental Health, Human Capital and Labor Market Outcomes*

Christopher J. Cronin†
Department of Economics, University of Notre Dame

Matthew P. Forsstrom‡
Department of Economics, Wheaton College

Nicholas W. Papageorge§
Department of Economics, Johns Hopkins University and IZA

November 26, 2018

This draft is preliminary. Please do not cite without permission from the corresponding author.

Abstract: There are two primary treatment alternatives available to those with mild to moderate depression or anxiety: psychotherapy and prescription medications. The medical literature and our analysis suggest 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. To explain this pattern, 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 it's productivity for a given individual; and (3) therapy is expensive. Preliminary results suggest that, while these factors affect treatment decisions, their role is small relative to the estimated utility cost of therapy, which may capture stigma, lack of information on the returns to therapy or other unobserved factors.

Keywords: Mental Health, Demand for Medical Care, Labor Supply, Structural Models.
JEL Classification: I10, I12, J22, J24

*First draft: October 31, 2016. We gratefully acknowledge helpful comments from: Bill Evans, David Bradford, Michael Dickstein, Donna Gilleskie, Michael Richards, Richard Frank, Robert Moffitt, Daniel Avdic, and Fabrice Etile along with seminar participants at ASHEcon Philadelphia, the Southeastern HESG Richmond, H2D2 Ann Arbor, SOLE Raleigh, Essen Health and Labour Conference, and EWEHE Prague.

†ccronin1@nd.edu.
‡matthew.forsstrom@wheaton.edu.
§papageorge@jhu.edu.

1 Introduction

1 The framework has been successfully 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, 2016), and links to labor market decisions and outcomes (Gilleskie, 1998).

Existing research has rarely applied the Grossman model to mental health.[1] As a result, our understanding of mental health treatment decisions is limited. For example, despite research suggesting that psychotherapy is more curative than medication, few individuals choose it and it is unclear why. This lack of understanding 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 depression or anxiety.[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 prescription 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: psychotherapy and prescription medications. The key empirical observation motivating our approach is that patients often opt for prescription medications even though medical literature has suggested that psychotherapy 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. In general, our preliminary analysis suggests that, while factors such as monetary and time costs can explain patient behavior to some degree, they play a fairly minor role when compared with unobserved factors captured as utility costs. In other words, it is difficult to find policy-variant factors that can credibly explain patient reluctance to use therapy in a model of dynamic decision-making and rational expectations. This opens the door to alternative explanations, for example, the possibility that individuals do not form correct beliefs about the returns to therapy or do not even consider therapy as an option (sometimes referred to as limited consideration sets) which will be explored in future analyses.

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 1996 - 2012 waves of 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. Our first contribution is to use observational data to provide causal estimates of the effect of various medical treatments on mental health. We identify these effects using an instrumental variables strategy that leverages plausibly exogenous variation in the supply of psychiatrists and medication prices over time within an individual's county of residence. The source of medication price variation has not been studied in the economics literature; Walmart's decision in 2006 to drop the price of virtually all monthly generic drug prescriptions to four dollars. Using this strategy, we show that psychotherapy is more effective than prescription medications, 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 psychotherapy in favor of prescription medications. To illustrate this point, Figure 1

[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 psychotherapy (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).

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 medication to treat mental illness has risen substantially (by roughly 50%) while the use of psychotherapy has fallen (see Figure 2). These patterns, coupled with estimates on the productivity of psychotherapy, suggest that the returns to therapy do not outweigh the costs of the investment for most patients.

To rationalize patient reluctance to use psychotherapy, we embed the estimated mental health production function into a dynamic model of mental health treatment, labor supply decisions, and labor market outcomes. 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. The model incorporates several downsides of psychotherapy, many of which are explicitly linked to employment.[5] This is motivated by previous literature demonstrating strong links between mental health and labor market outcomes (Frank and Gertler, 1991; Ettner, Frank, and Kessler, 1997; Stewart et al., 2003). First, psychotherapy is time consuming (Howard et al., 1986). Time costs could affect work hours conditional on employment.[6] Second, psychotherapy is expensive (Frank, Busch, and Berndt, 1998). Third, psychotherapy 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.[7] 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.[8]

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

[5] The current model does not explicitly model variation in time costs or prices. Time costs are captured as utility costs. The interaction between time costs and labor supply is captured by interactions in the utility function between labor and treatment consumption. Some variation in monetary costs of therapy is captured as utility costs.

[6] Moreover, the marginal cost of a lost leisure hour could be more costly for the employed if returns to leisure are diminishing. If these time costs are large enough, they could also imply that it is difficult to remain employed while consuming psychotherapy, which is particularly salient for individuals who sort into occupations with less flexible work schedules.

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

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

for psychotherapy 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 psychotherapy. 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. One unique feature of this data set is that it includes mental health treatment choices and outcomes for individuals who are unemployed, which allows us to explore links between mental health conditions, employment decisions and outcomes. In particular, we can identify how psychotherapy is difficult to consume for individuals who are employed full time.

The estimated model reveals that in addition to reducing utility directly, mental illness reduces wage offers and employment, both through lower wages and a higher disutility of work. These consequences of mental illness create a strong incentive for patients to use the best mental health treatment available, which is psychotherapy. To explain low usage of therapy, estimates suggest that mismatch is an important deterrent to psychotherapy use. Additional utility costs capture factors that are not explicitly modeled, such as social stigma and lost leisure time. Moreover, working while using mental health treatment is costly, likely reflecting that working while consuming mental health treatments can reduce productivity and increase the effort costs of working. Using the estimated model, we also perform a number of counterfactual policy simulations to assess how treatment consumption, mental health and labor outcomes change if the costs of treatment decline. For example, we find full time employment rises by about 5% relative to baseline predictions among individuals with low starting mental health if the employment utility costs of therapy are set to zero. This suggests that policies that facilitate remaining employed while consuming therapy could raise therapy consumption, employment and wages through improvements in mental health.

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 medication is 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 medication. 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 psychotherapy. 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 medications 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. Psychotherapy 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 psychotherapy. 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 of which 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. Surveyed individual's also have the ability to report specific mental illnesses (e.g., depression, schizophrenia, dementia, etc.), which are coded as three-digit ICD-9 Condition Codes. These illnesses are not modeled explicitly in our main analysis because the diagnosis of a mental illness requires that an individual visit a doctor, meaning diagnosis data measure mental illness with significant error. That said, we use both diagnostic data as well as the Kessler-6 (K6) index in several robustness checks to validate the subjective measure of mental health.[9]

The MEPS data also contain rich information on demographics (e.g., education, age,

[9]The K6 index, discussed in detail in Kessler et al. (2003), is a continuous measure of nonspecific mental distress used to discriminate cases of serious mental illness from non-cases. The index is calculated from responses to six questions relating to how an individual has felt over the past month (e.g., "During the past 30 days, about how often did you feel (nervous / hopeless / restless or fidgety / so depressed that nothing could cheer you up / that everything was an effort / worthless)?"). The K6 question are asked in rounds 2 and 4 for cohorts 2006 and later, thus preventing us from using the measure in our main analysis. However, two observations per person is enough to estimate a mental health production function using this measure.

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 as the number of 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 in order to 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 are 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 3 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: (i) drop any observation where length is less than 3 months; (ii) drop any observation where length is greater than 8 months; and (iii) drop any individual whose 2nd, 3rd, or 4th interview is dropped in (i) or (ii). Finally, we exclude those who have been diagnosed with a severe mental (i.e., ICD-9 codes 290, 293-299), as these individuals are less likely to make medical care treatment decision in a way that is consistent with our model.

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. In Appendix Table A.I, we present summary statistics for several specifications to show how these restrictions might effect the generalizability of our results.

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 differ by age. Column 1 shows average subjective mental health, where categories are 1 (poor), 2 (fair), 3 (good), 4 (very good) and 5 (excellent). Column 2 presents the proportion of individuals in each age bracket who report depression or anxiety; ICD-9 codes 311 and 300, respectively. Similar to results in Column 1, mental health is relatively worse for older individuals in the sample. For example, 2.6% of 22-year-olds are diagnosed with depression or anxiety, while the same is true of 11.8% of individuals between 60 and 64. Similarity in these age patterns also suggest that subjective mental health captures variation in the diagnosis of anxiety and depression.

Columns 3 and 4 of Table 2 show medication and psychotherapy usage, respectively, by age group. Three patterns emerge. One, use of mental health treatment rises with age. Two, patients are about three-to-five time more likely to use medication than psychotherapy. Three, this patterns holds across age groups, though psychotherapy becomes relatively less popular as individuals age. Rising mental health treatment across age groups may simply reflect selection into treatment due to worse mental health. However, it can also reflect additional costs of treatment, including missed work, stigma or, in the case of medicine, side effects, which could rise with age. Our subsequent analyses will thus incorporate reported mental health and other factors, such as age, to explain patient decisions.

Table 3 presents sample means for demographic and labor market variables by treatment choice. The statistics indicate that those individuals who use psychotherapy 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 psychotherapy 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 psychotherapy for other reasons. For example, psychotherapy may be more productive for those who are more verbal and those who are more educated may have more flexible work schedules. As expected, subjective mental health is worse for individuals in treatment in comparison to those who are not, which means that individuals select into treatment. Resulting potential bias in estimates of treatment effects is discussed in Section 3.

Table 4 presents sample means by level of subjective mental health. Those with poor mental health are more likely to be female, older, less educated, have lower wages, and are less likely to be employed 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 psychotherapy 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 psychotherapy). Given that our analysis focuses on the effect of *any* psychotherapy, 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 psychotherapy sessions spans two interview periods, then the individual is coded as having chosen therapy for two consecutive periods.[10] Results from the alternative specification, which are similar to those presented here, are available upon request from the 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.[11] The results suggest that both psychotherapy and medication *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

[10] Because prescription medications 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 medications in the sample period observed in the data.

[11] 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.

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. 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.[12] 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.[13] 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 medication 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 medications demanded. 90% of our sample lives in a county with a Walmart and, therefore, had access to these low cost medications.

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 psychotherapy. The estimates reveal that the number of psychiatrists per capita significantly increases psychotherapy 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 medications for mental illness. The estimates reveal that both psychiatrists per capital and low cost generic prescriptions through Walmart significantly increase prescription medication use.

[12] 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.

[13] 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.

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 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).[14] Moreover, Sanderson and Windmeijer (2016) develop an F-statistic for weak instruments with multiple endogenous variables that has the same interpretation as a traditional F-statistic in the typical single endogenous variable model - i.e., the bias of the IV estimate relative to the OLS estimate is approximately $1/F$.[15] Table 6 provides traditional joint F-statistics, as well as Kleibergen-Paap LM and Sanderson-Windmeijer 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.[16] In Column 3, the presence of psychiatrists significantly increases the use of psychotherapy for previously married and white individuals. Access to low cost medications through Walmart decreases psychotherapy use, presumably as individuals substitute therapy for prescription medications. In Column 4, the presence of psychiatrists increases prescription medication use, but for males only, as

[14]The Kleibergen-Paap LM statistic is a cluster-robust alternative to a similar statistic provided in Cragg and Donald (1993), which is only valid with homoskedastic errors.

[15]Sanderson and Windmeijer (2016) F-statistic is also is only valid assuming homoskedastic errors; however, a valid weak instrument test that relaxes this assumption with multiple endogenous variables has yet to be developed.

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

does the presence of a Walmart after the generic medication price drop. The Kleibergen-Paap LM statistic allows us to reject the null of underidentification at a 7% significance level, while the Sanderson-Windmeijer F-statistic suggest that our instruments are not weak, as the approximate relative bias is well under 10%.

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 medication and psychotherapy are found to be effective in improving an individual's mental health. Moreover, consistent with the medical literature cited above, psychotherapy is found to have a larger positive effect than prescription medications.[17] 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 psychotherapy is more productive than other treatment options. Yet, as alluded to earlier, psychotherapy 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 psychotherapy. Widespread reluctance to use psychotherapy suggests that there are important costs. Section 4.2 explores potential costs that help to explain why patients forgo psychotherapy. 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

[17] The parameters on any medication and any psychotherapy 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.

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 employment 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 Psychotherapy?

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

4.2.1 Therapist Mismatch

A striking feature of the data is that a large proportion of the individuals who attend psychotherapy do so only once or twice before stopping treatment. To show this, we define a psychotherapy *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 psychotherapy visits within each treatment episode. Notice that about 40% of these treatment episodes contain only one or two visits, meaning one or two psychotherapy sessions are attended without any sessions attended in the two months preceding or following these visits. Many of those who consume psychotherapy 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 psychotherapy 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 psychotherapy 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. The psychology literature refers to the latter as "therapeutic alliance" (Ardito and Rabellino, 2011). 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 psychotherapy 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.[18] That said, we are able to provide evidence that these mismatch episodes are unlikely to 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 medication and psychotherapy 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 psychotherapy on mental health, suggesting that mismatch visits are not efficacious.[19] As such, we believe that these sessions generally represent costly, unproductive medical care consumption that occurs primarily due to a lack of information.

4.2.2 Psychotherapy and Labor Supply

Table 8 presents results from regressions of employment and hours worked, conditional on employment, on mental health status, education, mental health treatment decisions, and interactions. Both regressions control for demographics and individual and time fixed effects. The results suggest that using psychotherapy is associated with lower employment and working fewer hours. Moreover, this association is the strongest for those with the lowest levels of education. Note that by controlling for subjective mental health and individual fixed effects, we insure that within-individual changes in psychotherapy use, conditional on changes in mental health, are leading individual to working less. These results also provide interpretable magnitudes. The first psychotherapy 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. Additional visits are also costly, but the effects are diminishing. According to estimates, attending therapy is also associated with lower employment, though there do not appear to be large

[18]Several conversations with practicing psychiatrists and psychologists have confirmed the prevalence of mismatch. Moreover, it was suggested that patients rarely visit once every few months and that such a treatment regimen is almost never recommended by a practitioner.

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

differences by education group on this extensive margin.

These patterns suggest the possibility of variation in work-time costs across education groups for attending psychotherapy. 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 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 psychotherapy 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 psychotherapy could play an important role.

4.2.3 Monetary Costs

The third reason that individuals may avoid psychotherapy is the relative financial expense. In Table 9 we provide summary statistics of the monetary costs associated with psychotherapy and prescription medications for an interview period across insurance groups. Note that the psychotherapy costs presented in the table exclude mismatch visits; thus, the psychotherapy prices reflect the total price of roughly eight visits on average. 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

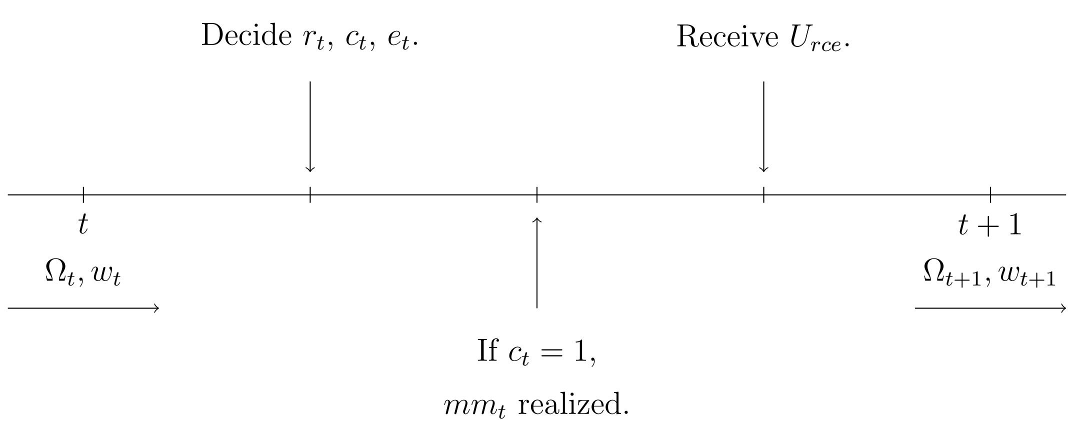
An individual decides among treatment and labor supply alternatives considering both the contemporaneous and the expected future utility associated with each alternative.[20] Current treatment decisions impact the distribution from which mental health is drawn in the next period, while labor supply decisions determine an individual's accumulation of work experience over time. The model has a finite time horizon with discrete time periods and

[20]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.

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 over four six-month periods. In the terminal period, T, the individual receives a continuation value that depends on her terminal mental health and accumulated human capital.

The timing of decisions and outcomes within a period is summarized in the figure below. Entering period t, the individual 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 (or “Rx”), r_t, and/or any psychotherapy (or “couch”), c_t, and also whether to be employed full time ($e_t = 2$), part time ($e_t = 1$), or not at all ($e_t = 0$). If psychotherapy is chosen, she then receives a draw that determines whether she experiences a therapist mismatch, $mm_t = 1$, or not, $mm_t = 0$. A mismatch results in an individual incurring the utility and financial costs associated with a single psychotherapy visit, but receiving no benefits in the form of improved mental health.

Figure 1: Timing



The total number of psychotherapy visits that occur within a decision period, in reality, accumulate as the solution to an intra-period problem. The problem can be understood as follows: on day 1, an individual decides whether to attend psychotherapy 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 psychotherapy. The simplest reason is that they realize a positive/negative utility draw for psychotherapy on any given day. Another reason, particularly for patients new to psychotherapy, 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 psychotherapy. We assume that this initial decision is 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 psychotherapy 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 psychotherapy.

5.1 State Vector and State Transitions

Human capital updates deterministically. Experience, K_t, has an initial value of zero and increases by one each period that the individual decides to be employed. 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:

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

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 psychotherapy sessions (through δ_3) has direct productive effects on mental health. Mismatch sessions are assumed to have no productive effects on mental health.[21]

[21]In future iterations of this paper, we will estimate the model and perform counterfactual simulations

Demographic variables, including age (A_t), sex (F_t), and race (r_t), also influence the decision problem. These variables are collected in the vector D_t.

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

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.

$$\begin{aligned}U_{rce} =& \frac{C_t^{1-\sigma-\theta M_t} - 1}{1-\sigma-\theta M_t} + r_t(\alpha_0 + \alpha_1 M_t + \alpha_2 F_t + \alpha_3 MSA_t) + c_t(\alpha_4 + \alpha_5 M_t + \alpha_6 F_t + \alpha_7 MSA_t) \\ &+ M_t(\alpha_8 M_t + \sum_{e=1}^{2} \alpha_{9,e} \mathbb{1}_{[e_t=e]}) + \sum_{e=1}^{2} \alpha_{10,e} \mathbb{1}_{[e_t=e]} \\ &+ \sum_{e=1}^{2} \alpha_{11,e} \mathbb{1}_{[e_t=e]} \mathbb{1}_{[c_t=1, mm_t=0]} + \sum_{e=1}^{2} \alpha_{12,e} \mathbb{1}_{[e_t=e]} r_t + \epsilon_t^{rce}\end{aligned} \tag{2}$$

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 psychotherapy visits to have different effects on utility across the income distribution. Preference parameters on prescription medication use, α_0, and psychotherapy use, α_4, capture preferences for these treatment alternatives net of explicitly modeled treatment costs and benefits (i.e., financial costs, improvements in mental health, etc.). We allow the preference for treatment to vary by demographic characteristics, including sex (F_t) and an indicator of whether or not the individual lives in a metropolitan statistical area (MSA_t). Preferences for medication and therapy also vary by mental health. As discussed in section 6.3, not including interactions of mental health with treatment leads to estimating large negative utility parameters that are aver in order to rationalize the very small percent of the total sample using treatment.

Mental health impacts the marginal utility of consumption via θ. Leisure preferences are captured by $\alpha_{10,1}$ and $\alpha_{10,2}$. The parameters $\alpha_{9,1}$ and $\alpha_{9,2}$ allow for employment to result in

across the range of possible productivities of mismatch sessions, from mismatch having no productive effect to having the same effect as psychotherapy 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 psychotherapy use results in a significantly lower productivity of therapy (see Columns 2 and 3 of Table 5).

more disutility as mental health worsens. The parameters $\alpha_{11,1}$, $\alpha_{11,2}$, $\alpha_{12,1}$, and $\alpha_{12,2}$ allow for an interaction between preferences for treatment and employment; these parameters could capture, for instance, that the time costs of a successful therapy match are more pertinent for those who are employed.

5.3 Budget Constraint

Consumption in a period is constrained as follows:

$$C_t = w_t h_t 26 + I_t - p_1 r_t - p_2 \mathbb{1}[c_t = 1, mm_t = 1] - p_3 \mathbb{1}[c_t = 1, mm_t = 0] \tag{3}$$

where w_t is the hourly wage offer, h_t is weekly hours worked, I_t is other family income, p_1 is the price of medication, and p_2 (p_3) is the price for the median number of sessions when mismatch does (not) occur.[22] Part-time and full-time employees are assumed to work twenty and forty hours per weeks, respectively. Weekly earnings are multiplied by twenty six, the number of weeks in each 6-month period.

Hourly wage offers are drawn from one of two distributions depending on whether the individual works in the part-time market or the full-time market. These distributions are modeled as a log-normal distribution that depends on education, wages in the period before the sample begins, an indicator that initial wages were zero, experience accumulated since first observation, current mental health, and demographic characteristics, including age, race, and sex (represented by D_t, below):

$$\begin{aligned} ln(w^e_t) = \gamma^e_0 + \sum_{k=1}^{3} \gamma^e_{1,k} \mathbb{1}[E_t = k] + w_0(\gamma^e_2 + \gamma^e_3 A_t) + \mathbb{1}_{w0=0}(\gamma^e_4 + \gamma^e_5 A_t) + \gamma^e_6 K_t \\ + \sum_{m=1}^{5} \gamma^e_{7,m} \mathbb{1}_{M_t=m} + \gamma_8 D_t + \epsilon^{w^e}_t \end{aligned} \tag{4}$$

$$\epsilon^{w^e}_t \sim N(0, \sigma^2_{w^e})$$

We control for initial wages and experience accumulated within the sample period to control for the fact that we have no measure of work experience in the data set.

[22]Currently, treatment prices are based on a hedonic regression of out-of-pocket costs on insurance type, year, and interactions of insurance type with year. The medium number of session associated with a successful visit is eight. In future iterations of the model we may estimate the price distribution and allow out-of-pocket costs to vary at the individual level.

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:[23]

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

When making her decision, the individual must integrate over the distributions of mm_t, M_{t+1}, and future wage offers, $f(w_{t+1})$, 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}:

$$EV(\Omega_{t+1}, w_{t+1}) = E_{\epsilon^{rce}}[\max_{rce} V_{rce}(\Omega_{t+1}, w_{t+1}, \epsilon_{t+1}^{rce})] \tag{6}$$

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.

[23] In period T, the expected future utility in this equation is replaced with V_f.

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 31 parameters using this nested algorithm: 13 preference parameters, 10 parameters in the wage offer function, 6 parameters in the terminal value function, and 2 mental health production function parameters. As discussed in Sections 3 and 6.2.1, 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. These parameters can be found in Column 3 of Table 5. The probability of mismatch, which is allowed to vary by an individual's race and sex, is likewise estimated outside of the model using a logit specification.

6.1 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,[24] 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:

$$V(\Omega_{t+1}, w_{t+1}) = \gamma + \log\left(\sum_{r=0}^{1}\sum_{c=0}^{1}\sum_{e=0}^{2} exp(\overline{V}_{rce}(\Omega_{t+1}, w_{t+1}))\right) \tag{7}$$

where γ is Euler's constant and $\overline{V}_{rce}(\Omega_{t+1}, w_{t+1})$ is the deterministic portion of the alternative

[24]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.

specific value function, $V_{rce}(\Omega_{t+1}, w_{t+1}, \epsilon^{rce}_{t+1})$. The assumption of a Type I Extreme Value and additively separable preference shock also yields the following choice probabilities:

$$P(d_t^{rce} = 1|\cdot_t) = \frac{exp(\overline{V}_{rce}(\Omega_t, w_t))}{\sum\limits_{r=0}^{1}\sum\limits_{c=0}^{1} exp(\overline{V}_{rce}(\Omega_t, w_t)} \tag{8}$$

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 to calculate expected future utility. All such 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 parameters Θ is:

$$L_{i,t}(\Theta|\Omega_t) = \left(\prod_{r=0}^{1}\prod_{c=0}^{1}\prod_{e=0}^{2} P(d_t^{rce} = 1|\Omega_t, w_t)^{d_t^{rce}}\right) f_w(w_t)^{\mathbb{1}[e_t>0]} \left(\prod_{m=1}^{5} P(M_t = m|\Omega_t, r_{t-1}, c_{t-1})^{\mathbb{1}[M_t=m]}\right) \tag{9}$$

6.2 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.

6.2.1 Exogeneity

There are two 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 psychotherapy and prescription medications 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.[25]

Second, 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 and lagged mental health serve this role.

6.2.2 Uniqueness of Structural Parameters

The likelihood function consists of choice probabilities for combinations of employment (e_t), medication (r_t), and psychotherapy (c_t), as well as contributions from observed wages. Preference parameters found in Equation (2) only impact the likelihood function through the 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 and α_1) are determined by 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.[26] For example, if individuals are more willing to consume treatment (i.e., take on the costs of consuming treatment) as their mental health worsens, it must be

[25] It is well understood that instrumental variables approaches, such as 2SLS, estimate 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.

[26] 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.

the case that these individuals dislike being in a poor mental health state. Preferences for employment (α_5), are identified by the popularity of working part-time and full-time given the amount of income that the 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 (θ) is identified by variation in treatment choices across the income distribution.

We jointly model the wage offer distribution and selection into employment. Hence, we identify the full distribution of wage offers rather than only the distribution of accepted offers. Currently, family income serves as an exclusion restriction that impacts the decision to work, but not one's own wage distribution conditional on education, work experience, and demographic characteristics. In the future, other family characteristics could enter the model, such as number of children, that impact the decision to work but do not impact wages directly.

6.3 Structural Parameter Estimates

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.

According to estimates, both medication and therapy are costly, with a greater cost associated with therapy compared to medication.[27] The utility costs also vary significantly by mental health status, sex, whether the individual lives in a metropolitan statistical area, and employment status. The large negative utility associated with treatment rationalizes the low probability that agents use mental health treatment given the returns to mental health. Better mental health not only improves earnings, which we discuss in greater detail below, and also shifts the curvature of the utility function. The utility cost of consuming therapy corresponds to roughly one-third or more of income with the cost depending on mental health status, employment, whether or not one lives in a metropolitan statistical area, sex, and insurance status. These costs should be interpreted as capturing a number of factors, beyond monetary costs, that render mental health treatment an unattractive option, including lost leisure due to time costs. Utility costs can also include the cost of searching for a therapist, fear of stigma or simply a distaste for spending time talking about psychological distress with a stranger. In the case of pharmaceuticals, costs can include various side effects, which include difficulty focusing, sexual problems and weight gain, among many others. A

[27] It is important to note that both treatments are interacted with the level of mental health, which takes an integer value from 1 to 5. These interactions are negative for both treatments.

further possibility is that the utility costs may be capturing the fact that many individuals are simply unaware that psychological treatment is a viable option for improving psychological health. If so, individuals may fail to consider the option of therapy even though they would benefit from it if they had correct information. Further, those with consistently good mental health may never consider treatment.[28] Our model would capture this lack of information or consideration as a high utility cost of therapy.

As expected, individuals receive disutility from working, which rationalizes the large portion of the sample that chooses not to work. Moreover, we estimate sizable interactions between mental health, treatments options and labor supply. We find that the utility cost of working is lower when individuals are in better mental health, which likely captures that working is particularly difficult for individuals suffering depression or anxiety. Moreover, estimates reveal that there is an additional utility cost of working while consuming mental health treatments.

Wage offer parameters are presented in Table 12. As expected, wages are increasing in education and accumulated experience.[29] Women and African Americans earn less than white males, which is consistent with findings in the literature. Importantly, wages increase with mental health, which is in line with Grossman (1972), who views health as a form of human capital that raises productivity. Higher wage offers due to better mental health constitute one of the reasons that mental health improvements raise lifetime utility. Thus, improved labor market outcomes coupled with low rates of therapy uses are responsible for high estimated utility costs of therapy.

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 psychotherapy would impact 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. We perform the following counterfactuals:

1. Eliminating all costs of therapy treatment, including utility and monetary costs.

[28] That those with good mental health are unlikely to consider mental health treatment is reflected in the large negative parameter on the interaction of mental health with therapy usage.

[29] Coefficients on part time and full time work experience should be interpreted in light of the human capital approximation described in subsection **??**.

2. Lowering the costs of therapy so that they are equal to those of medication.

3. Zeroing out the therapy-employment interaction in the utility function.

4. Equalizing the therapy-employment interaction with the medication-employment interaction.

5. Eliminating mismatch.

Results are presented in Table 13. In column (1), we use estimated model parameters to simulate averaged treatment choices, labor supply decisions, wages and mental health levels. Results in this column not only serve a baseline from which to compare counterfactual quantities. They also validate the model by allowing us to assess fit. The quantities are very close to the corresponding empirical moments.[30]

Column (2) of Table 13 presents simulated choices and outcomes under the counterfactual policy where all utility costs of medical treatment are eliminated. Therapy use rises dramatically (from about 2.1% usage to 49.7% usage, an increase of over 2,000%). Interestingly, medication usage falls from 8% to 6%. The reason is that if both therapy and medication costs are the same, there is little reason to use medication since therapy is more productive at improving mental health. Thus, medication usage under this counterfactual is driven in part by individuals who have a high utility draw for medication versus therapy. The drastic increase in therapy has implications for mental health and labor outcomes. Average mental health increases by about 9%, but for lower levels of initial mental health, the increase is larger. For individuals who start out with the lowest level of mental health, under simulation using model parameters, mental health ends on average at 2.476. Under the counterfactual, average mental health after 4 periods is 2.855, which is 15% higher. Moreover, labor supply (full-time) is 5% higher while wages are 1% higher under the counterfactual after 4 periods.

Next, we ask what happens to treatment consumption, mental health, labor supply and earnings if we set utility parameters so that the utility costs of therapy and medication are equal. Results are in Column (3). Under this counterfactual, treatment usage is roughly 7.5% for both treatments. In other words, therapy use rises and medication use falls slightly. The fact that usage is equal despite therapy being more productive is due to the risk of mismatch when patients consume therapy. Given these levels of usage, mental health is between 1% and 6% higher compared to the baseline.

If employment utility costs of mental health treatment are set to zero, we essentially eliminate any additional costs of therapy or medication consumption accruing solely to workers.

[30]We omit a table showing this explicitly. The empirical moments match up to the simulated moments nearly perfectly.

Simulated impacts are presented in Column (4) of Table 13. This counterfactual policy leads to more modest changes in behavior, including a 38% rise in therapy use and improvements in mental health of 0 to 1% relative to the baseline. For individuals starting with the lowest levels of mental health, full-time employment is about 32.9% after four periods versus 31.4% under the baseline, or about 5% higher. In Column (5), we present results where we equate the employment utility costs of treatment to those of medication. As a result, therapy rises by about 14%, which leads to minimal impacts on mental health, wages and labor supply.

8 Conclusion

Preliminary results suggest an important puzzle. Psychotherapy is a productive way to improve mental health. Mental health improvements, moreover, improve well-being as well as labor market outcomes. However, few individuals choose therapy. There are a number of candidate explanations. One explanation, which has been overlooked in earlier research, is a high likelihood of mismatch, where patients go once or twice and then do not return. We model this as an *ex post* shock rather than one of the treatment options that individuals face. This modeling decision is in line with evidence showing that 1-2 visits in isolation are generally not productive and are thus not likely to have been the intended amount. Rather, we model patients as going to therapy in the hope of completing an entire course. They pay a utility cost of seeking therapy, but may experience a mismatch and thus not enjoy the benefits.

While mismatch helps to explain patient reluctance to use therapy, other costs that have been discussed in earlier literature, including times costs or monetary costs, do not. Therefore, the model rationalizes patient behavior as a very high utility cost of therapy. What this utility cost captures, however, is not clear. It may capture a true utility cost, perhaps arising due to stigma. It may also capture the fact that most patients simply do not consider therapy as a viable option. Other candidate explanations include unobserved search costs for therapists in areas with few practitioners. Finally, the assumption of full information about the returns to therapy may be too restrictive if patents avoid therapy due to an under-estimation of its productivity. Future iterations of the model will explore these possible avenues to explain why the vast majority of patients avoid effective treatment.

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.

Ardito, Rita B and Daniela Rabellino. 2011. "Therapeutic alliance and outcome of psychotherapy: historical excursus, measurements, and prospects for research." *Frontiers in psychology* 2.

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.

Kessler, Ronald C, Peggy R Barker, Lisa J Colpe, Joan F Epstein, Joseph C Gfroerer, Eva Hiripi, Mary J Howes, Sharon-Lise T Normand, Ronald W Manderscheid, Ellen E Walters et al. 2003. "Screening for serious mental illness in the general population." *Archives of general psychiatry* 60 (2):184–189.

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. 2016. "Why Medical Innovation Is Valuable: Health, Human Capital, and the Labor Market." *Quantitative Economics* 7 (3):671–725.

Sanderson, Eleanor and Frank Windmeijer. 2016. "A weak instrument F-test in linear IV models with multiple endogenous variables." *Journal of Econometrics* 190 (2):212–221.

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.

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

	Subjective MH	Depression/Anxiety	Medication	Psychotherapy
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). "Subjective 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	Medication 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					
Psychotherapy	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	
Sanderson-Windmeijer F-Stat.	6.65		17.52		24.29		16.29		24.12	
(P-value)	(0.01)		(0.00)		(0.00)		(0.00)		(0.00)	
Kleibergen-Paap rk LM Stat.		2.36				6.91			7.72	
(P-value)		(0.13)				(0.07)			(0.05)	

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

	Employment				ln(Hourly Wage)				Hours			
	(1)		(2)		(3)		(4)		(5)		(6)	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Mental Health												
Very Good	-0.0033	0.0009	-0.0005	0.0009	-0.0037	0.0006	-0.0012	0.0007	-0.0261	0.0168	-0.0073	0.0177
Good	-0.0181	0.0010	-0.0048	0.0012	-0.0073	0.0008	-0.0020	0.0008	-0.0767	0.0203	-0.0326	0.0220
Fair	-0.0713	0.0018	-0.0268	0.0022	-0.0096	0.0016	-0.0017	0.0018	-0.1434	0.0417	-0.0269	0.0543
Poor	-0.1271	0.0034	-0.0519	0.0042	-0.0177	0.0039	-0.0080	0.0052	-0.5424	0.1049	-0.3638	0.1544
Year FE	X		X		X		X		X		X	
Individual FE			X				X				X	
Observations	561,883		561,883		353,342		353,342		417,487		417,487	

Notes: The excluded mental health group is "excellent". All models control for sex, age, race, marital status, MSA, family size, region, and education. Models for hours and hourly wage are estimated on those who are working. The difference in the number of observations in the log hourly wage and hours regression is due to the fact that some individual report their hours worked, but not their hourly wage.

Table 8: Regressing Labor Market Outcomes on Psychotherapy Sessions per Week

	Employed				Weekly Hours			
	(1)		(2)		(3)		(4)	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Any Therapy	-0.050	(0.007)	-0.016	(0.008)	-0.741	(0.269)	-0.518	(0.313)
“ × 12 to 15 years	0.017	(0.008)	0.007	(0.009)	0.477	(0.287)	0.345	(0.334)
“ × 16 or more years	0.045	(0.009)	0.009	(0.010)	0.461	(0.292)	0.324	(0.366)
Medication	-0.040	(0.002)	-0.014	(0.002)	0.002	(0.043)	0.080	(0.057)
Mental Health								
Very Good	-0.003	(0.001)	-0.000	(0.001)	-0.026	(0.017)	-0.007	(0.018)
Good	-0.017	(0.001)	-0.005	(0.001)	-0.075	(0.020)	-0.033	(0.022)
Fair	-0.067	(0.002)	-0.026	(0.002)	-0.133	(0.042)	-0.025	(0.054)
Poor	-0.119	(0.003)	-0.050	(0.004)	-0.519	(0.105)	-0.360	(0.155)
Year FE	X		X		X		X	
Individual FE			X				X	
Observations	561,883		561,883		417,487		417,487	

Notes: The sample for the employment regression has 297,300 observations. The sample for the hours regression is restricted to those who are working and for whom period length can be calculated at a weekly level (218,492 observations). The models condition 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.

Table 9: Out-of-Pocket Treatment Costs

	Any Consumption	
	Psychotherapy	Medication
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 Medications	α_0	0.045	0.090
$\times$ M_t	α_1	-0.849	0.019
$\times$ F_t	α_2	0.638	0.043
$\times$ MSA_t	α_3	-0.153	0.047
Any Therapy	α_4	-0.851	0.164
$\times$ M_t	α_5	-1.101	0.037
$\times$ F_t	α_6	0.314	0.076
$\times$ MSA_t	α_7	0.310	0.094
MH Squared	α_8	-0.016	0.011
Part Time	$\alpha_{10,1}$	-3.637	0.075
$\times$ MH	$\alpha_{9,1}$	0.395	0.018
$\times$ Therapy	$\alpha_{11,1}$	-0.475	0.166
$\times$ Medication	$\alpha_{12,1}$	-0.368	0.058
Full Time	$\alpha_{10,2}$	-3.473	0.090
$\times$ MH	$\alpha_{9,2}$	0.424	0.022
$\times$ Therapy	$\alpha_{11,2}$	-1.244	0.155
$\times$ Medication	$\alpha_{12,2}$	-0.698	0.050
CRRA	σ	0.878	0.003
MH in CRRA	θ	0.002	0.001

Table 11: Part-time Wage Offer Parameter Estimates

Variable	Parameter	Estimate	Std. Error
Constant	γ_0^1	2.198	0.084
High school education	$\gamma_{1,1}^1$	—	—
College education	$\gamma_{1,2}^1$	0.324	0.021
w_0	γ_2^1	0.014	0.001
$\times A_t$	γ_3^1	0.001	0.000
$\mathbb{1}_{w_0=0}$	γ_4^1	-1.642	0.053
$\times A_t$	γ_5^1	0.219	0.013
Experience	γ_6^1	0.146	0.008
MH=2	γ_7^1	-0.067	0.081
MH=3	γ_7^1	0.153	0.074
MH=4	γ_7^1	0.276	0.074
MH=5	γ_7^1	0.320	0.074
F_t	$\gamma_{8,1}^1$	0.041	0.018
A_t	$\gamma_{8,2}^1$	-0.086	0.011
Black	$\gamma_{8,3}^1$	-0.196	0.024
Other	$\gamma_{8,4}^1$	-0.022	0.032
Std. Dev.	σ_w^1	0.696	0.003

Notes: The excluded category for education are those with 16 or more years. The effect of a high school level of education on part-time wages is not identified.

Table 12: Full-time Wage Offer Parameter Estimates

Variable	Parameter	Estimate	Std. Error
Constant	γ_0^2	1.710	0.048
High school education	$\gamma_{1,1}^2$	0.217	0.009
College education	$\gamma_{1,2}^2$	0.363	0.010
w_0	γ_2^2	0.025	0.000
$\times A_t$	γ_3^2	-0.000	0.000
$\mathbb{1}_{w_0=0}$	γ_4^2	-0.092	0.029
$\times A_t$	γ_5^2	-0.088	0.007
Experience	γ_6^2	0.052	0.003
MH=2	γ_7^2	0.165	0.047
MH=3	γ_7^2	0.325	0.045
MH=4	γ_7^2	0.361	0.045
MH=5	γ_7^2	0.385	0.045
F_t	$\gamma_{8,1}^2$	-0.074	0.006
A_t	$\gamma_{8,2}^2$	0.040	0.004
Black	$\gamma_{8,3}^2$	-0.060	0.009
Other	$\gamma_{8,4}^2$	-0.005	0.012
Std. Dev.	σ_w^2	0.531	0.001

Notes: The excluded category for education are those with 16 or more years. The effect of a high school level of education on part-time wages i not identified.

Table 13: Counterfactual Policy Simulations

	Base		Sim 1		Sim 2		Sim 3		Sim 4		Sim 5		Sim 6	
	Level	% Δ	Level	% Δ	Level	% Δ	Level	%Δ	Level	% Δ	Level	% Δ	Level	% Δ
Therapy	0.021	0	0.497	2267	0.074	252	0.029	38	0.024	14	0.018	-14	0.022	5
Medication	0.080	0	0.060	-25	0.075	-6	0.079	-1	0.079	-1	0.080	0	0.080	0
Avg Wage	31.565	0	31.728	1	31.674	0	31.519	0	31.549	0	31.588	0	31.564.	0
Working PT	0.201	0	0.205	2	0.202	0	0.201	0	0.201	0	0.201	0	0.201	0
Working PT if $\text{MH}_1 = 1$	0.154	0	0.163	6	0.155	1	0.153	-1	0.152	-1	0.152	-1	0.154	0
Working FT	0.533	0	0.558	5	0.533	0	0.536	1	0.535	0	0.533	0	0.533	0
Working FT if $\text{MH}_1 = 1$	0.314	0	0.352	12	0.323	3	0.329	5	0.321	2	0.311	-1	0.315	0
Avg MH	3.959	0	4.326	9	4.010	1	3.967	0	3.961	0	3.956	0	3.959	0
Avg MH if $\text{MH}_1 = 1$	2.476	0	2.855	15	2.633	6	2.510	1	2.487	1	2.450	-1	2.479	0

Notes:

· Sim 1: Remove all costs of therapy (utility and monetary costs and mismatch probability)

· Sim 2: Equate therapy costs to medication costs (utility and monetary costs)

· Sim 3: Zero out employment and therapy utility interaction

· Sim 4: Equate therapy-employment interaction with medication-employment interaction

· Sim 5: Remove possibility of mismatch

· Sim 6: Remove monetary cost of therapy

Figure 1: MH Diagnosis Over Time

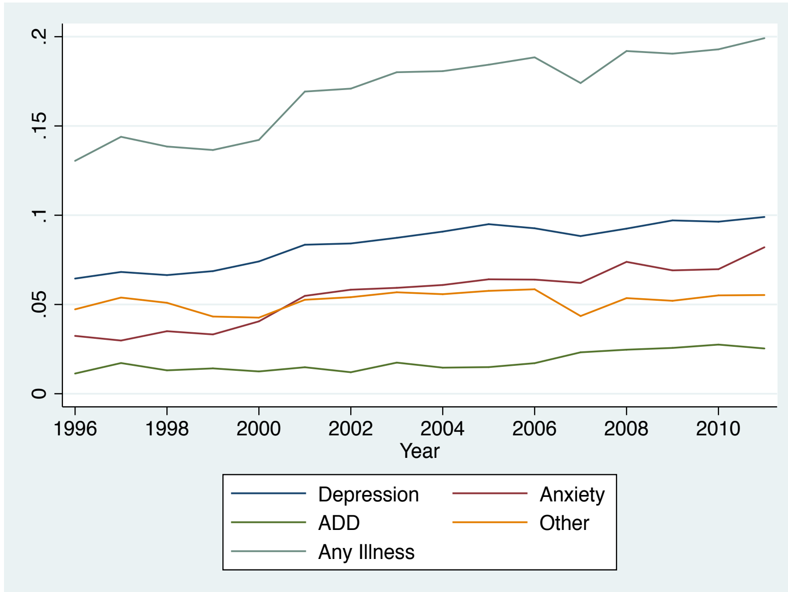


Figure 2: MH Treatment Choices Over Time

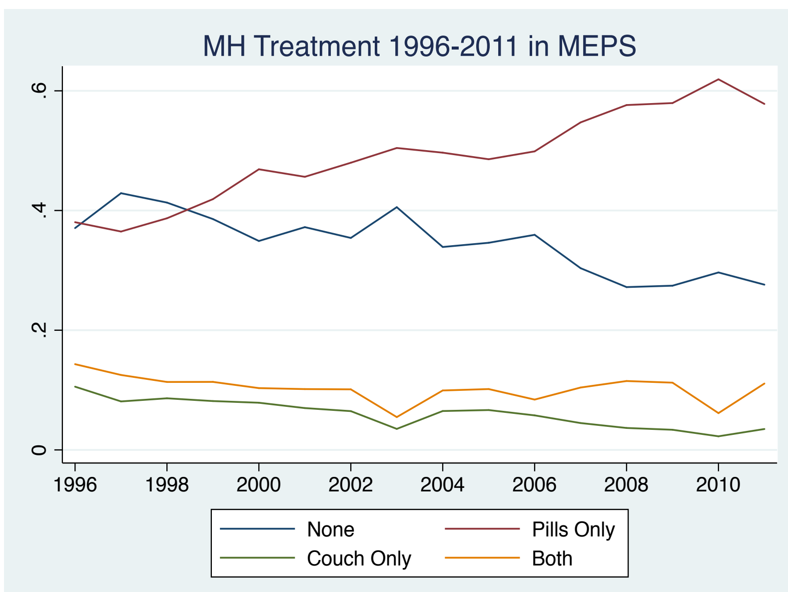


Figure 3: The Distribution of Period Lengths in MEPS

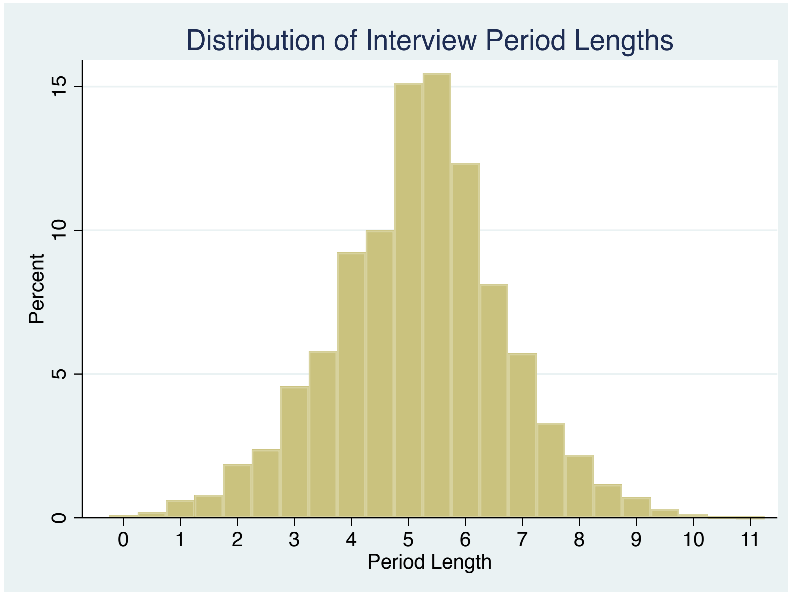


Figure 4: Therapist Mismatch

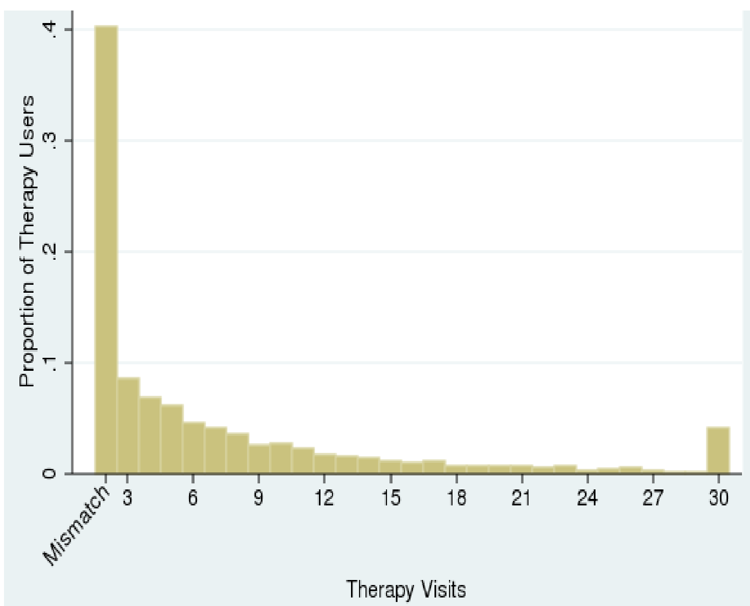


Figure 5: Human Capital Approximation

Appendix

Table A.I: Sample Statistics Across Limiting Samples

	Sample A	Sample B	Sample C
Demographics			
Male	0.457	0.456	0.457
Age	42.623	42.850	42.884
Live in M.S.A.	0.823	0.819	0.821
Married	0.616	0.614	0.629
Northeast	0.161	0.157	0.156
Midwest	0.203	0.203	0.202
South	0.376	0.379	0.380
West	0.260	0.262	0.262
Black	0.155	0.155	0.154
Other (non-white)	0.071	0.071	0.071
Schooling & Employment			
Years of School	12.728	12.679	12.693
Employed	0.736	0.727	0.742
Hourly Wage	17.184	17.198	17.357
Treatment Decisions			
Psychotherapy	0.018	0.019	0.013
Medication	0.073	0.077	0.065
Conditions			
Depression/Anxiety	0.081	0.084	0.079
Individuals	114,267	100,481	98,056
Observations	457,068	385,615	376,234

Notes: Sample A indicates MEPS participants from 1996-2012 between the ages of 22 and 64, excluding Round 1. Sample B eliminates from Sample A all periods with a length less than three months and greater than 8 months, as well as any individual with an excluded month in rounds two, three, or four. Sample C eliminates from Sample B individuals with severe mental disorders.