A Dynamic Factor Model of Skill Formation and Mental Health: Counterfactual Analysis of Interventions for Social Mobility*

Cecilia S. Diaz-Campo† M. Antonella Mancino‡ Salvador Navarro§

July 31, 2025

Abstract

This paper develops a dynamic latent factor framework to characterize the joint evolution of latent individual skills and outcomes under treatment and selection. We then specialize this framework to analyze mental health trajectories among justice-involved youth, focusing on the distributional impacts of incarceration. Leveraging rich longitudinal data from the Pathways to Desistance study, we recover the joint distribution of latent cognitive skills and mental health, estimate heterogeneous causal effects, and study how incarceration reshapes both levels and relative ranks within the mental health distribution. This approach provides a comprehensive distributional counterfactual analysis of incarceration's long-run effects.

Keywords: Bayesian Analysis, Dynamics, Mental Health, Incarceration, Counterfactuals, Mobility

JEL Code: C11, I10, K00

*The authors acknowledge funding from the Social Sciences and Humanities Research Council of Canada. Data used for this project were supported by the National Institute on Drug Abuse through a cooperative agreement that calls for scientific collaboration between the grantees and the National Institute on Drug Abuse staff.

†Olin Business School, Washington University in St. Louis. Email: cdiazcampo@wustl.edu.

‡Department of Economics, Wilfrid Laurier University. Email: amancino@wlu.ca.

§Department of Economics, University of Western Ontario. Email: snavarr@uwo.ca.

1 Introduction

Throughout his career, Steven Durlauf's influential body of work has profoundly shaped our understanding of inequality (e.g., Durlauf, 1996), social interactions (e.g., Brock and Durlauf, 2001; Durlauf, 2001; Durlauf and Ioannides, 2010), intergenerational mobility (e.g., Cholli and Durlauf, 2022; Durlauf et al., 2022), and crime (e.g., Durlauf et al., 2010; Durlauf and Nagin, 2010). This paper continues in his intellectual tradition by examining the distributional consequences of incarceration on mental health. In doing so, we build on Durlauf's legacy of integrating rigorous theoretical frameworks and careful empirical analysis to illuminate how complex social and economic processes shape heterogeneous individual outcomes.

The consequences of incarceration on individual well-being,[1] particularly on mental health (see, e.g., Turney et al., 2012; Haglund et al., 2014; Salazar, 2020; Bhuller et al., 2025), represent a critical area of social scientific inquiry and public policy concern. Mental health, in turn, is well known to have far-reaching effects on educational attainment and labor market outcomes (see, e.g., Currie and Stabile, 2006; Currie et al., 2010; Wang et al., 2023; Biasi et al., 2025; Diaz-Campo et al., 2025; Jolivet and Postel-Vinay, 2025). For the purpose of effective policy design and scientific understanding, it is necessary to move beyond simple correlations and descriptive summaries to the construction of counterfactuals (Heckman, 2001, 2005; Cunha et al., 2006; Imbens, 2022).

Counterfactual analysis allows us to ask what would have happened to an individual's mental health trajectory had they, for instance, not experienced incarceration, or had their pre-existing cognitive and mental health vulnerabilities been different. Such analyses are essential for disentangling the effects of incarceration itself from pre-existing individual heterogeneity and other life circumstances. Moreover, understanding the heterogeneous impacts of incarceration, i.e., recognizing that its effects may vary substantially across individuals with different baseline observed and *unobserved* characteristics, is crucial for evaluating the effects of incarceration policy. Traditional causal inference design-based approaches provide a clean, robust understanding of average causal effects, at the price of often limiting our ability to move beyond aggregate statistics to see who is

[1]For evidence on the consequences of incarceration on labor market outcomes, see e.g., Kling (2006); Mueller-Smith (2015); Gordon et al. (2023); Garin et al. (2025); Mancino (2025); on recidivism, see e.g., Kuziemko (2013); Bhuller et al. (2020); Rose and Shem-Tov (2021); and on overall health, see e.g., Binswanger et al. (2007, 2009); Hjalmarsson and Lindquist (2022); Norris et al. (2024).

most affected and how.

Following Cunha et al. (2006), this paper develops and applies a methodology to construct joint distributions of counterfactuals, and applies it to study how incarceration affects mental health for a sample of youth offenders. Our goal is to move beyond estimating average effects to identify how incarceration shifts the entire distribution of mental health outcomes for different individuals. By recovering the joint distribution of counterfactual outcomes, we can move beyond average summaries of effects and answer questions like which groups, as determined by their observed and unobserved characteristics, are affected by incarceration and how they are affected.

We introduce a dynamic factor model designed to address these challenges by analyzing the evolution of mental health among justice-involved youth from the Pathways to Desistance study. We explicitly model latent cognitive skills and mental health at baseline, their role in selection into incarceration over a seven-year period, and their influence on subsequent mental health outcomes. Our approach builds upon the framework of Cunha et al. (2010),[2] by incorporating multiple, interacting latent traits within a dynamic setting. A key model feature is a flexible mixture for the joint distribution of baseline latent skills and a subsequent shock to the mental health production function, allowing for rich patterns of unobserved heterogeneity and dependence.

By leveraging this econometric structure, we aim to: (1) estimate the distribution of causal impacts of incarceration on mental health at the seven-year follow-up, accounting for selection on observed and unobserved characteristics; (2) explore heterogeneous treatment effects, particularly how the impact of incarceration varies with baseline cognitive skills and mental health status; and (3) analyze the distributional dynamics of mental health by examining both positional mobility, how individuals shift their relative standing within the mental health distribution under incarceration versus no incarceration; and absolute growth, measuring changes in individuals' mental health levels within the overall distribution. In doing so, we seek to provide a comprehensive understanding of the interplay between individual vulnerabilities, justice system involvement, and mental health trajectories. The methodology allows us to move beyond simple descriptive statements to a more robust analysis of the distributional consequences of incarceration, in the spirit of constructing and evaluating policy counterfactuals as advocated by Cunha et al. (2006).

[2]See also Agostinelli and Wiswall (2025).

Our analysis reveals that incarceration has substantial and heterogeneous negative effects on mental health among justice-involved youth, even after accounting for selection on observed and unobserved factors. We uncover how incarceration shifts the entire distribution of mental health outcomes, with individuals experiencing incarceration generally showing worse mental health trajectories. We identify notable heterogeneity: the impact of incarceration varies with initial cognitive skills and mental health status, with certain subgroups disproportionately affected. Beyond average effects, we document significant mobility in mental health rankings induced by incarceration, with individuals in the middle of the mental health distribution exhibiting the greatest rank changes, while those at the extremes showing greater persistence. Our findings highlight how incarceration limits upward mobility in mental health, especially among lower-ranked individuals.

The paper proceeds as follows. Section 2 develops the general econometric framework for identifying joint distributions of latent factors, including the treatment assignment model and dynamic evolution across time. Section 3 tailors this general framework to the Pathways to Desistance data, specifying measurement systems for cognitive skills and mental health, and treatment rules relevant to incarceration and mental health dynamics. Section 4 presents estimation results and assesses model fit through graphical and statistical checks. Section 5 conducts counterfactual analyses of incarceration's distributional impacts on mental health, exploring heterogeneity, mobility, and growth in outcomes. Finally, Section 6 concludes with a discussion of the implications of our findings and avenues for future research.

2 Identifying Joint Distributions of Counterfactuals using Dynamic Factor Models

We begin by assuming the analyst has access to a collection of observed variables, such as psychometric scores or standardized tests, that depend on a smaller set of latent factors, for example, cognitive skills or socioemotional traits. Some of these measurements may be observed repeatedly over time. The analyst may be interested in the joint dynamics of these measurements and/or of the latent variables.

To simplify, we assume two periods, $t \in \{1, 2\}$, and two latent factors, $(\theta_{a,i,t}, \theta_{b,i,t})$. For individual i and period t, the analyst observes a vector of measurements $\{Y_{i,j,t}\}_{j=1}^{J_t}$. All measures are assumed continuous. We omit covariates for clarity; they can be reintroduced with standard modifications. Extensions to allow for additional factors, discrete or mixed outcomes, or richer panel structures can be done following the literature (see e.g., Jöreskog and Goldberger, 1975; Jöreskog, 1977; Carneiro et al., 2003; Cunha et al., 2010; Williams, 2020).

2.1 Measurement Model in Period 1

In period 1, each measurement satisfies:

$$Y_{i,j,1} = \gamma_{j,1} + \theta_{a,i,1}\alpha_{a,j,1} + \theta_{b,i,1}\alpha_{b,j,1} + \varepsilon_{i,j,1}, j = 1, ..., J_1. \tag{1}$$

The θ's are latent factors, α's are loadings, and ε's are measurement errors or "uniquenesses."

We invoke the following set of assumptions to identify the system

Assumption 1.

Location: $\mathbb{E}[\theta_{a,1}] = \mathbb{E}[\theta_{b,1}] = \mathbb{E}[\varepsilon_{j,1}] = 0, j = 1, ..., J_1$.

Dedicated Measures: $\alpha_{b,1,1} = \alpha_{b,2,1} = 0$; $\alpha_{a,3,1} = \alpha_{a,4,1} = 0$.

Scale and Sign: $\alpha_{a,1,1} = 1$, $\alpha_{b,3,1} = 1$.

Independence: $(\theta_{a,1}, \theta_{b,1}) \perp\!\!\!\perp \{\varepsilon_{j,1}\}_{j=1}^{J_1}$, *and* $\varepsilon_{j,1} \perp\!\!\!\perp \varepsilon_{j',1}$ *for* $j \neq j'$.

These assumptions are sufficient but not necessary. The location restrictions help identify the means $\gamma_{j,1}$. The dedicated measures condition ensures that at least two items load exclusively on a single factor; weaker versions are also valid (e.g., Williams, 2020). Scale and sign normalizations determine the magnitude and direction of the factors. For example, if measure 1 was an IQ score, more of factor $\theta_{a,1}$ would be associated with a higher IQ score (direction), and the factor would be measured in IQ units (magnitude). Independence between factors and measurement errors facilitates identification. Some form of independence between measures and the factors is needed, although it does not need to be as strong as the one we impose here (e.g., Cunha et al., 2010). Similarly, it is possible to allow for some forms of correlation between the ε's.

From observed covariances, we can identify the covariance between factors from

$$\text{cov}\left(Y_{1,1}, Y_{3,1}\right) = \text{cov}\left(\theta_{a,1}, \theta_{b,1}\right) \equiv \rho_1. \tag{2}$$

We can identify the loadings from

$$\alpha_{a,2,1} = \frac{\text{cov}\left(Y_{2,1}, Y_{3,1}\right)}{\rho_1}, \alpha_{b,4,1} = \frac{\text{cov}\left(Y_{1,1}, Y_{4,1}\right)}{\rho_1}, \tag{3}$$

and the factor variances

$$\sigma_a^2 = \frac{\text{cov}\left(Y_{1,1}, Y_{2,1}\right)}{\alpha_{a,2,1}}, \sigma_b^2 = \frac{\text{cov}\left(Y_{3,1}, Y_{4,1}\right)}{\alpha_{b,4,1}}. \tag{4}$$

Provided a rank condition holds, the remaining loadings ($j > 4$) are identified from

$$\begin{aligned} \text{cov}\left(Y_{1,1}, Y_{j,1}\right) &= \alpha_{a,j,1}\sigma_a^2 + \alpha_{b,j,1}\rho_1, \\ \text{cov}\left(Y_{3,1}, Y_{j,1}\right) &= \alpha_{a,j,1}\rho_1 + \alpha_{b,j,1}\sigma_b^2. \end{aligned} \tag{5}$$

Finally, we can recover the variance of the uniqueness terms from

$$\text{var}\left(Y_{j,1}\right) = \alpha_{a,j,1}^2\sigma_a^2 + \alpha_{b,j,1}^2\sigma_b^2 + 2\alpha_{a,j,1}\alpha_{b,j,1}\rho_1 + \sigma_{\varepsilon_{j,1}}^2. \tag{6}$$

2.1.1 Identification of Distributions

Once the parameters of the measurement system are identified, we turn to identifying the joint distribution of the latent factors $(\theta_{a,1}, \theta_{b,1})$ and the uniquenesses nonparametrically. We begin by applying Kotlarski's theorem (Kotlarski, 1967; Rao, 1992), which shows that the marginal distribution of a latent variable and two independent measurement errors can be recovered from their sum under independence assumptions.

To proceed, define the demeaned variables $\tilde{Y}_{j,1} = Y_{j,1} - \gamma_{j,1}$. From the measurement system and dedicated measures assumptions, we know that

$$\begin{aligned} \tilde{Y}_{i,1,1} &= \theta_{a,i,1} + \varepsilon_{i,1,1}, \\ \frac{\tilde{Y}_{i,2,1}}{\alpha_{a,2,1}} &= \theta_{a,i,1} + \frac{\varepsilon_{i,2,1}}{\alpha_{a,2,1}}. \end{aligned} \tag{7}$$

Since these expressions are two independent noisy measurements of the same latent variable $\theta_{a,1}$, and the noise terms are mutually independent and mean-zero, Kotlarski's result applies, and the

marginal distributions of $\theta_{a,1}, \varepsilon_{1,1}, \varepsilon_{2,1}$ are identified nonparametrically. The same reasoning applies for $\theta_{b,1}$ using measurements $\tilde{Y}_{3,1}$ and $\tilde{Y}_{4,1}$.

To identify the joint distribution of $(\theta_{a,1}, \theta_{b,1})$, take the characteristic function of $\left(\tilde{Y}_{1,1}, \tilde{Y}_{3,1}\right)$

$$\begin{aligned}\phi_{\tilde{Y}_{1,1},\tilde{Y}_{3,1}}(y_1, y_3) &= \mathbb{E}\left[e^{i\left((\theta_{a,1}+\varepsilon_{1,1})y_1+\left(\theta_{b,1}+\varepsilon_{3,1}\right)y_3\right)}\right] \\ &= \mathbb{E}\left[e^{i\left(\theta_{a,1}y_1+\theta_{b,1}y_3\right)}e^{i(\varepsilon_{1,1}y_1+\varepsilon_{3,1}y_3)}\right] \\ &= \phi_{\left(\theta_{a,1},\theta_{b,1}\right)}(y_1, y_3) \times \phi_{(\varepsilon_{1,1},\varepsilon_{3,1})}(y_1, y_3).\end{aligned} \tag{8}$$

The left-hand side is observable from data, and $\phi_{(\varepsilon_{1,1},\varepsilon_{3,1})}(y_1, y_3)$ is identified from marginal distributions via earlier steps. We recover the joint characteristic function of the latent variables by

$$\phi_{\left(\theta_{a,1},\theta_{b,1}\right)}(y_1, y_3) = \frac{\phi_{\tilde{Y}_{1,1},\tilde{Y}_{3,1}}(y_1, y_3)}{\phi_{(\varepsilon_{1,1},\varepsilon_{3,1})}(y_1, y_3)}, \tag{9}$$

which identifies the joint distribution of $(\theta_{a,1}, \theta_{b,1})$, assuming regularity conditions and non-vanishing characteristic functions.

Higher-order moments provide an alternative route. For example:

$$\begin{aligned}\mathbb{E}\left[\tilde{Y}_{1,1}^2\tilde{Y}_{2,1}\right] &= \alpha_{a,2,1}\mathbb{E}\left[\theta_{a,1}^3\right], \\ \mathbb{E}\left[\tilde{Y}_{3,1}^2\tilde{Y}_{4,1}\right] &= \alpha_{b,4,1}\mathbb{E}\left[\theta_{b,1}^3\right], \\ \mathbb{E}\left[\tilde{Y}_{1,1}^2\tilde{Y}_{3,1}\right] &= \mathbb{E}\left[\theta_{a,1}^2\theta_{b,1}\right], \\ \mathbb{E}\left[\tilde{Y}_{1,1}\tilde{Y}_{3,1}^2\right] &= \mathbb{E}\left[\theta_{a,1}\theta_{b,1}^2\right].\end{aligned} \tag{10}$$

Hence, if all moments exist, taking higher-order moments identifies the joint distribution of the latent factors (Billingsley, 1995).

This identification result is a core feature of the methodology, distinguishing it from others that make stronger distributional assumptions or rely on rank preservation or perfect dependence.

2.2 Treatment Assignment

We now consider a binary treatment that occurs between periods 1 and 2. The probability of being treated is modeled as a threshold-crossing rule driven by latent factors. Define the selection index

$$T = \gamma_T(X_i) + \theta_{a,i,1}\alpha_{a,T} + \theta_{b,i,1}\alpha_{b,T} + \varepsilon_{i,T}, \tag{11}$$

where X_i is a vector of observables that affect the probability of receiving treatment. The individual is treated if $T_i > 0$, so the treatment indicator is

$$D_i = \mathbb{1}\{T_i > 0\}. \tag{12}$$

We assume $\varepsilon_{i,T}$ is mean-zero and independent of everything else. To normalize the scale in this latent variable model, we impose $\text{Var}(\varepsilon_T) = 1$.

Under standard identification arguments from discrete choice models (Manski, 1988; Matzkin, 1992), $\gamma_T(X_i)$ and the distribution of the sum $(\theta_{a,1}\alpha_{a,T} + \theta_{b,1}\alpha_{b,T} + \varepsilon_T)$ are identified. Using conditional distributions of outcomes and treatment assignment

$$F_{Y_{j,1}|D=d}(y \mid X = x)\Pr(D = d \mid X = x) \tag{13}$$

and varying the evaluation points via y, x, we can identify the joint distribution of

$$\theta_{a,1}\alpha_{a,j,1} + \theta_{b,1}\alpha_{b,j,1} + \varepsilon_{j,1}, \text{ and } \theta_{a,1}\alpha_{a,T} + \theta_{b,1}\alpha_{b,T} + \varepsilon_T. \tag{14}$$

This identifies the cross-covariances of these random variables, which in turn identify $\alpha_{a,T}, \alpha_{b,T}$ (see e.g., Cunha et al., 2006; Fruehwirth et al., 2016).

2.3 Identification in Period 2: Dynamics and Evolving Latent Structure

Let $Y_{i,j,2}$ denote the measurements in the second period. The measurement system is assumed to be structurally similar to that in period 1

$$Y_{i,j,2} = \gamma_{j,2} + \theta_{a,i,2}\alpha_{a,j,2} + \theta_{b,i,2}\alpha_{b,j,2} + \varepsilon_{i,j,2}, j = 1, ..., J_2. \tag{15}$$

The latent factors evolve according to a first-order law of motion:[3]

$$\begin{aligned}\theta_{a,i,2} &= D_i\delta_{a,T} + \theta_{a,i,1}\lambda_{a,a} + \theta_{b,i,1}\lambda_{b,a} + \theta_{a,i,1}D_i\lambda_{a,a,T} + \theta_{b,i,1}D_i\lambda_{b,a,T} + U_{a,i}, \\ \theta_{b,i,2} &= D_i\delta_{b,T} + \theta_{a,i,1}\lambda_{a,b} + \theta_{b,i,1}\lambda_{b,b} + \theta_{a,i,1}D_i\lambda_{a,b,T} + \theta_{b,i,1}D_i\lambda_{b,b,T} + U_{b,i}.\end{aligned} \tag{16}$$

We maintain the following assumptions

Assumption 2.

Location: $\mathbb{E}[U_a] = \mathbb{E}[U_b] = \mathbb{E}[\varepsilon_{j,2}] = 0, j = 1, ..., J_2.$

[3]Notice that this model allows for what Heckman et al. (2006) refer to as essential heterogeneity.

Dedicated Measures: $\alpha_{b,1,2} = \alpha_{b,2,2} = 0$; $\alpha_{a,3,2} = \alpha_{a,4,2} = 0$.

Scale and Sign: $\alpha_{a,1,2} = 1$, $\alpha_{b,3,2} = 1$.

Independence: $(\theta_{a,1}, \theta_{b,1}) \perp\!\!\!\perp \{\varepsilon_{j,2}\}_{j=1}^{J_2}$, $(\theta_{a,1}, \theta_{b,1}) \perp\!\!\!\perp (U_a, U_b)$, $(U_a, U_b) \perp\!\!\!\perp \{\varepsilon_{j,2}\}_{j=1}^{J_2}$, $U_a \perp\!\!\!\perp U_b$, *and* $\varepsilon_{j,2} \perp\!\!\!\perp \varepsilon_{j',2}$ *for* $j \neq j'$.

Assumption 2 is very similar to assumption 1. The difference is that we no longer assume that the factors are mean zero in period 2. Instead, we impose the assumption on the innovations in the law of motion, U_a, U_b.

Given this structure, the measurement system in period 2 is identified by the exact same logic as in period 1. The loadings $(\alpha_{a,j,2}, \alpha_{b,j,2})$ and the distributions of the residuals and latent factors (up to their unknown mean) are all identified via cross-covariances.

2.3.1 Identifying the Law of Motion via Control Functions

The law of motion links period 2 latent factors with observables from period 1 and the treatment indicator. For identification, we exploit the link between outcomes and factor structure, and the fact that, given our results so far, the relation between $(\theta_{a,1}, \theta_{b,1})$ and D_i is known.

If we solve for $\theta_{a,i,2}$ from $Y_{i,1,2}$ and replace into the law of motion (16), it follows that

$$Y_{i,1,2} = D_i\delta_{a,T} + \theta_{a,i,1}\lambda_{a,a} + \theta_{b,i,1}\lambda_{b,a} + \theta_{a,i,1}D_i\lambda_{a,a,T} + \theta_{b,i,1}D_i\lambda_{b,a,T} + U_{a,i} + \varepsilon_{i,1,2}. \tag{17}$$

Next, consider the following expectation

$$\mathbb{E}\left[Y_{i,1,2} \mid D_i = 0, X_i = x\right] = \mathbb{E}\left[\theta_{a,i,1} \mid D_i = 0, X_i = x\right]\lambda_{a,a} + \mathbb{E}\left[\theta_{b,i,1} \mid D_i = 0, X_i = x\right]\lambda_{b,a}, \tag{18}$$

where we condition on X_i from equation (11), but more generally we could condition on $\gamma_T(X_i)$. The two expectations on the right-hand side are known from our analysis in the previous sections. Therefore, $\lambda_{a,a}$ and $\lambda_{b,a}$ are identified under a standard rank condition by varying the conditioning evaluation points given by x. Similarly, from

$$\begin{aligned}\mathbb{E}\left[Y_{i,1,2} \mid D_i = 1, X_i = x\right] - \mathbb{E}\left[\theta_{a,i,1} \mid D_i = 1, X_i = x\right]\lambda_{a,a} - \mathbb{E}\left[\theta_{b,i,1} \mid D_i = 1, X_i = x\right]\lambda_{b,a} = \\ \delta_{a,T} + \mathbb{E}\left[\theta_{a,i,1} \mid D_i = 1, X_i = x\right]\lambda_{a,a,T} + \mathbb{E}\left[\theta_{b,i,1} \mid D_i = 1, X_i = x\right]\lambda_{b,a,T},\end{aligned} \tag{19}$$

$\lambda_{a,a,T}$ and $\lambda_{b,a,T}$ are identified, and hence $\delta_{a,T}$ as well.

Having recovered all of the parameters of the law of motion, we can form $(Y_{i,1,2} - D_i\delta_{a,T})$. Since the distribution of $(\theta_{a,i,1}\lambda_{a,a} + \theta_{b,i,1}\lambda_{b,a} + \theta_{a,i,1}D_i\lambda_{a,a,T} + \theta_{b,i,1}D_i\lambda_{b,a,T})$ is now known, and it is independent of $(\varepsilon_{i,1,2} + U_{a,i})$, we can recover the distribution of $(\varepsilon_{i,1,2} + U_{a,i})$ by deconvolution.

An analogous procedure where we use $Y_{i,2,2}$ to solve for $\theta_{a,i,2}$ instead allows us to identify the distribution of $(\varepsilon_{i,2,2} + \alpha_{a,2,2}U_{a,i})$. We can then apply Kotlarksi to the distributions $\left(\varepsilon_{i,1,2} + U_{a,i}, \frac{\varepsilon_{i,2,2}}{\alpha_{a,2,2}} + U_{a,i}\right)$ and identify the marginal distributions of $U_{a,i}$, $\varepsilon_{i,1,2}$ and $\varepsilon_{i,2,2}$ nonparametrically.

The same sequence of steps applied to θ_b using $Y_{i,3,2}$ and $Y_{i,4,2}$, shows that we can also recover $\delta_{b,T}, \lambda_{a,b}, \lambda_{b,b}, \lambda_{a,b,T}, \lambda_{b,b,T}$, and the distributions of $\varepsilon_{3,2}, \varepsilon_{4,2}$ and U_b nonparametrically. It follows that the unknown means, $\mathbb{E}[\theta_{a,2}]$ and $\mathbb{E}[\theta_{b,2}]$, can now be calculated from the identified law of motion.

2.4 Counterfactual Analysis

With the assumptions and procedures above, we identify all factor loadings across periods and selection equations, all factor and uniqueness distributions nonparametrically, the law of motion for latent variables across periods, and the joint distributions of factors, treatment decisions, and outcomes. These results provide the full structure necessary for counterfactual analysis and recovery of potential outcomes. In the next section, we formalize this structure to generate and evaluate counterfactual outcomes, including nonparametric identification of standard and non-standard treatment effect parameters.

2.4.1 Potential Outcomes and Counterfactual Distributions

An alternative interpretation for the law of motion in equation (16) is that it describes a model of potential outcomes. If we explicitly write subscripts 0 and 1 to refer to the potential not treated and treated states, the factors in period 2 as functions of the treatment would be $(\theta_{a,i,2,0}, \theta_{a,i,2,1})$. It follows that the realized factor for individual i is

$$\theta_{a,i,2} = D_i\theta_{a,i,2,1} + (1 - D_i)\,\theta_{a,i,2,0}, \tag{20}$$

and similarly for $\theta_{b,2}$. This gives us a potential outcomes representation of observed measurements $(Y_{i,j,2,0}, Y_{i,j,2,1})$ as well.

Given nonparametric identification, one can go beyond the representation. We can fully charac-

terize the distribution of $(\theta_{a,2,0}, \theta_{a,2,1})$ and $(\theta_{b,2,0}, \theta_{b,2,1})$ and, by mapping through the measurement system, the joint distribution of $(Y_{j,2,0}, Y_{j,2,1})$.

2.4.2 Standard Treatment Parameters

Having recovered the joint distribution of potential outcomes, treatment effect parameters are readily computed. Let

$$\begin{aligned} \Delta_{\theta_{a,i,2}} &= \theta_{a,i,2,1} - \theta_{a,i,2,0} \\ &= \delta_{a,T} + \theta_{a,i,1}\lambda_{a,a,T} + \theta_{b,i,1}\lambda_{b,a,T}. \end{aligned} \tag{21}$$

For example, in terms of the latent factor $\theta_{a,i,2}$, we can recover the average treatment effect (ATE)

$$ATE\left(\theta_{a,i,2}\right) = \mathbb{E}\left[\Delta_{\theta_{a,i,2}}\right] = \delta_{a,T}, \tag{22}$$

or the average treatment effect on the treated (ATT) and on the untreated (ATU)

$$\begin{aligned} ATT\left(\theta_{a,i,2}\right) &= \mathbb{E}\left[\Delta_{\theta_{a,i,2}} \mid D_i = 1\right] \\ &= \delta_{a,T} + \mathbb{E}\left[\theta_{a,i,1} \mid D_i = 1\right]\lambda_{a,a,T} + \mathbb{E}\left[\theta_{b,i,1} \mid D_i = 1\right]\lambda_{b,a,T}, \\ ATU\left(\theta_{a,i,2}\right) &= \mathbb{E}\left[\Delta_{\theta_{a,i,2}} \mid D_i = 0\right] \\ &= \delta_{a,T} + \mathbb{E}\left[\theta_{a,i,1} \mid D_i = 0\right]\lambda_{a,a,T} + \mathbb{E}\left[\theta_{b,i,1} \mid D_i = 0\right]\lambda_{b,a,T}. \end{aligned} \tag{23}$$

We can also define the marginal treatment effect (MTE) of Heckman and Vytlacil (1999)

$$\begin{aligned} MTE\left(\theta_{a,i,2}\right) &= \mathbb{E}\left[\Delta_{\theta_{a,i,2}} \mid T_i = 0\right], \\ &= \delta_{a,T} + \mathbb{E}\left[\theta_{a,i,1} \mid T_i = 0\right]\lambda_{a,a,T} + \mathbb{E}\left[\theta_{b,i,1} \mid T_i = 0\right]\lambda_{b,a,T}, \end{aligned} \tag{24}$$

which defines the average return for individuals who are on the edge of being assigned to treatment.

Parallel expressions follow for parameters based on $\theta_{b,i,2}$ or observed measurements $Y_{i,j,2}$.

2.4.3 Distributional and Mobility Counterfactuals

Beyond means, nonparametrically recovering the joint distribution of potential outcomes enables analysis of more detailed questions. For example, we can estimate the fraction of the population

who would benefit from universal treatment

$$\Pr\left(\Delta_{\theta_{a,i,2}} > 0\right), \tag{25}$$

or the fraction amongst the treated who benefit from treatment

$$\Pr\left(\Delta_{\theta_{a,i,2}} > 0 \mid D_i = 1\right). \tag{26}$$

We can also calculate counterfactual mobility parameters. For example, we can calculate a mobility table where we look at the probability that an individual that starts in the p^{th}-percentile of the untreated distribution, $\theta_{a,2,0}$, ends in the q^{th}-percentile of the treated distribution, $\theta_{a,2,1}$.

While these examples illustrate the advantage of recovering the joint distribution of potential outcomes, they are not exhaustive (see e.g., Cunha et al., 2006). More generally, parameters specified as functions of $\theta_{a,2}$ may also be formulated in terms of $\theta_{b,2}$ or in terms of the observed measures $Y_{j,2}$.

2.4.4 Policy Counterfactuals

We are not restricted to working with the counterfactual distribution of the observed treatment. The structure also supports simulation and evaluation of alternative treatment assignment rules. For two distinct policies (I and II) resulting in different treatment rules,[4] label the respective treatment indicators D^I and D^{II}. The approach allows calculation of, for example, the proportion of individuals whose treatment status changes between policies, or of the proportion of these "switchers" who benefit from the alternative policy

$$\Pr\left(\Delta_{\theta_{a,i,2}} > 0 \mid D^I \neq D^{II}\right). \tag{27}$$

The framework we just described allows us to obtain a richer understanding of the inequality and social mobility consequences of existing policies, and of policy changes, than are available from measures based on mean treatment parameters. Furthermore, we can analyze these changes holding the period fixed, i.e., without the need to wait for time to elapse. Finally, we can use our analysis to generate various counterfactual states even if they have never been observed.

[4] For example, because there is a change in how one of the $X's$ is treated between policies.

3 Empirical Application: Incarceration and Mental Health in the Pathways to Desistance Data

In this section, we apply our methodology to examine how incarceration shapes mental health trajectories among young offenders. We begin by describing the data used in the analysis. We then specialize the model in Section 2 to account for specific features of our data.

3.1 Data

This study utilizes data from the Pathways to Desistance (PD) project, a longitudinal study that tracks the development and life trajectories of adolescent offenders as they transition into early adulthood. The sample comprises youth who were adjudicated for serious offenses in either juvenile or adult court in Maricopa County, Arizona, or Philadelphia County, Pennsylvania, between November 2000 and January 2003. Eligible participants were between 14 and 18 years old at the time of their offense and were required to provide informed assent or consent. A total of 1,354 individuals enrolled in the study, corresponding to a 67% enrollment rate.[5]

Participants completed a baseline interview shortly after their adjudication (within 75 days for those in the juvenile system, and within 90 days after decertification hearings or arraignments for those in the adult system, depending on the site). The study includes six follow-up interviews conducted every six months and four annual follow-ups thereafter. Interviews were typically held in participants' homes or secure facilities if they were incarcerated. The panel data span up to seven years for each individual. To encourage participation and reduce attrition, individuals received $50 for the baseline interview, with payments increasing for subsequent follow-ups (Monahan et al., 2009). The study maintained high retention rates, exceeding 90% for the first six waves and remaining at or above 83% for the later annual waves.

An important feature of the PD dataset is its detailed longitudinal information on criminal activity and incarceration. In addition, it includes rich data on mental health and a wide range of individual characteristics that may be relevant for predicting incarceration. These features are crucial for analyzing the relationship between incarceration and mental health outcomes.

[5]For more information on the PD study see Schubert et al. (2004).

The baseline survey includes basic demographic information, such as age, gender, ethnicity, and location. It also contains data on incarceration before the baseline. Specifically, the survey records whether individuals were ever detained in a juvenile facility or jail before the event that led them into the PD survey. Incarceration data are subsequently collected in each follow-up survey wave. These data are based on self-reports gathered during each wave of the survey.[6] At each follow-up, individuals report whether they were incarcerated during the recall period, the type of facility (e.g., prison, jail, or detention center), and the number of months spent in each facility.

The baseline survey includes a rich set of self-reported measures of mental health symptoms from the Brief Symptom Inventory (BSI). The BSI is a self-report tool where participants assess the degree to which they have been disturbed by any of 53 possible symptoms over the past week. Responses are rated on a scale of 0 ("Not at All"), 1 ("A Little Bit"), 2 ("Moderately"), 3 ("Quite a Bit"), and 4 ("Extremely"), with higher scores reflecting more severe mental health symptoms. The BSI includes nine subscales, each targeting a specific symptom group. These include: somatization, obsessive-compulsive behaviors, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. Each subscale is an average of between 4 and 7 individual symptoms, and hence can take more than 5 values for each subscale. The BSI measures are collected repeatedly in the follow-up surveys.

Lastly, the baseline survey also incorporates results from a comprehensive array of standardized psychometric tests administered to each individual to measure cognitive skills. Cognitive measures include scores from the Wechsler Abbreviated Scale of Intelligence (WASI), which provides an estimate of overall intellectual functioning (IQ) derived from two subtests: Vocabulary and Matrix Reasoning. In addition, the data include results from two neurological tests designed to assess cognitive dysfunction associated with impairment of the brain's frontal cortex: the Stroop Color-word test and the Trail Making Test. The Stroop Test examines cognitive interference in reading ability through three timed tasks involving color words, colored symbols, and mismatched word-color combinations. Participants are instructed to read words or name ink colors as quickly and accurately as possible. The Trail Making Test assesses overall brain function and potential neurological impair-

[6]To encourage accurate self-reporting, responses are kept confidential, and participants are provided a certificate of confidentiality from the U.S. Department of Justice.

ment through two parts. Part A requires individuals to connect a sequence of numbered circles in order, while Part B involves alternating between numbers and letters in sequential order. Relevant to our model, follow-up surveys do not collect information on these cognitive measures.

We restrict our sample to individuals who have not been treated before the beginning of the study (i.e., we exclude those who were ever detained before the baseline survey). The final sample comprises 686 individuals, all of whom have baseline information on at least one cognitive or mental health measure.

Table 1 presents descriptive statistics for our sample. Men make up the majority of the sample, with women comprising 17.1% of participants. The sample is almost evenly divided between Phoenix (44.6%) and Philadelphia (55.4%). A substantial share of the sample consists of minority individuals, with 42.3% identifying as Black and 30.5% as Hispanic. Seven years after the baseline survey, 55.2% of individuals have missing information on either incarceration or mental health, 23.0% have experienced at least one month of incarceration during that period, and 21.7% have never been incarcerated.

Table 1: Pathways to Desistance Study - Summary Statistics - Demographics

	Mean	SD	N
Age at Baseline			
Share Age 14	0.134	0.341	686
Share Age 15	0.203	0.402	686
Share Age 16	0.325	0.469	686
Share Age 17	0.265	0.442	686
Share Age 18	0.073	0.260	686
Share Female	0.171	0.376	686
Ethnicity			
Share White	0.226	0.419	686
Share Hispanic	0.305	0.461	686
Share Black	0.423	0.494	686
Share Other	0.047	0.211	686
Share Phoenix	0.446	0.497	686
Incarceration Indicator - 7 Years After			
Have Not Been Incarcerated	0.217	0.413	686
Have Been Incarcerated	0.230	0.421	686
Missing Information	0.552	0.498	686

Notes: The descriptive statistics reported in this table correspond to data from the baseline survey, except for the incarceration indicator, which is based on both baseline and follow-up surveys.

Table 2 presents descriptive statistics for self-reported mental health symptoms at baseline and

seven years later. At baseline, for each of the nine measures, only a minority of individuals report experiencing symptoms to at least a mild degree on average (i.e., a little bit, moderately, quite a bit, or extremely), with rates ranging from 8.5% for phobic anxiety to 33.5% for paranoid ideation. Seven years later, all nine measures indicate a modest reduction in the proportion of individuals reporting symptoms at or above a mild level on average, even after accounting for missing data. For instance, among individuals without missing information, the share reporting at least mild somatization symptoms decreases from 12.3% to 10.0%. This pattern holds across all measures, suggesting a slight improvement in mental health over time.

Table 3 reports descriptive statistics for the tests designed to measure cognitive skills. The average IQ score among participants is notably lower than the general population mean of 100, with only 10% of individuals scoring above that benchmark. For cognitive impairment, the Trail-Making assessments classify individuals into one of four categories, with the two lowest indicating some degree of impairment. In our sample, 20.3% show signs of mild, moderate, or severe impairment based on Trail-Making A, and 35.1% based on Trail-Making B. For each of the Stroop Measures, scores above 40 are typically regarded as within the normal range. At the baseline, 36.0%, 53.1%, and 20.6% have scores below normal for the Word, Color, and Color/Word tests, respectively.

3.2 An Empirical Model of Skills and Mental Health Dynamics

We now adapt the general framework developed in Section 2 to an empirical setting focused on the joint dynamics of cognitive skills and mental health. While generalizing the model along some dimensions, we also tailor it to the specific features of our data.

At baseline, each individual i is endowed with a level of cognitive skill, denoted by C_i, and a level of mental health, denoted by M_i. We also observe a set of background characteristics X_i, which are allowed to correlate with observed outcomes but are assumed to be independent of latent factors and measurement errors. Associated with these latent states are vectors of measurements: $\{\mathcal{C}_{i,j}\}_{j=1}^{N_C}$ for cognitive skill and $\{\mathcal{M}_{i,j}\}_{j=1}^{N_M}$ for mental health.

Table 2: Pathways to Desistance Study - Summary Statistics - Mental Health Measures

	Mean - Baseline	Mean - Year 7	N
BSI Somatization			
Not At All (0)	0.380	0.249	686
Between 0 and 1	0.421	0.153	686
1 or More	0.112	0.045	686
Missing	0.086	0.552	686
BSI Depression			
Not At All (0)	0.310	0.239	686
Between 0 and 1	0.417	0.137	686
1 or More	0.187	0.071	686
Missing	0.086	0.552	686
BSI Hostility			
Not At All (0)	0.236	0.140	686
Between 0 and 1	0.415	0.233	686
1 or More	0.262	0.074	686
Missing	0.086	0.552	686
BSI Interpersonal Sensitivity			
Not At All (0)	0.411	0.299	686
Between 0 and 1	0.353	0.093	686
1 or More	0.150	0.055	686
Missing	0.086	0.552	686
BSI Anxiety			
Not At All (0)	0.347	0.203	686
Between 0 and 1	0.411	0.187	686
1 or More	0.156	0.058	686
Missing	0.086	0.552	686
BSI Obsessive-Compulsive			
Not At All (0)	0.213	0.146	686
Between 0 and 1	0.443	0.203	686
1 or More	0.258	0.099	686
Missing	0.086	0.552	686
BSI Phobic Anxiety			
Not At All (0)	0.570	0.313	686
Between 0 and 1	0.259	0.108	686
1 or More	0.085	0.026	686
Missing	0.086	0.552	686
BSI Paranoid Ideation			
Not At All (0)	0.156	0.138	686
Between 0 and 1	0.423	0.187	686
1 or More	0.335	0.122	686
Missing	0.086	0.552	686
BSI Psychoticism			
Not At All (0)	0.334	0.255	686
Between 0 and 1	0.418	0.141	686
1 or More	0.162	0.051	686
Missing	0.086	0.552	686

Notes: This table reports descriptive statistics for mental health measures from the Brief Symptom Inventory (BSI) at the baseline survey and the last follow-up survey (i.e., seven years after baseline). The BSI comprises nine subscales, each corresponding to a specific grouping of symptoms: somatization, obsessive-compulsive behaviors, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. For each symptom, responses are rated on a scale of 0 ("Not At All"), 1 ("A Little Bit"), 2 ("Moderately"), 3 ("Quite a Bit"), and 4 ("Extremely"), with higher scores indicating more severe mental health symptoms.

Table 3: Pathways to Desistance Study - Summary Statistics - Cognitive Skills

(a) Panel A: WASI IQ

	Percentile	N
1%	55	681
5%	63	681
10%	67	681
25%	77	681
50%	86	681
75%	94	681
90%	102	681
95%	107	681
99%	116	681

(b) Panel B: Trail Making

	Share of Sample	N
Trail Making Part A		
Perfectly Normal	0.402	686
Normal	0.388	686
Mil/Moderately Impaired	0.134	686
Moderately/Severly Impaired	0.069	686
Missing	0.007	686
Trail Making Part B		
Perfectly Normal	0.362	686
Normal	0.281	686
Mil/Moderately Impaired	0.261	686
Moderately/Severly Impaired	0.090	686
Missing	0.006	686

(c) Panel C: Stroop Measures

	Share of Sample	N
Stroop Word		
Share Below or Equal 40	0.360	686
Share Above 40	0.624	686
Share Missing	0.016	686
Stroop Color		
Share Below or Equal 40	0.531	686
Share Above 40	0.453	686
Share Missing	0.016	686
Stroop Color/Word		
Share Below or Equal 40	0.206	686
Share Above 40	0.778	686
Share Missing	0.016	686

Notes: The descriptive statistics reported in this table correspond to data from the baseline survey. The estimate of general intellectual ability (IQ) is based on two subtests: Vocabulary and Matrix Reasoning. The Trail-Making Test measures general brain function and consists of two parts: A and B. Scores take one of four values, with the lowest two indicating mild/moderate or moderate/severe impairment. The Stroop Color/Word Test assesses the effects of interference on reading ability through three parts, measuring interference from words, colors, and combined word-color. Scores on the Stroop tests are continuous, with values above 40 considered within the normal range.

For mental health, we employ the nine subscales from the BSI ($N_M = 9$). All nine are treated as ordered discrete measures. For each subscale j, the number of values (K_j) corresponds to the number of distinct values where we group the "1 or More" into a single category. We use six measures of cognitive skills ($N_C = 6$). There are four continuous measures: the WASI IQ and the three Stroop scores, and two Trail-Making scores, which are measured on an ordered discrete scale.[7] The vector of individual characteristics, X, includes dummies for age, gender, ethnicity, and location.

We define a latent structure that links each measurement to the latent factors. For the cognitive measurements, we assume

$$\mathcal{C}^*_{i,j} = X_i'\gamma_{\mathcal{C},j} + C_i\psi_j + \varepsilon_{i,\mathcal{C},j}, j \in 1, ..., N_C. \tag{28}$$

Analogously, for the mental health measurements

$$\mathcal{M}^*_{i,j} = X_i'\gamma_{\mathcal{M},j} + M_i\mu_j + \varepsilon_{i,\mathcal{M},j}, j \in 1, ..., N_M. \tag{29}$$

When the corresponding observed measure is continuous, we set the measured outcome equal to the latent one, e.g., $\mathcal{C}_{i,j} = \mathcal{C}^*_{i,j}$. For discrete, ordered outcomes taking values $k = 1, ..., K_j$, we impose a standard ordered threshold structure:

$$\mathcal{M}_{i,j} = k \text{ if } o_{\mathcal{M},j,k-1} < \mathcal{M}^*_{i,j} \leq o_{\mathcal{M},j,k}, k = 1, ..., K_j, \tag{30}$$
$$o_{\mathcal{M},j,0} = -\infty, o_{\mathcal{M},j,1} = 1, o_{\mathcal{M},j,K_j} = \infty;$$

with $o_{\mathcal{M},j,0} = -\infty, o_{\mathcal{M},j,1} = 1$, and $o_{\mathcal{M},j,K_j} = \infty$. Similar conventions are applied to other measures.

Identification proceeds analogously to the system described in Section 2,[8] with factors now encompassing cognitive skill and mental health. The measurement errors ε represent idiosyncratic deviations from the factor structure and are assumed mean-zero and mutually independent. As before, we assume that C and M are independent of the measurement errors. For scale and sign normalization, we set $\psi_1 = 1$ and $\mu_1 = -1$. Namely, the cognitive skills factor is normalized to have a loading of one on the WASI IQ score, while the mental health factor is normalized to have a loading of negative one on the BSI hostility measure.[9]

[7] The Trail-Making scores are reverse-coded so that higher scores reflect less cognitive impairment, aligning them with the remaining cognitive measures.

[8] For the discrete case, identification follows from similar arguments to the ones we describe for the continuous case in Section 2. See Carneiro et al. (2003) for example.

[9] We normalize the mental health factor to negative one, rather than one, because higher BSI scores indicate

Treatment is defined as whether an individual goes to prison between the baseline period and the final observation, seven years later. Treatment operates through a latent index model

$$T_i = X_i'\gamma_T + C_i\psi_T + M_i\mu_T + \varepsilon_{i,T}, \tag{31}$$

with treatment indicator

$$D_i = \mathbb{1}\left\{T_i > 0\right\}. \tag{32}$$

We assume $\varepsilon_{i,T}$ is independent of C_i, M_i, and the measurement errors, and that X_i is independent of all latent variables and uniquenesses.

We observe repeated mental health measurements in period 7, denoted by $\mathcal{M}_{i,j,7}$, for $j = 1, ..., N_M$. Let the corresponding latent variables be

$$\mathcal{M}^*_{i,j,7} = X_i'\gamma_{\mathcal{M},j,7} + M_{i,7}\mu_{j,7} + \varepsilon_{i,\mathcal{M},j,7}, j \in \left\{1, 2, ..., N_M\right\}, \tag{33}$$

and

$$\mathcal{M}_{i,j,7} = k \text{ if } o_{\mathcal{M},j,7,k-1} < \mathcal{M}^*_{i,j,7} \leq o_{\mathcal{M},j,7,k}, k = 1, ..., K_j, \tag{34}$$

$$o_{\mathcal{M},j,7,0} = -\infty, o_{\mathcal{M},j,7,1} = 1, o_{\mathcal{M},j,7,K_j} = \infty;$$

where we normalize $\mu_{j,2} = -1$. Specifically, the mental health factor at year seven is normalized to have a loading of negative one on the BSI hostility measure.

The law of motion for mental health allows treatment effects to depend directly on baseline factors

$$M_{i,7} = D_i\delta_T + C_i\lambda_C + M_i\lambda_M + C_iD_i\lambda_{CT} + M_iD_i\lambda_{MT} + U_i. \tag{35}$$

U_i is a mean zero random variable that is independent of C_i, M_i, and the measurement errors. As in the previous sections, we assume all uniquenesses are mutually independent and independent of the factors. This dynamic formulation allows treatment effects to be heterogeneous, depending on both baseline cognitive skills and baseline mental health.

more severe mental health symptoms.

3.3 Estimation

Although the model is semi-parametrically identified, we adopt parametric assumptions to facilitate estimation. We postulate normality for all uniquenesses and use mixtures of normals to approximate the distributions of latent variables.

For continuous measurements, we assume

$$\varepsilon_{i,j} \sim N\left(0, \sigma^2_{\varepsilon_j}\right). \tag{36}$$

For discrete outcomes, including ordered measures and treatment, we assume standard normal errors with unit variance.

The joint distribution of baseline cognitive skill and mental health is approximated by a finite mixture of normals

$$f_{C,M}\left(z_C, z_M\right) = \sum_{h=1}^{6} \pi_{CM,h} \phi\left(z_C; m_{C,h}, \sigma^2_{C,h}\right) \phi\left(z_M; m_{M,h}, \sigma^2_{M,h}\right) \tag{37}$$

subject to $\sum_{h=1}^{6} \pi_{CM,h} m_{C,h} = 0; \sum_{h=1}^{6} \pi_{CM,h} m_{M,h} = 0$, where $\phi(z; m, \sigma^2)$ is the pdf of a normal random variable with mean m and variance σ^2 evaluated at z. The innovation U_i is modeled analogously using a three-component mixture

$$f_U\left(u\right) = \sum_{h=1}^{3} \pi_{U,h} \phi\left(u; m_{U,h}, \sigma^2_{U,h}\right) \tag{38}$$

subject to $\sum_{h=1}^{3} \pi_{U,h} m_{U,h} = 0$.

We estimate the model using Monte Carlo Markov Chain (MCMC) methods with a Gibbs sampler. Each iteration of the sampler sequentially draws from the full conditional distribution of every parameter and latent variable, conditioning on the current values of the other parameters. We either use non-informative priors or diffuse proper priors when a prior is called for. We briefly describe the sampler next, and leave a more detailed summary of it for Appendix OA.1.

For each continuous cognitive measurement, estimation is straightforward. Conditional on C_i, the equation becomes linear in parameters with a normally distributed error. The posterior distributions of regression coefficients and loadings are Gaussian, and the inverse error variances follow standard conjugate Gamma distributions.

The ordered models are completed by sampling the latent utilities, e.g., $\mathcal{M}^*_{i,j}$, which are con-

ditionally normally distributed given their observed categories. Thresholds are assigned diffuse uniform priors, and given these, latent utilities are sampled from truncated normal distributions. The treatment assignment equation is a special case.

In year 7, measurement equations for mental health are handled similarly, conditional on the constructed values of $M_{i,7}$ from the law of motion. The joint presence of observed measurements, latent traits, treatment, and interactions allows us to recover the parameters of the dynamic system, including the heterogeneous treatment effects and propagation mechanisms.

To sample the parameters governing the law of motion, $\delta_T, \lambda_C, \lambda_M, \lambda_{CT}, \lambda_{MT}$, we construct regressors from transformations of observed variables and latent states

$$\mathcal{M}^{**}_{i,j,7} = D_{i,j}\delta_T + C_{i,j}\lambda_C + M_{i,j}\lambda_M + CD_{i,j}\lambda_{CT} + MD_{i,j}\lambda_{MT} + U_{i,j} + \varepsilon_{i,\mathcal{M},j,7} \tag{39}$$

with $\mathcal{M}^{**}_{i,j,7} \equiv \mathcal{M}^{*}_{i,j,7} - X_i'\gamma_{\mathcal{M},j,7}$, and $D_{i,j} \equiv D_i\mu_{j,7}$, $CD_{i,j} \equiv C_iD_i\mu_{j,7}$, and so on. Each parameter appears across measurement equations indexed by j, and we estimate them by pooling across individuals and outcomes.

Posterior computations for the mixture distributions follow standard Bayesian mixture updating. For each latent variable, posterior sampling alternates between updating mixture component parameters, reassigning individuals to components, and drawing component weights from Dirichlet distributions. Sampling latent factors proceeds by conditioning on all observed and latent objects in the complete data likelihood. Posterior distributions are normal given the structure of the system.

This estimation strategy delivers full posterior distributions for all parameters and latent variables. With these recovered, the model permits recovery of the counterfactual distributions and treatment effect functionals introduced in the preceding section.

Our Gibbs sampling procedure offers substantial computational advantages over traditional estimation approaches. In our experience, the serial MCMC algorithm we employ achieves orders-of-magnitude speed improvements relative to optimization-based methods such as maximum likelihood, even when the latter are run with parallelized objective function evaluations. The iterative structure and tractability of the Gibbs sampler enable efficient exploration of the posterior distribution, making it particularly well-suited to the high-dimensional latent variable models considered here.

4 Results

In this section, we present estimation results obtained using the Gibbs sampler described above. To ensure thorough exploration of the posterior distribution, we run the algorithm for a total of 140,000 iterations, discarding the initial 40,000 draws as burn-in to mitigate the influence of starting values. To further reduce autocorrelation in the retained samples, we save every 20th draw from the remaining iterations for a total of 5,000 draws. This approach balances computational efficiency with the need for a representative sample from the posterior distribution of the model parameters.

4.1 Summarizing Posterior Distributions and Variance Decomposition

We start by presenting the results from our factor analysis, in which we project our baseline and future measurements onto three factors: one related to cognitive skills, one related to baseline mental health, and one related to mental health at year seven. Figure 1a plots the average estimated joint posterior distributions of C and M. There is a small positive correlation between cognitive skills and mental health at baseline. Figure 1b shows the average estimated posterior distribution of the law of motion innovation to mental health in year 7 (U). As can be seen, the distributions are highly non-normal. highlighting the importance of allowing for our flexible mixture models.

The means and standard deviations of the estimated posterior distributions of the factor measurement model are presented in Tables OA.2.1, OA.2.2, and OA.2.3. The results in Table OA.2.1 show that higher cognitive skills are associated with better scores across all cognitive measures we use. Likewise, Tables OA.2.2 and OA.2.3 show that better mental health is associated with fewer BSI symptoms.[10]

Figure 2 presents a variance decomposition to assess the extent to which measurement error influences our cognitive and mental health measures. Specifically, we break down the variance of the latent component of each measurement into two parts: the share explained by the underlying factor and the share attributable to measurement error. Figure 2a shows that the cognitive factor is more strongly related to the Stroop measures of cognitive dysfunction than to the WASI-IQ and Trail-Making measures. Nevertheless, even for the Stroop measures, cognitive skills explain on average

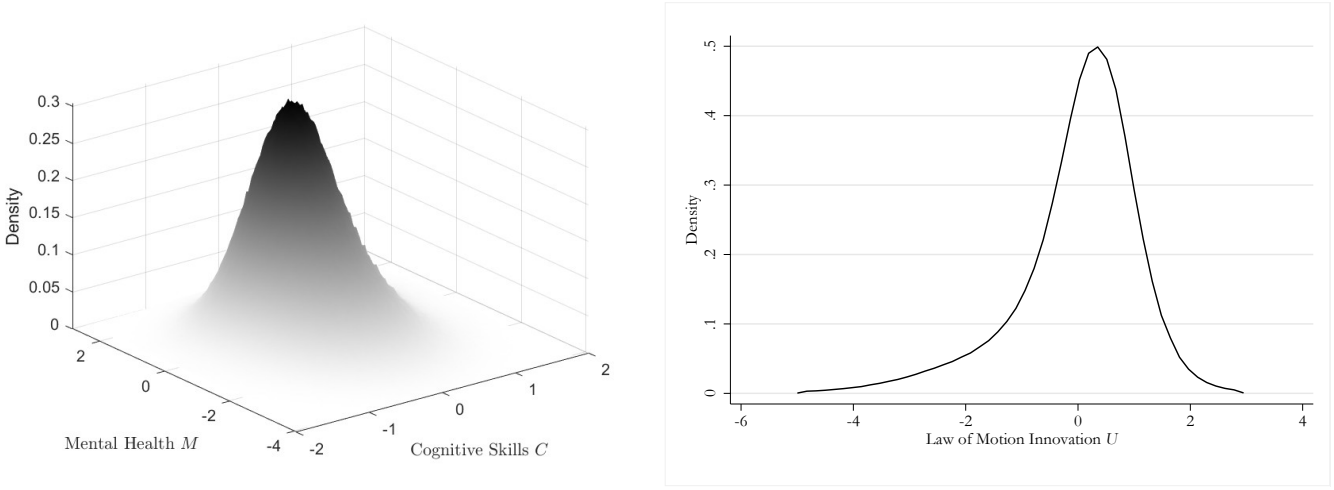
[10]The negative sign of the mental health factor in Tables OA.2.2 and OA.2.3 for all nine measures is consistent with the coding of BSI scores, where larger values reflect worse mental health.

(across posterior draws) only around 60% of the unobserved variance. Figure 2b shows that the baseline mental health factor explains more than 45% of the variance for all BSI measures. Moreover, our mental health factor is more closely related to obsessive-compulsive behaviors, anxiety, and depression than to phobic anxiety and interpersonal sensitivity. A similar pattern is observed for the mental health factor at year 7 (see Figure 2c).

Figure 1: Estimated Posterior Distributions of the Factors

(a) Joint Distribution of C and M

(b) Distribution of U



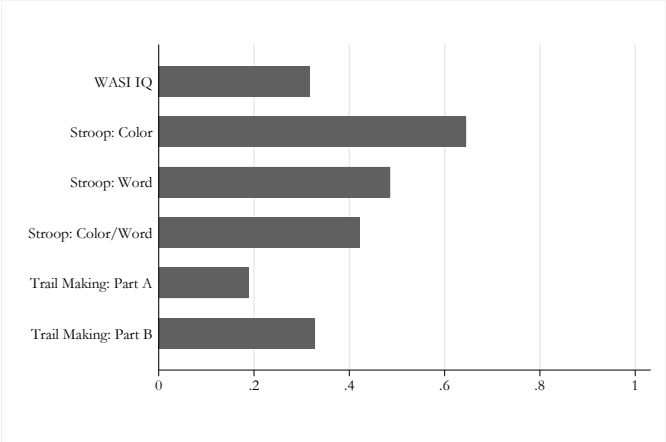
Notes: This figure reports the estimated posterior distributions of the factors. Figure 1a displays the estimated joint distribution of cognitive skills and mental health, while Figure 1b displays the estimated density of the law of motion innovation to mental health in year 7.

Table 4 summarizes the estimated posterior distributions for the treatment equation (31). As expected, higher cognitive skills and better mental health at baseline are negatively associated with spending any time in jail, prison, or detention within the seven-year window after the baseline survey. This selection is further illustrated in Figure 3, which displays the distributions of cognitive factor and baseline mental health conditional on treatment status.

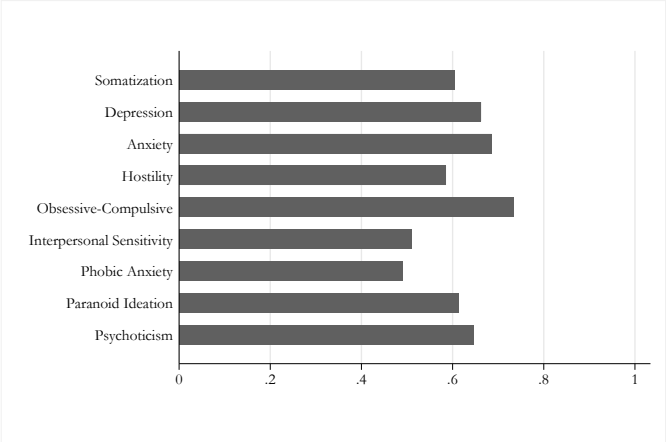
Lastly, Table 5 reports the estimates for the law of motion of mental health in equation (35). The results provide evidence of time dependence in mental health. On average, better baseline mental health is positively associated with better future mental health. The estimates also indicate that being incarcerated negatively affects mental health on average, with this effect being larger for individuals with higher cognitive skills and better baseline mental health, although the posterior distributions of these parameters have relatively large standard deviations.

Figure 2: Fraction of the Variance Explained by the Factor

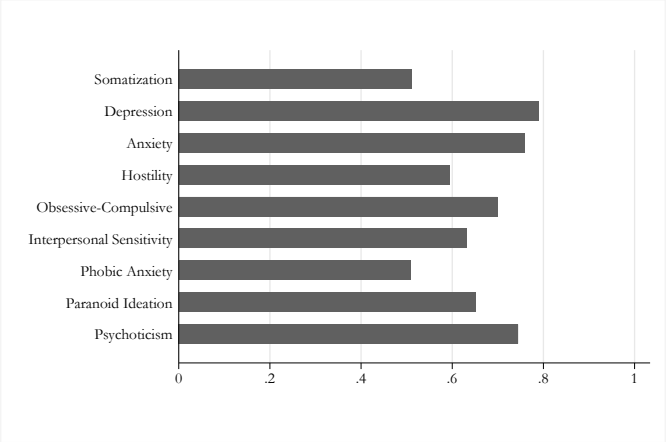
(a) Cognitive Skills: C



(b) Mental Health (Baseline): M



(c) Mental Health (Year 7): M_7



Notes: These figures present the average fraction of the variance of each cognitive and mental health measure explained by the cognitive and mental health factors. For example, 32% of the fraction of the variance of the residualized (against X) WASI IQ measure is explained by the cognitive skills C.

4.2 Model Fit

In this section, we assess how our estimated model fits the observed data. Figures 4, 5, and 6 present graphical assessments of model fit for the cognitive outcomes, baseline mental health measures, and year 7 mental health measures, respectively. The model performs well overall. For the continuous cognitive measures (Figures 4a-4d), the predicted distributions closely match the observed ones in terms of dispersion and general shape. The observed distributions tend to exhibit slightly higher peaks around the mode compared to the predicted ones, suggesting a modest underestimation of concentration near the central tendency. For the discrete variables, both cognitive and mental health, the predicted frequencies align closely with the empirical data.

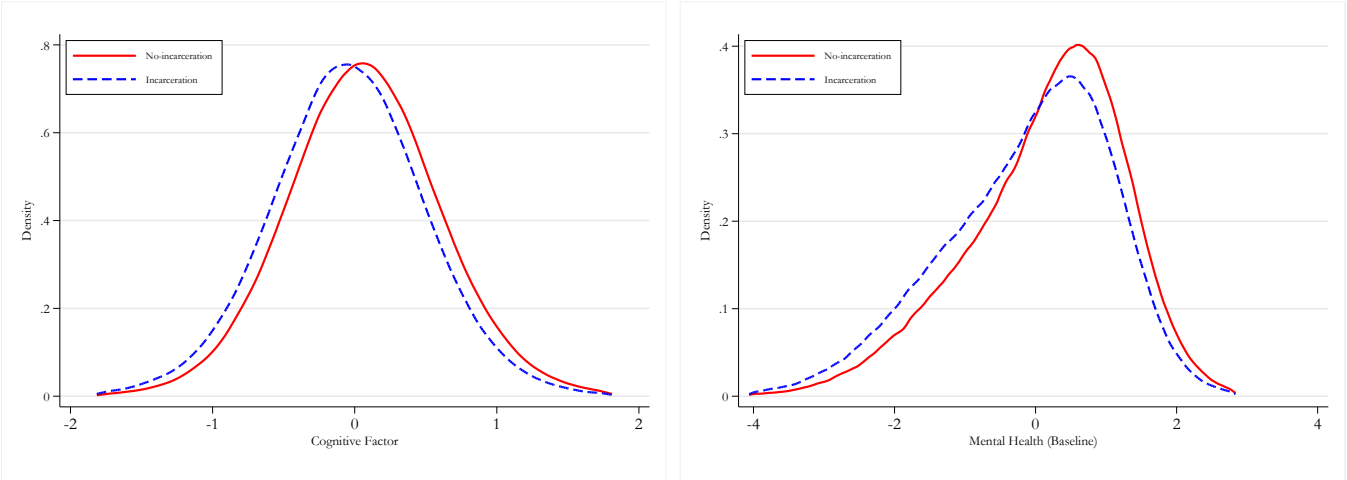
Table 4: Estimated Parameters from Model - Treatment Equation

	Mean	SD
Constant	-0.135	0.309
Age 15	0.259	0.201
Age 16	0.405	0.184
Age 17	0.181	0.192
Age 18	0.067	0.265
Female	-1.271	0.169
White	-0.118	0.297
Hispanic	0.083	0.285
Black	0.374	0.292
Phoenix	0.289	0.148
Cognitive Skills (ψ_T)	-0.271	0.129
Mental Health (μ_T)	-0.134	0.056

Notes: This table reports the mean and standard deviation of each parameter across 5,000 draws from the posterior distribution for the treatment equation (31).

Figure 3: Distribution of Mental Health and Cognitive Factors By Treatment

(a) Cognitive Skills: C (b) Mental Health (Baseline): M



Notes: These figures show the estimated densities of the cognitive and baseline mental health factors conditional on incarceration status. For example, the dashed-line in Figure 3a plots $f(C \mid D = 1)$.

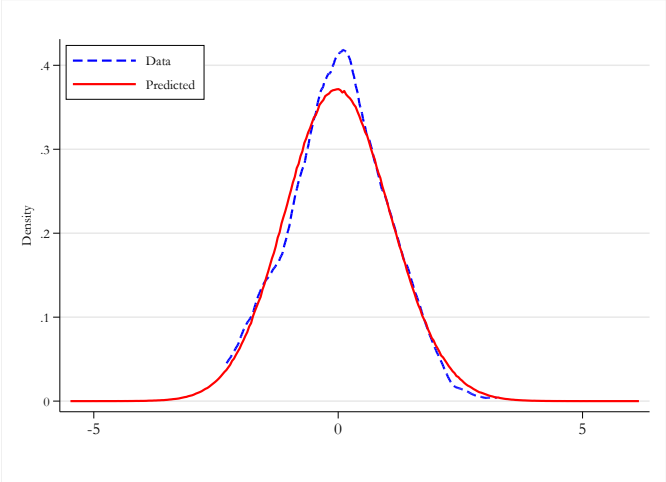
Table 5: Estimated Parameters from Model - Law of Motion Equation

	Mean	SD
Incarceration (δ_T)	-0.248	0.121
Cognitive Skills (λ_C)	-0.133	0.193
Mental Health (λ_M)	0.334	0.142
Incarceration $\times$ Cognitive Skills (λ_{CT})	-0.067	0.255
Incarceration $\times$ Mental Health (λ_{MT})	-0.155	0.110

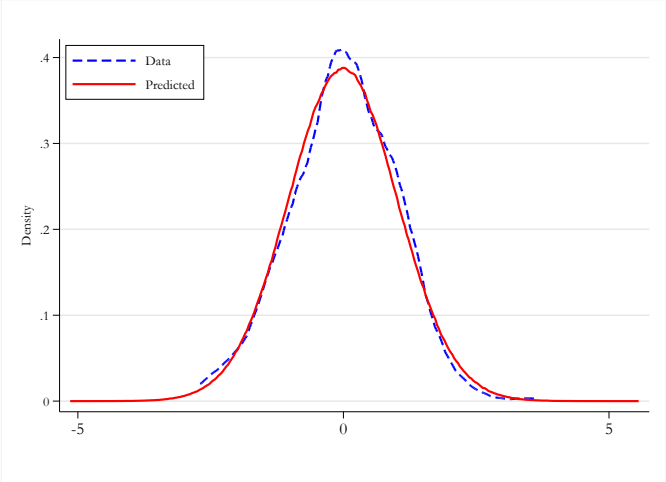
Notes: This table reports the mean and standard deviation of each parameter across 5,000 draws from the posterior distribution for the law of motion of mental health in equation (35).

Figure 4: Posterior Predictive Fit - Cognitive Measures

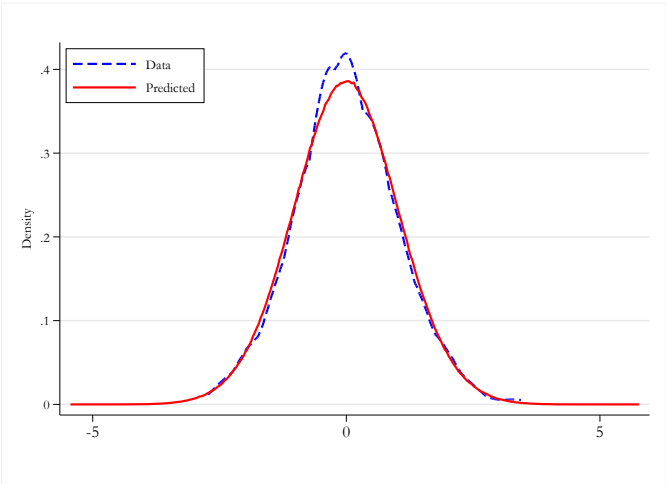
(a) WASI IQ



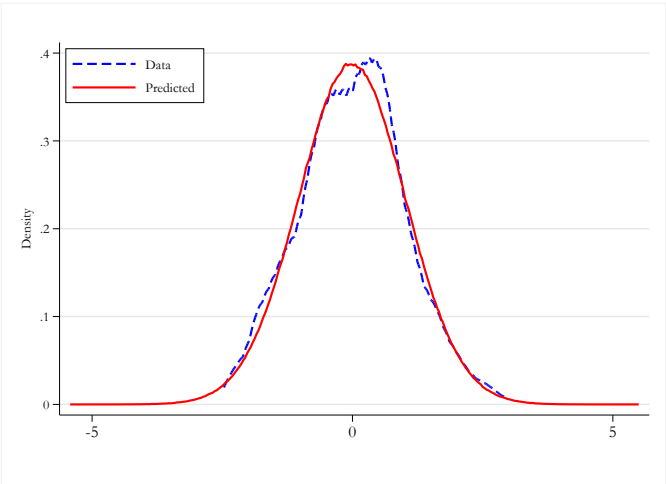
(b) Stroop Word



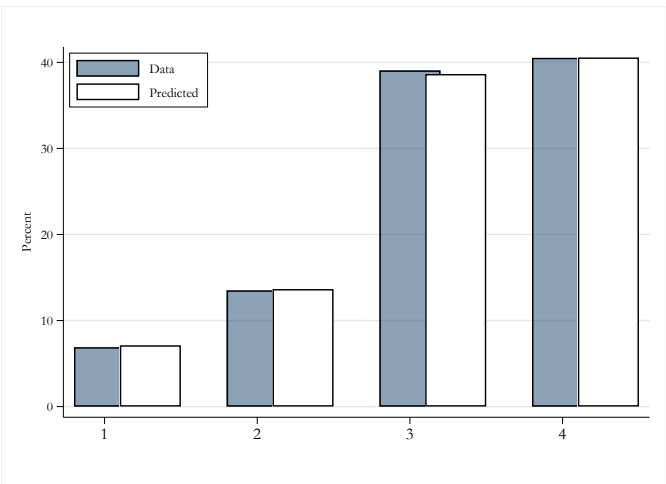
(c) Stroop Color



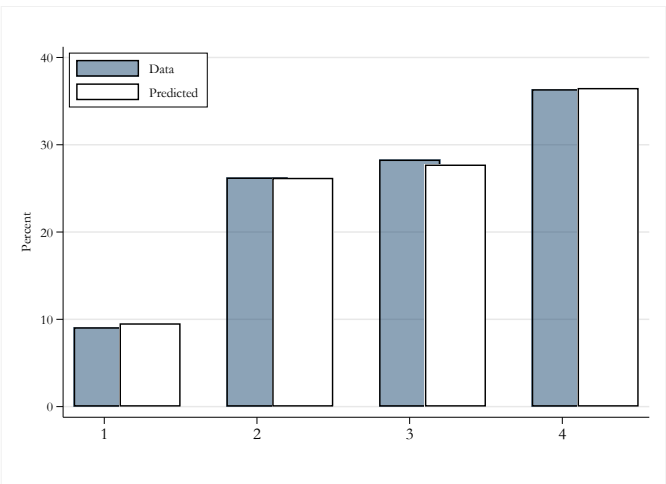
(d) Stroop Color/Word



(e) Trail Making A

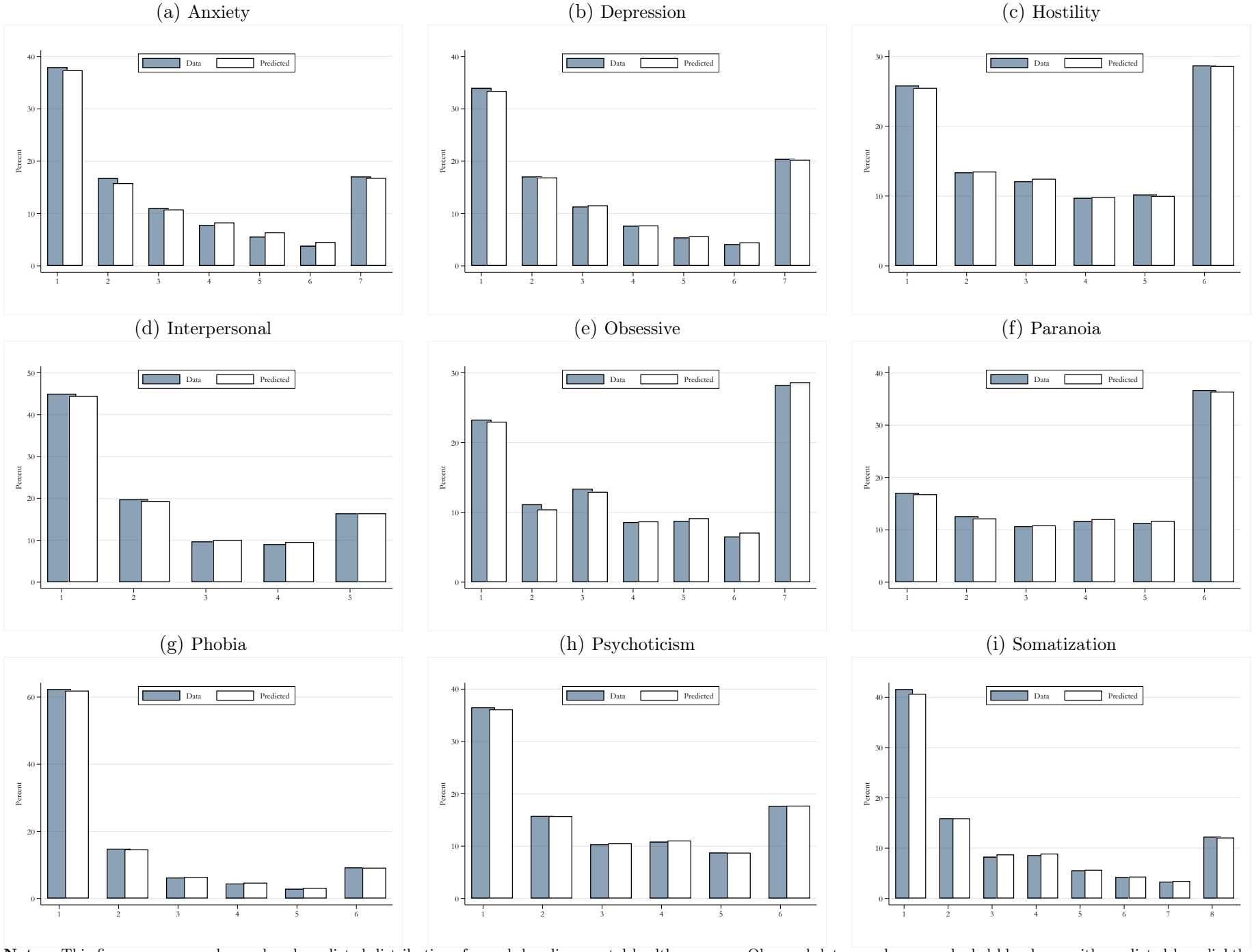


(f) Trail Making B



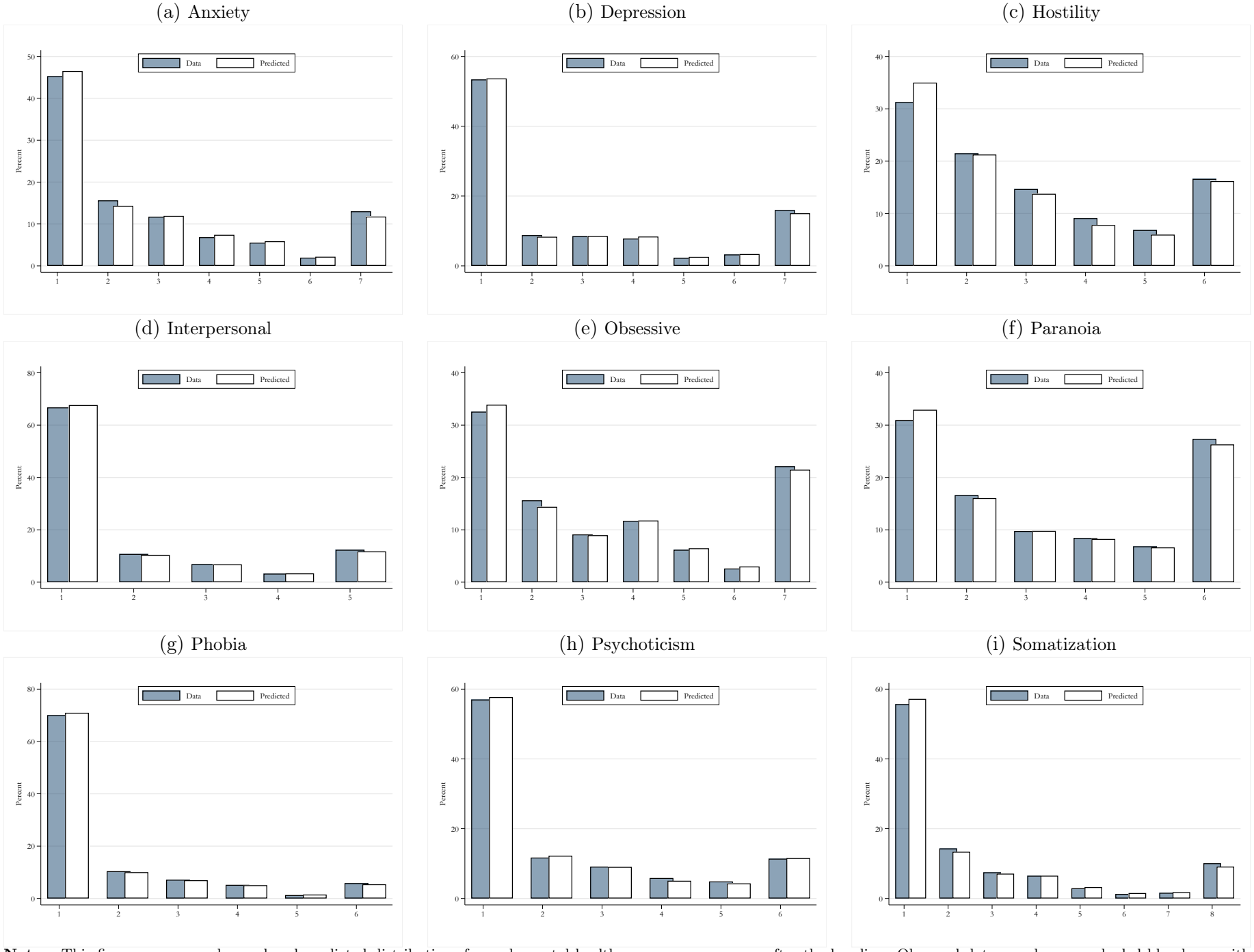
Notes: This figure compares observed and predicted distributions for each cognitive measure. For continuous variables (Figures 4a-4d), observed distributions appear as blue dashed lines and predicted as solid red lines. For discrete variables (Figures 4e and 4f), observed data are shown as shaded blue bars, with predicted bars lightly overlaid in white. Predicted values are based on 10 million simulation draws from the posterior distributions.

Figure 5: Posterior Predictive Fit - BSI Mental Health Measures (Baseline)



Notes: This figure compares observed and predicted distributions for each baseline mental health measure. Observed data are shown as shaded blue bars, with predicted bars lightly overlaid in white. Predicted values are based on 10 million simulation draws from the posterior distributions.

Figure 6: Posterior Predictive Fit - BSI Mental Health Measures (Year 7)



Notes: This figure compares observed and predicted distributions for each mental health measure seven years after the baseline. Observed data are shown as shaded blue bars, with predicted bars lightly overlaid in white. Predicted values are based on 10 million simulation draws from the posterior distributions.

We formally measure the discrepancy between model and data by defining three test quantities. For ease of comparison, for our first test quantity, we use a classical frequentist interpretation of the Chi-Square test. Therefore, we use the Chi-Square test statistic, defined as the sum of differences between observed and predicted outcome frequencies, each squared and divided by the prediction

$$\chi^2_m = \sum_{b=1}^{B^m} \frac{(O_b^m - E_b^m)^2}{E_b^m}, \tag{40}$$

where outcome measures are indexed by m, bins are indexed by b with total number of bins indexed by B_m, observed count for bin b is indexed by O_b^m, and predicted count for bin b is indexed by E_b^m. Column (1) in Table 6 reports the *Classical* p-value from the chi-square distribution for our test statistic for the cognitive and mental health measures displayed on the rows. Out of 24 measures, we reject the null (in a classical sense) of equality of frequencies in only 7 instances for a 5% significance level. In two of these cases, BSI Obsessive-Compulsive at baseline and BSI Anxiety at Year 7, the p-value is very close to the 5% threshold.

Let $Y_{i,j}$ denote the observed value of measure j for individual i, and $\overline{Y}_j$ its mean. Define Θ_s to be the collection of parameters evaluated using the s-th parameter draw of the estimated posterior of the model. Then, $\widehat{Y}_{i,j,s}$ represents the predicted value of measure j evaluated at Θ_s. Finally, let $\overline{\widehat{Y}}_j = \mathbb{E}[\widehat{Y}_{i,j,s}]$ where the expectation is taken over both i and s. The Posterior Predictive P-values (PPP) is defined as

$$\frac{1}{S \times N} \sum_s \sum_i \mathbb{1}[\|\widehat{Y}_{i,j,s}, \overline{\widehat{Y}}_j\| > \|\overline{Y}_j, \overline{\widehat{Y}}_j\|], \tag{41}$$

the probability that the predicted data could be more extreme than the observed data, for a given distance measure $\|\cdot\|$. A model is suspect if its discrepancy has a tail-area probability near 0 or 1, indicating that the observed pattern would be unlikely to be seen in replications of the data if the model were true.

Columns (2) and (3) present two versions of the PPP as a *Bayesian* measure of fit for the measures on the rows. Column (2) uses the Mahalanobis distance, while column (3) uses a joint Euclidean distance of means and standard deviations. Neither measure is close to 1, and in only a few instances (4 in each case) are these values smaller than 0.05.

Table 6: Goodness Of Fit

	Chi-Square Classical p-value (1)	Mahalanobis Bayesian PPP (2)	Euclidean Bayesian PPP (3)
Cognitive Skills: C			
WASI IQ†	0.108	0.074	0.152
Stroop Word†	0.781	0.548	0.644
Stroop Color†	0.018	0.397	0.496
Stroop CW†	0.091	0.597	0.677
Trail Making A	0.275	0.287	0.193
Trail Making B	0.186	0.249	0.183
Mental Health: M			
BSI Somatization	0.014	0.004	0.010
BSI Depresion	0.043	0.020	0.046
BSI Anxiety	0.706	0.759	0.864
BSI Hostility	0.105	0.086	0.043
BSI Obsessive	0.048	0.086	0.116
BSI Interpersonal	0.470	0.490	0.593
BSI Phobia	0.793	0.740	0.451
BSI Paranoia	0.427	0.478	0.347
BSI Psychoticism	0.256	0.264	0.146
Mental Health: M_7			
BSI Somatization	0.075	0.214	0.025
BSI Depresion	0.038	0.039	0.167
BSI Anxiety	0.045	0.117	0.310
BSI Hostility	0.543	0.824	0.681
BSI Obsessive	0.629	0.668	0.698
BSI Interpersonal	0.095	0.197	0.118
BSI Phobia	0.000	0.041	0.565
BSI Paranoia	0.434	0.590	0.242
BSI Psychoticism	0.639	0.745	0.733

Notes: This table presents three measures of Goodness of Fit for our cognitive and mental health measures. Column (1) displays the *Classical* p-values from the chi-square distribution. Columns (2) and (3) report *Bayesian* Posterior Predictive P-values using the Mahalanobis distance and the joint Euclidean distance of means and standard deviations, respectively. For discrete variables, the total number of bins is the number of categories in the empirical distribution. For continuous variables (indexed by a dag), we set the number of bins equal to 4 and the cutoffs equal to the quartiles of the simulated distribution.

5 Counterfactuals

Overall, our estimated model fit the data really well. It demonstrates a strong ability to replicate the observed data patterns. These results support the model's suitability for counterfactual policy analysis.

5.1 Treatment Effects

We now use the potential outcomes interpretation of the law of motion of mental health introduced in Section 2.4.1 to present estimates of standard treatment effect parameters. Table 7 presents summary measures of the posterior distributions of some estimands of interest: the Average Treatment Effect, the Average Treatment on the Treated, the Average Treatment on the Untreated, and the Marginal Treatment Effect.

Column (1) presents the posterior mean of the treatment effect parameters. In addition to point summaries, in columns (2) to (6), we report measures of posterior uncertainty. Columns (2) and (3) present the lower and upper bounds of the 95% posterior credibility interval, respectively. Column (4) presents the posterior probability that the parameter is positive. Columns (5) and (6) report the posterior probability that the estimated parameters are close to zero. Specifically, Column (5) shows the posterior probability that the absolute value of the parameter is less than 0.01. Column (6) reports the probability for the analogous test using a threshold equal to 1% of the parameter's range across the relevant posterior distribution. In all cases, there is strong evidence of an average negative effect of incarceration on mental health, with the effect being larger (in absolute value) for the individuals typically not incarcerated (i.e., $|ATT(M_7)| < |ATU(M_7)|$).

Table 7: Mean Treatment Effects

	Mean (1)	lb (2)	ub (3)	P($\cdot > 0$) (4)	P($|\cdot| < 0.01$) (5)	P($|\cdot| < 1\%$) (6)
ATE(M_7)	-0.248	-0.488	-0.009	0.022	0.007	0.006
ATT(M_7)	-0.229	-0.465	0.008	0.028	0.011	0.009
ATU(M_7)	-0.274	-0.514	-0.033	0.012	0.005	0.004
MTE(M_7)	-0.255	-0.510	-0.008	0.023	0.007	0.006

Notes: Let Θ_s be the collection of parameters evaluated using the s-th parameter draw of the estimated posterior of the model. Column (1) presents $\frac{1}{S}\sum_{s=1}^{S} ATE(M_7;\Theta_s)$. Columns (2) and (3) present the 2.5 and 97.5 percentiles of $\{ATE(M_7;\Theta_s)\}_{s=1}^{S}$. Column (4), (5), and (6) are $\frac{1}{S}\sum_{s=1}^{S}\mathbb{1}[ATE(M_7;\Theta_s)>0]$, $\frac{1}{S}\sum_{s=1}^{S}\mathbb{1}[|ATE(M_7;\Theta_s)|<0.01]$, and $\frac{1}{S}\sum_{s=1}^{S}\mathbb{1}[|ATE(M_7;\Theta_s)|<0.01\times Range(\{ATE(M_7;\Theta_s)\}_{s=1}^{S})]$, respectively. And similar definitions apply for the remaining parameters.

As discussed in Section 2.4.3, a key advantage of nonparametrically identifying the factor distri-

butions is the ability to analyze features of the outcome distribution beyond mean treatment effects. This point is illustrated in Figure 7, which presents the counterfactual distributions of outcomes and gains for mental health in period 7. Figure 7a and Figure 7b show that the potential mental health distribution absent incarceration, $(M_{i,7,0})$, first-order stochastically dominates the distribution under incarceration, $(M_{i,7,1})$, regardless of incarceration status (D). Nonetheless, as shown in Figure 7c, the estimated gains from incarceration are consistently larger for individuals who are actually incarcerated.

These results align with the estimated mean treatment effects reported in Table 7. For the majority of individuals, mental health deteriorates following incarceration. However, because we estimate the full distribution of gains, it is also possible to observe that a nontrivial fraction of the population experiences positive gains.

5.2 Mental Health: Mobility and Growth

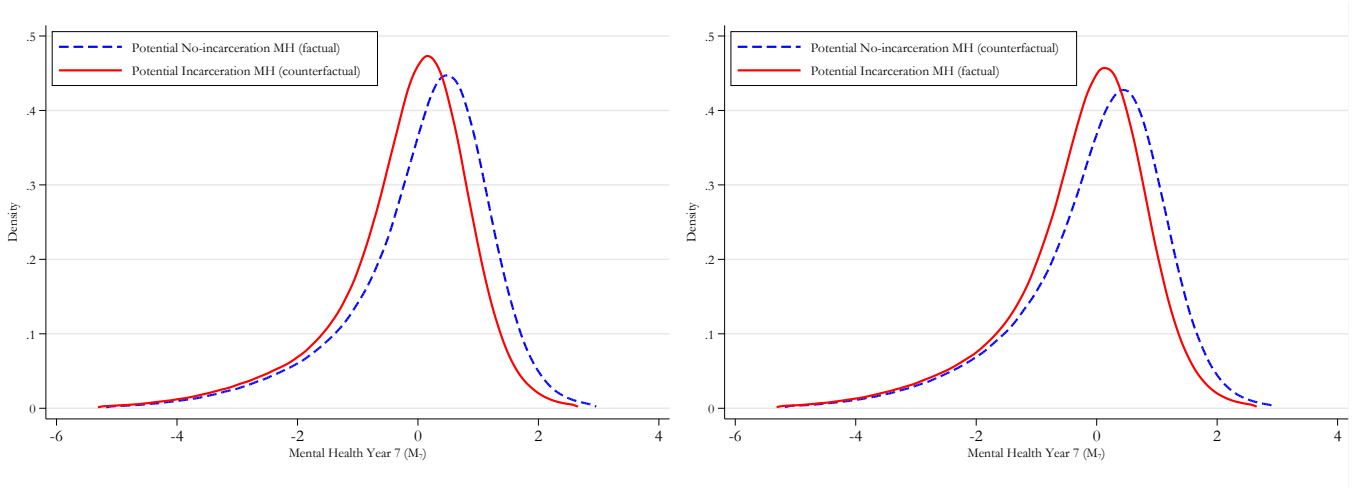
Understanding the impact of incarceration on individual well-being requires not only examining average changes in mental health, but also investigating how it alters individuals' positions within the overall distribution. In this section, we present both detailed mobility matrices and summary measures that provide a comprehensive view of mobility patterns and their heterogeneity across the cognitive skill distribution.

We measure mobility by adapting the concept of *positional movement* from Fields (2006). This approach measures changes in individuals' ranks across two counterfactual distributions, rather than changes in absolute levels. In our context, it captures how a person's position in the potential mental health distribution absent incarceration compares relative to their position in the potential distribution under incarceration. Positional mobility is particularly informative if individuals derive welfare from their relative standing in the mental health distribution rather than from absolute levels of mental health. Under this assumption, rank-based mobility provides a direct measure of changes in individual well-being.

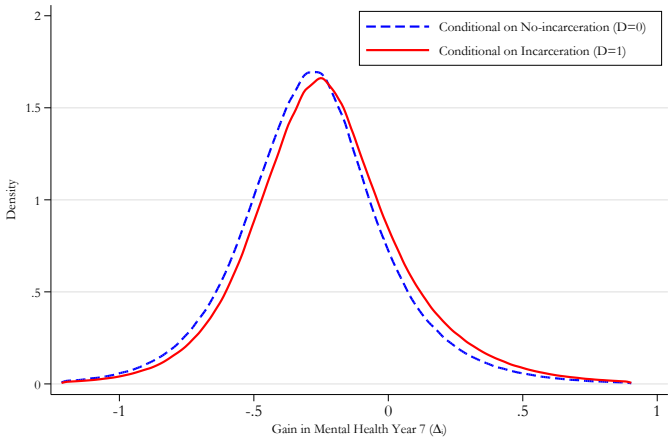
Figure 7: Factual and Counterfactual Distributions of Mental Health and Mental Health Gain (Year 7) conditional on Incarceration Status

(a) Mental Health Distribution (Year 7) Conditional on No-incarceration (D=0)

(b) Mental Health Distribution (Year 7) Conditional on Incarceration (D=1)



(c) Mental Health Gain from Incarceration (Year 7), by Incarceration Status (D)



Notes: Figure 7a displays the factual density $f(M_{i,7,0}|D_i = 0)$ and counterfactual density $f(M_{i,7,1}|D_i = 0)$ for people who do not go to prison. Figure 7b displays the factual density $f(M_{i,7,1}|D_i = 1)$ and counterfactual density $f(M_{i,7,0}|D_i = 1)$ for people who go to prison. Figure 7c displays the distribution of mental health gain from imprisonment, $\Delta_i = M_{i,7,1} - M_{i,7,0}$, conditional on no imprisonment ($f(\Delta_i|D_i = 0)$) and conditional on imprisonment ($f(\Delta_i|D_i = 1)$).

Figure 8 presents a decile-by-decile mobility matrix comparing the joint distribution of the two potential outcomes of mental health in year 7 $(M_{7,0}, M_{7,1})$.[11] Let $Q_{j,0}$ denote the jth quantile of the distribution of $M_{7,0}$, and let $Q_{k,1}$ denote the kth quantile of the distribution of $M_{7,1}$. The entry in cell (j, k) reports the probability

$$\Pr\left(M_{i,7,1} \in (Q_{k-1,1},\, Q_{k,1}] \mid M_{i,7,0} \in (Q_{j-1,0},\, Q_{j,0}]\right), \tag{42}$$

that is, each cell reports the probability that an individual who falls into a given decile of the mental health distribution under the no-incarceration scenario $(M_{7,0})$ would instead fall into a particular decile under the incarceration scenario $(M_{7,1})$.[12] For example, we read from Figure 8a, row 1, column 1, that an individual with potential mental health absent incarceration drawn from the first decile of the distribution of $M_{7,0}$ has a 89.2% probability of drawing a potential mental health under incarceration in the first decile of the distribution of $M_{7,1}$. In contrast, an individual in the sixth decile of the distribution of $M_{7,0}$ has only a 36.7% probability of being in the sixth decile of the distribution of $M_{7,1}$, indicating greater mobility in the middle of the distribution.

This pattern reflects a more general finding present in all panels of Figure 8: extreme deciles (such as the first and tenth) exhibit more *immobility* across counterfactual scenarios, while interior deciles are more fluid, independent of the cognition level. That is, individuals at the very top or bottom of the mental health distribution tend to retain their relative position regardless of incarceration and cognitive status, whereas individuals in the middle are more likely to shift ranks under the alternative scenario. This suggests that incarceration has more heterogeneous effects in the center of the distribution and relatively less impact on those already at the tails. Within the extreme deciles of the mental health distribution, the lowest deciles exhibit more *immobility* than the highest deciles, and this is true for all levels of cognitive skills. For example, going back to Figure 8a, individuals in the first decile of the distribution of $M_{7,0}$ have only a 10.8% probability of leaving their decile if they go to prison, while this probability is more than doubled (22.5%) for individuals in the tenth decile of the distribution of $M_{7,0}$.

Furthermore, beginning from the lowest deciles of the counterfactual mental health distribution

[11] Appendix Figures OA.3.1 and OA.3.2 replicate this mobility exercise on the potential mental health distributions conditional on incarceration status.

[12] Figures 8b-8d repeat the same exercise for individuals in a given tertile of the cognitive skills distribution, for Q calculated using all individuals. Formally, $\Pr\left(M_{i,7,1} \in (Q_{k-1,1},\, Q_{k,1}] \mid M_{i,7,0} \in (Q_{j-1,0},\, Q_{j,0}], C \in \text{Tertile } t\right)$.

absent incarceration, upward mobility at the intensive margin is notably constrained for those who experience incarceration. For example, among individuals initially in the first decile under the no-incarceration scenario, almost none transition above the third decile when subjected to incarceration. This pattern holds consistently across all strata of cognitive skills: even those with relatively high cognitive skills who occupy the bottom of the mental health distribution prior to incarceration exhibit very limited mobility within the incarceration distribution (see Figure 8d). In contrast, individuals in the tenth decile under no incarceration can experience downward mobility to as low as the fifth decile upon incarceration. Overall, relative rank persistence is more pronounced at the lower end than at the upper end of the distribution, regardless of cognitive skill level.

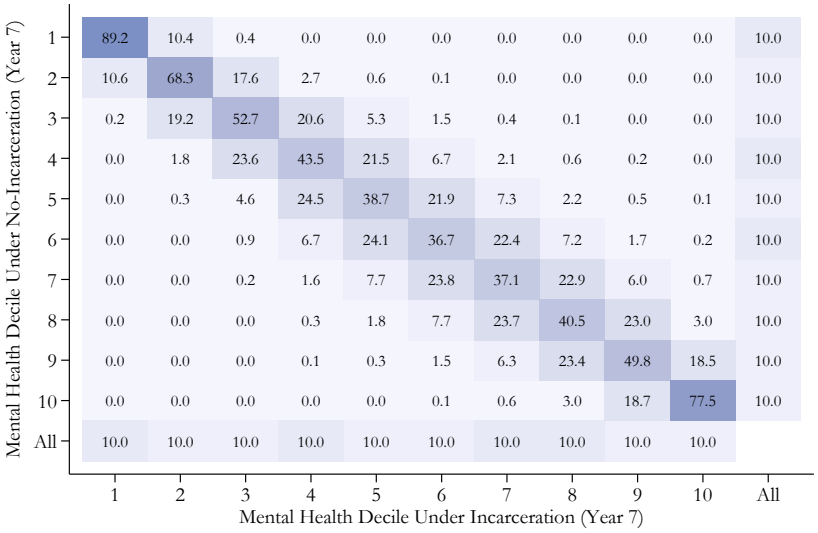
Nonetheless, higher cognitive skills increases mental health mobility at the top. For example, comparing the tenth row of Figure 8b (low cognition) to the tenth row of Figure 8d (high cognition), the share of individuals who remain in the top decile of the mental health distribution after incarceration decreases from 80.1% to 71.9% when we move from the first to the third tertile of the cognitive skills distribution. Focusing on the marginal distributions across the panels of Figure 8 reveals that individuals with relatively high cognitive skills tend to have lower positions in the mental health distribution relative to individuals with low and medium cognition levels. However, this reverses when we look at the potential mental health distributions conditional on individuals who do not go to prison (see Figure OA.3.1).

The mobility matrices in Figure 8 are useful for visualizing mobility patterns across mental health deciles within each group. However, to compare overall levels of mobility across different groups (i.e., across the various mobility matrices in Figure 8), we require a summary statistic. Table 8 addresses this need by presenting three alternative measures of positional movement, each condensing the full mobility matrix into a single index. Column (1) reports the Chi-Square Mobility Index, which measures how much the observed mobility matrix deviates from random mobility. Column (2) presents the Mobility Ratio Index, which measures the share of individuals who do *not* remain on the same decile across the two distributions. Column (3) shows the Mean Deciles Move Index, which captures the average number of deciles each individual moves between the two distributions.[13] In all three cases, higher values of the index correspond to greater mobility.

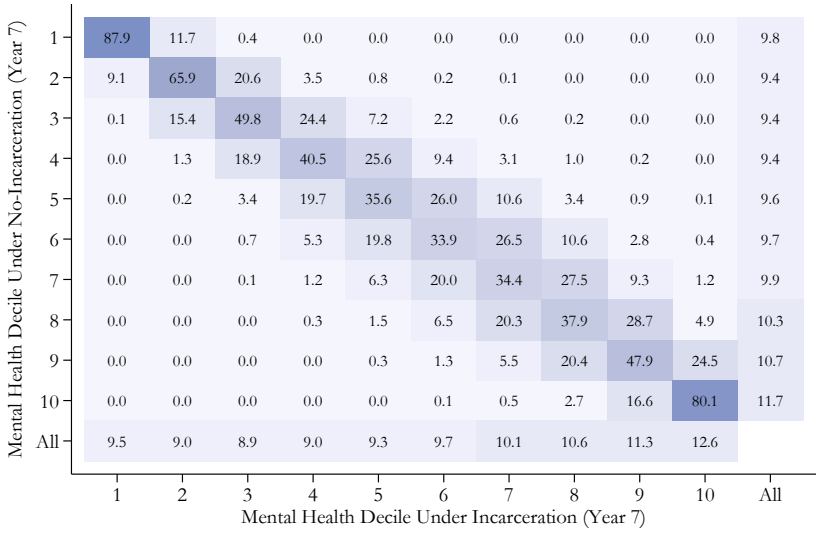
[13]Detailed formulas for each measure are provided in Appendix Section OA.3.1.

Figure 8: Positional Mobility Analyses (Cell %)

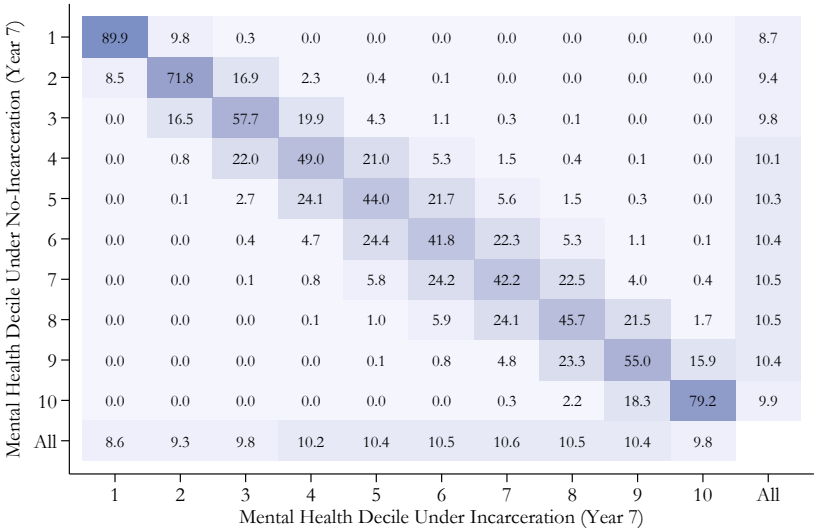
(a) Unconditional



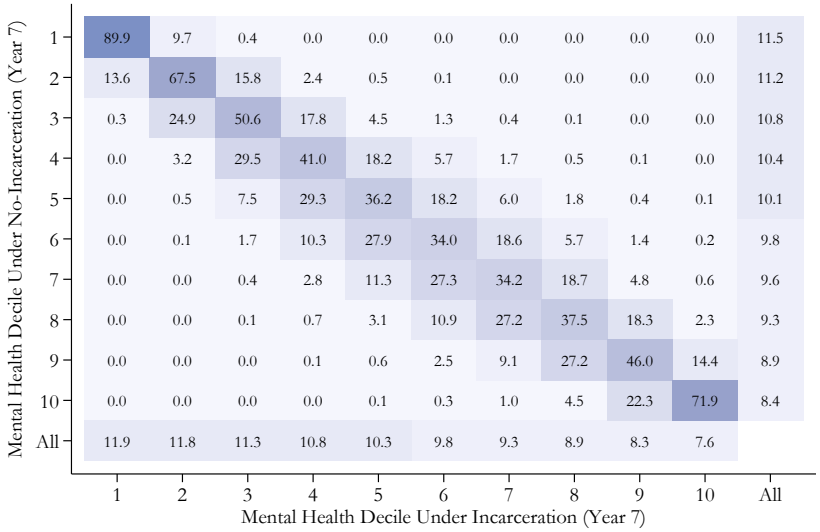
(b) For C (Cognitive) in Tertile 1



(c) For C (Cognitive) in Tertile 2



(d) For C (Cognitive) in Tertile 3



Notes: Let $Q_{j,0}$ be the jth quantile of $M_{7,0}$, and $Q_{k,1}$ be the kth quantile of $M_{7,1}$. Cell (j,k) in Figure 8a reports $\Pr\left(M_{i,7,1} \in (Q_{k-1,1},\, Q_{k,1}] \mid M_{i,7,0} \in (Q_{j-1,0},\, Q_{j,0}]\right)$. Figures 8b-8d report $\Pr\left(M_{i,7,1} \in (Q_{k-1,1},\, Q_{k,1}] \mid M_{i,7,0} \in (Q_{j-1,0},\, Q_{j,0}], C \in \text{Tertile } t\right)$. Shading ranges from white (0%) to dark blue (100%).

Each row of Table 8 corresponds to a different group defined by incarceration status and cognitive skill level. A consistent pattern emerges across all three metrics: mental health mobility follows a U-shaped relationship with cognitive skills—higher among individuals with either low or high cognitive skills, and lower among those in the middle of the cognitive distribution. This suggests that both ends of the cognitive spectrum may be more sensitive to the effects of incarceration on mental health rankings, while middle-skill individuals are relatively more anchored in their initial position.

Table 8: Mobility Measures - Mental Health Year 7 (M_7)

	Chi-Square Mobility Index (1)	Mobility Ratio Index (2)	Mean Deciles Move Index (3)
Unconditional on Imprisonment:			
No Prison to Prison	-3006.078	0.466	0.585
No Prison to Prison | Low Cognitive	-2894.640	0.481	0.619
No Prison to Prison | Med Cognitive	-3353.547	0.432	0.514
No Prison to Prison | High Cognitive	-2851.146	0.485	0.622
Conditional on D=0:			
No Prison to Prison	-3044.109	0.467	0.583
No Prison to Prison | Low Cognitive	-2953.909	0.478	0.605
No Prison to Prison | Med Cognitive	-3404.747	0.432	0.511
No Prison to Prison | High Cognitive	-2866.396	0.491	0.631
Conditional on D=1:			
No Prison to Prison	-2989.046	0.465	0.586
No Prison to Prison | Low Cognitive	-2867.578	0.484	0.627
No Prison to Prison | Med Cognitive	-3326.192	0.432	0.516
No Prison to Prison | High Cognitive	-2846.510	0.480	0.614

Notes: This table reports three alternative measures of *positional* movement: (1) the Chi-Square Mobility Index, which measures how much the observed transition matrix deviates from random mobility; (2) the Mobility Ratio Index, which measures the share of individuals who do *not* remain on the same decile across the two distributions; and (3) the Mean Deciles Move Index, which captures the average number of deciles each individual moves between the two distributions. Detailed formulas for each measure are provided in Appendix Section OA.3.1. In all three cases, higher values indicate greater mobility. Each row corresponds to a different subsample, defined by incarceration status and cognitive skill level.

While positional mobility captures changes in relative standing in the mental health distribution, it does not reveal whether individuals experience improvements or declines in absolute mental health. A downward shift in decile may still coincide with a mental health gain if the overall distribution shifts upward. To assess mental health *growth*, we fix the reference to the observed distribution of

mental health in year 7 (M_7). We then compare each individual's location in this distribution under the incarceration and no-incarceration scenarios, i.e., we summarize $(f_{M_7}(M_{i,7,0}),\ f_{M_7}(M_{i,7,1}))$.

Figure 9 presents a decile-by-decile mobility matrix summarizing this growth in individuals' outcomes under the two scenarios.[14] Let Q_j denote the jth quantile of the distribution of M_7. The entry in cell (j, k) reports the probability

$$\Pr\left(M_{i,7,1} \in (Q_{k-1},\ Q_k] \mid M_{i,7,0} \in (Q_{j-1},\ Q_j]\right), \tag{43}$$

that is, each cell reports the probability that an individual under the no-incarceration scenario ($M_{7,0}$) who falls into a given decile of the observed mental health distribution, would instead fall into a particular decile of the same distribution under the incarceration scenario ($M_{7,1}$).[15]

Unlike the positional mobility analysis, this matrix uses a single distribution to assign deciles and evaluates whether individuals move up or down in that common distribution based on their counterfactual outcome. The accompanying Table 9 quantifies growth using four indices: (1) The share of individuals who move at least one decile downward; (2) The average number of deciles lost, assigning zero to non-losers; (3) The average mental health loss among those who move at least one decile downward ("losers"); and (4) The average mental health gain among individuals who move at least one decile upward ("winners").

Compared to the positional mobility matrices, Figure 9 shows an even stronger pattern of persistence at the bottom of the distribution: individuals starting in the first decile almost always remain in the first decile. For example, we read from Figure 9a, row 1, column 1, that an individual with potential mental health absent incarceration drawn from the first decile of the distribution of M_7 has a 93.7% probability of drawing a potential mental health under incarceration in the first decile of the distribution of M_7. This pattern holds across all levels of cognition (see Figures 9b to 9d). Moreover, individuals who do not typically go to prison (i.e., conditional on $D = 0$) experience substantially larger declines in mental health if incarcerated than those who typically go to prison (i.e., conditional on $D = 1$), indicating that incarceration harms the

[14] Appendix Figures OA.3.3 and OA.3.4 replicate this growth exercise on the potential mental health distributions conditional on incarceration status.

[15] Figures 9b-9d repeat the same exercise for individuals in a given tertile of the cognitive skills distribution, for Q calculated using all individuals. Formally, $\Pr\left(M_{i,7,1} \in (Q_{k-1},\ Q_k] \mid M_{i,7,0} \in (Q_{j-1},\ Q_j], C \in \text{Tertile } t\right)$.

Figure 9: Growth Analyses (Cell %)

(a) Unconditional



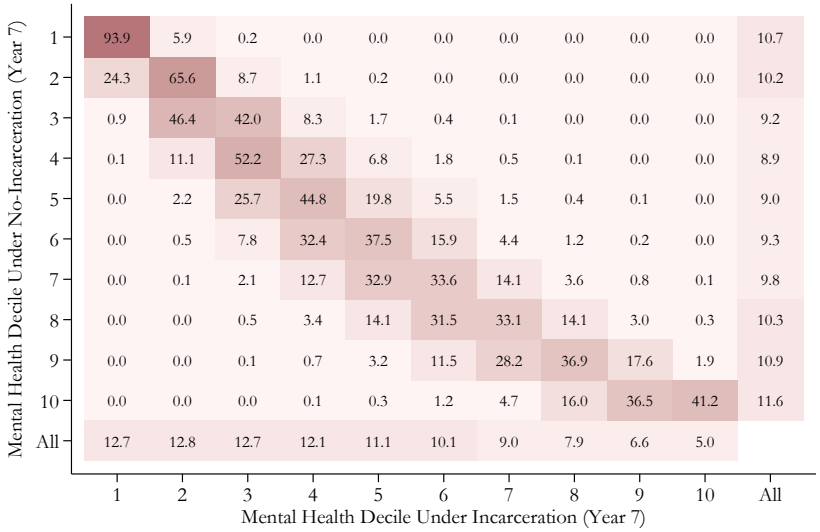
(b) For C (Cognitive) in Tertile 1



(c) For C (Cognitive) in Tertile 2



(d) For C (Cognitive) in Tertile 3



Notes: Let Q_j be the jth quantile of M_7. Cell (j,k) in Figure 9a reports $\Pr\left(M_{i,7,1} \in (Q_{k-1}, Q_k] \mid M_{i,7,0} \in (Q_{j-1}, Q_j]\right)$. Figures 9b-9d report $\Pr\left(M_{i,7,1} \in (Q_{k-1}, Q_k] \mid M_{i,7,0} \in (Q_{j-1}, Q_j], C \in \text{Tertile } t\right)$. Shading ranges from white (0%) to dark red (100%).

Table 9: Growth Measures - Mental Health Year 7 (M_7)

	Negative Growth Ratio (1)	Deciles Downward Move (2)	Mental Health Loss (Losers) (3)	Mental Health Gain (Winners) (4)
Unconditional on Imprisonment:				
No Prison to Prison	0.555	0.816	0.401	0.315
No Prison to Prison | Low Cognitive	0.496	0.715	0.390	0.316
No Prison to Prison | Med Cognitive	0.586	0.822	0.372	0.302
No Prison to Prison | High Cognitive	0.585	0.912	0.438	0.326
Conditional on D=0:				
No Prison to Prison	0.585	0.875	0.411	0.304
No Prison to Prison | Low Cognitive	0.525	0.766	0.397	0.300
No Prison to Prison | Med Cognitive	0.613	0.872	0.381	0.291
No Prison to Prison | High Cognitive	0.607	0.962	0.447	0.317
Conditional on D=1:				
No Prison to Prison	0.533	0.772	0.393	0.322
No Prison to Prison | Low Cognitive	0.478	0.684	0.386	0.323
No Prison to Prison | Med Cognitive	0.565	0.783	0.366	0.308
No Prison to Prison | High Cognitive	0.564	0.865	0.429	0.332

Notes: This table presents four measures of mental health growth. Column (1) reports the share of individuals who move down by at least one decile. Column (2) shows the average number of deciles lost among the full population, assigning zero to those who do not move down. Column (3) reports the average mental health loss for individuals who move at least one decile downward ("losers"), while Column (4) reports the average mental health gain for individuals who move at least one decile upward ("winners").

former more severely (see Figures OA.3.3 and OA.3.4). These patterns are confirmed by our indices in Table 9. For instance, individuals who typically do not go to prison ($D = 0$) have a negative growth ratio of 0.585 and an average downward move of 0.875 deciles, compared with 0.533 and 0.772, respectively, for individuals who typically go to prison ($D = 1$).

6 Conclusion: Towards a Richer Understanding of Policy Effects

This paper develops and estimates a dynamic factor model to analyze the distributional consequences of incarceration on mental health among justice-involved youth. By modeling the joint evolution of latent cognitive and mental health factors, we move beyond average effects to characterize heterogeneity and social mobility patterns in greater detail.

A central contribution is the nonparametric recovery of the joint distribution of unobserved skill factors without imposing restrictive parametric assumptions on their functional form. This feature enables a flexible and rich description of individual variation and policy impact heterogeneity. Nevertheless, this nonparametric identification and estimation rely on a set of carefully stated assumptions, such as the existence of multiple measurements (dedicated measures) and independence conditions among measurement errors and latent variables.

Methodologically, this work demonstrates the value and feasibility of combining latent factor models with dynamic structural equations to uncover distributional policy effects under a transparent set of identifying conditions. The introduction of MCMC methods for estimation dramatically reduces the computational complexity required for estimation.

Our empirical results illustrate substantial variation in mental health trajectories associated with incarceration status and baseline skill endowments. Counterfactual analyses reveal that policy effects differ across subpopulations defined by initial skill levels and mental health status, underscoring the limitations of mean effect summaries.

Our analysis is restricted to youth who have already engaged in relatively serious criminal behavior at an early age, a subgroup that accounts for a disproportionate share of overall youth crime. While this focus provides valuable insight into a high-impact segment of the population, it may not generalize to the broader population. In particular, the mental health consequences of incarceration documented here may not generalize to individuals without prior justice involvement. Whether similar patterns would arise in the broader youth population remains an open empirical question, and addressing this issue constitutes an important direction for future research.

References

F. Agostinelli and M. Wiswall. Estimating the technology of children's skill formation. *Journal of Political Economy*, 133(3):846–887, 2025.

M. Bhuller, G. B. Dahl, K. V. Løken, and M. Mogstad. Incarceration, recidivism, and employment. *Journal of Political Economy*, 128(4):1269–1324, 2020.

M. Bhuller, L. Khoury, and K. V. Løken. Mental health consequences of correctional sentencing. *American Economic Journal: Economic Policy*, 17(1):70–105, 2025.

B. Biasi, M. S. Dahl, and P. Moser. Career effects of mental health: Evidence from an innovation in treating bipolar disorder. *Journal of Political Economy: Microeconomics*, 2025. (forthcoming).

P. Billingsley. *Probability and Measure*. John Wiley & Sons, 1995.

I. A. Binswanger, M. F. Stern, R. A. Deyo, P. J. Heagerty, A. Cheadle, J. G. Elmore, and T. D. Koepsell. Release from prison - a high risk of death for former inmates. *New England Journal of Medicine*, 356(2):157–165, 2007.

I. A. Binswanger, P. M. Krueger, and J. F. Steiner. Prevalence of chronic medical conditions among jail and prison inmates in the usa compared with the general population. *Journal of Epidemiology & Community Health*, 63(11):912–919, 2009.

W. A. Brock and S. N. Durlauf. Discrete choice with social interactions. *The Review of Economic Studies*, 68(2):235–260, 2001.

P. Carneiro, K. T. Hansen, and J. J. Heckman. 2001 lawrence r. klein lecture estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review*, 44(2):361–422, 2003.

N. A. Cholli and S. N. Durlauf. Intergenerational mobility. 2022.

F. Cunha, J. J. Heckman, and S. Navarro. Counterfactual analysis of inequality and social mobility. *Mobility and Inequality: Frontiers of Research in Sociology and Economics*, pages 290–348, 2006.

F. Cunha, J. J. Heckman, and S. M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

J. Currie and M. Stabile. Child mental health and human capital accumulation: The case of adhd. *Journal of Health Economics*, 25(6):1094–1118, 2006.

J. Currie, M. Stabile, P. Manivong, and L. L. Roos. Child health and young adult outcomes. *Journal of Human Resources*, 45(3):517–548, 2010.

C. S. Diaz-Campo, B. H. Hamilton, M. Luccioni, and H. S. Suh. Mental health and the early career dynamics of young men. *Working Paper*, 2025.

S. N. Durlauf. A theory of persistent income inequality. *Journal of Economic Growth*, 1(1):75–93, 1996.

S. N. Durlauf. A framework for the study of individual behavior and social interactions. *Sociological Methodology*, 31(1):47–87, 2001.

S. N. Durlauf and Y. M. Ioannides. Social interactions. *Annual Review of Economics*, 2(1):451–478, 2010.

S. N. Durlauf and D. S. Nagin. The deterrent effect of imprisonment. In *Controlling Crime: Strategies and Tradeoffs*, pages 43–94. University of Chicago Press, 2010.

S. N. Durlauf, S. Navarro, and D. A. Rivers. Understanding aggregate crime regressions. *Journal of Econometrics*, 158(2):306–317, 2010.

S. N. Durlauf, A. Kourtellos, and C. M. Tan. The great gatsby curve. *Annual Review of Economics*, 14(1):571–605, 2022.

G. S. Fields. The many facets of economic mobility. In *Inequality, Poverty and Well-Being*, pages 123–142. Springer, 2006.

J. C. Fruehwirth, S. Navarro, and Y. Takahashi. How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects. *Journal of Labor Economics*, 34 (4):979–1021, 2016.

A. Garin, D. Koustas, C. McPherson, S. Norris, M. Pecenco, E. K. Rose, Y. Shem-Tov, and J. Weaver. The impact of incarceration on employment, earnings, and tax filing. *Econometrica*, 93(2):503–538, 2025.

G. Gordon, J. B. Jones, U. Neelakantan, and K. Athreya. Incarceration, employment, and earnings: Dynamics and differences. *Review of Economic Dynamics*, 51:677–697, 2023.

A. Haglund, D. Tidemalm, J. Jokinen, N. Långström, P. Lichtenstein, S. Fazel, and B. Runeson. Suicide after release from prison: A population-based cohort study from sweden. *Journal of Clinical Psychiatry*, 75(10):1047, 2014.

J. J. Heckman. Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy*, 109(4):673–748, 2001.

J. J. Heckman. The scientific model of causality. *Sociological Methodology*, 35(1):1–97, 2005.

J. J. Heckman and E. J. Vytlacil. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences*, 96 (8):4730–4734, 1999.

J. J. Heckman, S. Urzua, and E. Vytlacil. Understanding instrumental variables in models with essential heterogeneity. *The review of Economics and Statistics*, 88(3):389–432, 2006.

R. Hjalmarsson and M. J. Lindquist. The health effects of prison. *American Economic Journal: Applied Economics*, 14(4):234–270, 2022.

G. W. Imbens. Causality in econometrics: Choice vs chance. *Econometrica*, 90(6):2541–2566, 2022.

G. Jolivet and F. Postel-Vinay. A structural analysis of mental health and labour market trajectories. *The Review of Economic Studies*, 92(3):1920–1954, 2025.

K. G. Jöreskog. Structural equation models in the social sciences: Specification, estimation and testing. *Applications of Statistics*, pages 265–287, 1977.

K. G. Jöreskog and A. S. Goldberger. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70(351a): 631–639, 1975.

J. R. Kling. Incarceration length, employment, and earnings. *American Economic Review*, 96(3): 863–876, 2006.

I. Kotlarski. On characterizing the gamma and the normal distribution. *Pacific Journal of Mathematics*, 20(1):69–76, 1967.

I. Kuziemko. How should inmates be released from prison? an assessment of parole versus fixed-sentence regimes. *The Quarterly Journal of Economics*, 128(1):371–424, 2013.

M. A. Mancino. Rehabilitating futures: Assessing the effects of correctional employment-focused programs on recidivism and employment. *European Economic Review*, 173:104954, 2025.

C. F. Manski. Identification of binary response models. *Journal of the American Statistical Association*, 83(403):729–738, 1988.

R. L. Matzkin. Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica*, pages 239–270, 1992.

K. C. Monahan, L. Steinberg, E. Cauffman, and E. P. Mulvey. Trajectories of antisocial behavior and psychosocial maturity from adolescence to young adulthood. *Developmental Psychology*, 45 (6):1654, 2009.

M. Mueller-Smith. The criminal and labor market impacts of incarceration. *Working Paper*, 18, 2015.

S. Norris, M. Pecenco, and J. Weaver. The effect of incarceration on mortality. *Review of Economics and Statistics*, 106(4):956–973, 2024.

B. P. Rao. *Identifiability in Stochastic Models*. Academic Press, 1992.

E. K. Rose and Y. Shem-Tov. How does incarceration affect reoffending? estimating the dose-response function. *Journal of Political Economy*, 129(12):3302–3356, 2021.

D. F. Salazar. *Essays on Criminal Behaviour, Human Capital Formation, and Mental Health*. PhD thesis, The University of Western Ontario (Canada), 2020.

C. A. Schubert, E. P. Mulvey, L. Steinberg, E. Cauffman, S. H. Losoya, T. Hecker, L. Chassin, and G. P. Knight. Operational lessons from the pathways to desistance project. *Youth Violence and Juvenile Justice*, 2(3):237–255, 2004.

K. Turney, C. Wildeman, and J. Schnittker. As fathers and felons: Explaining the effects of current and recent incarceration on major depression. *Journal of Health and Social Behavior*, 53(4): 465–481, 2012.

B. Wang, R. Frank, and S. Glied. Lasting scars: The impact of depression in early adulthood on subsequent labor market outcomes. *Health Economics*, 32(12):2694–2708, 2023.

B. Williams. Identification of the linear factor model. *Econometric Reviews*, 39(1):92–109, 2020.

Online Appendix

OA.1 Priors and Posterior Derivations Used in Estimation

In this appendix, we provide a detailed summary of the estimation steps for our Bayesian model. Because the derivations follow standard approaches in the Bayesian literature, we present the final forms of the derived posteriors and explicitly state the priors and updating rules used. For inference, we implement a Gibbs sampler that cycles through sampling from the conditional posterior distributions of parameters and latent variables, completing the likelihood at each step by conditioning on latent factors C, M, U.

Priors and Posteriors for Baseline Continuous Cognitive Measures

For any continuous baseline cognitive measure j, we assume the regression model:

$$\mathcal{C}_j = X\gamma_{\mathcal{C},j} + C\psi_j + \varepsilon_{\mathcal{C},j}, j \in 1, ..., N_C,\ \varepsilon_{i,\mathcal{C},j} \sim N\left(0, \sigma^2_{\varepsilon_{\mathcal{C},j}}\right). \tag{O1}$$

- Priors
  - Coefficients on covariates, $\gamma_{\mathcal{C},j}$: non-informative prior.
  - Loading, $\psi_j \sim N\left(0, \sigma^2_{0,\psi,j}\right)$ with $\sigma^2_{0,\psi,j} = 10$ (diffuse).
  - Uniqueness variance, non-informative prior on $\frac{1}{\sigma^2_{\varepsilon_{\mathcal{C},j}}}$.
- Posteriors
  - Conditional posterior for $\gamma_{\mathcal{C},j}$:

$$\gamma_{\mathcal{C},j} \mid \centerdot \sim N\left(\widehat{\gamma}_{\mathcal{C},j}, \Sigma_{\gamma,\mathcal{C},j}\right),\ \ \widehat{\gamma}_{\mathcal{C},j} = \left(X'X\right)^{-1} X'\left(\mathcal{C}_j - C\psi_j\right),\ \ \Sigma_{\gamma,\mathcal{C},j} = \sigma^2_{\varepsilon_{\mathcal{C},j}}\left(X'X\right)^{-1}. \tag{O2}$$

  - Conditional posterior for ψ_j:

$$\psi_j \sim N\left(\widehat{\psi}_j, \Sigma_{\psi,j}\right),\ \ \Sigma_{\psi,j} = \frac{1}{\sigma^2_{\varepsilon_{\mathcal{C},j}}} C'C + \frac{1}{\sigma^2_{0,\psi,j}},\ \ \widehat{\psi}_j = \Sigma^{-1}_{\psi,j} \frac{1}{\sigma^2_{\varepsilon_{\mathcal{C},j}}} C'\left(\mathcal{C}_j - X\gamma_{\mathcal{C},j}\right). \tag{O3}$$

– Conditional posterior for uniqueness precision:

$$\frac{1}{\sigma^2_{\varepsilon_{\mathcal{C},j}}} \sim Gamma\left(\frac{n_j}{2}, \frac{\varepsilon'_{\mathcal{C},j}\varepsilon_{\mathcal{C},j}}{2}\right), \tag{O4}$$

where n_j is the number of observations for this equation.

The above procedure is repeated for each continuous cognitive measure.

Priors and Posteriors for Ordered Discrete Measures

For discrete ordered measures (e.g., baseline mental health M), the model assumes latent indices

$$\mathcal{M}^*_{i,j} = X'_i\gamma_{\mathcal{M},j} + M_i\mu_j + \varepsilon_{i,\mathcal{M},j}, \;\; \varepsilon_{i,\mathcal{M},j} \sim N(0,1). \tag{O5}$$

We use $\mathcal{M}^*_{i,j}$ for data completion.

- Priors
  - Coefficients on covariates, $\gamma_{\mathcal{M},j}$: non-informative prior.
  - Loading, $\mu_j \sim N\left(0, \sigma^2_{0,\mu,j}\right)$ with $\sigma^2_{0,\mu,j} = 10$ (diffuse).
  - Cutoffs, $o_{\mathcal{M},j,k} \sim \text{Unif}(\underline{o}, \overline{o})$ with $\underline{o} = -1000$ and $\overline{o} = 1000$.
- Posteriors
  - Coefficients $\gamma_{\mathcal{M},j}, \mu_j$ sampled like in a linear regression (variance fixed to 1).
  - Cutoffs updated via uniform distributions constrained by observed latent categories:

$$o_{\mathcal{M},j,k} \sim \text{Unif}\left(\max\left(o_{\mathcal{M},j,k-1}, \max_{i:\mathcal{M}_{i,j}=k} \mathcal{M}^*_{i,j,}, \underline{o}\right), \min\left(o_{\mathcal{M},j,k+1}, \min_{i:\mathcal{M}_{i,j}=k+1} \mathcal{M}^*_{i,j,}, \overline{o}\right)\right). \tag{O6}$$

  - Latent indices $\mathcal{M}^*_{i,j}$ sampled from truncated normal distributions:

$$\mathcal{M}^*_{i,j} \sim TN_{(o_{\mathcal{M},j,\mathcal{M}_{i,j}-1}, o_{\mathcal{M},j,\mathcal{M}_{i,j}}]}\left(X'_i\gamma_{\mathcal{M},j} + M_i\mu_j, 1\right). \tag{O7}$$

Priors and Posteriors for Treatment Model

Notice that the treatment model is a special case of the sampler for ordered discrete measures. The treatment variable D_i is modeled similarly with a single cutoff fixed at zero, so the latent index samples from truncated normals either above or below zero.

Mental Health System in Period 7

Conditioned on (C_i, M_i, U_i) and all the other parameters, the latent factor $M_{i,7}$ is known from the law of motion equation. Thus, the measurement model at period 7 follows the same ordered discrete setup described above, and sampling proceeds analogously.

Posteriors for the Law of Motion Parameters

Incorporating the law of motion into the measurement model, define:

$$\mathcal{M}^*_{i,j,7} = X_i'\gamma_{\mathcal{M},j,7} + (\delta_T D_i + \lambda_C C_i + \lambda_M M_i + \lambda_{CT} C_i D_i + \lambda_{MT} M_i D_i + U_i)\,\mu_{j,7} + \varepsilon_{i,\mathcal{M},j,7}. \tag{O8}$$

Rearranging and defining:

$$\mathcal{M}^{**}_{i,j,7} = \delta_T D_{i,j} + \lambda_C C_{i,j} + \lambda_M M_{i,j} + \lambda_{CT} C D_{i,j} + \lambda_{MT} M D_{i,j} + U_{i,j} + \varepsilon_{i,\mathcal{M},j,7}, \tag{O9}$$

with notation such as $D_{i,j} = D_i \mu_{j,7}$, $C_{i,j} = C_i \mu_{j,7}$, and so on. This is a standard linear regression model like the ones we have been describing, except that the same parameters δ_T, λ_C, etc, show up in all period 7 equations. As a consequence, the sample moves over both i and j. In other words, if I have a non-informative prior on δ_T for example, the posterior would be

$$\delta_T \sim N\left(\mu_{\delta_T}, \Sigma_{\delta_T}\right) \tag{O10}$$

with

$$Z^{\delta_T}_{ij} = \mathcal{M}^{**}_{i,j,7} - \lambda_C C_{i,j} - \lambda_M M_{i,j} - \lambda_{CT} C D_{i,j} - \lambda_{MT} M D_{i,j} - U_{i,j}, \tag{O11}$$

$$\Sigma_{\delta_T} = \left(\sum_{i,j} D_{i,j} D_{i,j}'\right)^{-1} \tag{O12}$$

and

$$\mu_{\delta_T} = \Sigma_{\delta_T} \left(\sum_{i,j} D_{i,j} Z^{\delta_T}_{i,j}\right). \tag{O13}$$

An equivalent posterior follows for the remaining parameters as in the case of the measurement systems we described before. The only difference being that the sum is done over both i and j.

Next is the derivation of the posterior for the parameters describing the distributions of the latent factors C, M and U. We begin by describing the posterior for the parameters of the mixture for the univariate case of U. We complete the likelihood by defining h_{Ui} to be a variable that takes value h

if the element of the mixture U_i is drawn from is element h. Let $I_{U,h} = \{i : h_{Ui} = h\}, n_{U,h} = |I_{U,h}|$.

Given a normal prior for the h^{th} component mean such that

$$m_{U,h} \sim N\left(m_{U0}, s_{U0}\right), \tag{O14}$$

the posterior for the h^{th} component means is:

$$m_{U,h} \sim N\left(\bar{m}_{U,h}, V_{U,h}^{-1}\right), \quad V_{U,h} = s_{U0} + \frac{n_{U,h}}{\sigma_{U,h}^2}, \quad \bar{m}_{U,h} = \frac{s_{U0} m_{U0} + \frac{1}{\sigma_{U,h}^2} \sum_{i \in I_{U,h}} U_i}{V_{U,h}}. \tag{O15}$$

In order to have a diffuse prior, we set $m_{U0} = 0$ and $s_{U0} = 10$.

Given a Gamma prior for the h^{th} component inverse of the variance such that

$$\frac{1}{\sigma_{U,h}^2} \sim Gamma\left(a_0, b_0\right) \tag{O16}$$

The variances are updated according to the posterior:

$$\frac{1}{\sigma_{U,h}^2} \sim \text{Gamma}\left(a_0 + \frac{n_{U,h}}{2},\ b_0 + \frac{1}{2} \sum_{i \in I_{U,h}} \left(U_i - m_{U,h}\right)^2\right). \tag{O17}$$

We set $a_0 = b_0 = 2$.

Given a Dirichlet prior for the mixture weights such that

$$\pi_U \sim Dirichlet\left(a_U, ..., a_U\right), \tag{O18}$$

$$\pi_U \sim \text{Dirichlet}(a_U + n_{U,1}, \ldots, a_U + n_{U,3}). \tag{O19}$$

In order to update the mixture component indicators h_{Ui}, we notice that the conditional likelihood is given by

$$P(h_{Ui} = h \mid U_i, \{m_{U,h}, \sigma_{U,h}^2, \pi_{U,h}\}) \propto \pi_{U,h} \cdot \phi\left(U_i; m_{U,h}, \sigma_{U,h}^2\right). \tag{O20}$$

This defines a discrete distribution over h, which can be sampled by computing log-weights:

$$\log w_{U,h} = \log \pi_{U,h} - \frac{1}{2}\left[\frac{\left(U_i - m_{U,h}\right)^2}{\sigma_{U,h}^2} + \log \sigma_{U,h}^2\right]. \tag{O21}$$

Then, defining $\rho_{i,U,h} \propto e^{\log w_{U,h}}$ and normalize $\sum_h \rho_{i,U,h} = 1$ we can sample from

$$h_{Ui} \sim \text{Categorical}(\rho_{i,U,1}, \ldots, \rho_{i,U,3}). \tag{O22}$$

Next, we notice that, conditional on the mixture component indicators, sampling the parameters

of the distributions for C_i, M_i follows the exact same process as that for U_i. The only difference arises when sampling the mixture component indicators, as they need to consider both factors. In this case, the log-weights are computed as

$$\log w_{CM,h} = \log \pi_{CM,h} - \frac{1}{2}\left[\frac{(C_i - m_{C,h})^2}{\sigma^2_{C,h}} + \log \sigma^2_{C,h} + \frac{(M_i - m_{M,h})^2}{\sigma^2_{M,h}} + \log \sigma^2_{M,h}\right]. \tag{O23}$$

Defining $\rho_{i,CM,h} \propto e^{\log w_{CM,h}}$ and normalizing $\sum_h \rho_{i,CM,h} = 1$ we can sample from the posterior

$$h_{CMi} \sim Categorical\left(\rho_{i,CM,h}, ..., \rho_{i,CM,6}\right). \tag{O24}$$

The final step is to describe the Gibbs sampler steps to sample the latent factors that we use for the completion. Take any of the factors, say C_i. The relevant terms of the posterior are

$$\begin{aligned} f\left(C_i \mid \cdot\right) \propto & \left[\prod_{j=1}^{N_C} \phi\left(\mathcal{C}^*_{i,j}; X_i'\gamma_{\mathcal{C},j} + C_i\psi_j, \sigma^2_{\varepsilon_{\mathcal{C},j}}\right)\right] \times \phi\left(C_i; m_{C0}, \sigma^2_{C0}\right) \\ & \times \phi\left(T_i; X_i'\gamma_T + C_i\psi_T + M_i\mu_T, 1\right) \\ & \times \prod_{j=1}^{N_{M_7}} \phi\left(\mathcal{M}^{**}_{i,j,7}; \delta_T D_{i,j} + \lambda_C C_{i,j} + \lambda_M M_{ij} + \lambda_{CT} CD_{i,j} + \lambda_{MT} MD_{i,j} + U_{i,1j}, 1\right) \end{aligned} \tag{O25}$$

This is the product of a normal likelihood and a normal prior, so the posterior is normal. Let us define:

$$\begin{aligned} \frac{1}{\sigma^2_{C|i}} &= \sum_{j=1}^{N_C} \frac{\psi_j^2}{\sigma^2_{\varepsilon_{\mathcal{C},j}}} + \psi_T^2 + \sum_{j=1}^{N_{M_7}} \left(\lambda_C \mu_{j,7} + \lambda_{CT} D_i \mu^2_{j,7}\right) + \frac{1}{\sigma^2_{C0}} \\ \eta_{C|i} &= \begin{pmatrix} \sum_{j=1}^{N_C} \frac{\psi_j}{\sigma^2_{\varepsilon_{\mathcal{C},j}}} \left(\mathcal{C}^*_{i,j} - X_i'\gamma_{\mathcal{C},j}\right) + \psi_T\left(T_i - X_i'\gamma_T - M_i\mu_T\right) + \frac{m_{c0}}{\sigma^2_{C0}} \\ \sum_{j=1}^{N_{M_7}} \left(\lambda_C \mu_{j,7} + \lambda_{CT} D_i \mu_{j,7}\right)\left(\mathcal{M}^{**}_{i,j,7} - \delta_T D_{i,j} - \lambda_M M_{i,j} - \lambda_{MT} MD_{i,j} - U_{i,j}\right) \end{pmatrix} \end{aligned} \tag{O26}$$

Then the posterior for C_i is:

$$C_i \mid \cdot \sim N\left(m_{C|i}, \sigma^2_{C|i}\right), \quad \text{where} \quad m_{C|i} = \sigma^2_{C|i}\eta_{C|i}. \tag{O27}$$

The posteriors for M_i and U_i can be derived in a similar fashion.

OA.2 Empirical Application - Additional Results

OA.2.1 Additional Model Results

Table OA.2.1: Estimated Parameters from Factor Model — Cognitive Skills

	(1) WASI IQ	(2) Stroop Color	(3) Stroop Word	(4) Stroop Color/Word	(5) Trail Making Part A	(6) Trail Making Part B
Constant	-0.223	-0.335	-0.223	-0.508	0.826	1.122
	(0.208)	(0.219)	(0.214)	(0.213)	(0.258)	(0.277)
Age 15	-0.011	0.237	0.150	0.308	0.700	0.419
	(0.133)	(0.140)	(0.137)	(0.135)	(0.164)	(0.175)
Age 16	-0.077	0.322	0.272	0.411	0.807	0.428
	(0.124)	(0.131)	(0.128)	(0.127)	(0.153)	(0.164)
Age 17	0.082	0.398	0.207	0.479	0.754	0.540
	(0.126)	(0.134)	(0.129)	(0.130)	(0.156)	(0.168)
Age 18	0.111	0.117	0.253	0.166	0.902	0.529
	(0.177)	(0.187)	(0.182)	(0.184)	(0.219)	(0.234)
Female	-0.148	0.129	0.216	0.069	0.238	0.133
	(0.101)	(0.106)	(0.105)	(0.100)	(0.128)	(0.135)
White	0.491	0.162	0.123	0.281	0.447	0.289
	(0.197)	(0.203)	(0.202)	(0.199)	(0.240)	(0.259)
Hispanic	-0.223	-0.051	-0.141	-0.005	0.070	-0.132
	(0.193)	(0.199)	(0.197)	(0.191)	(0.233)	(0.252)
Black	-0.067	-0.008	-0.210	-0.042	-0.134	-0.199
	(0.196)	(0.201)	(0.200)	(0.195)	(0.237)	(0.252)
Phoenix	0.519	0.044	0.207	0.248	0.374	0.387
	(0.096)	(0.100)	(0.099)	(0.098)	(0.122)	(0.131)
Variance	0.679	0.388	0.533	0.582	1.000	1.000
	(0.040)	(0.039)	(0.037)	(0.039)	(0.000)	(0.000)
Cognitive Factor	1.000	1.499	1.266	1.162	0.854	1.244
	(0.000)	(0.105)	(0.095)	(0.091)	(0.114)	(0.137)
Cutoff 1					0.000	0.000
					(0.000)	(0.000)
Cutoff 2					0.789	1.198
					(0.072)	(0.083)
Cutoff 3					2.057	2.100
					(0.094)	(0.101)

Notes: This table reports the mean and standard deviation of each parameter across 5,000 draws from the posterior distribution for the cognitive measure system. Standard errors are shown in parentheses below the mean point estimates. IQ and the Stroop components are modeled using a linear-in-parameters specification, while the Trail-Making tests are estimated using an ordered threshold model. The cognitive skills factor is normalized to have a loading of one on the WASI IQ score.

Table OA.2.2: Estimated Parameters from Factor Model — Mental Health Measures (Baseline)

	(1) Somatization	(2) Depression	(3) Anxiety	(4) Hostility	(5) Obsessive Compulsive	(6) Interpersonal Sensitivity	(7) Phobic Anxiety	(8) Paranoid Ideation	(9) Psychoticism
Constant	0.585	0.794	0.926	0.842	1.405	0.336	-0.242	1.745	0.980
	(0.349)	(0.368)	(0.390)	(0.338)	(0.422)	(0.327)	(0.333)	(0.353)	(0.364)
Age 15	0.147	0.091	-0.156	0.260	0.014	-0.026	-0.073	0.102	0.070
	(0.221)	(0.241)	(0.252)	(0.218)	(0.263)	(0.212)	(0.225)	(0.221)	(0.234)
Age 16	-0.090	0.153	-0.068	0.236	0.207	-0.067	-0.088	0.024	0.153
	(0.208)	(0.219)	(0.232)	(0.202)	(0.239)	(0.194)	(0.208)	(0.201)	(0.215)
Age 17	0.142	0.467	0.397	0.509	0.530	0.146	0.197	0.449	0.304
	(0.222)	(0.232)	(0.241)	(0.213)	(0.255)	(0.203)	(0.215)	(0.212)	(0.226)
Age 18	0.247	0.952	0.362	0.416	0.878	0.123	0.135	1.200	1.041
	(0.294)	(0.322)	(0.329)	(0.293)	(0.351)	(0.278)	(0.295)	(0.306)	(0.309)
Female	0.501	0.404	0.283	0.587	0.339	0.419	0.171	0.239	0.275
	(0.175)	(0.190)	(0.197)	(0.178)	(0.215)	(0.164)	(0.166)	(0.179)	(0.185)
White	0.119	-0.195	-0.250	0.057	-0.174	-0.230	-0.287	-0.442	-0.712
	(0.322)	(0.347)	(0.354)	(0.312)	(0.382)	(0.306)	(0.318)	(0.324)	(0.346)
Hispanic	0.055	0.038	0.029	-0.131	-0.175	-0.103	0.009	-0.239	-0.304
	(0.315)	(0.340)	(0.352)	(0.309)	(0.374)	(0.299)	(0.309)	(0.321)	(0.336)
Black	-0.504	-0.464	-0.679	-0.229	-0.612	-0.308	-0.238	-0.279	-0.631
	(0.323)	(0.344)	(0.358)	(0.310)	(0.384)	(0.303)	(0.315)	(0.321)	(0.337)
Phoenix	-0.226	-0.148	-0.075	-0.078	0.266	0.099	-0.047	-0.320	-0.081
	(0.165)	(0.173)	(0.183)	(0.162)	(0.194)	(0.155)	(0.157)	(0.163)	(0.171)
Variance	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mental Health	-1.042	-1.182	-1.246	-1.000	-1.401	-0.860	-0.825	-1.061	-1.141
	(0.085)	(0.095)	(0.106)	(0.000)	(0.117)	(0.075)	(0.074)	(0.090)	(0.097)
Cutoff 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cutoff 2	0.645	0.734	0.711	0.567	0.554	0.716	0.618	0.617	0.668
	(0.058)	(0.063)	(0.069)	(0.057)	(0.064)	(0.059)	(0.062)	(0.066)	(0.061)
Cutoff 3	1.026	1.257	1.229	1.056	1.184	1.150	0.962	1.082	1.138
	(0.072)	(0.076)	(0.086)	(0.076)	(0.092)	(0.072)	(0.076)	(0.082)	(0.080)
Cutoff 4	1.466	1.646	1.682	1.457	1.613	1.665	1.276	1.576	1.703
	(0.088)	(0.084)	(0.104)	(0.082)	(0.109)	(0.090)	(0.091)	(0.095)	(0.103)
Cutoff 5	1.799	1.966	2.091	1.912	2.098		1.539	2.081	2.264
	(0.100)	(0.094)	(0.122)	(0.091)	(0.124)		(0.104)	(0.108)	(0.123)
Cutoff 6	2.097	2.255	2.433		2.520				
	(0.111)	(0.105)	(0.138)		(0.140)				
Cutoff 7	2.382								
	(0.124)								

Notes: This table reports the mean and standard deviation of each parameter across 5,000 draws from the posterior distribution for the mental health measure system. Standard errors are shown in parentheses below the mean point estimates. Each component of the BSI is modeled using an ordered threshold model. For each measure, the number of values (K) corresponds to the number of distinct values between zero and one, including zero and one. Thus, the number of cutoffs varies across measures. The mental health factor is normalized to have a loading of minus one on the BSI hostility measure.

Table OA.2.3: Estimated Parameters from Factor Model — Mental Health Measures (Year 7)

	(1) Somatization	(2) Depression	(3) Anxiety	(4) Hostility	(5) Obsessive Compulsive	(6) Interpersonal Sensitivity	(7) Phobic Anxiety	(8) Paranoid Ideation	(9) Psychoticism
Constant	0.071	1.033	0.466	0.281	1.473	-0.678	-0.785	1.093	-0.185
	(0.538)	(0.811)	(0.765)	(0.577)	(0.689)	(0.661)	(0.573)	(0.629)	(0.748)
Age 15	0.065	0.607	0.169	-0.158	0.211	0.292	0.432	0.551	0.691
	(0.291)	(0.405)	(0.367)	(0.283)	(0.331)	(0.350)	(0.326)	(0.311)	(0.386)
Age 16	-0.029	0.268	-0.438	-0.123	0.076	0.254	0.187	0.181	0.387
	(0.263)	(0.378)	(0.336)	(0.263)	(0.300)	(0.327)	(0.308)	(0.284)	(0.350)
Age 17	0.069	0.536	0.347	0.393	0.497	0.123	0.624	0.197	0.305
	(0.269)	(0.381)	(0.343)	(0.265)	(0.305)	(0.334)	(0.304)	(0.286)	(0.361)
Age 18	0.047	0.164	-0.232	0.555	0.467	-0.324	0.511	0.546	0.630
	(0.382)	(0.531)	(0.491)	(0.372)	(0.423)	(0.511)	(0.419)	(0.395)	(0.491)
Female	0.389	0.066	0.136	0.471	0.497	0.392	-0.091	-0.132	-0.162
	(0.209)	(0.310)	(0.274)	(0.213)	(0.246)	(0.250)	(0.241)	(0.234)	(0.290)
White	-0.125	-1.579	-0.270	0.535	-0.842	-0.007	-0.179	-0.653	-0.935
	(0.526)	(0.808)	(0.748)	(0.570)	(0.662)	(0.631)	(0.569)	(0.608)	(0.743)
Hispanic	-0.482	-1.354	-0.481	0.199	-1.150	-0.601	0.102	-0.695	-0.333
	(0.519)	(0.788)	(0.740)	(0.560)	(0.653)	(0.631)	(0.557)	(0.605)	(0.722)
Black	-0.340	-1.402	-0.203	0.331	-1.273	-0.143	-0.270	-0.363	-0.232
	(0.519)	(0.788)	(0.744)	(0.562)	(0.660)	(0.627)	(0.547)	(0.608)	(0.720)
Phoenix	-0.286	-0.471	0.013	-0.459	-0.062	-0.132	-0.598	-0.327	-0.307
	(0.222)	(0.311)	(0.280)	(0.218)	(0.250)	(0.266)	(0.244)	(0.239)	(0.295)
Variance	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mental Health	-0.845	-1.604	-1.470	-1.000	-1.262	-1.084	-0.843	-1.131	-1.414
	(0.101)	(0.201)	(0.177)	(0.000)	(0.153)	(0.138)	(0.109)	(0.135)	(0.177)
Cutoff 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cutoff 2	0.519	0.437	0.678	0.819	0.610	0.551	0.503	0.651	0.644
	(0.072)	(0.081)	(0.093)	(0.086)	(0.083)	(0.086)	(0.082)	(0.087)	(0.103)
Cutoff 3	0.851	0.947	1.351	1.405	0.997	1.029	0.984	1.052	1.247
	(0.092)	(0.122)	(0.134)	(0.103)	(0.106)	(0.122)	(0.119)	(0.105)	(0.146)
Cutoff 4	1.230	1.574	1.890	1.810	1.564	1.337	1.514	1.422	1.687
	(0.114)	(0.171)	(0.171)	(0.116)	(0.134)	(0.146)	(0.165)	(0.119)	(0.171)
Cutoff 5	1.469	1.810	2.463	2.199	1.937		1.738	1.760	2.179
	(0.129)	(0.189)	(0.217)	(0.132)	(0.152)		(0.189)	(0.133)	(0.204)
Cutoff 6	1.602	2.169	2.734		2.135				
	(0.136)	(0.216)	(0.235)		(0.163)				
Cutoff 7	1.777								
	(0.148)								

Notes: This table reports the mean and standard deviation of each parameter across 5,000 draws from the posterior distribution for the mental health measure system in year 7. Standard errors are shown in parentheses below the mean point estimates. Each component of the BSI is modeled using an ordered threshold model. For each measure, the number of values (K) corresponds to the number of distinct values between zero and one, including zero and one. Thus, the number of cutoffs varies across measures. The mental health factor is normalized to have a loading of minus one on the BSI hostility measure.

OA.3 Counterfactuals - Additional Definitions and Figures

OA.3.1 Measures of Positional Mobility

Chi-Square Mobility Index (CSMI): The Chi-Square Mobility Index measures how much the observed mobility matrix deviates from random mobility. It compares the observed mobility matrix to a benchmark of random mobility. That is, a hypothetical scenario in which an individual's position in the distribution under incarceration would be randomly assigned, conditional on his position absent incarceration, and we would expect a uniform distribution across deciles. Specifically, each cell in the mobility matrix would contain 10% of the mass. The CSMI quantifies deviations from this uniform benchmark by summing the squared differences between observed and predicted probabilities across all cells. Larger deviations imply more dependence between the two distributions and therefore *less* mobility. Formally, the index is computed as $\text{CSMI}(M_{7,0}, M_{7,1}) = (-1) \times \sum_{i=1}^{10} \sum_{j=1}^{10} \frac{(O_{ij} - 10)^2}{10}$, where O_{ij} is the probability of being in the ith decile of the distribution of $M_{7,0}$ and the jth decile of the distribution of $M_{7,1}$. By definition, the CSMI is negative, with higher values implying greater mobility.

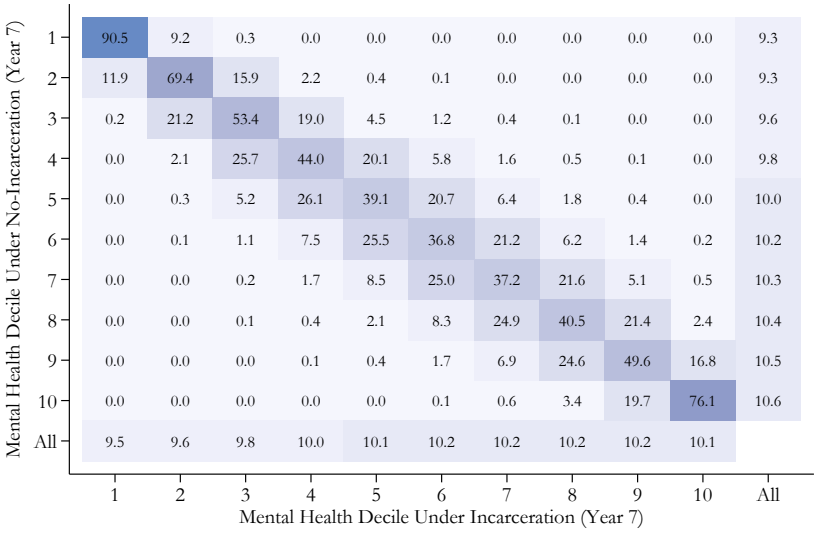
Mobility Ratio Index (MRI):, which measures the share of individuals who do *not* remain on the same decile across the two distributions. The Mobility Ratio Index is computed as $\text{MRI}(M_{7,0}, M_{7,1}) = 1 - \frac{\sum_{i=1}^{10} N_{ii}}{\sum_{i=1}^{10} \sum_{j=1}^{10} N_{ij}}$, where N_{ij} counts the number of individuals that are simultaneously in the ith decile of the distribution of $M_{7,0}$ and the jth decile of the distribution of $M_{7,1}$. The MRI ranges from 0 to 1, with higher values implying greater mobility.

Mean Deciles Move Index (MDMI):, which captures the average number of deciles each individual moves between the two distributions. The Mean Deciles Move Index is computed as $\text{MDMI}(M_{7,0}, M_{7,1}) = \frac{1}{N} \sum_{n=1}^{N} \left| Decile_n(M_{7,1}) - Decile_n(M_{7,0}) \right|$. Higher values of the MDMI imply greater mobility.

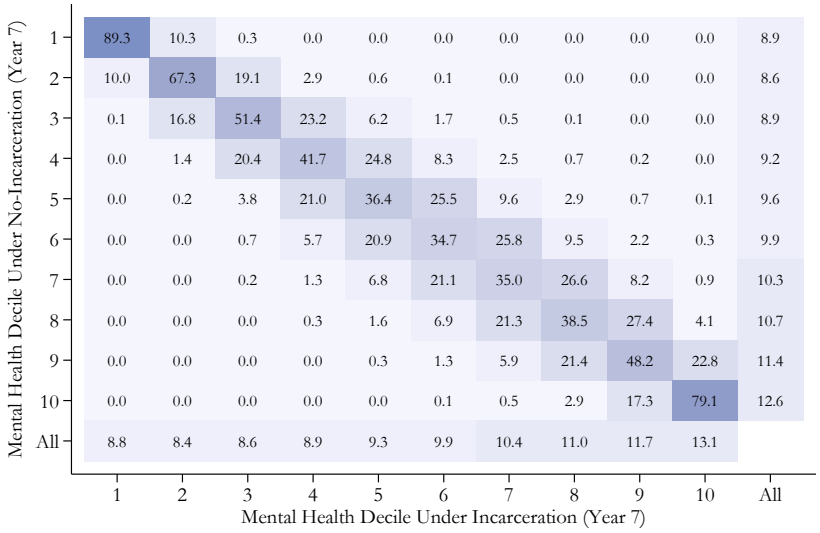
OA.3.2 Additional Figures

Figure OA.3.1: Positional Mobility Analyses (Cell %), for individuals who do not go to prison ($D = 0$)

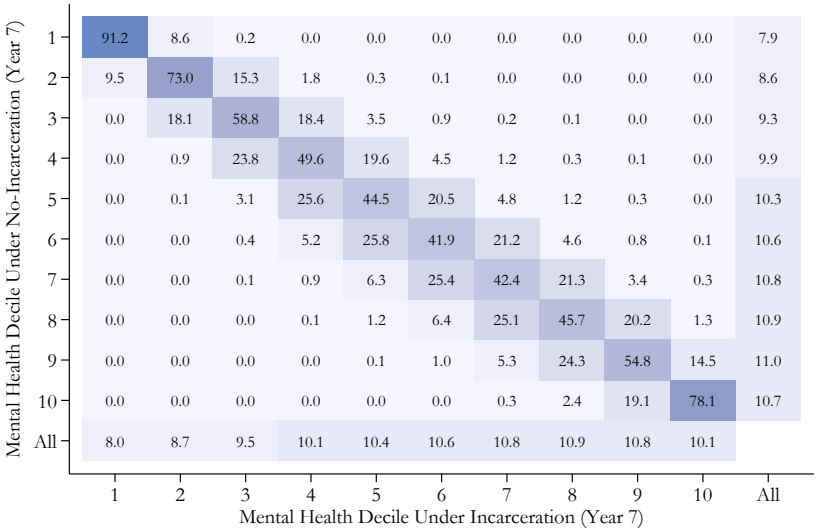
(a) Unconditional



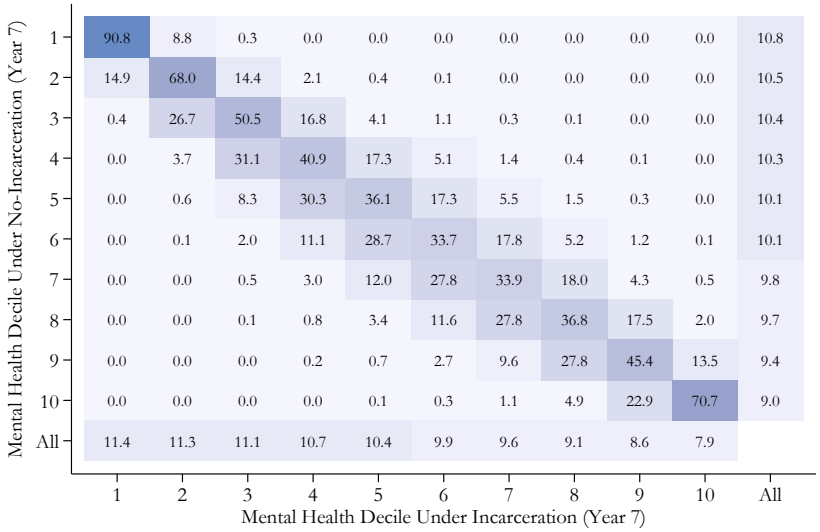
(b) For C (Cognitive) in Tertile 1



(c) For C (Cognitive) in Tertile 2



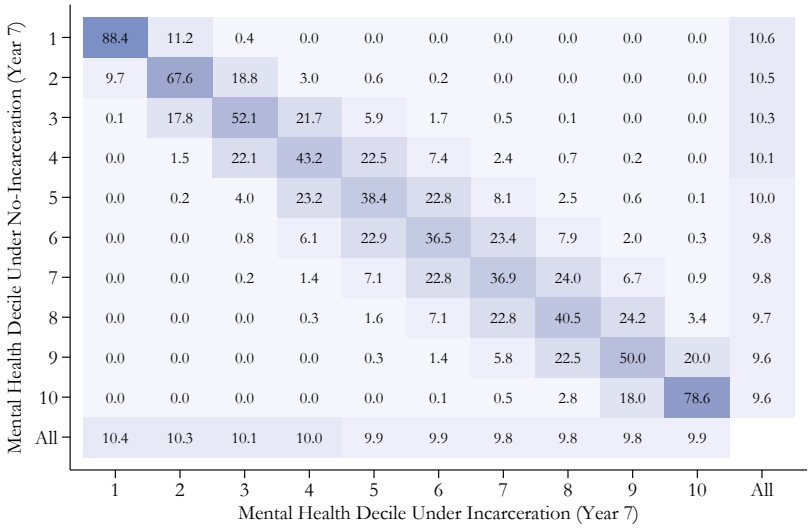
(d) For C (Cognitive) in Tertile 3



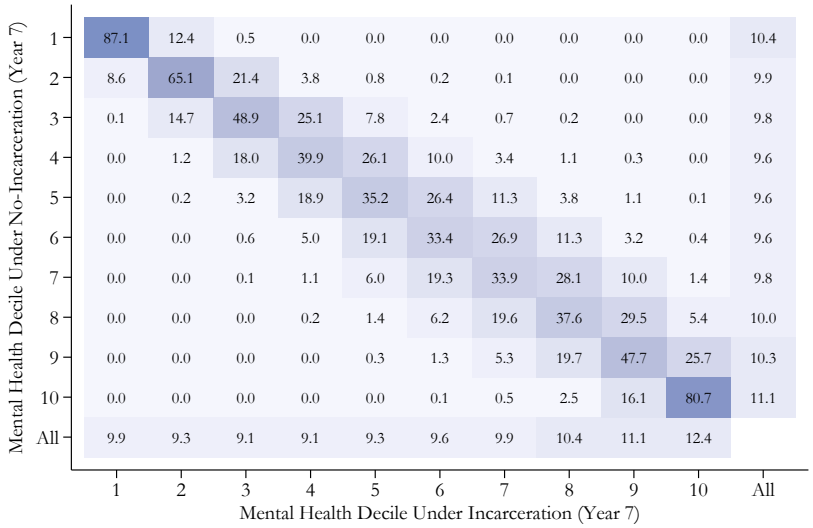
Notes: Let $Q_{j,0}$ be the jth quantile of $M_{7,0}$, and $Q_{k,1}$ be the kth quantile of $M_{7,1}$. Cell (j,k) in Figure OA.3.1a reports $\Pr\left(M_{i,7,1} \in (Q_{k-1,1},\, Q_{k,1}] \mid M_{i,7,0} \in (Q_{j-1,0},\, Q_{j,0}], D_i = 0\right)$. Figures OA.3.1b-OA.3.1d report $\Pr\left(M_{i,7,1} \in (Q_{k-1,1},\, Q_{k,1}] \mid M_{i,7,0} \in (Q_{j-1,0},\, Q_{j,0}], D_i = 0, C \in \text{Tertile } t\right)$. Shading ranges from white (0%) to dark blue (100%).

Figure OA.3.2: Positional Mobility Analyses (Cell %), for individuals who go to prison ($D = 1$)

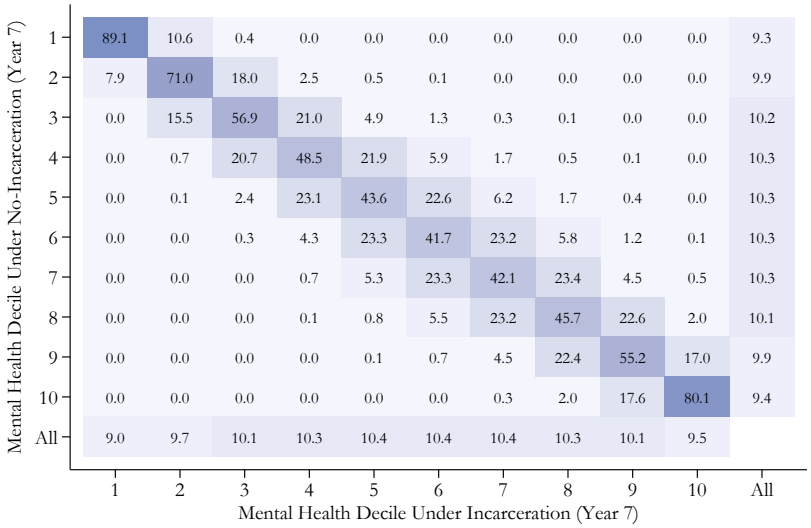
(a) Unconditional



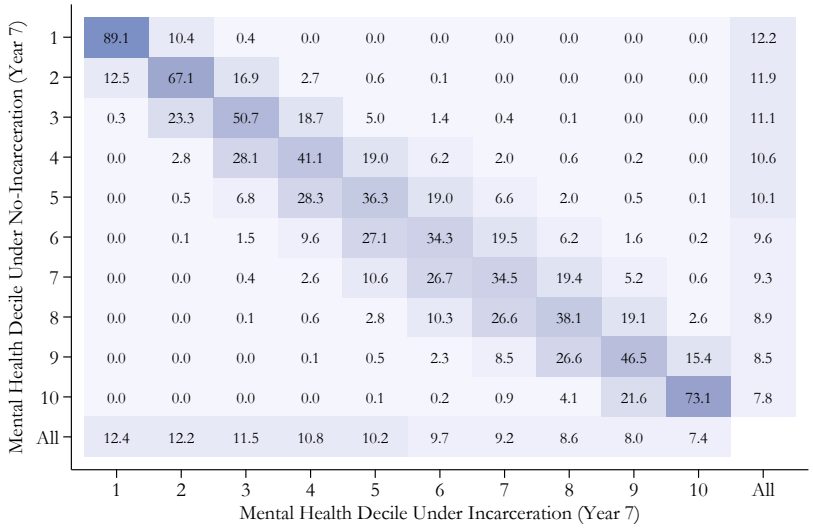
(b) For C (Cognitive) in Tertile 1



(c) For C (Cognitive) in Tertile 2



(d) For C (Cognitive) in Tertile 3



Notes: Let $Q_{j,0}$ be the jth quantile of $M_{7,0}$, and $Q_{k,1}$ be the kth quantile of $M_{7,1}$. Cell (j,k) in Figure OA.3.2a reports $\Pr\left(M_{i,7,1} \in (Q_{k-1,1},\, Q_{k,1}] \mid M_{i,7,0} \in (Q_{j-1,0},\, Q_{j,0}], D_i = 1\right)$. Figures OA.3.2b-OA.3.2d report $\Pr\left(M_{i,7,1} \in (Q_{k-1,1},\, Q_{k,1}] \mid M_{i,7,0} \in (Q_{j-1,0},\, Q_{j,0}], D_i = 1, C \in \text{Tertile } t\right)$. Shading ranges from white (0%) to dark blue (100%).

Figure OA.3.3: Growth Analyses (Cell %), for individuals who do not go to prison ($D = 0$)

(a) Unconditional



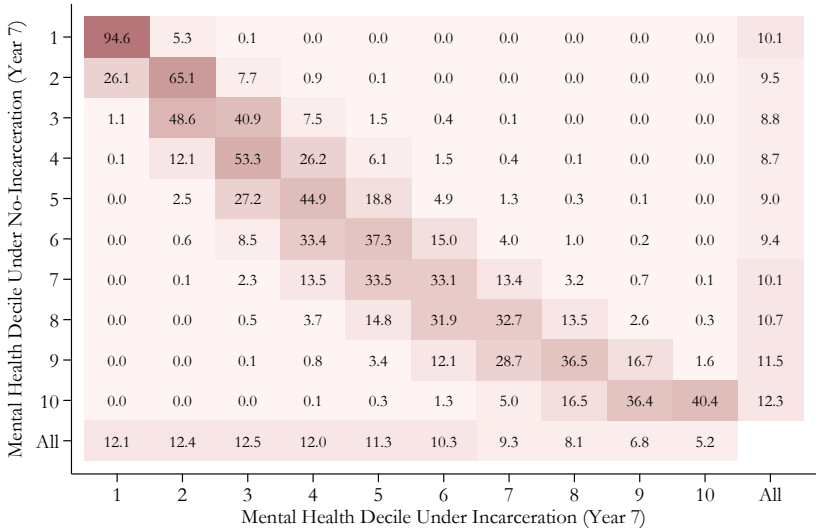
(b) For C (Cognitive) in Tertile 1



(c) For C (Cognitive) in Tertile 2



(d) For C (Cognitive) in Tertile 3



Notes: Let Q_j be the jth quantile of M_7. Cell (j,k) in Figure OA.3.3a reports $\Pr\left(M_{i,7,1} \in (Q_{k-1}, Q_k] \mid M_{i,7,0} \in (Q_{j-1}, Q_j], D_i = 0\right)$. Figures OA.3.3b-OA.3.3d report $\Pr\left(M_{i,7,1} \in (Q_{k-1}, Q_k] \mid M_{i,7,0} \in (Q_{j-1}, Q_j], D_i = 0, C \in \text{Tertile } t\right)$. Shading ranges from white (0%) to dark red (100%).

Figure OA.3.4: Growth Analyses (Cell %) for individuals who go to prison ($D = 1$)

(a) Unconditional



(b) For C (Cognitive) in Tertile 1



(c) For C (Cognitive) in Tertile 2



(d) For C (Cognitive) in Tertile 3



Notes: Let Q_j be the jth quantile of M_7. Cell (j,k) in Figure OA.3.4a reports $\Pr\left(M_{i,7,1} \in (Q_{k-1}, Q_k] \mid M_{i,7,0} \in (Q_{j-1}, Q_j], D_i = 1\right)$. Figures OA.3.4b-OA.3.4d report $\Pr\left(M_{i,7,1} \in (Q_{k-1}, Q_k] \mid M_{i,7,0} \in (Q_{j-1}, Q_j], D_i = 1, C \in \text{Tertile } t\right)$. Shading ranges from white (0%) to dark red (100%).