Testing a Macroeconomic Model of Mental Health: Evidence from Micro-Level Causal Inference*

Chihiro Watanabe

Hitotsubashi University

January, 2026 (revised: April, 2026)

Abstract

This Master's thesis replicates a macroeconomic structural model of mental health and evaluates its theoretical implications using micro-level data. The model treats mental health as an individual state variable that shapes life-cycle decisions through cognitive distortions. The model is replicated in Julia, clarifying key numerical solution methods. Using data from the Panel Study of Income Dynamics (PSID), the empirical analysis exploits variation in the timing of state-level mental health parity laws as a policy shock within a staggered difference-in-differences framework. The reduced-form results provide partial and nuanced evidence. Specifically, while estimated effects on labor supply and risky investment are broadly consistent with the model's quantitative implications, other outcomes remain less clear. Overall, the findings highlight both the empirical relevance of key structural implications and the difficulty of detecting reduced-form responses to parity law policies using observational data.

Keywords: mental health; heterogeneous-agent macroeconomic model; staggered DiD

*I am grateful to my advisors, Motohiro Sato, Takashi Oshio, and Shinsuke Suzuki, for their guidance and support. I also thank Masayuki Sawada, Takahiro Toriyabe, and the participants of a workshop organized by Reo Takaku for helpful comments and discussions.

Contents

1 Introduction

Mental disorders, including depression, schizophrenia, and dementia, constitute a major risk that individuals may face over the course of their lives. These disorders have long been recognized as medical illnesses and have been studied primarily within psychiatry and neuroscience. Accordingly, a substantial body of research has focused on identifying their biological mechanisms and developing treatments based on neurobiological processes.

At the same time, there has been growing recognition of the limitations of attempts to understand mental disorders solely through biological reductionism. Mental disorders are expressed not only as neurobiological abnormalities but also through complex interactions with psychological states and social environments. From this perspective, Engel (1977) proposed the Bio-Psycho-Social (BPS) model, which emphasizes the integrated roles of biological, psychological, and social factors in the underlying mechanisms of mental illness. This framework has been widely discussed internationally, particularly since the 1980s (Brückner, 2023), and it highlights the importance of analyzing mental disorders in the context of everyday social life.

While the BPS model provides an influential framework, analyzing how biological, psychological, and social factors jointly shape individual health status or behavior requires a more formal, theory-based approach. One natural candidate is the use of mathematical models, which allow the interaction of multiple factors to be studied in a disciplined and transparent way.

From the perspective of quantitative modeling of mental disorders, computational psychiatry has emerged as an important methodological approach to formalizing psychiatric phenomena. This field conceptualizes the brain and mind as information-processing systems and seeks to describe how external inputs are internally processed and translated into behavior using formal mathematical models. By matching behavioral data, reaction times, and neural measurements to computational models, researchers aim to estimate latent parameters governing cognitive processes, such as perception, learning, inference, and valuation. By comparing the estimated parameters across individuals, computational psychiatry seeks to identify and quantify cognitive and computational abnormalities that characterize mental disorders.

Despite its methodological sophistication, much of the existing work in computational psychiatry has focused on explaining behavior observed in relatively simple, laboratory-based experimental tasks. As a result, direct extensions of this approach to the analysis of long-term, complex decision-making in real-world social contexts remain limited.

In this respect, economic structural models can play a complementary role. Economics provides a formal framework for analyzing decision-making in social environments by explicitly modeling constraints, expectations about the future, and intertemporal trade-offs. Consumption, labor supply, savings, and risk-taking decisions can be jointly analyzed within a dynamic setting. In particular, the frameworks developed by Abramson et al. (2024) and Cronin et al. (2025), which incorporate mental health as an explicit state variable and allow cognitive distortions to affect economic choices, offer a theoretical foundation for extending the insights of computational psychiatry to social and economic behavior beyond the laboratory environment.

Building on this literature, this Master's thesis takes a first step toward understanding the relationship between mental disorders and socioeconomic behavior by focusing on the structural model proposed by Abramson et al. (2024). The thesis replicates their model using Julia, with careful attention to the numerical implementation and underlying algorithms. Furthermore, using micro-level data and causal inference methods, it empirically assesses mechanisms suggested by the theoretical model. This study is positioned as a foundational exercise in bridging computational psychiatry and economics, with the aim of analyzing how mental disorders shape decision-making in social and economic life.

The remainder of the thesis is organized as follows. Chapter 2 reviews the related literature. Chapter 3 presents the theoretical model of Abramson et al. (2024) and its replication in Julia.

Chapter 4 evaluates the implications of the model using causal inference methods. Chapter 5 concludes and discusses directions for future research. Appendix A reports robustness checks based on an alternative definition of treatment variables, while Appendix B provides a conceptual discussion of Karl Jaspers' distinction between Erklären (explanation) and Verstehen (understanding), illustrating the broader methodological context of the analysis.

The Julia and Stata code for this research will be available on my GitHub repository: `https://github.com/Chihiro2000GitHub/mentalhealth-macro-did`

2 Literature

Research on health using economic structural models can be traced back to Grossman (1972), which established what is now known as the Grossman model as one of the canonical frameworks in health economics. The Grossman model conceptualizes an individual's health status as a form of *capital*, providing a foundation for modeling the demand for health and medical services. In this framework, the accumulation of health capital directly increases utility, while deterioration in health reduces the amount of time available for work and leisure. The model assumes that individuals face a minimum threshold of health capital, below which survival is no longer possible. Individuals can invest in their health capital, mainly through medical utilization, in order to offset natural health depreciation over time. Individuals solve a dynamic optimization problem that jointly determines consumption, health investment, and time allocation, subject to both budget and time constraints, while taking into account the expected future path of health. Chapter 3 of Zweifel et al. (2009) provides a clear and comprehensive introduction to the Grossman model.

Cronin et al. (2025) is the first study to explicitly apply the Grossman framework to the context of mental health. Their analysis focuses on the contrast between talk therapy and antidepressant use. Despite evidence from clinical studies suggesting that talk therapy yields higher average treatment effects, antidepressants are far more commonly utilized in practice. To account for this discrepancy, their theoretical framework models treatment choice by explicitly allowing individuals to choose between talk therapy and antidepressants. The model further incorporates a non-monetary disutility associated with talk therapy, potentially reflecting stigma or psychological burdens, as a key factor discouraging its utilization. Using data from the Medical Expenditure Panel Survey (MEPS), the authors structurally estimate a dynamic discrete choice model of treatment and other economic variables.

In macroeconomics, a growing body of research has recently emerged under the label of *macro-health*, which employs life-cycle structural models incorporating individual health status to analyze the relationship between health and economic behavior. Representative studies include French and Jones (2011), Capatina (2015), and De Nardi et al. (2025). In general, this literature primarily emphasizes the role of individual heterogeneity in health status and medical expenditure risk within life-cycle models, and studies their interactions with labor supply, savings, retirement behavior, and social insurance systems.

From a macro-health perspective, Abramson et al. (2024) focus explicitly on mental illness. In their model, an agent's mental health state gives rise to distortions in decision-making. In particular, the model captures the loss of time available for work and leisure due to rumination, as well as a form of *negative thinking*, whereby agents evaluate future risks using pessimistically biased subjective probability distributions. In particular, this Master's thesis centers on this study and replicates its model in Julia. A detailed description of the theoretical framework and the numerical solution methods is presented in the next chapter.

Related to this line of research, a growing literature known as computational psychiatry seeks to formalize psychiatric phenomena using mathematical and computational models of cognitive and neural processes, including perception, learning, inference, and decision-making. As an overview of this field, Huys et al. (2016) provide a comprehensive introduction, while textbooks such as Redish and Gordon (2016) and Series (2020) offer a systematic presentations of its theoretical foundations. Broadly, computational psychiatry encompasses a range of modeling approaches, including biophysical models, neural network models, reinforcement learning models, and Bayesian inference models, which differ in their levels of abstraction and explanatory targets.

3 The Model and its Replication

This chapter presents the economic model of mental health developed by Abramson et al. (2024), and describes its numerical replication.[1] The aim of this chapter is to accomplish two related objectives. First, it provides a clear description of the theoretical framework, highlighting how mental health affects individual decision-making through time constraints and distorted expectations. Second, it explains the computational approach used to solve and replicate the model, making explicit several implementation details that are not fully described in the original paper. The chapter begins with an overview of the model and then moves on to the individual optimization problem, the mental health dynamics, the recursive formulation of the model, the equilibrium concept, the numerical solution method, and the replication results.

3.1 Overview of the Model

The model developed by Abramson et al. (2024) is a life-cycle structural model in which mental health is incorporated as an individual state variable that directly affects economic decision-making. Individuals face uncertainty over labor productivity, future mental health states, and returns from risky asset investment over the life cycle. They make forward-looking decisions regarding consumption, labor supply, savings, medical utilization, and risky asset investment. The key innovation of the model lies in its characterization of mental illness not merely as a source of reduced well-being, but as a factor that systematically distorts decision-making through both time constraints and subjective beliefs about future risks.

Mental health affects individual behavior through two main channels. First, poorer mental health reduces the amount of effective time available for work and leisure due to rumination.

[1] The Julia code used for this replication will be available on my GitHub repository: https://github.com/Chihiro2000GitHub/mentalhealth-macro-did

This time loss directly constrains individual choices for work, leisure, and medical utilization. Second, mental illness is assumed to distort individuals' expectations about future outcomes. In particular, agents with worse mental health evaluate future risks using pessimistically biased subjective probability distributions, leading them to overestimate adverse events relative to the objective probabilities.

The evolution of mental health over the life cycle is modeled as a stochastic process that depends on prior mental health status and medical treatment decisions. Individuals may choose whether to utilize mental health treatment, which affects the transition probabilities of future mental health states. Importantly, treatment decisions are made under uncertainty and are themselves influenced by distorted beliefs. By underestimating treatment effectiveness, individuals are less likely to seek care, which in turn worsens mental health and further reinforces pessimistic beliefs.

The model is set in an incomplete-markets environment with borrowing constraints and uninsurable idiosyncratic risk, following the canonical framework in the heterogenous-agent macro literature (e.g., Aiyagari, 1994). However, aggregate prices are treated as given in order to abstract from general equilibrium feedbacks and focus on individual behavior. By combining a life-cycle framework with state-dependent belief distortions, the model provides a unified structure for analyzing how mental illness shapes economic outcomes over time.

3.2 Individual's Problem

3.2.1 Timing and State Variables

At the beginning of age t, an individual is characterized by wealth a_t, lagged idiosyncratic probability component ν_{t-1}, mental health status m_t, exogenous treatment-access type ω, and age t.[2] Within a period, individuals make decisions under uncertainty. In particular, they choose a job j prior to the realization of idiosyncratic labor productivity. After productivity is realized, individuals choose consumption c_t, labor supply n_t, savings s_t, portfolio allocation

[2] This ω governs whether treatment is available in particular for mild illness.

k_t, and whether to utilize medical treatment τ_t $(= 0, 1)$.

Given this timing structure, labor income is determined as follows. Before idiosyncratic productivity is realized, individuals implicitly choose a job characterized by an earning threshold y_j. Each job j requires a minimum level of effective productivity to generate positive income. Labor income is therefore given by the following threshold earning function:

$$y(zn, j) = \begin{cases} y_j & \text{if } zn \geq y_j, \\ 0 & \text{otherwise.} \end{cases}$$

Here, z denotes idiosyncratic labor productivity, incorporating the effects of mental health, effective labor input, age-related productivity profiles, and an idiosyncratic productivity component ν.[3] n denotes hours worked, implying that zn corresponds to effective productivity. Importantly, j indexes a discrete earning opportunity characterized by the income requirement y_j. This formulation captures the idea that jobs differ in their minimum productivity requirements, and that individuals earn positive income only when their realized effective productivity exceeds the corresponding threshold. Although individuals do not explicitly choose jobs in the optimization problem, this formulation is equivalent to an implicit selection of an earning threshold prior to productivity realization.

3.2.2 Effective Time and Time Loss

Mental illness reduces effective time available for work and leisure through rumination, while treatment utilization also consumes time. Total available time within a period is normalized to one. The effective time endowment is therefore given by

$$\bar{n}(m_t, \tau_t) = 1 - n_r(m_t) - n_\tau \tau_t,$$

where $n_r(m_t)$ denotes time lost due to rumination, which depends on the individual's

[3] This idiosyncratic productivity component ν is a stochastic shock that evolves over time and contributes to changes in an individual's productivity-related state.

mental health status and increases as mental health deteriorates. Treatment utilization $\tau_t \in \{0, 1\}$ represents a discrete decision, and time n_τ is spent on treatment only when $\tau_t = 1$. Leisure is then given by $l_t = \bar{n}(m_t, \tau_t) - n_t$ where n_t denotes hours worked.

3.2.3 Asset Accumulation and Portfolio Return

Given savings s_t, and portfolio share $k_t \in [0, 1]$, next period wealth a_{t+1} satisfies

$$a_{t+1} = s_t \cdot R_t^s(k_t), \quad \text{and} \qquad R_t^s(k_t) = k_t \cdot R_t + (1 - k_t)R_f,$$

where R_t is the risky return and R_f is the risk-free return. The risky return contains a stochastic component: letting $r_t \equiv \log R_t$, the log return decomposes as

$$r_t = r_f + r_p + \upsilon_t,$$

where $r_f \equiv \log R_f$ is the log return on the risk-free asset, r_p is the expected risk premium over the risk-free asset, and υ_t is an innovation drawn from a discretized normal distribution, $\upsilon_t \sim \mathcal{N}(0,\, \sigma_\upsilon^2)$.

Borrowing is limited by

$$s_t \geq \underline{s}$$

where $\underline{s}$ is the borrowing limit.

3.2.4 Budget Constraint

After the realization of idiosyncratic productivity, the individual chooses consumption c_t, work hours n_t, savings s_t, risky-asset participation k_t, and treatment τ_t. The budget constraint is

$$c_t + \varphi_\tau \tau_t + \varphi_k \mathbf{1}_{k_t} + s_t \leq a_t + y_t(z_t n_t, j_t) \tag{1}$$

where φ_τ is the treatment cost and φ_k is a fixed participation cost for investing in risky assets.

3.2.5 Flow Utility

Per-period utility depends on consumption c_t and leisure $\bar{n}(m_t, \tau_t) - n_t$, while mental illness and treatment generate additional utility cost:

$$u(c_t, \bar{n}(m_t, \tau_t) - n_t) - \xi_m(m_t) - \xi_\tau \tau_t$$

where ξ_m captures utility cost of mental illness and ξ_τ captures the stigma cost of treatment.

3.3 Mental Health Dynamics

This section describes the evolution of mental health over the life cycle. Mental health is modeled as a discrete stochastic state that follows a Markov process, with transition probabilities that depend on the individual's current mental health status and treatment decisions. This formulation captures the persistence of mental illness and the role of medical treatment in shaping future mental health outcomes.

3.3.1 Mental Health States and Transitions

Let $m_t \in \mathcal{M}$ denote an individual's mental health status at age t, where $\mathcal{M}$ is a finite set of mental health states ordered by severity. Mental health evolves according to a first-order Markov chain. Conditional on current mental health status, treatment status, and idiosyncratic productivity, the probability of transitioning to the next period's mental health state is given by

$$Pr(m_{t+1} = m' \mid m_t = m, \tau_t = \tau, \nu_t = \nu) = \Gamma_m(\tau, \nu)$$

where $\Gamma_m(\cdot)$ denotes the transition probabilities conditional on current mental health m, treatment τ, and idiosyncratic productivity ν.

Importantly, treatment affects mental health dynamics by shifting transition probabilities toward better future states, without eliminating the possibility of relapse.

3.3.2 State-Dependent Beliefs and Subjective Transition Probabilities

While mental health evolves according to objective transition probabilities, individuals form subjective beliefs about future mental health dynamics. These beliefs depend on current mental health status. In particular, worse mental health is associated with pessimistically biased expectations regarding future transitions. This form of biased belief is referred to as *negative thinking* in the model.

To illustrate the mechanism, consider a random variable χ taking values in a finite outcome space Ω_χ. Let $w(\chi)$ denote the value associated with outcome χ, and let $q(\chi)$ denote the corresponding objective probability distribution. Individuals exhibiting negative thinking are assumed to form expectations using a subjective probability distribution. Specifically, they solve:

$$\min_{p} \mathbb{E}_p w(\chi) = \min_{p(\chi)} \sum_{\chi \in \Omega_\chi} p(\chi) w(\chi) \tag{2}$$

Without any restriction, this minimization problem would lead individuals to assign all probability mass to the worst possible outcome. To rule out such extreme pessimism, we impose a total variation constraint that restricts subjective beliefs to remain close to the objective distribution:

$$\frac{1}{2} \sum_{\chi \in \Omega_\chi} |p(\chi) - q(\chi)| \leq \kappa(m) \tag{3}$$

The parameter $\kappa(m)$ governs the extent to which subjective beliefs may deviate from objective probabilities and thus captures the severity of negative thinking. When $\kappa(m) = 0$, subjective and objective probabilities coincide, and no belief distortions is present. As $\kappa(m)$ increases, individuals are allowed to place greater probability mass on adverse outcomes by shifting probability away from more favorable states. In the solution to the minimization problem, individuals assign as much probability as permitted to the lowest-value outcomes, consistent with pessimistically biased expectations.

In the quantitative implementation, the degree of negative thinking is calibrated to vary systematically with mental health status. Mental health is assumed to take a finite number of discrete states, corresponding to healthy, mild, and serious conditions. The parameter $\kappa(m)$ is specified to be increasing in the severity of mental illness, so that individuals in worse mental health states are allowed to form more pessimistically distorted beliefs.

3.4 Recursive Formulation and Bellman Equation

This section provides the recursive formulation of the individual's problem. Given the current state $(a_t, \nu_t, m_t, \omega)$ at age t, individuals choose consumption, labor supply, savings, portfolio allocation, and treatment to maximize current utility plus the discounted next-period value.

3.4.1 Bellman Equation with State-Dependent Subjective Beliefs

The individual's value function satisfies the following Bellman equation:

$$
\begin{aligned}
&V_t(a_t, \nu_t, m_t, \omega) \\
&= \max_{c_t, n_t, s_t, k_t, \tau_t} \left\{ u\left(c_t, \bar{n}(m_t, \tau_t) - n_t\right) - \xi_m(m_t) - \xi_\tau \tau_t + \beta \min_{p_t} \mathbb{E}_{p_t}\left[V_{t+1}(a_{t+1}, \nu_{t+1}, m_{t+1}, \omega)\right] \right\}
\end{aligned}
$$

Current-period utility depends on consumption and leisure, where effective available time is reduced by rumination and treatment. Mental illness directly lowers utility through the disutility term $\xi_m(m_t)$, and treatment ($\tau_t = 1$) entails a stigma cost ξ_τ.

Future wealth evolves according to the asset accumulation equation defined in Section 3.2, where labor productivity, risky asset returns, and mental health evolve stochastically.

3.4.2 Interpretation

The key non-standard element of the Bellman equation is the expectation operator. Unlike the standard formulation with objective beliefs, individuals in this model evaluate future utility under distorted, state-dependent subjective beliefs. In particular, subjective beliefs distort the perceived transition probabilities governing next-period assets a_{t+1}, idiosyncratic labor productivity ν_{t+1}, and mental health m_{t+1}. Negative thinking therefore affects all these stochastic components.

This formulation implies that mental illness affects behavior not only through utility and time constraints, but also through distorted expectations about future outcomes. Individuals experiencing worse mental health think pessimistically about both the evolution of mental health and the returns to economic choices such as treatment, labor supply, and risky investment.

As a result, mental illness may reinforce itself through behavior. Pessimistic beliefs about treatment effectiveness reduce treatment take-up, which in turn worsens mental health and further reinforces pessimistic beliefs. Furthermore, rumination reduces effective time available to seek care and work, while pessimism about productivity and investment returns leads individuals to choose less demanding jobs and safer portfolios. These behavioral responses reduce financial resources available for treatment, thereby reinforcing adverse mental health dynamics over the life cycle.

3.5 Equilibrium Concept

This section defines the stationary equilibrium of the model. The equilibrium concept follows the standard definition used in heterogeneous-agent life-cycle models, with mental health explicitly incorporated as an individual state variable that affects preferences, beliefs, and time allocation.

Definition (stationary equilibrium)

A stationary equilibrium consists of a set of value functions $\{V_t\}_{t=1}^T$, policy functions for consumption, labor supply, savings, portfolio choice, and treatment decisions $\{g_{c,t}, g_{n,t}, g_{s,t}, g_{k,t}, g_{\tau,t}\}_{t=1}^T$, a law of motion for mental health states Γ_m, a sequence of age-specific distributions over individual states μ_t, and a vector of prices, such that

- Given prices, value and policy functions solve households' optimization problem under subjective beliefs defined in Section 3.3.
- Mental health evolves according to the Markov process: $m' = \Gamma_m(m' \mid m, \tau, \nu)$.
- Markets clear in a partial-equilibrium sense.
- The distribution of agents in each age μ_t is stationary in calendar time despite age-dependent individual decisions.

3.6 Numerical Solution and Replication Strategy

This section describes the numerical method used to solve the model and to compute the stationary equilibrium. The solution strategy follows standard approaches in heterogeneous-agent life-cycle models, combining backward induction for individual decision problems with forward simulation of the cross-sectional distribution. The numerical procedure consists of three main steps:

- discretization of the state space and parameter specification,
- solution of the individual optimization problem by backward induction, and
- forward simulation of the distribution of agents until stationarity is achieved.

3.6.1 Discretization and Initialization

Continuous state variables are discretized using finite grids. Idiosyncratic labor productivity follows a stochastic process that is discretized using the Tauchen method (Tauchen, 1986), while asset holdings, savings, portfolio shares, and labor supply are approximated on fixed grids.

The initial distribution of agents is specified exogenously. Newborn agents enter the economy with zero initial assets ($k_0 = 0$) and an initial mental health distribution is given by π_m. Initial productivity is drawn from the stationary distribution implied by the discretized productivity process.

All structural parameters are calibrated prior to the solution of the model and held fixed throughout the computation.

3.6.2 Backward Induction

Given prices and the discretized state space, the individual's dynamic optimization problem is solved by backward induction.

At the terminal age $t = T$, individuals maximize current-period utility only, as continuation values beyond the terminal age are set to zero. This determines the terminal value function V_T and the associated policy functions.

For each preceding age $t = T-1, \cdots, 1$, value and policy functions are computed recursively. For every combination of state variables, all feasible choices are evaluated on the discretized grids. Current-period utility is computed for each choice, and the expected next-period value is obtained using the two-step expectation operator described in Section 3.4, which combines objective expectations over productivity shocks with pessimistically distorted beliefs over future wealth and mental health.

The maximization of the Bellman equation yields age-dependent policy functions for consumption, labor supply, portfolio choice, and treatment decisions.

3.6.3 Forward Simulation and Stationarity

Using the policy functions obtained from backward induction, the distribution of agents is simulated forward in age. Starting from the initial distribution of newly entering cohorts, individual states evolve according to the policy functions and the stochastic processes governing productivity, risky asset returns, and mental health.

The cross-sectional distribution of agents is updated iteratively until convergence. Stationarity is achieved when the distance between successive distributions falls below a predetermined tolerance level ϵ, according to a chosen norm. In particular, convergence is declared when

$$||\mu_{t+1} - \mu_t|| < \epsilon$$

where t here denotes the iteration index, rather than calendar time or age.

In the replication, convergence is assessed using the maximum norm, which provides a conservative criterion for stationarity:

$$\max_{(a,\nu,m)\in\mathcal{X}} |\ \mu_{t+1}(a,\nu,m) - \mu_t(a,\nu,m)\ | < \epsilon$$

where $\mathcal{X}$ denotes the discretized state space.

Once convergence is achieved, the resulting age-specific distributions constitute the stationary equilibrium distribution of the economy.

3.7 Replication Results

This section reports the results of the Julia-based replication of Abramson et al. (2024). All code used for the replication is publicly available in the following GitHub repository: https://github.com/Chihiro2000GitHub/mentalhealth-macro-did

I begin by assessing the quantitative plausibility of the replicated model. Table 1 compares key life-cycle outcomes in the data from Panel Study of Income Dynamics (PSID) with their counterparts implied by the stationary equilibrium of the model. This exercise serves as a validation check, evaluating whether the model reproduces major cross-sectional patterns observed in the data.

In both the data and the model, individuals are classified into three mental health states: *Healthy, Mild,* and *Serious*. On the data side, these categories are constructed from the K6 psychological distress scale, which is discretized into three groups. Details of this classification will be provided in a footnote in Section 4.3. On the model side, mental health is modeled as a three-state Markov process, corresponding directly to these categories.

Consumption, Income, and *Wealth* are reported in units of 1,000 dollars. *Hours* denote weekly hours worked, measured as the fraction of total available time normalized to one. *Risky investment share* represents the fraction of total assets allocated to risky assets (see Abramson et al. (2024) for the precise definition). *Risky participation rate* measures the share of individuals within each group who hold any risky assets. All entries in the table are group means.

Overall, the replicated model reproduces the broad empirical patterns observed in the data. Across mental health states, worse mental health is associated with lower consumption, labor supply, income and wealth, as well as reduced engagement in risky asset markets. The magnitudes implied by the model are generally close to those observed in the PSID.

At the same time, some discrepancies remain. In particular, for the Mild and Serious groups, both the risky investment share and the risky participation rate are noticeably lower in the model than in the data. This gap may reflect remaining differences in variable definitions or aggregation procedures, or it may indicate that some aspects of the calibration or the modeling of risky investment behavior require further refinement.

Table 2 and Table 3 investigate how the two central psychological mechanisms in Abramson et al. (2024), which are *negative thinking* and *rumination*, shape economic and behavioral outcomes.[4] Each table compares the benchmark economy with counterfactual version in which the corresponding mechanism is shut down. In Table 2, negative thinking is eliminated by setting $\kappa(m) = 0$ for all mental health states, so that subjective beliefs coincide with objective transition probabilities. In Table 3, rumination is removed by setting the time-loss parameter to zero ($n_r = 0$), so that mental health no longer reduces the effective time endowment.

[4]In this replication, I omit the summary statistics for the "Δ Hours worked," "Consumption coefficient $\hat{\gamma_c}$," and "Investment coefficient $\hat{\gamma_k}$" reported in Abramson et al. (2024) from Table 2 and Table 3. This choice reflects two considerations. First, the precise definitions of these variables are not fully transparent in the original paper and could not be unambiguously recovered for implementation. Second, these coefficients are not central to the main questions of this Master's thesis.

Table 1: Validation: Averages

	Data			Model		
	Healthy	Mild	Serious	Healthy	Mild	Serious
Consumption	51	48	45	59	47	37
Hours	0.403	0.380	0.357	0.449	0.396	0.357
Income	65	57	48	60	48	38
Wealth	312	232	208	286	256	231
Risky investment share	0.581	0.512	0.466	0.548	0.379	0.241
Risky participation rate	0.662	0.576	0.529	0.514	0.380	0.269

Note: This table compares key life-cycle outcomes in the PSID data with their counterparts implied by the stationary equilibrium of the replicated model. Individuals are classified into three mental health states: Healthy, Mild, and Serious. In the data, these categories are constructed from the K6 psychological distress scale, discretized into three groups (see a footnote in Section 4.3 for details). In the model, mental health is represented as a three-state Markov process corresponding to these categories. Consumption, Income, and Wealth are reported in units of 1,000 dollars. Hours denote weekly labor supply, measured as the fraction of total available time normalized to one. Risky investment share is the fraction of total assets allocated to risky assets, Risky participation rate is the share of individuals within each group who hold any risky assets. All entries are group means.

These counterfactuals isolate the quantitative role of each channel by holding all other aspects of the environment fixed. The comparisons therefore clarify which dimensions of behavior are primarily driven by pessimistically distorted beliefs and which arise from time loss due to rumination.

Table 2 reveals that eliminating negative thinking leads to substantial changes in behavior, although some of these responses differ from those reported in Abramson et al. (2024). In the present replication, hours worked decline slightly when negative thinking is removed, whereas the original paper reports a small increase in labor supply. Similarly, while income is largely unchanged in Abramson et al. (2024), the replication produces a noticeable decline, particularly for individuals in the Serious mental health state. These discrepancies suggest that the labor-income channel is quantitatively sensitive to details of the numerical implementation and calibration.

At the same time, several core implications of the original model are preserved. First, in the absence of negative thinking, the motive for precautionary saving weakens, leading to lower wealth across mental health states. Second, because individuals no longer evaluate

Table 2: Effects of Negative Thinking

	Benchmark			No negative thinking $\kappa = 0$		
	Healthy	Mild	Serious	Healthy	Mild	Serious
Hours worked	0.449	0.396	0.357	0.452	0.391	0.342
Income (in thousands)	60	48	38	61	44	30
Wealth (in thousands)	286	256	231	278	236	194
Risky investment share	0.548	0.379	0.241	0.544	0.442	0.340
Risky participation rate	0.514	0.380	0.269	0.502	0.426	0.350
Treatment shares	0.000	0.551	0.807	0.000	0.326	0.477

Note: This table reports a counterfactual experiment that isolates the effects of negative thinking. The Benchmark columns correspond to the baseline economy, while the No negative thinking columns set $\kappa(m) = 0$ for all mental health states, so that subjective beliefs coincide with objective transition probabilities. All other parameters and model features are held fixed. Outcomes are reported by mental health state (Healthy, Mild, and Serious). Hours worked denote weekly labor supply as a fraction of total available time. Income and Wealth are measured in thousands of dollars. Risky investment share is the fraction of total assets allocated to risky assets, and Risky participation rate is the share of individuals who hold any risky assets. Treatment shares report the fraction of individuals who utilize treatment. All entries are group means in the stationary equilibrium.

investment returns pessimistically, both the risky investment share and the risky participation rate rise, especially among the Mild and Serious groups. Third, eliminating negative thinking reduces the perceived cost of remaining untreated, weakening the incentive to seek care. As a result, treatment shares fall markedly in the counterfactual economy.

In summary, these patterns confirm that, despite quantitative differences in labor and income responses, the replication captures the central quantitative role of negative thinking emphasized in Abramson et al. (2024): pessimistically distorted beliefs generate precautionary saving, discourage risk-taking, and sustain treatment demand, thereby shaping economic behavior through expectations about future outcomes.

Table 3, in contrast, isolates the effects of rumination. Removing time loss equalizes labor supply across mental health states: hours worked become nearly identical for Healthy, Mild, and Serious individuals. As a consequence, income and wealth gaps across mental health states shrink substantially. Individuals especially in worse mental health become markedly wealthier than in the benchmark economy. This increase in wealth, in turn, raises engagement in risky asset markets along both margins: the risky investment share (the intensive margin)

Table 3: Effects of Rumination

	Benchmark			No rumination $n_r = 0$		
	Healthy	Mild	Serious	Healthy	Mild	Serious
Hours worked	0.449	0.396	0.357	0.449	0.449	0.450
Income (in thousands)	60	48	38	60	55	50
Wealth (in thousands)	286	256	231	286	263	243
Risky investment share	0.548	0.379	0.241	0.545	0.388	0.252
Risky participation rate	0.514	0.380	0.269	0.513	0.387	0.277
Treatment shares	0.000	0.551	0.807	0.000	0.438	0.618

Note: This table reports a counterfactual experiment that isolates the effects of rumination. The Benchmark columns correspond to the baseline economy, while the No rumination columns set $n_r = 0$ for all mental health states, so that no individuals in worse mental health states lose effective available time. All other parameters and model features are held fixed. Outcomes are reported by mental health state (Healthy, Mild, and Serious). Hours worked denote weekly labor supply as a fraction of total available time. Income and Wealth are measured in thousands of dollars. Risky investment share is the fraction of total assets allocated to risky assets, and Risky participation rate is the share of individuals who hold any risky assets. Treatment shares report the fraction of individuals who utilize treatment. All entries are group means in the stationary equilibrium.

and the risky participation rate (the extensive margin) both increase.

Treatment behavior also changes. In the no-rumination economy, treatment shares fall relative to the benchmark. Although eliminating time loss mechanically frees up time that could be devoted to treatment, the dominant effect appears to operate through the cost side: when rumination is absent, the economic and psychological costs associated with poor mental health are substantially reduced. As a result, the incentive to seek treatment weakens, and fewer individuals choose to utilize care.

Importantly, these responses are broadly consistent with the findings reported in Abramson et al. (2024): the replication exhibits no major quantitative abnormalities or sign reversals in the effects of rumination across outcomes.

Table 4 presents a counterfactual simulation in which out-of-pocket costs are eliminated by setting $\varphi_\tau = 0$ in the equation (1). The largest effect of this intervention appears in treatment behavior. Treatment shares rise substantially for individuals in the Mild and Serious mental health states, increasing from 0.551 to 0.639 and from 0.807 to 0.936, respectively. By contrast, treatment remains absent among Healthy individuals in both environments.

Table 4: Counterfactual Simulation: Eliminating Out-of-Pocket Treatment Costs

	Benchmark			Treatment costs $\varphi_\tau = 0$		
	Healthy	Mild	Serious	Healthy	Mild	Serious
Mental health shares	0.876	0.091	0.032	0.881	0.089	0.030
Treatment shares	0.000	0.551	0.807	0.000	0.639	0.936
Hours worked	0.449	0.396	0.357	0.449	0.392	0.346
Income (in thousands)	60	48	38	60	49	41
Wealth (in thousands)	286	256	231	285	262	249
Risky investment share	0.548	0.379	0.241	0.544	0.391	0.261
Risky participation rate	0.514	0.380	0.269	0.511	0.390	0.290

Note: This table reports a counterfactual simulation in which out-of-pocket treatment costs are eliminated by setting $\varphi_\tau = 0$. The Benchmark columns correspond to the baseline economy, while the Treatment costs $\varphi_\tau = 0$ columns report outcomes under zero treatment costs. All other parameters and model features are held fixed. Outcomes are reported by mental health state (Healthy, Mild, and Serious). Mental health shares denote the stationary population shares in each health state. Treatment shares report the fraction of individuals who utilize treatment. Hours worked denote weekly labor supply as a fraction of total available time. Income and Wealth are measured in thousands of dollars. Risky investment share is the fraction of total assets allocated to risky assets, and Risky participation rate is the share of individuals who hold any risky assets. All entries are group means in the stationary equilibrium.

Changes in mental health composition are present but modest. The share of Healthy individuals increases slightly, while the shares of Mild and Serious states decline correspondingly, indicating a limited improvement in aggregate mental health.

Economic outcomes respond only weakly. Hours worked remain almost unchanged for the Healthy group and decline slightly for Mild and Serious individuals. Income and wealth exhibit small increases among those in worse mental health states, but the magnitudes are quantitatively modest. Similarly, both the risky investment share and the risky participation rate increase slightly for Mild and Serious individuals, while remaining nearly unchanged for the Healthy group.

Overall, the table shows that eliminating treatment costs generates a large response in treatment utilization, accompanied by only minor adjustments in labor supply, income, wealth, and portfolio choice. The outcomes of this counterfactual simulation are treated as quantitative implications of the structural model and will be evaluated in the next chapter through micro-level causal inference.

4 Model Assessment by Causal Inference

This chapter evaluates the implications of the structural model constructed and replicated in Chapter 3 using microeconometric methods of causal inference. While structural models allow researchers to explicitly describe behavioral mechanisms, they inevitably rely on modeling assumptions and parameter choices, and therefore involve the risk of model misspecification. Accordingly, this chapter conducts a causal analysis based on exogenous policy shocks and examines the extent to which the direction and magnitude implied by the structural model are supported by empirical data.[5]

4.1 Motivation: Structural Models and Their Limitations

The primary advantage of employing a structural model lies in its ability to decompose the difficulties faced by individuals with mental disorders into factors attributable to symptoms themselves and those arising from social constraints. Moreover, by counterfactually altering parameters or model settings, structural models enable the evaluation of the effects of events such as policy interventions, advances in medical technology, and changes in social stigma at the level of underlying mechanisms.

At the same time, the outcomes of such counterfactual simulations depend critically on whether the model assumptions adequately capture reality. For this reason, as a complementary approach to validating the structural model, it is important to estimate causal effects in a manner that relies as little as possible on structural assumptions, and to assess the consistency between theoretical predictions and empirical evidence.

4.2 Structural Implications to Be Assessed

Before turning to the empirical strategy, it is useful to clarify which implications of the structural model are being assessed. In the framework of Abramson et al. (2024), a natural

[5] The Stata code used for this analysis will be available on my GitHub repository: https://github.com/Chihiro2000GitHub/mentalhealth-macro-did

policy experiment is a reduction in the private out-of-pocket costs of mental health treatment. Such a policy directly targets affordability and corresponds to institutional changes that reduce patients' financial burden when seeking mental health care.

The structural implications to be assessed in this chapter are directly drawn from the counterfactual simulation reported in Table 4 of Chapter 3. In particular, the simulation analyzes an affordability-based intervention through a counterfactual experiment that sets the out-of-pocket treatment cost to zero ($\varphi_\tau = 0$; see the equation (1) in Chapter 3). This counterfactual analysis yields clear quantitative implications. While eliminating treatment costs substantially increases treatment utilization, particularly among individuals with mild and serious mental illness, the effects on other economic variables, including labor supply, income, wealth accumulation, and risky investment behavior, are small or even unchanged.

The structural model therefore implies that policies operating primarily through reduction in treatment costs are unlikely to generate large average responses in economic behavior with the exception of health expenditure. The empirical analysis that follows is designed to assess these implications using micro-level causal inference.

4.3 Empirical Strategy: Staggered Difference-in-Differences

To empirically assess the implications of reducing treatment costs discussed above, this study uses the introduction of mental health parity laws across U.S. states as an exogenous policy shock. Before turning to the details of the parity laws, this section first outlines the econometric methodology used to identify their causal effects.

Causal effects are estimated using the staggered difference-in-differences approach proposed by Callaway and Sant'Anna (2021). This method is designed to consistently estimate average treatment effects on the treated (ATT) even when the timing of treatment adoption differs across units. In contrast to the structural life-cycle model analyzed in the previous chapter, which necessarily imposes a number of parametric assumptions and may therefore be vulnerable to model misspecification, the empirical strategy adopted here places fewer restrictions on the analysis. Accordingly, the staggered difference-in-differences analysis serves as a complementary and transparent validation exercise, providing reduced-form evidence that helps assess the empirical plausibility of the structural model under weaker identifying assumptions.

As the treatment variable, I use the introduction of mental health parity laws across U.S. states, exploiting variation in the timing of policy implementation across states for identification. This adoption of state-level mental health parity laws regulates the extent to which health insurance plans are required to provide coverage for mental health services. Broadly speaking, these laws aim to reduce disparities between mental and physical health coverage by limiting differential treatment in benefit design, including visit limits, copayments, deductibles, and annual or lifetime caps. State mental health legislation varies substantially in scope and content. Some states have enacted full parity laws, which require mental health benefits to be provided under conditions similar to those for physical health conditions. Other states have adopted more limited regulations, such as minimum mandated benefits or mandated offering laws, which require insurers to offer mental health coverage or to meet minimum coverage standards but leave insurers with more freedom in setting the details of benefit coverage.

Information on the timing of state-level mental health parity laws is obtained from The National Conference of State Legislatures (2015).[6] The National Conference of State Legislartures (NCSL) provides detailed summaries of state legislation concerning mental health insurance coverage, including parity laws. The definition of treatment timing for state mental health parity laws is not straightforward, as parity legislation varies substantially in scope and content across states and over time. In many cases, laws with different degrees of coverage were enacted in different years within the same state, making it difficult to identify a single, uniquely defined treatment year. Considering this, the baseline analysis adopts a relatively inclusive definition of treatment, following the information provided by the NCSL.

[6] The URL is: https://www.ncsl.org/health/mental-health-benefits

Specifically, a state is classified as treated in the year in which either (i) a law explicitly described as a "Mental Health Parity" (full parity) became effective, or (ii) a law mandating or regulating mental health insurance benefits, such as "Mandated Offering" or "Minimum Mandated Benefits" came into force, as long as the law is not limited solely to alcoholism, drug, or substance abuse coverage.[7] This broader definition helps preserve the availability of not-yet-treated states as valid comparison groups over time, thereby mitigating the loss of usable observations in the staggered difference-in-differences framework. Table 5 summarizes the timing of mental health parity treatment adoption across states in the baseline analysis. The geographic distribution of parity law adoption is illustrated in Figure 1, which displays a map of the United States in which each state is shaded according to its adoption year, with the corresponding year printed within each state. States shaded in similar colors adopted parity laws in similar years. Notably, the map reveals no clear regional clustering in adoption timing, suggesting that treatment timing is unlikely to be driven by geographic patterns. In contrast, as a robustness check, Appendix A adopts a more restrictive definition of treatment timing. Under this alternative specification, a state is classified as treated only if the NCSL explicitly labels the legislation as a "Mental Health Parity" (full parity). States for which no such designation is available are excluded from the treatment group in this specification, resulting in a smaller but more tightly defined sample.

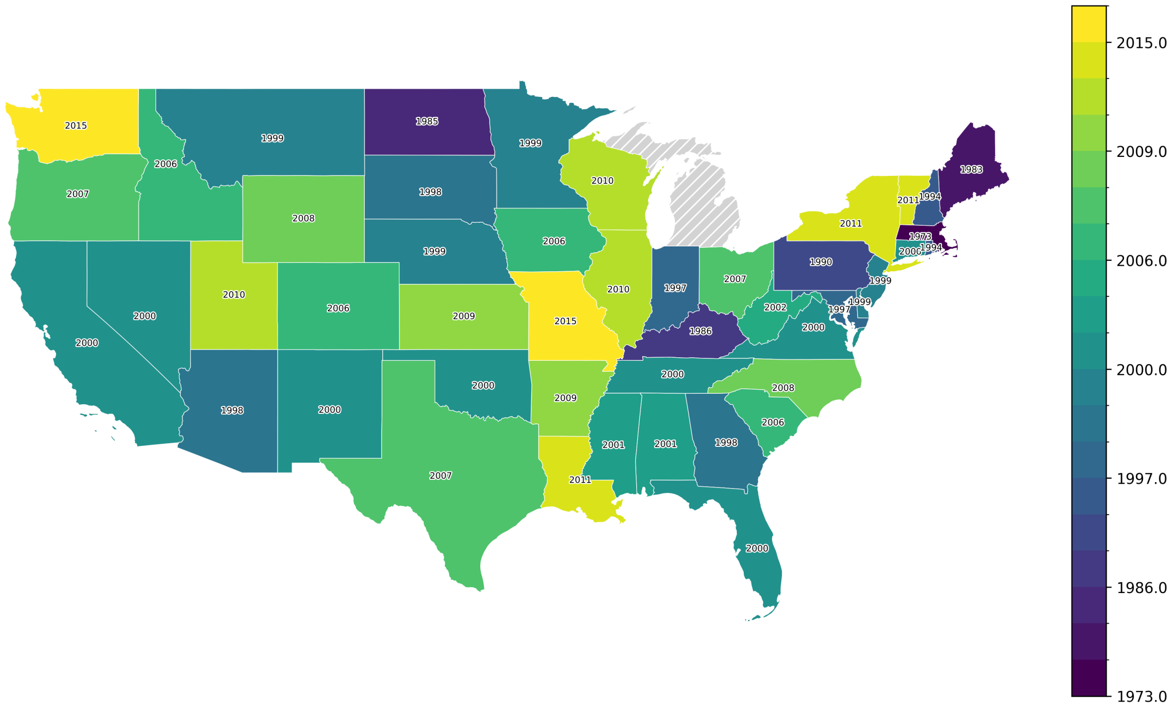
[7] Laws focusing solely on substance use disorders are excluded, as the empirical analysis here is motivated by the theoretical framework of Abramson et al. (2024), which primarily models mental illness in terms of depression and anxiety rather than substance use disorders.

Table 5: Implementation Year of State Mental Health Parity Laws

State	Year	State	Year
Alabama	2001	Montana	1999
Alaska	2009	Nebraska	1999
Arizona	1998	Nevada	2000
Arkansas	2009	New Hampshire	1994
California	2000	New Jersey	1999
Colorado	2006	New Mexico	2000
Connecticut	2000	New York	2011
Delaware	1999	North Carolina	2008
Florida	2000	North Dakota	1985
Georgia	1998	Ohio	2007
Hawaii	1988	Oklahoma	2000
Idaho	2006	Oregon	2007
Illinois	2010	Pennsylvania	1990
Indiana	1997	Rhode Island	1994
Iowa	2006	South Carolina	2006
Kansas	2009	South Dakota	1998
Kentucky	1986	Tennessee	2000
Louisiana	2011	Texas	2007
Maine	1983	Utah	2010
Maryland	1997	Vermont	2011
Massachusetts	1973	Virginia	2000
Michigan	—	Washington	2015
Minnesota	1999	West Virginia	2002
Mississippi	2001	Wisconsin	2010
Missouri	2015	Wyoming	2008

Note: The table reports the first year in which a state-level mental health parity, mandated offering, or minimum mandated benefits became effective, based on information collected from The National Conference of State Legislatures (2015). Michigan did not enact a qualifying parity law during the sample period.

Figure 1: Geographic Distribution of Parity Law Adoption in the U.S. (Full Definition)



Note: Each state is shaded according to the year in which a state-level mental health parity law first became active, using an ordered discrete color scale. The implementation year is printed within each state. Michigan did not enact a qualifying parity law during the sample period and is shown in gray. Alaska and Hawaii are excluded for visual clarity. The figure reveals no apparent regional clustering in adoption timing, supporting the plausibility of random treatment timing in the staggered DiD design.

The analysis is mainly based on the Panel Study of Income Dynamics (PSID), a nationally representative survey of U.S. individuals. Outcome variables are constructed from the PSID and include measures of mental health, namely the Kessler Psychological Distress Scale (K6),[8] three-category mental health classification based on K6 score,[9] and the number of hopeless days in the past 30 days, as well as economic outcomes such as log consumption, log

[8]This scale, developed by Kessler et al. (2002), is a widely used screening instrument for non-specific psychological distress in epidemiological and psychiatric research. The K6 score captures the frequent of negative emotional states experienced over the 30 days, with higher values indicating greater psychological distress. The score is constructed as the sum of responses to six questions asking how often, during the past 30 days, respondents felt (i) nervous, (ii) hopeless, (iii) restless or fidgety, (iv) so depressed that nothing could cheer them up, (v) that everything was an effort, and (vi) worthless. Each item is rated on a five-point scale ranging from 0 (”none of the time”) to 4 (”all of the time”), yielding a total score between 0 and 24.

[9]This category is constructed based on the K6 score, classifying individuals with scores of 0-7 as Healthy, 8-12 as Mild, and 13-24 as Serious. This categorization follows the approach adopted by Abramson et al. (2024) in their empirical analysis.

medical expenditure, hours worked, and the share of risky investment in household financial portfolios.[10] The sample covers the period from 2001 to 2023. Data are collected every two years, following the PSID survey structure. The starting year is determined by data availability, as key mental health measures, including the K6 score, are only available from 2001 onward. The analysis is restricted to household heads in order to maintain consistency with the theoretical framework of Abramson et al. (2024), which models economic decision-making at the household level. Individuals who change their state of residence over time are not excluded from the sample.

Table 6 presents pre-treatment summary statistics by treatment timing, comparing individuals in not-yet-treated and just-treated periods using PSID data. Individuals are classified according to their relative timing of exposure to state mental health parity laws. The not-yet-treated group consists of observations from periods that are at least two relative times (four years) prior to treatment adoption ($t \leq -2$), while the just-treated group corresponds to observations from the period immediately preceding treatment adoption ($t = -1$). Summary statistics are computed using only pre-treatment observations to describe baseline characteristics across treatment-timing groups. Overall, the two groups display broadly similar demographic characteristics, mental health measures, and economic outcomes prior to treatment. This descriptive evidence suggests a high degree of comparability between states approaching treatment adoption and those that have not yet been treated. Formal tests of the parallel trends assumption are conducted in subsequent sections using pre-trends tests. The bottom rows report the number of distinct individuals and person-year observations in each group. Individuals who moved across states during the sample period are excluded from the analysis.

Finally, for the purpose of addressing the issue of multiple hypothesis testing arising from

[10] In the empirical analysis, "Share of Risky Assets" follows the construction in Abramson et al. (2024). Household balance sheets are decomposed into risky assets (stocks, farm and business equity, and home equity), safer assets (vehicles, liquid accounts, and other non-risky assets), and intermediate assets (annuities and IRAs). Let R, S, and M denote these components. "Share of Risky Assets" is defined as $(|R| + 0.5|M|) \,/\, (|R| + |S| + |M| + |\text{liabilities}|)$.

Table 6: Summary statistics by treatment timing (PSID)

	Not-yet-treated mean	Just-treated mean
Age	39.96	41.86
Sex (Male = 1, Female = 2)	1.27	1.23
High School Graduate	0.52	0.53
College Graduate	0.62	0.58
Number of Adults in Household	2.10	2.19
Number of Children in Household	1.92	1.87
K6 Score	3.25	3.16
Mental Health 3 Categories (Healthy, Mild, Serious)	0.16	0.17
Number of Hopeless Days in 30 Days	4.76	4.54
Log Consumption	10.30	10.52
Log Health Expenditure	6.94	7.11
Weekly Hours Worked	37.23	36.33
Share of Risky Assets	0.41	0.38
N individuals	3811	1281
N person-years	10589	1281

Note: This table shows pre-treatment summary statistics by treatment timing using data from the PSID data, restricting the sample to individuals who are ever exposed to the policy during the sample period (i.e., ever-treated). Individuals are classified according to their relative time to treatment. "Not-yet-treated" refers to observations from pre-treatment periods that occur at least two relative times (four years) before policy adoption ($t \leq -2$). "Just-treated" corresponds to observations in the relative time immediately preceding treatment adoption ($t = -1$). Treatment timing is defined at the state level based on the adoption year of parity laws. Summary statistics are calculated using only pre-treatment observations in order to describe baseline characteristics across treatment-timing groups. The bottom rows report the number of distinct individuals and person-year observations in each group.

multiple testing of outcome variables, I will report p-values adjusted for the False Discovery Rate (FDR) following Benjamini and Hochberg (1995).

4.4 Results

4.4.1 Aggregate ATTs

Table 7 reports aggregate (simple average) treatment effects on the treated (ATTs) of state mental health parity laws on mental health and economic outcomes.

For the K6 score, the point estimate is positive but not statistically significant, with a 95 percent confidence interval that includes zero. Similarly, no statistically significant effects are detected for the three-category mental health classification. In contrast, the number of

Table 7: Aggregate ATT and FDR-adjusted p-values

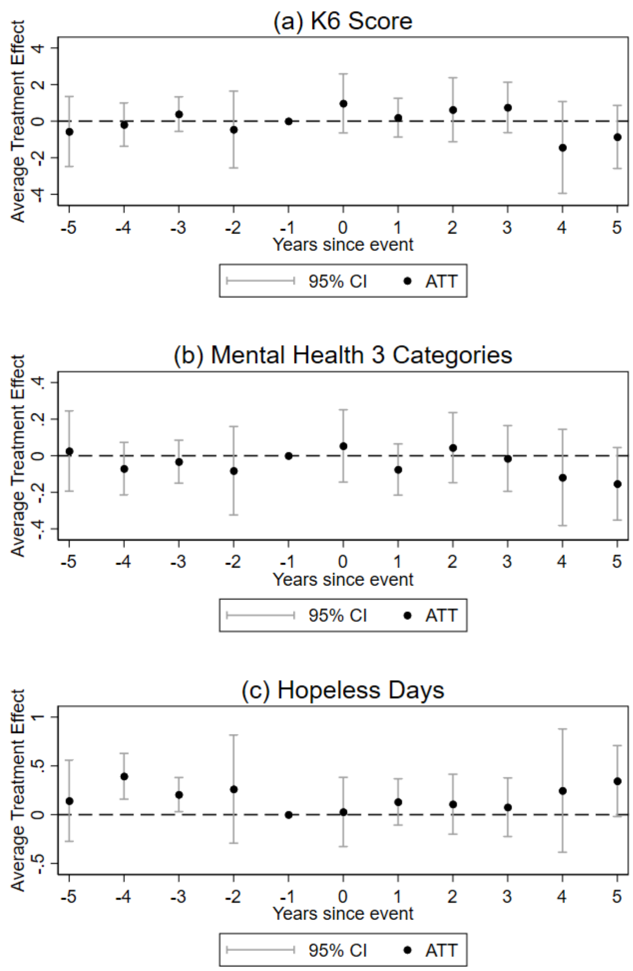
	ATT	SE	95% CI	p	Adjusted p
K6 Score	0.746	0.493	$[-0.220, 1.712]$	0.130	0.455
Mental Health Category	0.049	0.062	$[-0.073, 0.171]$	0.430	0.923
Hopeless Days	-0.222	0.103	$[-0.424, -0.021]$	0.031	0.215
Log Consumption	0.040	0.063	$[-0.083, 0.163]$	0.527	0.923
Log Health Expenditure	0.011	0.139	$[-0.262, 0.283]$	0.939	0.939
Weekly Hours Worked	-0.193	1.368	$[-2.875, 2.488]$	0.888	0.939
Share of Risky Assets	0.010	0.050	$[-0.089, 0.108]$	0.846	0.939

Note: This table reports aggregate (simple) average treatment effects on the treated (ATTs) estimated using the staggered difference-in-differences estimator of Callaway and Sant'Anna (2021). Standard errors are clustered at the state level, and 95% confidence intervals are reported in brackets. Reported p-values test the null hypothesis that the aggregate ATT equals zero. Considering multiple hypothesis testing across outcome variables, p-values are adjusted using the false discovery rate (FDR) procedure of Benjamini and Hochberg (1995). All regressions are weighted by sample weights.

hopeless days in the past 30 days exhibits a negative and statistically significant estimate at the 5 percent level in the unadjusted specification. However, this result does not survive adjustment for multiple hypothesis testing using the false discovery rate (FDR) procedure, suggesting that the evidence for an aggregate effect on hopelessness is not robust once multiple testing is taken into account.

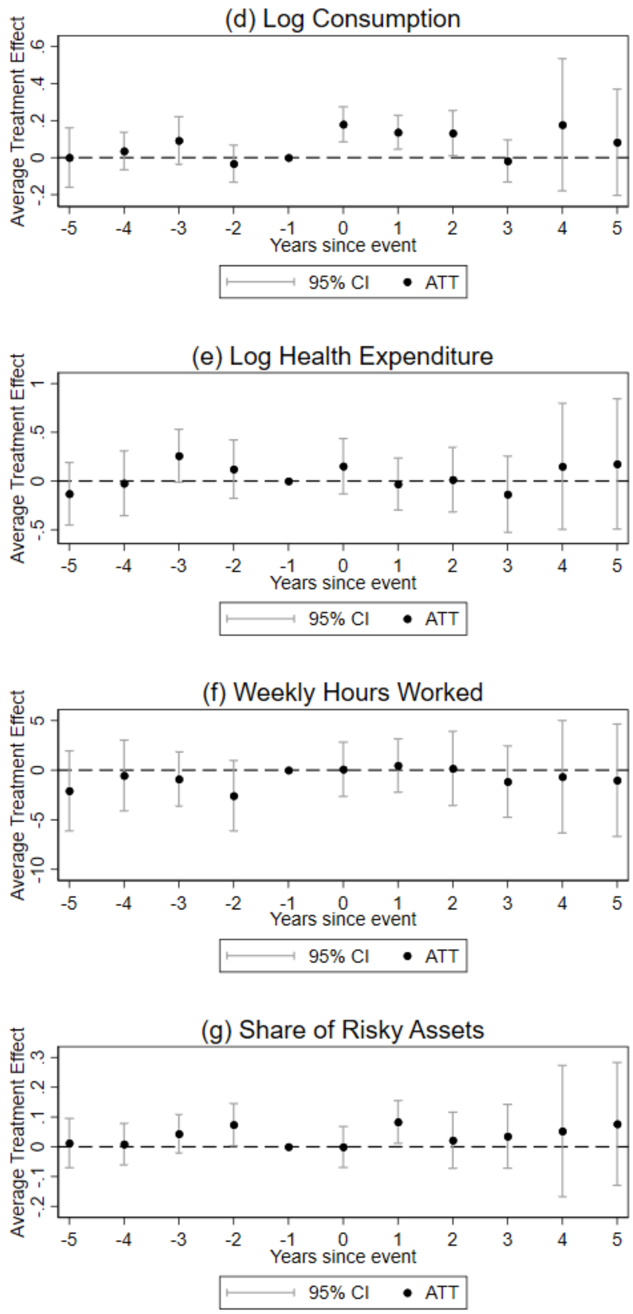
For economic outcomes, including log consumption, log health expenditure, weekly hours worked, and the share of risky assets, the estimated ATTs are statistically insignificant. The corresponding confidence intervals are wide and centered close to zero, indicating no clear evidence of economically meaningful average responses following parity law adoption.

Figure 2: Mental Health Responses to Treatment Adoption



Note: Panels (a)-(c) report event-study estimates of the average treatment effect on the treated (ATT) on mental health outcomes (K6 Score, mental health 3 categories, and hopelessness) using the staggered difference-in-differences estimator of Callaway and Sant'Anna (2021). Event time is defined relative to the year of treatment adoption, and the period immediately before treatment (event time -1) serves as the reference period. All regressions are weighted by sample weights. Points represent point estimates, and vertical bars indicate 95% confidence intervals. Standard errors are clustered at the state level.

Figure 3: Economic Outcome Responses to Treatment Adoption



Note: Panels (d)-(g) report event-study estimates of the average treatment effect on the treated (ATT) on economic outcomes (log consumption, log health expenditure, weekly hours worked, and the share of risky assets) using the staggered difference-in-differences estimator of Callaway and Sant'Anna (2021). Event time is defined relative to the year of treatment adoption, and the period immediately before treatment (event time -1) serves as the reference period. All regressions are weighted by sample weights. Points represent point estimates, and vertical bars indicate 95% confidence intervals. Standard errors are clustered at the state level.

Table 8: Pre-trends Tests

	Pre-trends p-value
K6 Score	0.766
Mental Health 3 Categories (Healthy, Mild, Serious)	0.917
Number of Hopelessness Days	0.515
Log Consumption	0.129
Log Health Expenditure	0.366
Weekly Hours Worked	0.547
Share of Risky Assets	0.096

Note: This table reports p-values from pre-trends tests based on the staggered difference-in-differences estimator of Callaway and Sant'Anna (2021). The null hypothesis is that there are no systematic differences in pre-treatment trends between treated cohorts and not-yet-treated units. Rejection of the null indicates the presence of systematic differences in trends prior to treatment adoption, suggesting that the parallel trends assumption may be less plausible for the corresponding outcome.

4.4.2 Event Study

Figure 2 and Figure 3 present event-study estimates of treatment effects on mental health and economic outcomes, respectively. Across all outcomes, pre-treatment coefficients are close to zero and do not exhibit systematic trends prior to policy adoption.

For mental health outcomes shown in Figure 2, post-treatment estimates fluctuate around zero, with no clear persistent pattern following the adoption of parity laws.

Similarly, Figure 3 shows no systematic post-treatment responses in economic outcomes. Estimates for consumption, health expenditure, labor hours, and portfolio risk-taking remain relatively stable around zero throughout the post-adoption period, and confidence intervals consistently include zero.

4.4.3 Pre-trends Test

Table 8 reports formal pre-trends test p-values for each outcome. The null hypothesis is that there are no systematic differences in pre-treatment trends between treated cohorts and not-yet-treated units. For all outcomes, this null cannot be rejected at 5% significance level. These results support the plausibility of the parallel trends assumption underlying the staggered difference-in-differences design.

4.4.4 Summary

Interpreted through the lens of the structural model, some findings are consistent with the model's implications, while others remain less clear.

The absence of statistically significant effects on labor supply and risky investment behavior can be viewed as broadly consistent with the model's implications that reducing out-of-pocket treatment costs alone should have limited effects on these economic outcomes. In this sense, the reduced-form estimates align with the quantitative predictions of the structural framework.

At the same time, the lack of a statistically significant response in health expenditure may be viewed as a possible mismatch with the model, which allows for increased treatment utilization following a reduction in treatment costs. However, given that none of the outcome variables exhibit clear post-treatment effects, this discrepancy should be interpreted with caution. It may reflect the fact that state-level parity laws constitute a relatively weak affordability shock.

Overall, the results suggest that the empirical design does not provide strong evidence against the structural model, but rather highlights the difficulty of detecting the reduced-form implications of parity law policies using observational data.

These baseline findings are robust to alternative definitions of treatment timing: adopting a stricter definition of mental health parity laws yields qualitatively similar results (see Appendix A).

5 Conclusion and Future Directions

5.1 Summary of the Thesis

This thesis studies the economic consequences of mental health through a structural modeling approach combined with microeconometric causal inference. In Chapter 3, I construct and replicate a life-cycle structural model in which mental health is explicitly incorporated as an individual state variable that affects consumption, labor supply, risky asset investment, and medical utilization through time loss and negative thinking. The model is solved numerically using backward induction and forward simulation, closely following the original framework while clarifying the underlying assumptions and computational procedures.

Building on this structural framework, Chapter 4 assesses the empirical relevance of the model using micro-causal inference. Exploiting variation in the timing of mental health parity law adoption across U.S. states, I implement a staggered difference-in-differences design using individual-level panel data from the Panel Study of Income Dynamics. The analysis examines both mental health outcomes and economic behaviors, while considering multiple hypothesis testing.

Overall, the empirical analysis does not provide strong evidence of statistically significant ATTs of mental health parity laws on either mental health or economic outcomes. Event-study estimates similarly do not reveal clear post-treatment effects. The pre-trends test provides no evidence of difference pre-treatment trends at 5% significance level, supporting the plausibility of the identifying assumptions underlying the staggered difference-in-differences design.

Taken together, these results suggest that, within the scope of the available data and empirical specification, the effects of parity law adoption are limited or difficult to detect. Although no statistically significant effects are detected, this study contributes by illustrating a concrete empirical strategy for evaluating structural models of mental health. The analysis here may serve as a reference point for future work that seeks to bridge structural modeling and reduced-form causal inference.

5.2 Contribution to the Study of Mental Health

This thesis contributes to the study of mental health by proposing an economic framework that explicitly incorporates social dimensions of mental illness into a unified, dynamic model of individual behavior. By treating mental health as an endogenous state variable, the model captures how psychological conditions interact with decisions over consumption, labor supply, risk-taking, and medical utilization, while remaining compatible with established clinical and biological perspectives.

A central contribution of this approach lies in its ability to formalize aspects of mental health that are difficult to capture within purely biomedical models. While biological mechanisms and clinical diagnoses play a fundamental role in shaping mental states, individuals' observed behaviors are also shaped by social norms, stigma, and constraints arising from their perceived social environments. In the proposed framework, these factors are incorporated as state variables or constraints, allowing their interaction with psychological conditions to be analyzed in a disciplined and transparent manner.

From this perspective, the framework developed in this thesis can be interpreted as broadly consistent with the Bio-Psycho-Social (BPS) perspective proposed by Engel (1977), which emphasizes the joint roles of biological, psychological, and social factors in mental health. Importantly, however, the present study does not seek to provide a direct or comprehensive formalization of the BPS model. Rather, it offers a tractable economic representation that provides mathematical and dynamic coherence to the social dimension of mental health, while abstracting from many clinical and biological details.

The contribution of this framework may also be interpreted in light of Karl Jaspers' distinction between *Erklären* (explanation) and *Verstehen* (understanding).[11] In Jaspers' formulation, *Erklären* refers to causal explanation in terms of underlying mechanisms, whereas *Verstehen* emphasizes understanding behavior through the subjective meanings and experiences of individuals. Whereas biomedical models primarily aim at causal explanation in terms of underlying mechanisms, the economic approach adopted here focuses on viewing behavior as the outcome of individuals' subjective beliefs and understanding of the situations they face, given the constraints under which decisions are made. In this limited sense, economic modeling may be regarded as offering a formal language that approaches *Verstehen*, by

[11] See, for example Jaspers (1913/1997), although I consulted the Japanese translation (Jaspers, 1971).

clarifying how individuals make sense of their environments and act on those interpretations, without claiming to fully capture the lived experience of mental states. A more detailed discussion of the distinction between *Erklären* and *Verstehen*, illustrated with clinical case examples, is provided in Appendix B.

The BPS model has the advantage of rejecting dogmatic approaches that rely solely on medical or biological explanations. At the same time, however, it is important to acknowledge that the BPS model itself has been subject to critique within psychiatry. Nassir Ghaemi criticizes the model for offering little more than the general claim that social and psychological dimensions matter alongside biological factors, without providing concrete guidance for clinical practice (Ghaemi, 2004). He also characterizes the BPS model as ”eclecticism”, in other words, an unprincipled approach. He argues that it encourages the indiscriminate use of biological and psychosocial interventions in treatment, ultimately resulting in an ”anything-goes” approach to clinical decision-making.

The economic model used in this thesis can be interpreted as broadly consistent with the BPS perspective. However, critiques of the BPS model, such as those advanced by Nassir Ghaemi, highlight the limitations of adopting it uncritically as theoretical foundation. In light of these critiques, future research should aim to develop economic models that more explicitly reflect ongoing debate within psychiatry, thereby providing a basis for more meaningful interdisciplinary discussion between psychiatry and economics.

5.3 Future Directions

Several promising directions for future research naturally emerge from this thesis. Broadly speaking, these extensions aim to relax key assumptions of existing models and to deepen the analysis of mental health by incorporating political, social, and behavioral dimensions that lie beyond the scope of the present study.

A first direction is to endogenize mental health policy within the structural framework. Existing models of mental health, including those of Abramson et al. (2024) and Cronin et al.

(2025), typically treat policy environment as exogenous. An extension would be to build a political economy model in which agents who differ in their mental health states participate in collective decision-making about policy, in the spirit of Corbae et al. (2009), who study redistribution policy through voting under income heterogeneity. Adapting this framework to heterogeneity in mental health could help explain why policies that appear disadvantageous to individuals with mental illness can emerge under particular and social conditions, as illustrated by extreme historical episodes such as the Nazi T4 program.[12] Moreover, counterfactual simulations could be used to assess whether alternative parameterizations or policy shocks might have prevented such episodes. At the same time, this approach raises concerns about computational feasibility, as models of this class rely on computationally intensive algorithms similar to those used by Krusell and Smith (1998).

A second direction is to model the dynamics of stigma and social prejudice explicitly. The psychiatric literature has identified stigma as an important barrier to seeking psychiatric care (e.g., Corrigan, 2004). The model replicated in this thesis adopts an overlapping-generations (OLG) structure, which provides a natural framework for studying intergenerational transmission. While inheritance is typically modeled in OLG settings as the transmission of wealth, a closely related extension would be to model the intergenerational transmission of stigma or beliefs about mental illness. In addition, socially prevailing stigma could be represented as an aggregate outcome of individual-level beliefs, which in turn affects individuals' decisions to seek psychiatric care. This framework would allow the analysis of how social prejudice feeds back into mental health outcomes and treatment decisions over time, as well as for the evaluation of policy interventions aimed at reducing stigma through counterfactual simulations. Such models may exhibit multiple equilibria, which can complicate the interpretation of the model and its implications.

A third direction concerns the behavioral foundations of individual decision-making. The present framework, like much of the traditional macroeconomic literature, relies on dynamic

[12]For a detailed discussion of this episode from the perspective of psychiatric history, see Brückner (2023).

programming (DP). This implicitly assumes that agents have full knowledge of the probability distributions governing future risks, evaluate expected outcomes accordingly, and make decisions based on those evaluations. While this assumption may be a useful abstraction for studying macroeconomy, it is arguably less appropriate when the object of interest is the behavior of individuals with mental illness, who may face cognitive constraints. Empirical evidence supports this concern. Daw et al. (2011) show that human decision-making cannot be adequately described as arising from complete knowledge of the environment's probabilistic structure. Instead, their results suggest that behavior reflects a mixture of model-based decision-making, where individuals make decisions based on an explicit understanding of the probability distributions in the environment, and model-free learning, in which choices are guided primarily by past reward experiences without exploiting the probability structure. These findings motivate the use of reinforcement learning (RL)[13] as an alternative to DP. In RL models, agents do not possess full knowledge of environmental risk distributions; rather, they learn about their environment structure through trial and error while searching for better actions over time. Importantly, this RL perspective is well established in the literature on computational psychiatry,[14] where it has been widely used to model learning and decision-making processes in both healthy individuals and those with mental disorders. Incorporating RL into an economic model of mental health would pose substantial challenges, but recent work, such as Ma, Stachurski, and Toda (2022), demonstrates that RL-inspired formulations, including Q-function-based representations, can be successfully embedded within macroeconomic dynamic programming problems.

To summarize, these extensions suggest a broader research agenda in which mental health is studied not only as an individual condition, but also as an outcome shaped by political processes, social dynamics, and boundedly rational behavior. Pursuing these directions would further strengthen the connection between economic modeling, empirical evidence, and clinical

[13] Sutton and Barto (2018) provides a comprehensive introduction to the reinforcement learning.

[14] Schultz et al. (1997) provide empirical evidence that dopaminergic activity in the brain encodes a signal consistent with the temporal-difference learning error used in reinforcement learning, thereby motivating the application of reinforcement learning frameworks in computational psychiatry.

insights, and would help clarify the role of mental health in shaping economic and social outcomes over time.

References

Abramson, B., J. Boerma, and A. Tsyvinski (2024): "Macroeconomics of Mental Health," Working Paper 32354, National Bureau of Economic Research.

Aiyagari, S. R. (1994): "Uninsured idiosyncratic risk and aggregate saving," *The Quarterly Journal of Economics*, 109, 659–684.

Benjamini, Y. and Y. Hochberg (1995): "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, 57, 289–300.

Brückner, B. (2023): *Nyūmon seishin igaku no rekishi [Introduction to the history of psychiatry]*, Seiwa Shoten, translated by Hiroyuki Hattori & Keiichi Yamamoto; Supervising editors Toshiya Murai & Takashi Kawashima. Original work: Brückner, B. (2010). *Geschichte der Psychiatrie*. Psychiatrie Verlag.

Callaway, B. and P. H. Sant'Anna (2021): "Difference-in-differences with multiple time periods," *Journal of econometrics*, 225, 200–230.

Capatina, E. (2015): "Life-cycle effects of health risk," *Journal of Monetary Economics*, 74, 67–88.

Corbae, D., P. D' Erasmo, and B. Kuruscu (2009): "Politico-economic consequences of rising wage inequality," *Journal of Monetary Economics*, 56, 43–61.

Corrigan, P. (2004): "How stigma interferes with mental health care." *American psychologist*, 59, 614.

Cronin, C. J., M. P. Forsstrom, and N. W. Papageorge (2025): "What good are treatment effects without treatment? mental health and the reluctance to use talk therapy," *Review of Economic Studies*, 92, 1699–1737.

Daw, N. D., S. J. Gershman, B. Seymour, P. Dayan, and R. J. Dolan (2011): "Model-based influences on humans' choices and striatal prediction errors," *Neuron*, 69, 1204–1215.

De Nardi, M., S. Pashchenko, and P. Porapakkarm (2025): "The lifetime costs of bad health," *Review of Economic Studies*, 92, 1987–2026.

Engel, G. L. (1977): "The need for a new medical model: a challenge for biomedicine," *Science*, 196, 129–136.

French, E. and J. B. Jones (2011): "The effects of health insurance and self-insurance on retirement behavior," *Econometrica*, 79, 693–732.

Ghaemi, S. N. (2004): *The concepts of psychiatry: a pluralistic approach to the mind and mental illness*, JHU Press.

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

Huys, Q. J., T. V. Maia, and M. J. Frank (2016): "Computational psychiatry as a bridge from neuroscience to clinical applications," *Nature neuroscience*, 19, 404–413.

Jaspers, K. (1971): *Seishin Byōrigaku Genron*, Misuzu Shobō, (Japanese translation of *Allgemeine Psychopathologie*, originally published in 1913, translated by Nishimaru, Shihō).

——— (1997): *General Psychopathology*, The Johns Hopkins University Press, (English translation of *Allgemeine Psychopathologie*, originally published in 1913, translated by Hoenig, J. and Hamilton, Marian W.).

Kessler, R. C., G. Andrews, L. J. Colpe, E. Hiripi, D. K. Mroczek, S.-L. Normand, E. E. Walters, and A. M. Zaslavsky (2002): "Short screening scales to monitor population prevalences and trends in non-specific psychological distress," *Psychological medicine*, 32, 959–976.

Krusell, P. and A. A. Smith, Jr (1998): "Income and wealth heterogeneity in the macroeconomy," *Journal of political Economy*, 106, 867–896.

Ma, Q., J. Stachurski, and A. A. Toda (2022): "Unbounded dynamic programming via the Q-transform," *Journal of Mathematical Economics*, 100, 102652.

Ozaki, N., M. Mimura, M. Mizuno, and T. Murai (2021): *Standard Psychiatry (in Japanese)*, Tokyo: Igaku-Shoin, (Japanese textbook).

Redish, A. and J. Gordon (2016): *Computational Psychiatry: New Perspectives on Mental Illness*, Strüngmann Forum reports, MIT Press.

Schultz, W., P. Dayan, and P. R. Montague (1997): "A neural substrate of prediction and reward," *Science*, 275, 1593–1599.

Series, P. (2020): *Computational Psychiatry: A Primer*, MIT Press.

Sutton, R. S. and A. G. Barto (2018): *Reinforcement Learning: An Introduction*, The MIT Press, second ed.

Tauchen, G. (1986): "Finite state markov-chain approximations to univariate and vector autoregressions," *Economics letters*, 20, 177–181.

The National Conference of State Legislatures (2015): "Mental Health Benefits: State Laws Mandating or Regulating," https://www.ncsl.org/health/mental-health-benefits, updated December 30, 2015.

Zweifel, P., F. Breyer, and M. Kifmann (2009): *Health Economics*, Springer Berlin Heidelberg.

Appendix A: Robustness Check with Different Definition of Parity Law

This appendix examines the robustness of the baseline results by adopting a more restrictive definition of treatment timing for state mental health parity laws. Whereas the baseline analysis in Chapter 4 classifies a state as treated when either a full parity law, a mandated offering law, or a minimum mandated benefits law becomes effective, the alternative specification considered here defines the treatment more narrowly. Specifically, a state is classified as treated only if the legislation is explicitly identified as ”Mental Health Parity” by The National Conference of State Legislatures (2015). States without such explicit designation are treated as untreated under this definition. As a consequence, the set of treated states is similar, and the sample size is reduced relative to the baseline analysis.

Table 9 reports the implementation year of state mental health parity laws under this strict definition. Compared to the baseline classification (Table 5), fewer states are identified as treated, reflecting the narrower scope of definition. This change in treatment classification mechanically reduces the number of treated observations available for estimation. Figure 4 displays the geographic distribution of parity law adoption under this strict definition. Compared to the baseline map (Figure 1), a far larger share of states appears in gray, indicating that they did not enact a qualifying parity law. Notably, untreated states are disproportionately concentrated in the South and West, suggesting that whether a state adopted a strict parity law may be geographically correlated. This geographic pattern raises a potential concern about the validity of the parallel trends assumption, as adoption under the strict definition is less evenly distributed across regions than in the baseline analysis.

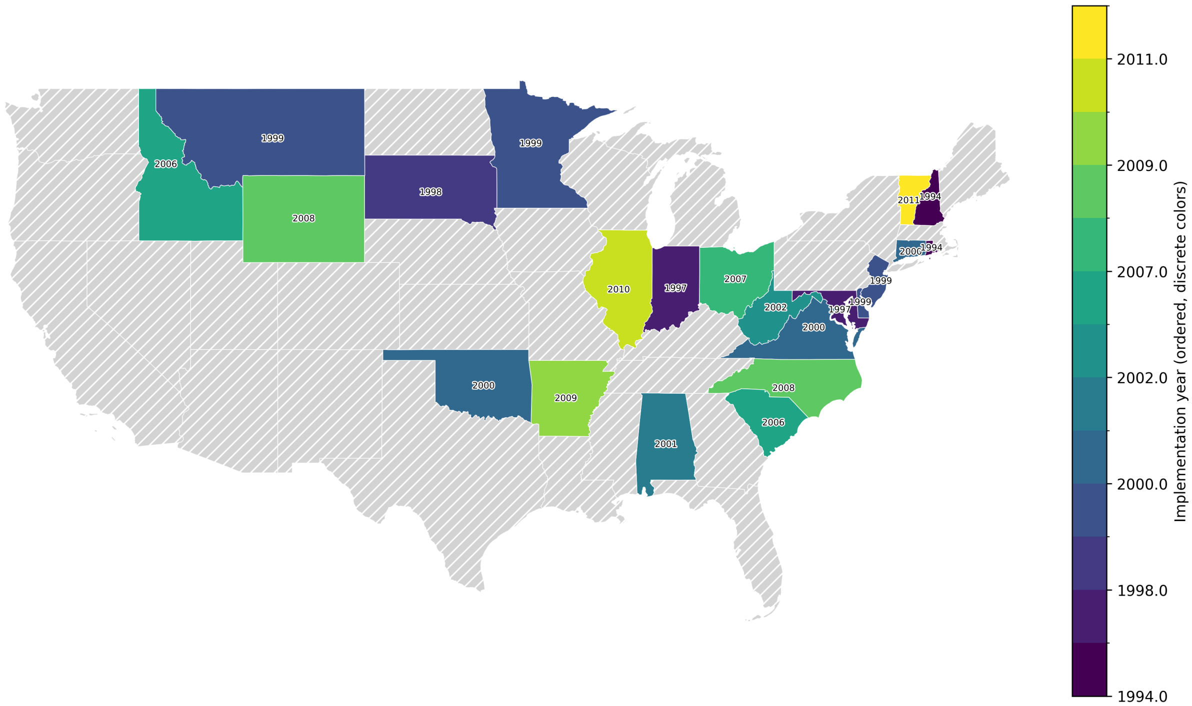
Table 10 presents pre-treatment summary statistics by treatment timing under the strict definition. As in the baseline analysis, individuals in not-yet-treated and just-treated states exhibit broadly similar demographic characteristics, mental health measures, and economic outcomes prior to treatment adoption. While the sample size is smaller than in the baseline

Table 9: Implementation Year of State Mental Health Parity Laws (Strict Definition)

State	Year	State	Year
Alabama	2001	Montana	1999
Alaska	—	Nebraska	—
Arizona	—	Nevada	—
Arkansas	2009	New Hampshire	1994
California	—	New Jersey	1999
Colorado	—	New Mexico	—
Connecticut	2000	New York	—
Delaware	1999	North Carolina	2008
Florida	—	North Dakota	—
Georgia	—	Ohio	2007
Hawaii	1988	Oklahoma	2000
Idaho	2006	Oregon	—
Illinois	2010	Pennsylvania	—
Indiana	1997	Rhode Island	1994
Iowa	—	South Carolina	2006
Kansas	—	South Dakota	1998
Kentucky	—	Tennessee	—
Louisiana	—	Texas	—
Maine	—	Utah	—
Maryland	1997	Vermont	2011
Massachusetts	—	Virginia	2000
Michigan	—	Washington	—
Minnesota	1999	West Virginia	2002
Mississippi	—	Wisconsin	—
Missouri	—	Wyoming	2008

Note: This table reports the implementation year of state mental health parity laws under a strict definition, in which a state is classified as treated only if the legislation is explicitly identified as a “Mental Health Parity Law” by the National Conference of State Legislatures (NCSL). States without a clearly designated parity law under this definition are marked as not treated.

Figure 4: Geographic Distribution of Parity Law Adoption in the U.S. (Strict Definition)



Note: Each state is shaded according to the year in which a state-level mental health parity law first became active, using an ordered discrete color scale. The implementation year is printed within each state. States that did not enact a qualifying law under this strict definition during the sample period are shown in gray. Alaska and Hawaii are excluded for visual clarity. In contrast to the baseline definition, adoption is concentrated in a limited subset of states, highlighting the more selective and uneven diffusion of strict parity legislation.

specification, there is no clear evidence of systematic differences in observable characteristics between treatment-timing groups before policy adoption.

Table 11 reports aggregate average treatment effects (ATTs) estimated using the staggered difference-in-differences estimator of Callaway and Sant'Anna (2021). Consistent with the baseline results reported in Table 7, the estimated ATTs for both mental health outcomes and economic outcomes are statistically insignificant. Confidence intervals are wide and centered close to zero, and none of the estimates are statistically significant after adjusting p-values for multiple hypothesis testing using the false discovery rate (FDR) procedure. Overall, the aggregate effects under the strict definition are qualitatively similar to those obtained in the baseline analysis.

Table 10: Robustness Check: Summary statistics by treatment timing (PSID)

	Not-yet-treated mean	Just-treated mean
Age	39.66	42.55
Sex (Male = 1, Female = 2)	1.25	1.23
High School Graduate	0.51	0.54
College Graduate	0.60	0.56
Number of Adults in Household	2.07	2.21
Number of Children in Household	1.96	1.83
K6 Score	3.38	3.37
Mental Health 3 Categories (Healthy, Mild, Serious)	0.16	0.18
Number of Hopeless Days in 30 Days	4.70	4.46
Log Consumption	10.30	10.58
Log Health Expenditure	7.05	7.16
Weekly Hours Worked	37.91	35.49
Share of Risky Assets	0.42	0.45
N individuals	1684	915
N person-years	4069	915

Note: This table shows pre-treatment summary statistics by treatment timing using data from the PSID data, restricting the sample to individuals who are ever exposed to the policy during the sample period (i.e., ever-treated). Individuals are classified according to their relative time to treatment. "Not-yet-treated" refers to observations from pre-treatment periods that occur at least two relative times (four years) before policy adoption ($t \leq -2$). "Just-treated" corresponds to observations in the relative time immediately preceding treatment adoption ($t = -1$). Treatment timing is defined at the state level based on the adoption year of parity laws. Summary statistics are calculated using only pre-treatment observations in order to describe baseline characteristics across treatment-timing groups. The bottom rows report the number of distinct individuals and person-year observations in each group.

Figure 5 and Figure 6 display event-study estimates for mental health and economic outcomes, respectively. As in the baseline analysis, pre-treatment coefficients are generally close to zero and do not exhibit systematic trends prior to treatment adoption. Post-treatment estimates fluctuate around zero, with no clear pattern following policy implementation. In some event-time periods, treatment effects are not estimated. This reflects the reduced sample size under the strict definition, which limits the availability of treated observations in certain relative-time periods.

Finally, Table 12 reports formal pre-trends test p-values based on the staggered difference-in-differences framework. As in the baseline specification (Table 8), the null hypothesis of no differential pre-treatment trends cannot be rejected at 5% significance level for any outcome.

Table 11: Robustness Check: Aggregate ATT and FDR-adjusted p-values

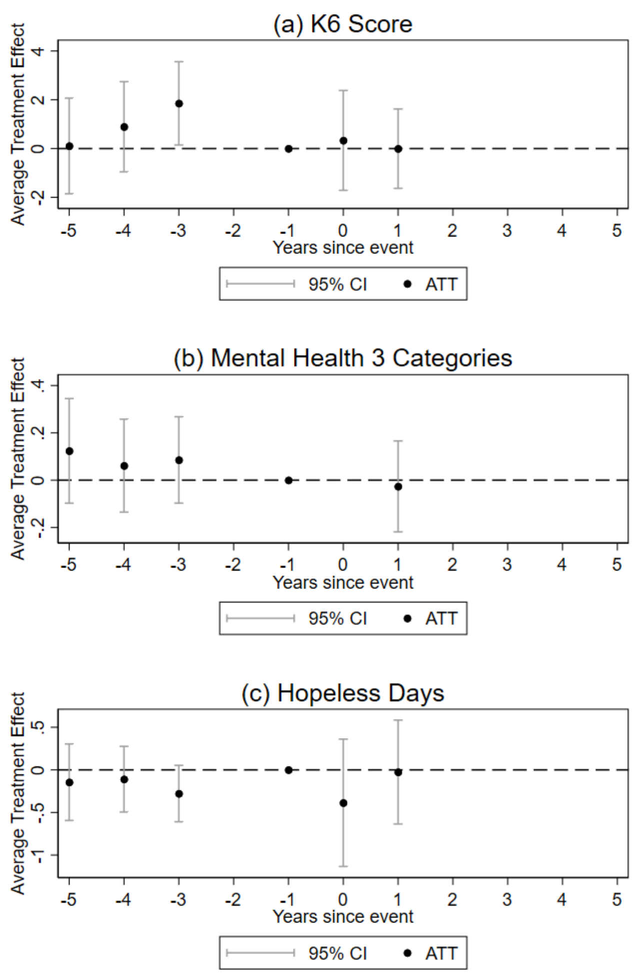
	ATT	SE	95% CI	p	Adjusted p
K6 Score	0.077	0.771	[−1.434, 1.587]	0.921	0.940
Mental Health Category	-0.026	0.098	[−0.219, 0.166]	0.788	0.940
Hopeless Days	-0.116	0.292	[−0.688, 0.456]	0.691	0.940
Log Consumption	0.086	0.081	[−0.072, 0.245]	0.287	0.940
Log Health Expenditure	-0.043	0.192	[−0.419, 0.333]	0.822	0.940
Weekly Hours Worked	0.156	2.053	[−3.868, 4.180]	0.940	0.940
Share of Risky Assets	0.027	0.049	[−0.069, 0.123]	0.585	0.940

Note: This table reports aggregate (simple) average treatment effects (ATTs) estimated using the staggered difference-in-differences estimator of Callaway and Sant'Anna (2021). Standard errors are clustered at the state level, and 95% confidence intervals are reported in brackets. Reported p-values test the null hypothesis that the aggregate ATT equals zero. Considering multiple hypothesis testing across outcome variables, p-values are adjusted using the false discovery rate (FDR) procedure of Benjamini and Hochberg (1995). All regressions are weighted by sample weights.

These results support the plausibility of the parallel trends assumption under the stricter treatment definition.

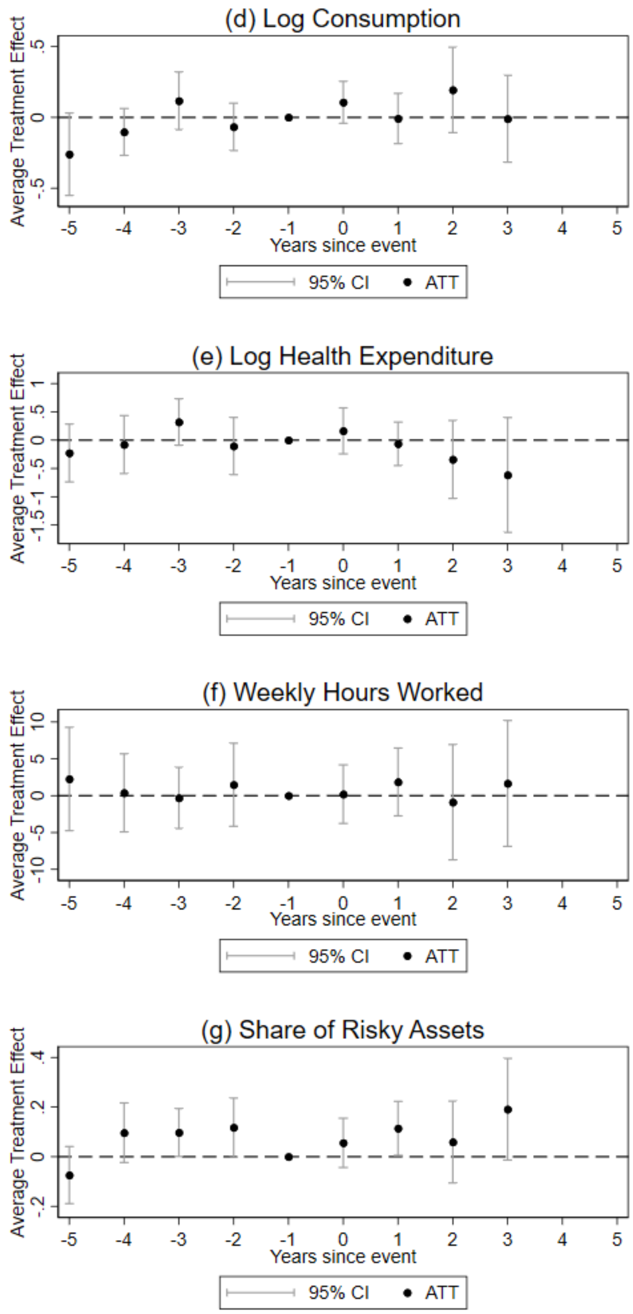
In summary, the robustness checks in this appendix indicate that the main findings of the baseline analysis are not sensitive to the definition of treatment timing. While adopting a stricter definition of parity laws reduces the sample size, the overall pattern of results remains unchanged.

Figure 5: Robustness Check: Mental Health Responses to Treatment Adoption



Note: Panels (a)-(c) report event-study estimates of the average treatment effect (ATT) on mental health outcomes (K6 Score, mental health 3 categories, and hopelessness) using the staggered difference-in-differences estimator of Callaway and Sant'Anna (2021). Event time is defined relative to the year of treatment adoption, and the period immediately before treatment (event time -1) serves as the reference period. All regressions are weighted by sample weights. Points represent point estimates, and vertical bars indicate 95% confidence intervals. Standard errors are clustered at the state level.

Figure 6: Robustness Check: Economic Outcome Responses to Treatment Adoption



Note: Panels (d)-(g) report event-study estimates of the average treatment effect (ATT) on economic outcomes (log consumption, log health expenditure, weekly hours worked, and the share of risky assets) using the staggered difference-in-differences estimator of Callaway and Sant'Anna (2021). Event time is defined relative to the year of treatment adoption, and the period immediately before treatment (event time -1) serves as the reference period. All regressions are weighted by sample weights. Points represent point estimates, and vertical bars indicate 95% confidence intervals. Standard errors are clustered at the state level.

Table 12: Robustness Check: Pre-trends Test

	Pre-trends p-value
K6 Score	0.118
Mental Health 3 Categories (Healthy, Mild, Serious)	0.423
Number of Hopeless Days	0.299
Log Consumption	0.176
Log Health Expenditure	0.427
Weekly Hours Worked	0.370
Share of Risky Assets	0.088

Note: This table reports p-values from pre-trends test based on the staggered difference-in-differences estimator of Callaway and Sant'Anna (2021). The null hypothesis is that all pre-treatment coefficients are equal to zero. Rejection of the null indicates the presence of systematic differences in trends prior to treatment adoption, suggesting that the parallel trends assumption may be less plausible for the corresponding outcome.

Appendix B: Jaspers' "Erklären" and "Verstehen"

This appendix briefly illustrates the distinction between *Erklären* (explanation) and *Verstehen* (understanding) proposed by Jaspers (1913/1997), as discussed in Section 5.2. We begin by considering two illustrative cases drawn from a standard psychiatry textbook (Ozaki et al., 2021), which are presented in the following:

An Example of *Erklären* (Explanation)

Male, in his 50s, office worker

While returning home from work, he became disoriented and was found wandering into a roadway in a direction different from that of his home. Upon encountering a police officer who attempted to assist him, he screamed in intense fear, shouting phrases such as "Help me!" and "Please don't kill me!" He also exhibited physical symptoms, including fever. He was admitted to a psychiatric unit, where further examinations were conducted. Based on the results of cerebrospinal fluid testing and other investigations, he was diagnosed with delirium due to viral encephalitis.

An Example of *Verstehen* (Understanding)

Female, in her 20s, university student

For two days, she had stopped responding entirely when spoken to by family members with whom she lived. Although it appeared that she was eating small amounts of snacks in her room, the details were unclear. It was also unclear whether she was sleeping adequately. Her family brought her to a psychiatric clinic for evaluation. Blood tests, electroencephalography, and brain MRI were performed, but no abnormalities were found. Subsequent information gathered from people involved revealed that she had recently and unexpectedly experienced the withdrawal of a job offer from a company from which she had already received an informal acceptance.

In the first case, the clinician understands the patient's symptom by identifying a specific medical cause, namely delirium due to viral encephalitis. This mode of understanding closely resembles that of general medicine: symptoms are explained by reference to an underlying pathological mechanism. In this sense, *Erklären* can be understood as a form of natural-scientific explanation of mental phenomena.

In contrast, the second case illustrates *Verstehen*. Here, no clear biological abnormality is detected through standard medical examinations. Instead, the clinician comes to understand the patient's mental state by identifying a meaningful life event, which is the sudden withdrawal of a previously secured job offer. By empathically grasping how this event might be experienced by the patient, the clinician attempts to make sense of her withdrawal and distress. This form of understanding relies not on causal mechanisms, but on reasons and subjective meanings. Jaspers argued that such forms of *Verstehen* should be consciously and systematically employed as a methodological tool in understanding in patients.

Jaspers, however, also emphasized what he called *methodologisches Bewusstsein* (methodological consciousness). That is, neither explanation nor understanding alone is sufficient for comprehending mental phenomena. To illustrate the limitations of relying exclusively on *Verstehen*, consider the following example. Suppose a patient suffers a traumatic brain injury in which a metal rod penetrates the skull and damages the frontal lobe.[15] Following the injury, characteristics that previously defined him, such as initiative, vitality, and the ability to plan and carry out work, are remarkably diminished. A common psychiatric manifestation in such cases is apathy, which is understood as a general reduction in goal-directed behavior.

One might attempt to understand this condition through *Verstehen* by attributing the loss of motivation to the patient's psychological response to the sudden injury, such as an inability to accept the traumatic event. However, this interpretation proves insufficient. In apathy following frontal lobe damage, observable behavior is markedly reduced, while mood itself is not necessarily depressed. Rather, while unable to initiate actions independently,

[15] This example is often discussed in connection with the historical case of Phineas Gage, a 19th-century patient who sustained frontal lobe damage following a railroad accident.

the patient can sustain behavior when prompted by others. Explaining this pattern requires reference to neurobiological mechanisms, for example, the role of the frontal lobe in initiating and activating behavior. In this case, an explanatory approach grounded in brain function becomes indispensable, highlighting the limits of relying solely on understanding.

Conversely, there are situations in which explanation alone fails. Consider an individual with a history of childhood abuse who has managed to live independently despite recurrent depressive episodes. Suppose that following the death of a pet cat with whom the individual had lived for many years, the person becomes unable to eat or function in daily life. An exclusively explanatory approach might seek abnormalities at the level of brain function or molecular biology. Yet such explanation offer little help in grasping the patient's suffering in this context. Here, an approach grounded in *Verstehen*, empathically appreciating the emotional significance of the loss, is essential for understanding the patient's mental state.

Taken together, these examples illustrate Jaspers' central methodological insight. That is, understanding mental phenomena requires a reflective balance between explanation and understanding, with careful attention to the limits and appropriate scope of each approach.