LUND UNIVERSITY

SCHOOL OF ECONOMICS AND MANAGEMENT

The Impact of Parental Education on the Early Detection of Autism Spectrum Disorder in Children.

by

Farahim Suleymanli

NEKP01

Master Essay II

May 2023

Supervisor: Simon Reese

Table of Contents

ABSTRACT

Utilizing the Double Machine Learning (DML) framework with both Partially Linear Regression (PLR) and Interactive Regression Models (IRM), this study examined the crucial role of parental education in the early detection of Autism Spectrum Disorder (ASD). Analysis was conducted on comprehensive data from the National Survey of Children's Health spanning the years 2017 to 2021. A significant negative association was found between parental education and the age at ASD diagnosis. This relationship was most notable in cases of moderate ASD severity, indicating that higher parental education levels tend to expedite ASD detection in this group. For mild and severe ASD cases, however, the relationship was less clear, suggesting that parental education's influence varies across the spectrum of ASD severity. The study's findings underscore the potential of parental education as a key socio-economic factor in ASD detection, highlighting a critical area for future research to further elucidate the role of socio-demographic factors across different ASD severity levels.

Keywords: ASD, DML, PLR, IRM, early detection, parental education, causal inference

To Chilanay and Zeynab

1. Introduction

Autism Spectrum Disorder (ASD) constitutes a multifaceted and pervasive neurodevelopmental disorder characterized by a broad spectrum of symptoms impacting an individual's social communication, interaction capabilities, and behaviour patterns (ed. American Psychiatric Association, 2013). The term ASD encompasses a diverse range of disorders, including autistic disorder, Asperger's syndrome, and pervasive developmental disorder-not otherwise specified, reflecting the considerable variability in symptom severity and manifestation among affected individuals (Volkmar, Reichow & McPartland, 2012).

In the economic analysis of health outcomes, identifying determinants of early detection, particularly in complex disorders like ASD, is integral, given its impact on intervention efficacy and subsequent cost implications. Parental education appears to be a salient factor, as parents' earliest observations often signal developmental anomalies. However, untangling the intricate relationship between parental education and early ASD detection is a challenging endeavour, complicated by a myriad of confounding variables. The principal purpose of the present study is to elucidate this intricate relationship, leveraging a sophisticated Double Machine Learning (DML) approach to yield robust insights. The structure and objectives of the research have been explicitly outlined in this section, setting a clear path for further investigation.

1.1 Prevalence, complexity, and importance of ASD

The prevalence of ASD among adults in the United States is noteworthy, with more than 5.4 million individuals, or 2.2% of the population, identified as being on the autism spectrum (Diament, 2020). Consequently, this escalating prevalence has prompted an increased emphasis on early detection and intervention strategies, as research consistently underscores the positive long-term outcomes that early intervention facilitates for individuals with ASD (Dawson et al., 2010). The latest estimates indicate that roughly 1 in 36 children aged eight years in the US were diagnosed with ASD in 2020 (CDC’s Autism and Developmental Disabilities Monitoring Network, 2023). Importantly, ASD transcends ethnic and socioeconomic boundaries; however, minority groups often experience delayed or incorrect diagnoses. The report by CDC’s ADDM Network (2023) revealed that in 2020, for the first time among 8-year-olds, Black, Hispanic, and Asian or Pacific Islander children exhibited a higher prevalence of ASD than White children. It further revealed that boys were found to be approximately four times more likely to be identified with ASD than girls. Additionally,

approximately one-third of these children had intellectual disabilities. The data indicated that children born in 2016 were 1.6 times more likely to receive an ASD diagnosis or special education classification by the age of four, compared to children born in 2012.

The aetiology of ASD remains enigmatic; however, current evidence points to a complex interplay of genetic and environmental factors (Geschwind, 2011). Advances in molecular genetics have identified several candidate genes and chromosomal loci associated with ASD, corroborating the hypothesis of a robust genetic foundation for the disorder (Gaugler et al., 2014). Environmental factors, such as prenatal exposure to certain medications, maternal infections, and advanced parental age, have also been implicated in ASD risk (Modabbernia, Velthorst & Reichenberg, 2017). Hence, the diagnosis of ASD is not a standardized process like many other diseases, as each child with ASD exhibits unique characteristics. Detecting ASD can be challenging due to the complexity of the disorder and the variability in the presentation of symptoms. Symptoms can vary widely in type and severity, and they may not become evident until a child reaches school age. Additionally, ASD often co-occurs with other neurodevelopmental or psychiatric disorders, which can complicate the diagnostic process (Matson & Goldin, 2013). It has been estimated that approximately 78% of children with ASD have at least one co-occurring mental health condition (Kerns, Rast & Shattuck, 2020).

The average age of ASD diagnosis is often later than it could be. According to the CDC (2022), while ASD can sometimes be diagnosed as early as 2 years of age, many children are not diagnosed until much older. This delay can be attributed to multiple factors including a lack of awareness or knowledge about ASD among parents and healthcare providers, variability in clinical practices, or limited access to diagnostic services, particularly among disadvantaged or minority populations (Zuckerman, Lindly & Chavez, 2017). The period from initial suspicion to final diagnosis is substantial, often in the range of 2.7-3.7 years, and can be even longer for children from minority or low-income families (Mandell et al., 2002; Shattuck et al., 2009).

Research has demonstrated that late detection of ASD can lead to a considerable economic burden. The costs associated with ASD, including healthcare, education, and supportive services, are significant and predicted to increase (Buescher et al., 2014; Leigh & Du, 2015). This financial impact extends to family households, influencing parental employment and earnings, often requiring adjustments in work schedules or compelling parents to leave the

workforce (Cidav, Marcus & Mandell, 2012). It has been shown that early detection of ASD can significantly influence these future costs. Chasson et al. (2007) and Ganz (2007) demonstrated that early detection and subsequent intervention can substantially reduce the societal costs of ASD by improving the trajectory of the disorder. This is corroborated by Peters-Scheffer et al. (2011), who show that comprehensive, early intervention programs for children with ASD can enhance outcomes, indirectly suggesting a potential reduction in future costs. Beyond the financial aspects, ASD also significantly affects the quality of life of individuals and their families, leading to increased health service utilization and additional stress (Weiss et al., 2018). Thus, these economic and quality-of-life impacts underscore the importance of early detection and intervention in ASD, which can potentially mitigate some of these challenges.

1.2 Aim of the study and its structure

The early detection of ASD holds substantial significance, not only because it can potentially alleviate future economic burdens on society, families, and the individuals with ASD themselves, but also because it can lead to improved developmental outcomes for those affected. However, the inherent complexity of ASD, characterized by a few core symptoms and co-occurring conditions, makes its detection particularly challenging. Thus, understanding the factors that could expedite detection becomes a critical element in the quest for timely diagnosis and intervention. Among these factors, parental education emerges as a cornerstone, given that parents often serve as the first line of detection for potential developmental anomalies. Hence, the objective of this study is to conduct a comprehensive investigation of the intricate and multifaceted relationship between age at diagnosis of ASD and parental education.

Prior research, including Delobel-Ayoub et al. (2015), Fountain et al. (2011), and Shattuck et al. (2009), has employed classical econometric techniques to estimate the relationship between age at diagnosis and parental education. Nevertheless, these approaches grapple with numerous challenges such as confounding variables, selection bias, overfitting, and model misspecification, which hinder their capacity to accurately explain the relationship. In contrast, this study embraces DML as a superior methodology, utilizing machine learning techniques to account for a broad range of observed and unobserved confounding variables, effectively mitigating issues like selection bias. Additionally, DML is less vulnerable to

overfitting and model misspecification, as it abstains from imposing rigid assumptions about the functional form of the relationship between variables.

By employing DML, this study endeavours to furnish more precise and dependable insights into the causal relationship between parental education and early detection of ASD across various severity levels. As a result, it contributes to the extant literature by enhancing our comprehension of the complexities linked to ASD detection and the significance of parental education, ultimately guiding targeted interventions and support programs for affected families.

The remainder of this study is organized into several sections, each addressing a crucial facet of the research. The second section offers a comprehensive literature review, examining previous research on ASD detection, parental education, and socio-economic status, as well as the application of DML methodology in analogous contexts. The third section delineates the methodology employed in the study, with a focus on the DML approach and its detailed implementation. The fourth section presents the data and variables utilized in the analysis, elaborating on the sources, measurements, and descriptive statistics. The fifth section reports the results and discusses their implications, encompassing both the overarching effect of parental education on ASD detection and the specific effects across different severity levels. Finally, the study concludes by summarizing the principal findings.

2. Literature review

In unravelling the complex aetiology and progression of ASD, it becomes vital to consider not just the core symptoms, but also the diverse comorbidities that often accompany this condition. The inclusion of these comorbidities in any analytical framework is crucial, as they may hold consequential insights into early detection and provide a more comprehensive understanding of the disorder's manifestation. Concurrently, parental education stands as a critical socio-economic factor that can greatly influence the timing of ASD diagnosis. The level of parental education, potentially impacting awareness of developmental milestones, access to healthcare, and advocacy skills, has been associated with earlier detection of ASD. Hence, to accurately discern the causal effect of parental education on the early detection of ASD, a plethora of relevant explanatory factors necessitates an analytical framework capable of managing high-dimensional data and addressing potential confounding factors. In this context, DML offers a robust and flexible approach, combining the power of machine learning with causal inference, which has been successful in estimating treatment effects in

various domains. This literature review will thus delve into the symptoms and comorbidities of ASD, underline the significance of parental education in its early detection, and elucidate the relevance of incorporating these elements in the DML framework.

2.1 Clinical signs and comorbidities

The core symptoms of ASD involve impairments in social interaction and communication, alongside the presence of restricted, repetitive behaviours, and interests (Baio, 2018). Such symptoms typically emerge during early childhood and endure throughout an individual's lifespan, frequently leading to substantial functional impairments across multiple domains, encompassing academic, occupational, and social settings (Lai, Lombardo & Baron-Cohen, 2014). CDC (2023) delineates additional related characteristics, including delayed language, motor, cognitive, or learning skills; hyperactive, impulsive, and/or inattentive behaviour; epilepsy or seizure disorders; atypical eating and sleeping habits; gastrointestinal issues; unusual mood or emotional responses; anxiety, stress, or excessive worry; and either a lack of fear or heightened fear responses.

Moreover, numerous studies have reported associations between ASD and various health-related conditions[1]. Xu et al. (2018) identified a significant positive association between common allergic conditions and ASD, while Chua et al. (2021) investigated potential links between allergies and neurodevelopmental disorders. Beesley (2020) discovered that children with Juvenile Idiopathic Arthritis (JIA) or those with a first-degree relative with JIA are significantly more likely to receive an ASD diagnosis. Studies by Wang et al. (2021) and Zhao et al. (2021) presented evidence for potential associations between ASD and specific blood disorders. Hahn et al. (2020) concluded that early signs of ASD-associated behaviour are present and detectable in new-borns with Down Syndrome. Sigmon et al. (2019) conducted a case-control study to examine the relationship between Congenital Heart Disease (CHD) and ASD, concluding that children with CHD have a higher likelihood of developing ASD. Several studies substantiated the notion that ASD is frequently accompanied by a diverse array of neurodevelopmental (Höglund Carlsson et al., 2013; Levy et al., 2010) and psychiatric comorbidities. These include, but are not limited to, anxiety disorders (White et al., 2009), epilepsy (Canitano, 2007), attention deficit hyperactivity disorder (Joshi et al., 2010), as well as Tourette's syndrome and tics (Canitano & Vivanti, 2007; Simonoff et al., 2008). Furthermore, Galvan et al. (2020) found significant links between ASD and birth

[1] These confounders were controlled in the study.

order, increasing gaps in parental ages, and birth order within the family. The findings of Alvarez et al. (2021) demonstrated that higher birth order in children with ASD correlates with increased functional and cognitive disabilities. Prematurity has been ascertained as a contributory determinant of ASD. The research conducted by Allen et al. (2020) showed that the susceptibility to autism associated with prenatal prematurity is more pronounced in female individuals.

2.2 Parental education and early detection of ASD

The escalating prevalence of ASD has created a global impetus to scrutinize the multifaceted contexts surrounding the care of autistic children, with particular attention directed toward the impact of socioeconomic status (SES). There is a growing literature in this direction where various studies have attempted to determine the relationship between ASD diagnosis and various factors such as household SES, ethnic groups, spatial factors, parental characteristics, etc. Research has shown that parental education, as a proxy for SES, is associated with the timing and likelihood of ASD diagnosis (Durkin et al., 2010). Children from families with higher parental education levels are more likely to be diagnosed with ASD at an earlier age (Bhasin & Schendel, 2007). This may be due to greater awareness of developmental milestones, increased access to healthcare services, and better advocacy for their children (Mandell et al., 2009).

Hrdlicka et al. (2016) conducted an empirical study, involving a sample of 160 children and their parents, intending to delineate the relationship between the age at ASD diagnosis and family SES. Their findings revealed a positive correlation between early ASD diagnosis and higher parental age at birth and higher parental education levels, while no discernible association was observed with family SES or the availability of informational resources accessible to the family. Fountain et al. (2011) examined temporal variations in characteristics associated with the time of diagnosis in ten cohorts of children diagnosed with ASD between 1992 and 2001 in California. According to the study, children with highly educated parents were diagnosed with ASD earlier, and this effect grew stronger with time. Furthermore, whereas autism was first identified disproportionately among the rich and educated, the distribution of diagnoses gradually became more evenly distributed throughout the community. Frenette et al. (2013) employed adjusted linear regression techniques to discern factors influencing the age of ASD diagnosis within a Canadian population-based cohort study from 1992 to 2005. The analysis revealed that maternal age, county of residence,

and attention-deficit/hyperactivity disorder diagnostic status emerged as salient determinants explaining the variation in age at ASD diagnosis. In their study, Elsabbagh et al. (2012) undertook a rigorous systematic review of an extensive collection of 600 epidemiological surveys. In contrast to the previous studies, their analysis led to the cogent conclusion that the empirical evidence gathered did not substantiate any notable disparities in the prevalence of ASD across geographic location, ethnicity, culture, and socioeconomic factors.

2.3 Application of double machine learning in health economics

DML is a novel approach that amalgamates the flexibility of machine learning with causal inference, thereby enabling the precise estimation of causal effects (Chernozhukov et al., 2018). It has been used to estimate treatment effects in various settings, including medicine, economics, and education (Chuang & Chen, 2023; Dube et al., 2020; Knaus, 2021; McConnell & Lindner, 2019). While DML has not yet been applied in any kind of ASD research, as per the extant literature, it has been successfully used to study other health-related outcomes. For example, Sanchez et al. (2022) incorporated Alzheimer's disease as a paradigm to elucidate the merits of causal machine learning in clinical contexts. Peet et al. (2023) exploited DML to quantify the average treatment effect of The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) on infant health outcomes. The researchers posited that the WIC program yields maximum benefits for those at the highest risk of adverse infant health outcomes. Oyenubi and Kollamparambil (2023) used a DML approach in conjunction with nationally representative data from South Africa, to explore the nexus between depressive symptoms and perceived nonadherence to COVID-19 preventive measures. Their empirical findings unveil a causal relationship between the perception of neighbours' noncompliance and self-reported depressive symptoms, thereby accentuating the deleterious ramifications of noncompliance on the psychological well-being of individuals in the context of a virulent pandemic. Drawing upon data from the 2016–17 Multiple Indicator Cluster Survey in Nigeria, Skoufias and Vinha (2021) conducted a comprehensive investigation into the associations between child stature, maternal education, and early childhood development (ECD) indicators, juxtaposing the conventional ad-hoc variable selection with the DML approach. Their analysis substantiated the significant influence of maternal education and chronic malnutrition on ECD measures. Moreover, the DML methodology unveiled a more intricate contextual understanding, elucidating divergent effects of maternal education on ECD and child nutrition in rural and urban settings, thereby

providing a pragmatic means of bolstering internal validity for policy design and inferences grounded in observational data.

The extant body of literature substantiates a discernible association between parental education and the early detection of ASD. Given the successful implementation of DML in an array of health-related domains, it is plausible to contend that it may hold substantial promise for augmenting our understanding of the role parental education plays in the early detection of ASD, through the provision of precise and unbiased estimations of causal effects. Consequently, this enhanced comprehension could bear significant implications for the formulation of targeted interventions and policy frameworks, thereby facilitating early detection and intervention for children with ASD across a broad spectrum of populations.

3. Methodology

3.1 Double Machine Learning

At its essence, DML is an advanced methodology for approximating treatment effects in the presence of a multitude of observed, high-dimensional potential confounders. DML is predicated upon a foundational econometric theorem, the Frisch-Waugh-Lovell (FWL) theorem (Frisch & Waugh, 1933; Lovell, 1963). The FWL theorem, an integral component of econometric theory, posits that the influence of a particular variable, once controlled for other predictors in a linear regression model, can be calculated by "partialling out" the linear impact of the other variables from both the outcome and the variable of interest. This process is carried out by calculating the residuals of the variable of interest and the outcome variable, and then performing a regression of the former on the latter. DML extrapolates this principle to a machine learning context, adeptly handling high-dimensional data and complex, non-linear interrelationships among variables. This study restates the estimation of the parameter of interest in terms of the moment conditions that satisfy the so-called Neyman orthogonality condition, within the FWL framework.

The underlying principle of DML involves partitioning the estimation process into two primary phases: (1) estimating nuisance functions through the application of machine learning techniques, and (2) approximating the causal effect via orthogonalized moment equations. Chernozhukov et al. (2018) demonstrated, in their seminal work, that the

implementation of Neyman-orthogonal moments/scores and cross-fitting[2] effectively counteracts the regularization and overfitting biases that may arise from the integration of machine learning algorithms into causal analysis. The authors specifically devised a series of straightforward results for achieving root-N consistent estimation, where N denotes the sample size, and for making legitimate inferential statements pertaining to a low-dimensional parameter of interest, θ_0, in the context of a high-dimensional or "highly complex" nuisance function. DML utilizes a method-of-moments estimator for the target parameter θ_0, which is founded on the empirical counterpart of the moment condition:

$$E[\psi(W; \theta_0, \eta_0)] = 0 \tag{1}$$

Where ψ is a score function, W is an observation vector that includes outcome (Y), treatment (D) and confounding (X) variables, and $\eta_0 = (g_0, m_0)$ is a population value of nuisance functions $\eta = (g, m)$. Nuisance functions are auxiliary functions that, even though they're not the primary concern, must be estimated to identify the causal effect of interest. These functions typically encompass the unknown components g_0 and m_0, which represent the non-linear associations between covariates and the outcome variable. Score functions (ψ) serve to form the orthogonalized moment conditions for the causal effect of interest after the nuisance functions have been estimated. Score functions constitute the partial derivatives of the likelihood function with respect to the model's parameters and are employed to gauge the sensitivity of the likelihood function to alterations in these parameters. They are the preliminary key elements of the inference procedure that fulfil the moment condition as specified in equation (1), where θ_0 identified as the unique solution. Score functions satisfy the Neyman orthogonality condition:

$$\partial_\eta E[\psi(W; \theta_0, \eta_0)]|_{\eta-\eta_0} = 0 \tag{2}$$

Where ∂_η denotes the Gateaux derivative operator. The DML framework applies the Neyman orthogonality condition and the Neyman orthogonal score function to ensure that moment conditions, used to identify the causal parameter of interest, θ_0, are locally insensitive to small perturbations of the nuisance function around its true value. Consequently, this permits the integration of estimates of the nuisance functions, even those characterized by a degree of

[2] Cross-fitting is a key step in the DML procedure that helps to mitigate overfitting and to achieve valid statistical inference. In the cross-fitting procedure, the data is divided into several folds or partitions (K folds). The nuisance functions are estimated on one part of the data (e.g., k-th fold), and these estimates are then used in the estimation of the parameter of interest on the remaining data ($K-1$ folds). This ensures that the estimation of the nuisance functions and the parameter of interest are performed on different data, reducing the risk of overfitting, and achieving a valid statistical inference.

error — a phenomenon frequently referred to as 'noisy' (Chernozhukov et al., 2018). Crucially, this integration does not substantially compromise the accuracy of θ_0 estimation and it remains root-N consistent. Such a characteristic paves the way for a more pragmatic and achievable estimation procedure, notably in contexts where procuring precise estimates of nuisance functions is inherently challenging. This understanding presents significant implications for the implementation of statistical models in practical economic research, providing opportunities for more resilient and reliable parameter estimation despite inherent noise in the data.

The estimation of causal parameters of interest is facilitated by two DML algorithms, denoted as DML1 and DML2 (Chernozhukov et al., 2018). The principal difference between these two algorithms stems from the need to perform cross-fitting to remove the overfitting bias. DML1 initially computes individual estimates, $\hat{\theta}_{0,k}$ for each fold I_k (partition of data, where k is the index of the fold), subsequently averaging these estimates to yield the final $\hat{\theta}_0$ estimate. In contrast, DML2 directly estimates the causal parameter $\hat{\theta}_0$ by solving one single equation incorporating the empirical expectation spanning all K folds[3]. This research adheres to the recommendation by Chernozhukov et al. (2018) to utilize the DML2 algorithm over DML1, as it potentially yields superior performance in small sample contexts, owing to its simultaneous consideration of information from all folds during the causal parameter estimation process. Consequently, the causal parameter $\hat{\theta}_0$ is calculated as the solution to the following equation:

$$\frac{1}{N}\sum_{k=1}^{K}\sum_{i\in I_k}\psi\left(W_i;\hat{\theta}_0,\hat{\eta}_{0,k}\right)=0 \tag{3}$$

In this equation N represents the total number of observations, K is the number of folds, W_i denotes the i-th observation, and $\hat{\eta}_{0,k}$ signifies the already provided machine learning estimator of η_0. Given that the previously discussed score functions for the DML models are linear in θ, equation (3) can be rewritten as:

$$\begin{aligned}\frac{1}{N}\sum_{k=1}^{K}\sum_{i\in I_k}\psi\left(W_i;\hat{\theta}_0,\hat{\eta}_{0,k}\right) &= \frac{1}{N}\sum_{k=1}^{K}\sum_{i\in I_k}\left(\psi_a(W;\eta)\theta+\psi_b(W;\eta)\right)\\ &= \frac{1}{N}\sum_{k=1}^{K}\sum_{i\in I_k}\psi_a(W;\eta)\theta+\frac{1}{N}\sum_{k=1}^{K}\sum_{i\in I_k}\psi_b(W;\eta)=0\end{aligned}$$

[3] Although the authors maintain that the choice of K holds no asymptotic consequences under the prescribed conditions, empirical examples, and simulations indicate that 4 or 5 folds often outperform 2 folds. In the present study, the primary K value is set at 2, with 4 and 5 folds employed for additional robustness assessments.

Where ψ_a and ψ_b are the components of the linear score function ψ. To solve for θ, we can rearrange the equation above as:

$$\frac{1}{N}\sum_{k=1}^{K}\sum_{i\in I_k}\psi_a(W;\eta)\theta = -\frac{1}{N}\sum_{k=1}^{K}\sum_{i\in I_k}\psi_b(W;\eta)$$

$$\theta = -\frac{\frac{1}{N}\sum_{k=1}^{K}\sum_{i\in I_k}\psi_b(W;\eta)}{\frac{1}{N}\sum_{k=1}^{K}\sum_{i\in I_k}\psi_a(W;\eta)}$$

Finally, relying on the empirical expectation[4] of the linear score functions, the estimated causal parameter can be expressed as:

$$\hat{\theta}_0 = -\frac{E_N[\psi_b(W;\eta)]}{E_N[\psi_a(W;\eta)]} \tag{4}$$

The asymptotic variance of $\hat{\theta}_0$ is calculated as:

$$var(\hat{\theta}_0) = \hat{J}_0^{-2}\frac{1}{N}\sum_{k=1}^{K}\sum_{i\in I_k}[\psi(W_i;\hat{\theta}_0,\hat{\eta}_{0,k})]^2 \tag{5a}$$

$$\hat{J}_0 = \frac{1}{N}\sum_{k=1}^{K}\sum_{i\in I_k}\psi_a(W_i;\hat{\eta}_{0,k}) \tag{5b}$$

In the present study, two DML models — Partially Linear Regression Model and the Interactive Regression Model — have been utilized to estimate the causal effect of parental education on the early detection of ASD in children.

3.2 Partially Linear Regression Model (PLR)

The PLR model stands as a flexible model structure that broadens the premises of traditional linear regression models. Deviating from the conventional practice of presuming specific functional forms for the relationships among confounders, the output, and the treatment variables, the PLR model adopts a general approach, accommodating more diverse, undefined forms. By abstaining from the imposition of these specific assumptions, the PLR model facilitates a more precise and realistic representation of variable interactions. This, in turn, augments the credibility of the causal estimates extracted from the model. However, the model assumes an additive separability between the treatment and the covariates. This means that the effect of the treatment on the outcome is assumed to be constant across different values of the covariates, an assumption that might not always hold in practice. The PLR model is mathematically articulated as follows:

[4] Empirical expectation is defined as: $E_{n,k}[\psi(W)] := \frac{1}{n}\sum_{i\in I_k}\psi(W_i)$, where n is the size of each fold I_k, and $k \in [K] = \{1, \dots, K\}$.

$$Y = D\theta_0 + g_0(X) + U, \quad E[U|X, D] = 0 \tag{6a}$$

$$D = m_0(X) + V, \quad E[V|X] = 0 \tag{6b}$$

In this context, Y signifies the dependent variable, while D embodies the policy or intervention variable of interest. Other control variables are contained in the vector $X = (X_1, \ldots, X_p)$, and U and V are respective stochastic error terms. The primary equation encompasses the first equation, with θ_0 as the parameter targeted for estimation. Provided that D is exogenous conditional on controls X, θ_0 can be interpreted as the average treatment effect parameter. D and Y have a unidirectional causal relationship (Figure 1). X confounds the identification of the causal effect, hence identification can be achieved through V, which encapsulates the variation in D that remains independent of X. The second equation characterizes the relationship between the treatment and control variables, which helps to remove regularization bias.

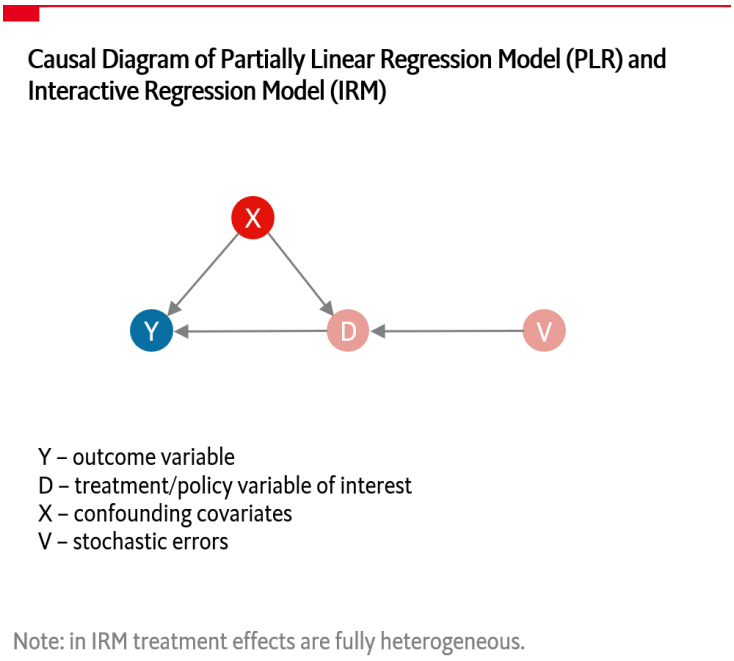

Figure 1

The score function for the PLR model is derived from the Neyman orthogonal score function combined with Robinson's (1988) partialling-out approach:

$$\begin{aligned}\psi(W;\theta,\eta) &:= \big(Y - \theta\big(D - m(X)\big) - l(X)\big)\big(D - m(X)\big) \\ &= -\big(D - m(X)\big)\big(D - m(X)\big)\theta + \big(Y - l(X)\big)\big(D - m(X)\big) \\ &= \psi_a(W;\eta)\theta + \psi_b(W;\eta), \quad \eta = (l, m)\end{aligned} \tag{7}$$

$$\psi_a = -\big(D - m(X)\big)\big(D - m(X)\big) \tag{8a}$$

$$\psi_b = \big(Y - l(X)\big)\big(D - m(X)\big) \tag{8b}$$

Where $W = (Y, D, X)$, l and m are P-square-integrable functions[5] mapping the support of X to $\mathbb{R}$. Chernozhukov et al. (2018) showed that in this case, θ_0 satisfies the equations (1) and (2) for $\eta_0 = (l_0, m_0)$, where $l_0(X) = E(Y|X)$[6] and $m_0(X) = E(D|X)$.

3.3 Interactive Regression Model (IRM)

The PLR model, despite its weak assumptions concerning the influence of control variables X on treatment and outcome variables, demands an additive separability of these two categories of variables, as stipulated by Equation (6a). This rigidity in the model's specification is attenuated in the IRM, which introduces the capacity for incorporating more complex, undefined interactions between the treatment variable D and control variables X within the conditional mean of the outcome variable Y. Nevertheless, such increased flexibility is not without its trade-offs. Specifically, the IRM is inherently constrained to the context of binary treatment variables. IRM can be expressed in the following form:

$$Y = g_0(D, X) + U, \quad E[U|X, D] = 0 \tag{9a}$$

$$D = m_0(X) + V, \quad E[V|X] = 0 \tag{9b}$$

IRM provides a robust approach for estimating the average treatment effect (ATE) and the average treatment effect on treated (ATET). Only the estimation of ATE as a target parameter has been performed in this study:

$$\theta_0 = \mathrm{E}[g_0(1, X) - g_0(0, X)] \tag{10}$$

D and Y are affected by confounding factors X through propensity score $m_0(X)$ and the function $g_0(D, X)$ respectively. Hence, the complexity and unknown nature of both functions make machine learning methods an attractive option. The score function for ATE estimation is:

$$\begin{aligned} \psi(W; \theta, \eta) &:= g(1, X) - g(0, X) + \frac{D\big(Y - g(1,X)\big)}{m(X)} - \frac{(1-D)\big(Y - g(0,X)\big)}{1 - m(X)} - \theta \\ &= \psi_a(W; \eta)\theta + \psi_b(W; \eta) \end{aligned} \tag{11}$$

$$\eta(X) = \big(g(0, X), g(1, X), m(X)\big), \quad \eta_0(X) = (g_0(0, X), g_0(1, X), m_0(X))$$

[5] The term 'P-square-integrable' refers to a function that is square-integrable with respect to the probability measure P. In the context of probability theory and statistics, a function is considered square-integrable if the integral of the square of the absolute value of the function is finite.

[6] Note that $l_0(X)$ is different from $g_0(X)$ in a way that it is not conditional on D.

$$\psi_a = -1 \tag{8a}$$

$$\psi_b = g(1,X) - g(0,X) + \frac{D(Y-g(1,X))}{m(X)} - \frac{(1-D)(Y-g(0,X))}{1-m(X)} \tag{8b}$$

The nuisance function comprises P-square-integrable functions g and m, which map the support of (D,X) to $\mathbb{R}$ and the support of X to the interval $(\varepsilon, 1-\varepsilon)$, respectively, with $\varepsilon \in (0, 0.5)$. The true values of functions g and m are given by:

$$g_0(D,X) = E(Y|D,X), \qquad m_0(X) = P(D=1|X)$$

The propensity score $m_0(X)$, denoting the conditional probability of exposure to the treatment given a vector of observed covariates, serves as a crucial statistical device for mitigating confounding in observational research. Propensity scores facilitate covariate balance, thereby curtailing bias in the estimate of the treatment effect. However, the score is susceptible to large variations induced by extreme values, particularly in instances of insufficient overlap in covariates across treated and control groups. Consequently, Chernozhukov et al. (2018) advance the notion of propensity score trimming, a method of excluding observations with propensity scores at extreme ends to ameliorate the issue of excessive weights that inflate variance and potentially bias. In the present analysis, trimming bounds are applied at 0.05 and 0.95, respectively.

3.4 Nuisance function estimation

Given the available sample size and extant literature on DML estimation, this study employs four machine learning methodologies: Least Absolute Shrinkage and Selection Operator (LASSO), Random Forests (RF), Extreme Gradient Boosting (XGBoost), and an Ensemble model. The Ensemble framework amalgamates previously mentioned and already tuned three machine learning models by estimating the nuisance functions as weighted averages of estimates derived from LASSO, RF, and XGBoost. The estimation procedure encompasses both regression and classification tasks for each machine learning methodology, as the treatment variable is binary.

In the machine learning models examined herein, a meticulous hyperparameter optimization process was employed, leveraging the grid search algorithm in conjunction with 10-fold cross-validation. This methodological decision primarily stems from the fact that grid search requires a low computational burden. It provides a thorough exploration of the hyperparameter space, ensuring the most suitable combination of values is identified. The 10-

fold cross-validation technique was utilized to balance the computational complexity and reliability of model performance estimates, effectively mitigating overfitting while maintaining robust generalization performance. It provides a more robust estimate of the model's generalization performance compared to techniques like leave-one-out cross-validation, which can be computationally expensive and produce high variance estimates (Efron, 1983; Kohavi, 1995). Evaluation metrics incorporated Mean Squared Error (MSE) for regression-based models and Misclassification Error (CE) for classification-oriented models. These metrics are both straightforward and interpretable, thereby precluding the imposition of additional computational complexity. The comprehensive tuning procedure for each model is delineated in Appendix 2.

4. Data

4.1 Data source and considerations

This study is based on data from the National Survey of Children's Health (NSCH), carried out by the United States Census Bureau. The NSCH provides a representative sample of the nation's population and is designed to offer valuable insights into children's health and well-being. It focuses specifically on factors that affect children's physical, emotional, and behavioural growth. The survey collects data on various aspects of children's lives, including their health status, access to healthcare services, family environment, and parental characteristics.

The data spans the years 2017 to 2021 and includes 175,231 observations of children based on a random sampling process. Within this dataset, 5,112 observations are associated with children diagnosed with ASD, of which 3,706 data points remain after the data cleaning. A detailed description of the variables is given in Appendix 1. The dataset encompasses a total of 52 distinct variables[7], and primarily, it consists of binary variables that embody all the principal ASD signs/symptoms discussed in Section 2. In employing this dataset, two cardinal aspects are contemplated: parental educational attainment, and ASD severity.

4.1.1 Parental educational attainment

The analysis utilizes the highest education level among all household members as the dependent variable, rather than solely parents' education, recognizing the broader family

[7] For the implementation of the LASSO model, all categorical variables have been transformed into dummy variables. Furthermore, second and third-order polynomial expansions of non-binary variables, along with their respective interaction terms, have been incorporated into the analysis.

environment's impact on a child's well-being and early detection of ASD. This approach accounts for the critical role played by various family members, such as grandparents and siblings, in a child's upbringing, sharing knowledge and resources, and supporting parents. By incorporating a more comprehensive measure of household education, the study captures a wider array of influences on ASD's early detection, reflecting the intricate interplay of social, economic, and environmental factors. This leads to a more accurate depiction of the relationship between education and early detection of ASD.

4.1.2 ASD severity

The relationship between parental education and early detection of ASD is initially estimated utilizing the entire dataset, without distinguishing between the varying levels of its severity. This approach serves as a foundational analysis to establish a baseline understanding of the association between parental education and early ASD detection across the entire spectrum of the disorder.

Subsequently, the data is stratified by ASD severity[8], allowing for a more nuanced investigation of the relationship between parental education and early detection within each severity category. By examining these relationships separately for mild (1,820 observations), moderate (1,488 observations), and severe[9] (398 observations) ASD cases, the study acknowledges the potential disparity in the impact of parental education on early detection across the distinct severity strata.

Higher parental education is expected to positively influence early ASD detection (as elaborated in Section 2), as better-educated parents may possess a greater awareness of developmental milestones and access to healthcare resources. This effect may vary across the ASD severity spectrum, with higher parental education potentially more influential in moderate and severe cases. Conversely, the relationship between parental education and early detection may be less pronounced for mild ASD cases, as the subtler manifestations of the condition might be more challenging to identify, even for well-educated parents. The stratification approach employed in this study is useful and more informative, as it accommodates variations across the ASD severity continuum. Nonetheless, the study

[8] Considering the small sample size for each severity strata, only 3-fold cross-validation has been performed during the tuning process for all machine learning models discussed in section 3.4.

[9] It should be noted that, in the estimation process pertaining to severe cases of ASD, only a two-fold cross-fitting procedure was employed, given the constraints posed by the limited sample size within this specific subpopulation.

concedes data constraints, particularly with respect to severe ASD cases, which may affect the robustness of the findings and necessitate cautious interpretation of the results.

4.2 Confounding factors

This study considers a variety of factors that could potentially serve as confounding variables in the relationship between parental education and the early detection of ASD in children[10].

Research suggests that the age of the mother at birth could play a significant role. Younger mothers may possess lower education levels due to having children earlier in life, which could potentially impact the early detection of ASD (Mirowsky, 2017). Conversely, older mothers may be more educated and experienced, potentially leading to earlier detection of developmental concerns (Goin-Kochel, Mackintosh & Myers, 2006).

The mental and physical health of the parents may also act as confounding factors. Parents experiencing mental health issues might exhibit lower educational attainment (Ettner, Frank & Kessler, 1997) and encounter challenges in recognizing early signs of autism in their children (Osborne & Reed, 2008). Similarly, parents in poor physical health may have their education and capacity to detect autism early affected by their health conditions (Case & Paxson, 2006).

The nativity of parents is another factor considered in this study. Immigrant parents, with potentially diverse educational backgrounds, could face language and cultural barriers, thus impacting their access to health care services for their children (Weathers et al., 2008). This could, in turn, influence their ability to seek early diagnosis and support for children with ASD.

The presence of multiple children with special needs in a household can be a double-edged sword. On one hand, it may heighten awareness of developmental milestones and early signs of autism (Chavira et al., 2014). On the other, it could strain family resources, possibly affecting parents' education and the early detection of autism in other children.

Geographic location can also influence both parental education (Panizzon, 2015) and early autism detection, as studies have identified several spatial ASD hotspots in the US (Bakian et al., 2015; Bradshaw et al., 2023). States with better access to educational opportunities and autism-related services may exhibit a higher prevalence of well-educated parents and earlier autism detection rates. Additionally, differences in state policies, funding, and awareness

[10] All of those confounding factors have been included as explanatory variables in the estimation.

campaigns can also impact the relationship between parental education and early detection of ASD.

The study also includes the factor of premature birth, as it can be associated with various health and developmental issues (Johnson et al., 2010), potentially complicating parents' ability to identify early signs of autism. Moreover, parents of premature babies may have encountered challenges during pregnancy that could have affected their education and capacity to recognize early ASD signs. Okui (2023) and Ruiz et al. (2015) have demonstrated this negative relationship between premature birth and parental education.

Lastly, the size of a household might exert some influence on the early detection of ASD. Larger households may possess more diverse educational backgrounds and a broader range of experiences, which could influence the early detection of ASD (Sicherman et al., 2018). However, larger households might also place greater demands on parents' time and resources, affecting their ability to pursue education (Blaabæk, Jæger & Molitoris, 2019) and closely monitor their children's development.

4.3 Descriptive statistics

The sample under investigation comprises 3,706 children diagnosed with ASD and their respective parents or caregivers. As portrayed in Figure 2, a striking contrast in the conditional probability of diagnosis across the three levels of ASD severity emerges. Regardless of the educational background of the parents[11], children with severe ASD are likely to receive a diagnosis earlier than their counterparts diagnosed with moderate and mild ASD. The average ages at diagnosis for mild, moderate, and severe ASD stand at 70, 56, and 43 months, respectively. Further examination reveals only a marginal discrepancy in the age at detection between mild and severe cases based on

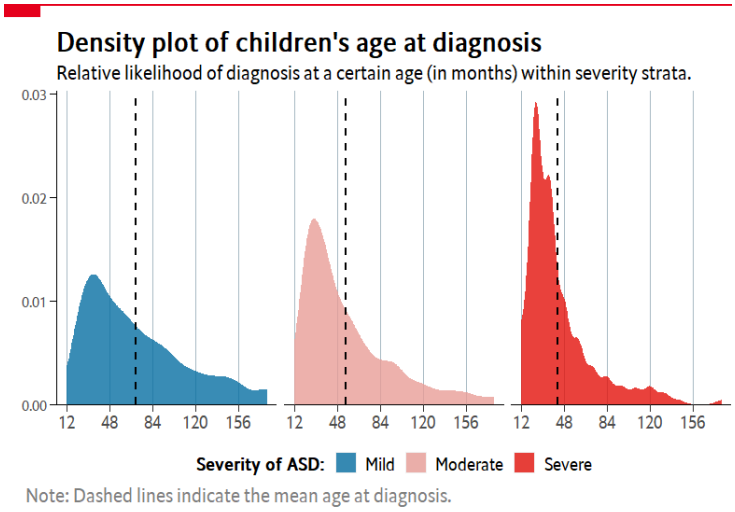

Figure 2

[11] The educational attainment of parents should be understood as the highest level of education among household members.

parental education. However, in moderate cases, a substantial difference of approximately 7 months becomes evident, with children of highly educated parents diagnosed at 53 months and those with less educated parents at 60 months. These observations corroborate the assertions in section 4.1.2 and suggest that severe cases might be easier for parents of all education levels to recognize due to pronounced symptoms, whereas mild cases could present diagnostic challenges without professional assistance. For moderate cases, heightened parental awareness possibly underlies earlier detection.

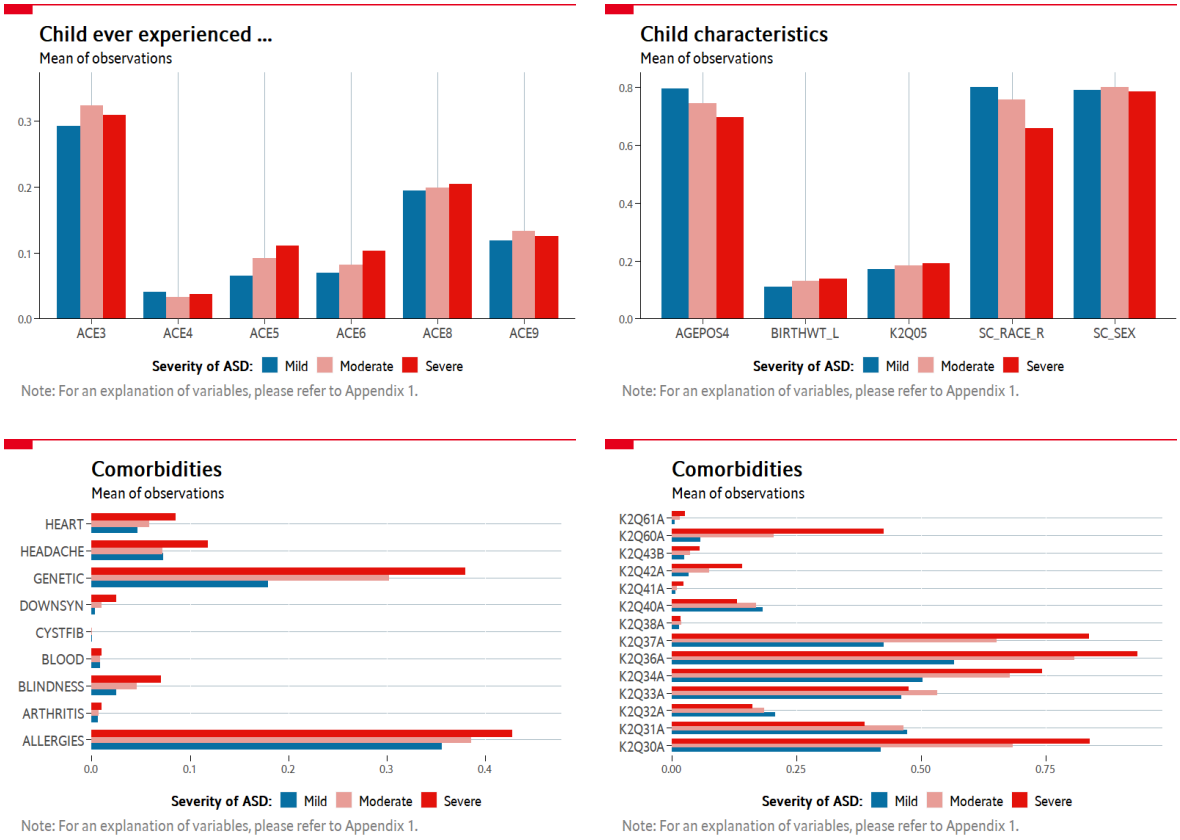

Figure 3

In Figure 3, the general characteristics of children are graphed according to ASD severity and the presence of co-occurring conditions. Notably, there are no significant differences in experienced life events across all severity strata. However, marked variations appear between mild and severe ASD cases concerning parental incarceration (ACE5) and exposure to interparental violence (ACE6). These factors could potentially impinge on parental attentiveness and timely intervention for autistic children. Gender (SC_SEX) does not seem to influence the severity of ASD, while lower birth weight (BIRTHWT_L) and premature birth (K2Q05) are slightly more common in severe ASD cases. Interestingly, children from non-white racial backgrounds tend to exhibit severe ASD cases more frequently.

Additionally, being the only or the oldest child in the family (AGEPOS4) is a more common feature of children with mild ASD than with severe condition.

In the realm of comorbidities, children with severe ASD generally manifest a larger number of co-occurring conditions than those with mild ASD. Exceptions are observed in the prevalence of asthma (K2Q40A), depression (K2Q32A), and attention deficit disorder or attention-deficit/hyperactivity disorder (K2Q31A) which appear to be more common among children with mild ASD. Anxiety (K2Q33A) is the sole condition that exhibits a higher prevalence in moderate ASD cases. The presence of these comorbidities, influencing the frequency of medical consultations, may facilitate an earlier diagnosis for severe cases.

Finally, it is noted that the mean age of mothers at childbirth is approximately 29 years across all severity levels. While the number of siblings does not significantly differ across ASD severity levels, children with severe ASD tend to have more siblings with special healthcare needs in the household. This observation suggests a potential genetic influence and heightened parental alertness that might contribute to earlier detection.

5. Empirical results and discussion

5.1 Results

The results from the DML estimation process, employing both PLR and IRM models, are presented in Table 1. The table showcases causal parameter estimates (θ) for the effect of parental education (having a higher education or not) on the early detection of ASD in children (age at diagnosis in months), considering different severity levels (mild, moderate, and severe cases) and using Ensemble models based on LASSO, RF, and XGBoost. A standard OLS estimate of the relationship has been included to serve as a benchmark.

The evidence presented in Table 1 underscores the role of parental education as a determinant of early ASD detection in children. This relationship is particularly pronounced when examining the data through the lens of the OLS and Ensemble-PLR models without considering the stratification of ASD severity (θ_{all}), revealing a statistically significant negative relationship at the 5% level. This finding suggests that higher parental education can potentially lead to earlier ASD detection. However, with the introduction of heterogeneity in the Ensemble-IRM model, the impact of parental education on early detection of ASD appears to exhibit greater complexity. In this context, the data does not provide enough statistical evidence to definitively reject the null hypothesis, which posits that the effect of

parental education on early detection of ASD could be insignificant. This nuanced relationship underscores the multifaceted nature of ASD detection and the integral role played by severity classifications in understanding the dynamics of early diagnosis.

Table 1. DML estimates against OLS benchmark

	OLS	Ensemble-PLR	Ensemble-IRM
		2-fold cross-fitting	
θ_{all}	-3.379**	-3.212**	-2.094
	(1.550)	(1.532)	(2.908)
θ_{mild}	-1.140	-0.479	-0.796
	(2.711)	(2.795)	(4.143)
$\theta_{moderate}$	-6.396***	-6.843***	-7.799***
	(2.175)	(2.139)	(2.278)
θ_{severe}	2.411	-0.698	-1.202
	(3.457)	(2.817)	(5.097)

Significance level: ** 5%; *** 1%

Upon disaggregating these findings by severity levels, a nuanced picture emerges. For mild and severe ASD cases (θ_{mild} and θ_{severe}), all model findings lack statistical significance, suggesting that parental education does not have a substantial influence on the early detection of mild and severe ASD cases. Intriguingly, the OLS model shows a positive correlation, potentially suggesting its inability to accurately depict the connection between parental education and early ASD detection, possibly due to the complex nature of this relationship or constraints stemming from insufficient data. This inconsistency suggests that this relationship remains an unresolved area of inquiry requiring further exploration.

Conversely, moderate ASD cases ($\theta_{moderate}$) display a robust and statistically significant negative correlation across all models at the 1% level. This implies that parental education exerts a significant influence on the early detection of moderate ASD cases, with this effect being slightly more pronounced in the Ensemble - IRM (-7.799). This pronounced influence illuminates the potent role parental education plays, particularly when confronted with moderate manifestations of ASD.

5.2 Robustness check

In order to affirm the validity and reliability of the primary findings of this study, a series of robustness checks have been implemented. The robustness checks aim to scrutinize the consistency of the relationship between parental education and the early detection of ASD across varying severity levels. This is achieved by applying three different cross-fitting folds and utilizing distinct machine learning models described in section 3.4. The intention is to

reinforce the findings by examining their stability under varied model configurations and machine learning methods. By exploring the extent to which the observed effects endure across diverse specifications, the robustness checks offer a more nuanced and comprehensive understanding of the results.

The robustness-checking procedure initiates with the examination of the PLR model estimates in Table 2. The primary focus here is the overarching impact of parental education on the early detection of ASD (θ_{all}), viewed through the lens of various machine learning techniques and cross-fitting procedures. The estimates derived from these explorations consistently reveal a negative and statistically significant relationship, fortified at the 5% significance level. This underscores the potential association between higher parental education and advanced detection of ASD, approximated between 3.2 and 3.5 months. Such findings bolster the assertion that heightened levels of parental education catalyse the earlier detection of ASD in children.

When considering the severity level, a divergence in the results is observed. For mild ASD cases (θ_{mild}), the estimates across all machine learning techniques and cross-fitting procedures are not statistically significant. This indicates an absence of compelling evidence to substantiate a relationship between parental education and early detection for this subset of ASD cases.

In stark contrast, moderate ASD cases ($\theta_{moderate}$) consistently yield a negative and statistically significant relationship at the 1% level across all machine learning techniques and cross-fitting procedures. These estimates are greater in magnitude compared to the overall effect, underscoring the pivotal role of parental education in the early detection of moderate ASD cases. Indeed, the estimates suggest that higher parental education could expedite the detection of moderate ASD cases by approximately 6.5 to 7 months.

As for severe ASD cases (θ_{severe}), the estimates, derived from the 2-fold cross-fitting procedure, do not manifest a statistically significant relationship across all machine learning techniques. This precludes the drawing of any firm conclusions concerning the relationship between parental education and the early detection of severe ASD cases based on the available data.

In contrast to the PLR estimates, the IRM estimates (Table 3) for the overall effect of parental education on the early detection of ASD (θ_{all}) present less consistent findings. While some

Table 2. DML causal parameter estimates - PLR

	LASSO	Random Forest	Gradient Boosting	Ensemble
		2-fold cross-fitting		
θ_{all}	-3.358**	-3.466**	-3.330**	-3.212**
	(1.537)	(1.527)	(1.557)	(1.532)
θ_{mild}	-0.042	-0.385	-1.434	-0.479
	(2.766)	(2.778)	(2.816)	(2.795)
$\theta_{moderate}$	-6.943***	-6.920***	-6.598***	-6.843***
	(2.108)	(2.160)	(2.116)	(2.139)
θ_{severe}	1.954	-0.215	-1.826	-0.698
	(2.643)	(2.939)	(2.642)	(2.817)
		4-fold cross-fitting		
θ_{all}	-3.358**	-3.432**	-3.342**	-3.238**
	(1.537)	(1.530)	(1.551)	(1.533)
θ_{mild}	-0.042	-1.063	-0.674	-0.621
	(2.766)	(2.738)	(2.772)	(2.763)
$\theta_{moderate}$	-6.943***	-6.685***	-6.750***	-6.772***
	(2.108)	(2.110)	(2.099)	(2.100)
θ_{severe}	-	-	-	-
		5-fold cross-fitting		
θ_{all}	-3.358**	-3.194**	-3.379**	-3.189**
	(1.537)	(1.526)	(1.548)	(1.526)
θ_{mild}	-0.042	-0.841	-0.731	-0.415
	(2.766)	(2.755)	(2.831)	(2.781)
$\theta_{moderate}$	-6.943***	-7.101***	-6.474***	-6.752***
	(2.108)	(2.100)	(2.112)	(2.103)
θ_{severe}	-	-	-	-

Significance level: ** 5%; *** 1%

machine learning techniques and cross-fitting procedures suggest a negative relationship, a majority of the estimates do not achieve statistical significance. An exception to this trend is observed in the XGBoost technique, utilizing 5-fold cross-fitting. Here, a statistically significant relationship emerges at the 5% level, linking higher parental education to earlier detection of ASD by approximately 4.7 months.

Akin to the PLR findings, the IRM estimates for mild ASD cases (θ_{mild}) across all machine learning techniques and cross-fitting procedures are not statistically significant, suggesting an absence of a clear relationship between parental education and early detection for mild ASD cases.

Table 3. DML causal parameter estimates - IRM

	LASSO	Random Forest	Gradient Boosting	Ensemble
		2-fold cross-fitting		
θ_{all}	-3.148	-2.105	-3.248	-2.094
	(2.433)	(1.718)	(2.172)	(2.908)
θ_{mild}	-0.225	-1.510	-2.117	-0.796
	(3.354)	(3.275)	(3.991)	(4.143)
$\theta_{moderate}$	-6.330***	-7.211***	-7.018**	-7.799***
	(2.225)	(2.557)	(3.304)	(2.278)
θ_{severe}	4.385	-0.914	-2.927	-1.202
	(2.913)	(3.697)	(8.285)	(5.097)
		4-fold cross-fitting		
θ_{all}	-3.148	-1.915	-4.078*	-3.454
	(2.433)	(1.697)	(2.143)	(2.458)
θ_{mild}	-0.225	-1.256	-0.375	-1.414
	(3.354)	(3.049)	(3.836)	(3.916)
$\theta_{moderate}$	-6.330***	-7.365***	-7.032**	-6.480***
	(2.225)	(2.245)	(2.904)	(2.334)
θ_{severe}	-	-	-	-
		5-fold cross-fitting		
θ_{all}	-3.148	-3.080*	-4.688**	-2.103
	(2.433)	(1.690)	(2.137)	(2.368)
θ_{mild}	-0.225	-1.771	-2.940	-0.519
	(3.354)	(3.011)	(3.982)	(4.320)
$\theta_{moderate}$	-6.330***	-7.455***	-6.608**	-6.768***
	(2.225)	(2.213)	(2.808)	(2.312)
θ_{severe}	-	-	-	-

Significance level: * 10%, ** 5%; *** 1%

Moreover, for moderate ASD cases ($\theta_{moderate}$), the IRM estimates consistently show a negative and statistically significant relationship at the 1% level, mirroring the PLR findings. The estimates propose that children with moderate ASD and highly educated parents are diagnosed approximately 6.3 - 7.8 months earlier than their counterparts who do not have parents holding higher degrees.

Lastly, just as with the PLR models, the IRM estimates for severe ASD cases (θ_{severe}) from the 2-fold cross-fitting procedure reveal no statistically significant relationship across all machine learning techniques. Thus, the available data does not permit us to draw definitive conclusions about the relationship between parental education and early detection for this subset of cases.

In conclusion, our robustness checks, employing both the PLR and IRM models across a variety of machine learning techniques and cross-fitting procedures, corroborate the significant role of parental education in the early detection of ASD, especially in moderate cases. However, the relationship between parental education and the early detection of ASD in mild and severe cases remains ambiguous. These findings, though not entirely uniform, add credence to the primary results, bolstering the robustness of the study. Future research, with perhaps broader or more specific data, may illuminate these areas further, providing additional insights into the nuanced role of parental education in the early detection of ASD across varying severity levels.

5.3 Discussion

The findings of this study, derived from both PLR and IRM models, indicate that parental education has a significant impact on the early detection of ASD, particularly for moderate cases. The lack of statistically significant results for mild cases, as observed in both models, may be attributed to the subtler symptoms associated with this severity level, which could make it challenging for parents to identify the early signs of ASD, irrespective of their education level.

The absence of statistically significant results for severe cases, specifically when employing 2-fold cross-fitting in both PLR and IRM models, may be due to the limited sample size for this subpopulation, which could reduce the power of the analysis. Consequently, further research with larger sample sizes may be required to elucidate the relationship between parental education and early detection of severe ASD cases. An alternative explanation for the lack of statistically significant results for severe cases of ASD could be that the symptoms associated with severe ASD are more apparent and easier to identify than those of mild and moderate cases. Consequently, parents, regardless of their education level, may be able to spot the signs of severe ASD in their children more or less at the same time. This could result in a relatively equal likelihood of early detection among both highly educated and less educated parents, leading to the observed lack of a significant relationship between parental education and early detection of severe ASD cases. Lastly, the higher prevalence of comorbidities associated with ASD, which often leads to increased medical consultations, could also potentially play a role in its detection irrespective of parental education.

It is crucial to note that the interpretation of these effects as causal (as average treatment effect) is contingent on certain assumptions. Both the PLR and IRM models used in this study

assume that there is no unobserved confounding, meaning all variables that could affect both parental education and early detection of ASD have been accounted for in the models. Additionally, the IRM model assumes an additive error structure, meaning that the effect of unobserved factors is the same across different levels of parental education. Therefore, while the evidence suggests a causal relationship, the possibility of bias due to unobserved confounding or violation of the additive error structure cannot be entirely ruled out.

The profound implications of the current research findings become more discernible when contextualized alongside pertinent studies that illuminate the economic implications of ASD. Cidav et al.'s (2013) research on ASD-associated expenditures reveals that adjusted mean total expenditures grow significantly with each year of age, with a marked 23% increase between the age brackets of 3–5 and 6–11. These findings, when juxtaposed with our results, gain economic significance. The early detection of moderate ASD cases facilitated by increased parental education, estimated to hasten diagnosis by 6-7 months, could potentially mitigate these rising costs. Considering that the average age of diagnosis in the present study for moderate ASD cases is 56 months (approximately 4.7 years), the impact of early detection can be particularly salient in mitigating the significant cost jump associated with the transition from 3-5 years to 6-11 years. Additionally, a persuasive argument for Early Intensive Behavioural Intervention (EIBI) is presented in the study by Chasson et al. (2007), with projections indicating savings of $208,500 per child over an eighteen-year educational span in Texas. The implications of our findings, which emphasize the pivotal role of parental education in facilitating early ASD detection, suggest an indirect contribution to these savings via the potential for earlier access to EIBI. The advancement of diagnosis by 6-7 months could lead to diminished special education costs, thus aligning with the substantial savings proposed by Chasson et al. (2007). Moreover, these are direct cost savings and do not account for the additional societal benefits and private savings that may result from earlier detection and intervention, such as enhanced productivity for parents, improved quality of life for the children, and potential future contributions to society by these children. These benefits, although harder to quantify, are nonetheless vital considerations in appreciating the full economic value of this study’s findings.

5.4 Limitations

While this research offers compelling insights into the influence of parental education on the early detection of ASD, it is not devoid of caveats that must be considered. At the outset, it

should be noted that the PLR and IRM models employed in the study come with inherent assumptions such as the exclusion of unobserved confounders and the presupposition of an additive error structure. Secondly, the conclusions drawn for severe ASD cases rest on a relatively limited sample size, which might call into question the generalizability and statistical power of these findings. Thirdly, while the focus on parental education is justified by the findings, it is important to acknowledge that this research does not sufficiently probe into the roles of other socio-demographic factors or specific parental awareness about ASD. These elements are assumed to be subsumed within the location and education variables. Furthermore, the study does not consider the stratification within different subtypes of ASD, such as autistic disorder, Asperger's syndrome, and pervasive developmental disorder-not otherwise specified. Each subtype may present distinct characteristics and analysing them separately could provide more nuanced insights into the relationship between parental education and early detection of ASD. Finally, the binary representation of parental education — delineating whether a parent has obtained a university-level education or not — poses another limitation. This reductionist approach aids in the binary distinction of education levels but fails to capture the inherent heterogeneity within these categories. The varying impacts associated with different levels of higher education are thereby overlooked. However, this binary treatment was dictated by the IRM model's constraints employed in our study. Future research could consider exploring a more detailed stratification of parental education or alternative model structures that permit a continuous treatment variable. Incorporating these considerations could pave the way toward a more nuanced understanding of ASD's early detection, contributing to the depth and breadth of this research area.

6 Conclusion

Utilizing the DML framework and applying both PLR and IRM methods, this study conducts a thorough exploration of the relationship between parental education and the timing of ASD diagnosis.

The findings unveil a noteworthy inverse association between general parental education and the age of ASD diagnosis, particularly pronounced in moderate ASD cases. Higher levels of parental education — obtained university degree (bachelor’s degree or higher) — lead to earlier detection of ASD, underscoring the instrumental role of parental education in catalysing early interventions and facilitating appropriate support systems for children with ASD.

Contrarily, for mild and severe cases of ASD, the evidence presents a less transparent picture of the relationship between parental education and early detection. The absence of consistent statistical significance in these categories suggests a degree of complexity and variability that warrants further investigation. Thus, future research is encouraged to delve into the intricacies of how parental education may differentially influence the timing of diagnosis across the ASD severity spectrum, specifically with a larger and more detailed data set.

By highlighting the substantial role of parental education in the early diagnosis of moderate ASD cases, this study sheds light on the importance of general education in informing parents' ability to facilitate timely diagnosis and intervention. The results of this study substantiate the value of broad-based educational attainment as a key factor influencing early ASD detection and affirm the utility of deploying both PLR and IRM models in analysing complex relationships within health-related research. As efforts to improve early detection and treatment of ASD persist, this study contributes a valuable piece to the broader puzzle, helping refine our understanding of the influence of parental education on ASD diagnosis timing.

7 References

Allen, L., Leon-Attia, O., Shaham, M., Shefer, S. & Gabis, L. V. (2020). Autism Risk Linked to Prematurity Is More Accentuated in Girls, *PLOS ONE*, vol. 15, no. 8, p.e0236994

Alvares, G. A., Licari, M. K., Stevenson, P. G., Bebbington, K., Cooper, M. N., Glasson, E. J., Tan, D. W., Uljarević, M., Varcin, K. J., Wray, J. & Whitehouse, A. J. O. (2021). Investigating Associations between Birth Order and Autism Diagnostic Phenotypes, *Journal of Child Psychology and Psychiatry*, vol. 62, no. 8, pp.961–970

American Psychiatric Association (ed.). (2013). Diagnostic and Statistical Manual of Mental Disorders: DSM-5, 5th ed., Washington, D.C: American Psychiatric Association

Andoy Galvan, J. A., Ramalingam, P. N., Patil, S. S., bin Shobri, M. A. S., Chinna, K., Sahrir, M. S. & Chidambaram, K. (2020). Mode of Delivery, Order of Birth, Parental Age Gap and Autism Spectrum Disorder among Malaysian Children: A Case-Control Study, *Heliyon*, vol. 6, no. 10, p.e05068

Baio, J. (2018). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States,

2014, *MMWR. Surveillance Summaries*, [e-journal] vol. 67, Available Online: https://www.cdc.gov/mmwr/volumes/67/ss/ss6706a1.htm [Accessed 8 May 2023]

Bakian, A. V., Bilder, D. A., Coon, H. & McMahon, W. M. (2015). Spatial Relative Risk Patterns of Autism Spectrum Disorders in Utah, *Journal of Autism and Developmental Disorders*, vol. 45, no. 4, pp.988–1000

Beesley, R. (2020). Thu0507 Association Between Juvenile Idiopathic Arthritis and Autism, *Annals of the Rheumatic Diseases*, vol. 79, no. Suppl 1, pp.492–492

Bhasin, T. K. & Schendel, D. (2007). Sociodemographic Risk Factors for Autism in a US Metropolitan Area, *Journal of Autism and Developmental Disorders*, vol. 37, no. 4, pp.667–677

Blaabæk, E. H., Jæger, M. M. & Molitoris, J. (2019). Family Size and Educational Attainment: Cousins, Contexts, and Compensation, *European Journal of Population = Revue Européenne de Démographie*, vol. 36, no. 3, pp.575–600

Bradshaw, J., Eberth, J. M., Zgodic, A., Federico, A., Flory, K. & McLain, A. C. (2023). County-Level Prevalence Estimates of Autism Spectrum Disorder in Children in the United States, *Journal of Autism and Developmental Disorders*, [e-journal], Available Online: https://doi.org/10.1007/s10803-023-05920-z [Accessed 9 May 2023]

Buescher, A. V. S., Cidav, Z., Knapp, M. & Mandell, D. S. (2014). Costs of Autism Spectrum Disorders in the United Kingdom and the United States, *JAMA Pediatrics*, vol. 168, no. 8, pp.721–728

Canitano, R. (2007). Epilepsy in Autism Spectrum Disorders, *European Child & Adolescent Psychiatry*, vol. 16, no. 1, pp.61–66

Canitano, R. & Vivanti, G. (2007). Tics and Tourette Syndrome in Autism Spectrum Disorders, *Autism*, vol. 11, no. 1, pp.19–28

Case, A. & Paxson, C. (2006). Children's Health and Social Mobility, *The Future of Children*, vol. 16, no. 2, pp.151–173

CDC's Autism and Developmental Disabilities Monitoring Network. (2023). 2023 ADDM Factsheet, Factsheet, Centers for Disease Control and Prevention, Available Online:

https://www.cdc.gov/ncbddd/autism/pdf/2023-ADDM-Factsheet_508.pdf [Accessed 8 May 2023]

Chasson, G. S., Harris, G. E. & Neely, W. J. (2007). Cost Comparison of Early Intensive Behavioral Intervention and Special Education for Children with Autism, *Journal of Child and Family Studies*, vol. 16, no. 3, pp.401–413

Chavira, D. A., Drahota, A., Garland, A. F., Roesch, S., Garcia, M. & Stein, M. B. (2014). Feasibility of Two Modes of Treatment Delivery for Child Anxiety in Primary Care., [e-journal], Available Online: https://escholarship.org/uc/item/21604757 [Accessed 9 May 2023]

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. & Robins, J. (2018). Double/Debiased Machine Learning for Treatment and Structural Parameters, *The Econometrics Journal*, vol. 21, no. 1, pp.C1–C68

Chua, R. X. Y., Tay, M. J. Y., Ooi, D. S. Q., Siah, K. T. H., Tham, E. H., Shek, L. P.-C., Meaney, M. J., Broekman, B. F. P. & Loo, E. X. L. (2021). Understanding the Link Between Allergy and Neurodevelopmental Disorders: A Current Review of Factors and Mechanisms, *Frontiers in Neurology*, vol. 11, p.603571

Chuang, H.-C. & Chen, J. (2023). Exploring Industry-Distress Effects on Loan Recovery: A Double Machine Learning Approach for Quantiles, *Econometrics*, vol. 11, no. 6, p.6

Cidav, Z., Lawer, L., Marcus, S. C. & Mandell, D. S. (2013). Age-Related Variation in Health Service Use and Associated Expenditures Among Children with Autism, *Journal of Autism and Developmental Disorders*, vol. 43, no. 4, pp.924–931

Cidav, Z., Marcus, S. C. & Mandell, D. S. (2012). Implications of Childhood Autism for Parental Employment and Earnings, *Pediatrics*, vol. 129, no. 4, pp.617–623

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., Donaldson, A. & Varley, J. (2010). Randomized, Controlled Trial of an Intervention for Toddlers With Autism: The Early Start Denver Model, *Pediatrics*, vol. 125, no. 1, pp.e17–e23

Delobel-Ayoub, M., Ehlinger, V., Klapouszczak, D., Maffre, T., Raynaud, J.-P., Delpierre, C. & Arnaud, C. (2015). Socioeconomic Disparities and Prevalence of Autism Spectrum Disorders and Intellectual Disability, *PLOS ONE*, vol. 10, no. 11, p.e0141964

Diament, M. (2020). CDC Researchers: Over 5 Million US Adults Have Autism, *Disability Scoop*, Available Online: https://www.disabilityscoop.com/2020/05/13/cdc-researchers-over-5-million-adults-autism/28327/ [Accessed 8 May 2023]

Dube, A., Jacobs, J., Naidu, S. & Suri, S. (2020). Monopsony in Online Labor Markets, *American Economic Review: Insights*, vol. 2, no. 1, pp.33–46

Durkin, M. S., Maenner, M. J., Meaney, F. J., Levy, S. E., DiGuiseppi, C., Nicholas, J. S., Kirby, R. S., Pinto-Martin, J. A. & Schieve, L. A. (2010). Socioeconomic Inequality in the Prevalence of Autism Spectrum Disorder: Evidence from a U.S. Cross-Sectional Study, *PLoS ONE*, vol. 5, no. 7, p.e11551

Efron, B. (1983). Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation, *Journal of the American Statistical Association*, vol. 78, no. 382, pp.316–331

Elsabbagh, M., Divan, G., Koh, Y.-J., Kim, Y. S., Kauchali, S., Marcín, C., Montiel-Nava, C., Patel, V., Paula, C. S., Wang, C., Yasamy, M. T. & Fombonne, E. (2012). Global Prevalence of Autism and Other Pervasive Developmental Disorders: Global Epidemiology of Autism, *Autism Research*, vol. 5, no. 3, pp.160–179

Ettner, S. L., Frank, R. G. & Kessler, R. C. (1997). The Impact of Psychiatric Disorders on Labor Market Outcomes, *Industrial and Labor Relations Review*, vol. 51, no. 1, pp.64–81

Fountain, C., King, M. D. & Bearman, P. S. (2011). Age of Diagnosis for Autism: Individual and Community Factors across 10 Birth Cohorts, *Journal of Epidemiology & Community Health*, vol. 65, no. 6, pp.503–510

Frenette, P., Dodds, L., MacPherson, K., Flowerdew, G., Hennen, B. & Bryson, S. (2013). Factors Affecting the Age at Diagnosis of Autism Spectrum Disorders in Nova Scotia, Canada, *Autism*, vol. 17, no. 2, pp.184–195

Frisch, R. & Waugh, F. V. (1933). Partial Time Regressions as Compared with Individual Trends, *Econometrica*, vol. 1, no. 4, pp.387–401

Ganz, M. L. (2007). The Lifetime Distribution of the Incremental Societal Costs of Autism, *Archives of Pediatrics & Adolescent Medicine*, vol. 161, no. 4, pp.343–349

Gaugler, T., Klei, L., Sanders, S. J., Bodea, C. A., Goldberg, A. P., Lee, A. B., Mahajan, M., Manaa, D., Pawitan, Y., Reichert, J., Ripke, S., Sandin, S., Sklar, P., Svantesson, O., Reichenberg, A., Hultman, C. M., Devlin, B., Roeder, K. & Buxbaum, J. D. (2014). Most Genetic Risk for Autism Resides with Common Variation, 8, *Nature Genetics*, vol. 46, no. 8, pp.881–885

Geschwind, D. H. (2011). Genetics of Autism Spectrum Disorders, *Trends in Cognitive Sciences*, vol. 15, no. 9, pp.409–416

Goin-Kochel, R. P., Mackintosh, V. H. & Myers, B. J. (2006). How Many Doctors Does It Take to Make an Autism Spectrum Diagnosis?, *Autism*, vol. 10, no. 5, pp.439–451

Hahn, L. J., Hamrick, L. M., Kelleher, B. L. & Roberts, J. E. (2020). Autism Spectrum Disorder-Associated Behaviour in Infants with Down Syndrome, *Journal of health science & education*, vol. 4, no. 2, p.180

Höglund Carlsson, L., Norrelgen, F., Kjellmer, L., Westerlund, J., Gillberg, C. & Fernell, E. (2013). Coexisting Disorders and Problems in Preschool Children with Autism Spectrum Disorders, *The Scientific World Journal*, vol. 2013, p.e213979

Hrdlicka, M., Vacova, M., Oslejskova, H., Gondzova, V., Vadlejchova, I., Kocourkova, J., Koutek, J. & Dudova, I. (2016). Age at Diagnosis of Autism Spectrum Disorders: Is There an Association with Socioeconomic Status and Family Self-Education about Autism?, *Neuropsychiatric Disease and Treatment*, vol. 12, pp.1639–1644

Johnson, S., Hollis, C., Kochhar, P., Hennessy, E., Wolke, D. & Marlow, N. (2010). Autism Spectrum Disorders in Extremely Preterm Children, *The Journal of Pediatrics*, vol. 156, no. 4, pp.525-531.e2

Joshi, G., Petty, C., Wozniak, J., Henin, A., Fried, R., Galdo, M., Kotarski, M., Walls, S. & Biederman, J. (2010). The Heavy Burden of Psychiatric Comorbidity in Youth with

Autism Spectrum Disorders: A Large Comparative Study of a Psychiatrically Referred Population, *Journal of Autism and Developmental Disorders*, vol. 40, no. 11, pp.1361–1370

Kerns, C. M., Rast, J. E. & Shattuck, P. T. (2020). Prevalence and Correlates of Caregiver-Reported Mental Health Conditions in Youth With Autism Spectrum Disorder in the United States, *The Journal of Clinical Psychiatry*, vol. 82, no. 1, p.11637

Knaus, M. C. (2021). A Double Machine Learning Approach to Estimate the Effects of Musical Practice on Student's Skills, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 184, no. 1, pp.282–300

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection., in *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95)*, Vol. vol 2, International Joint Conferences on Artificial Intelligence, Montreal, Quebec, Canada, 20 August 1995, Montreal, Quebec, Canada, pp.1137–1143, Available Online: https://www.ijcai.org/Proceedings/95-2/Papers/016.pdf

Lai, M.-C., Lombardo, M. V. & Baron-Cohen, S. (2014). Autism, *The Lancet*, vol. 383, no. 9920, pp.896–910

Leigh, J. P. & Du, J. (2015). Brief Report: Forecasting the Economic Burden of Autism in 2015 and 2025 in the United States, *Journal of Autism and Developmental Disorders*, vol. 45, no. 12, pp.4135–4139

Levy, S. E., Giarelli, E., Lee, L.-C., Schieve, L. A., Kirby, R. S., Cunniff, C., Nicholas, J., Reaven, J. & Rice, C. E. (2010). Autism Spectrum Disorder and Co-Occurring Developmental, Psychiatric, and Medical Conditions Among Children in Multiple Populations of the United States, *Journal of Developmental & Behavioral Pediatrics*, vol. 31, no. 4, p.267

Lovell, M. C. (1963). Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis, *Journal of the American Statistical Association*, vol. 58, no. 304, pp.993–1010

Mandell, D. S., Listerud, J., Levy, S. E. & Pinto-martin, J. A. (2002). Race Differences in the Age at Diagnosis Among Medicaid-Eligible Children With Autism, *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 41, no. 12, pp.1447–1453

Mandell, D. S., Wiggins, L. D., Yeargin-Allsopp, M., Carpenter, L. A., Daniels, J., Thomas, K. C., DiGuiseppi, C., Durkin, M. S., Giarelli, E., Pinto-Martin, J. A., Morrier, M. J., Nicholas, J. S., Shattuck, P. T. & Kirby, R. S. (2009). Racial/Ethnic Disparities in the Identification of Children With Autism Spectrum Disorders, *American Journal of Public Health*, vol. 99, no. 3, pp.493–498

Matson, J. L. & Goldin, R. L. (2013). Comorbidity and Autism: Trends, Topics and Future Directions, *Research in Autism Spectrum Disorders*, vol. 7, no. 10, pp.1228–1233

McConnell, K. J. & Lindner, S. (2019). Estimating Treatment Effects with Machine Learning, *Health Services Research*, vol. 54, no. 6, pp.1273–1282

Mirowsky, J. (2017). Education, Social Status, and Health, New York: Routledge

Modabbernia, A., Velthorst, E. & Reichenberg, A. (2017). Environmental Risk Factors for Autism: An Evidence-Based Review of Systematic Reviews and Meta-Analyses, *Molecular Autism*, vol. 8, no. 1, p.13

Okui, T. (2023). Analysis of an Association between Preterm Birth and Parental Educational Level in Japan Using National Data, *Children (Basel, Switzerland)*, vol. 10, no. 2, p.342

Osborne, L. A. & Reed, P. (2008). Parents' Perceptions of Communication with Professionals during the Diagnosis of Autism, *Autism*, vol. 12, no. 3, pp.309–324

Oyenubi, A. & Kollamparambil, U. (2023). Does Noncompliance with COVID-19 Regulations Impact the Depressive Symptoms of Others?, *Economic Modelling*, vol. 120, p.106191

Panizzon, D. (2015). Impact of Geographical Location on Student Achievement: Unpacking the Complexity of Diversity, in A. Bishop, H. Tan, & T. N. Barkatsas (eds), *Diversity in Mathematics Education: Towards Inclusive Practices*, [e-book] Cham: Springer

International Publishing, pp.41–61, Available Online: https://doi.org/10.1007/978-3-319-05978-5_3 [Accessed 9 May 2023]

Peet, E. D., Schultz, D., Lovejoy, S. & Tsui, F. (Rich). (2023). Variation in the Infant Health Effects of the Women, Infants, and Children Program by Predicted Risk Using Novel Machine Learning Methods, *Health Economics*, vol. 32, no. 1, pp.194–217

Peters-Scheffer, N., Didden, R., Korzilius, H. & Sturmey, P. (2011). A Meta-Analytic Study on the Effectiveness of Comprehensive ABA-Based Early Intervention Programs for Children with Autism Spectrum Disorders, *Research in Autism Spectrum Disorders*, vol. 5, no. 1, pp.60–69

Robinson, P. M. (1988). Root-N-Consistent Semiparametric Regression, *Econometrica*, vol. 56, no. 4, p.931

Ruiz, M., Goldblatt, P., Morrison, J., Kukla, L., Švancara, J., Riitta-Järvelin, M., Taanila, A., Saurel-Cubizolles, M.-J., Lioret, S., Bakoula, C., Veltsista, A., Porta, D., Forastiere, F., van Eijsden, M., Vrijkotte, T. G. M., Eggesbø, M., White, R. A., Barros, H., Correia, S., Vrijheid, M., Torrent, M., Rebagliato, M., Larrañaga, I., Ludvigsson, J., Olsen Faresjö, Å., Hryhorczuk, D., Antipkin, Y., Marmot, M. & Pikhart, H. (2015). Mother's Education and the Risk of Preterm and Small for Gestational Age Birth: A DRIVERS Meta-Analysis of 12 European Cohorts, *Journal of Epidemiology and Community Health*, vol. 69, no. 9, pp.826–833

Sanchez, P., Voisey, J. P., Xia, T., Watson, H. I., O'Neil, A. Q. & Tsaftaris, S. A. (2022). Causal Machine Learning for Healthcare and Precision Medicine, *Royal Society Open Science*, vol. 9, no. 8, p.220638

Screening and Diagnosis | Autism Spectrum Disorder (ASD) | NCBDDD. (2022). *Centers for Disease Control and Prevention*, Available Online: https://www.cdc.gov/ncbddd/autism/screening.html [Accessed 14 May 2023]

Shattuck, P. T., Durkin, M., Maenner, M., Newschaffer, C., Mandell, D. S., Wiggins, L., Lee, L.-C., Rice, C., Giarelli, E., Kirby, R., Baio, J., Pinto-Martin, J. & Cuniff, C. (2009). Timing of Identification Among Children With an Autism Spectrum Disorder:

Findings From a Population-Based Surveillance Study, *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 48, no. 5, pp.474–483

Sicherman, N., Loewenstein, G., Tavassoli, T. & Buxbaum, J. D. (2018). Grandma Knows Best: Family Structure and Age of Diagnosis of Autism Spectrum Disorder, *Autism*, vol. 22, no. 3, pp.368–376

Sigmon, E. R., Kelleman, M., Susi, A., Nylund, C. M. & Oster, M. E. (2019). Congenital Heart Disease and Autism: A Case-Control Study, *Pediatrics*, vol. 144, no. 5, p.e20184114

Signs & Symptoms | Autism Spectrum Disorder (ASD) | NCBDDD | CDC. (2023). *Centers for Disease Control and Prevention*, Available Online: https://www.cdc.gov/ncbddd/autism/signs.html [Accessed 8 May 2023]

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T. & Baird, G. (2008). Psychiatric Disorders in Children With Autism Spectrum Disorders: Prevalence, Comorbidity, and Associated Factors in a Population-Derived Sample, *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 47, no. 8, pp.921–929

Skoufias, E. & Vinha, K. (2021). Child Stature, Maternal Education, and Early Childhood Development in Nigeria, *PLoS ONE*, vol. 16, no. 12, p.e0260937

Volkmar, F. R., Reichow, B. & McPartland, J. (2012). Classification of Autism and Related Conditions: Progress, Challenges, and Opportunities, *Dialogues in Clinical Neuroscience*, vol. 14, no. 3, pp.229–237

Wang, N., Zhao, Y. & Gao, J. (2021). Association Between Peripheral Blood Levels of Vitamin A and Autism Spectrum Disorder in Children: A Meta-Analysis, *Frontiers in Psychiatry*, [e-journal] vol. 12, Available Online: https://www.frontiersin.org/articles/10.3389/fpsyt.2021.742937/full [Accessed 8 May 2023]

Weathers, A. C., Novak, S. P., Sastry, N. & Norton, E. C. (2008). Parental Nativity Affects Children's Health and Access to Care, *Journal of Immigrant and Minority Health*, vol. 10, no. 2, pp.155–165

Weiss, J. A., Isaacs, B., Diepstra, H., Wilton, A. S., Brown, H. K., McGarry, C. & Lunsky, Y. (2018). Health Concerns and Health Service Utilization in a Population Cohort of Young Adults with Autism Spectrum Disorder, *Journal of Autism and Developmental Disorders*, vol. 48, no. 1, pp.36–44

White, S. W., Oswald, D., Ollendick, T. & Scahill, L. (2009). Anxiety in Children and Adolescents with Autism Spectrum Disorders, *Clinical Psychology Review*, vol. 29, no. 3, pp.216–229

Xu, G., Snetselaar, L. G., Jing, J., Liu, B., Strathearn, L. & Bao, W. (2018). Association of Food Allergy and Other Allergic Conditions With Autism Spectrum Disorder in Children, *JAMA Network Open*, vol. 1, no. 2, p.e180279

Zhao, H., Zhang, H., Liu, S., Luo, W., Jiang, Y. & Gao, J. (2021). Association of Peripheral Blood Levels of Cytokines With Autism Spectrum Disorder: A Meta-Analysis, *Frontiers in Psychiatry*, [e-journal] vol. 12, Available Online: https://www.frontiersin.org/articles/10.3389/fpsyt.2021.670200/full [Accessed 8 May 2023]

Zuckerman, K., Lindly, O. J. & Chavez, A. E. (2017). Timeliness of Autism Spectrum Disorder Diagnosis and Use of Services Among U.S. Elementary School–Aged Children, *Psychiatric Services*, vol. 68, no. 1, pp.33–40

Appendix 1: Variables used in the study.

Table 4

Description of features

No	Feature	Description	Type
1	ACE3	Child ever experienced parent or guardian divorce or separation.	Binary (1 - yes; 0 - no)
2	ACE4	Child ever experienced parent or guardian death.	Binary (1 - yes; 0 - no)
3	ACE5	Child's parent or guardian served time in jail or prison (after child's birth).	Binary (1 - yes; 0 - no)
4	ACE6	Child ever saw or heard parents or adults slap, hit, kick, punch one another in the home.	Binary (1 - yes; 0 - no)
5	ACE8	Child ever lived with anyone who was mentally ill, suicidal, or severely depressed.	Binary (1 - yes; 0 - no)
6	ACE9	Child ever lived with anyone who had a problem with alcohol or drugs.	Binary (1 - yes; 0 - no)
7	AGEPOS4	Birth order of selected child in household.	Binary (1 - first/only; 0 - other)
8	SC_SEX	Sex of selected child.	Binary (1 - male; 0 - female)
9	SC_RACE_R	Race of selected child.	Binary (1 - white; 0 - other)
10	SC_AGE_YEARS	Age of a child	Integer
11	BIRTHWT_L	Low birth weight (<2500g)	Binary (1 - yes; 0 - no)
12	K2Q05	Child was born more than 3 weeks before their due date.	Binary (1 - yes; 0 - no)
13	BLINDNESS	Child has blindness or problems with seeing, even when wearing glasses.	Binary (1 - yes; 0 - no)
14	ALLERGIES	A doctor or other health care provider ever told you that this child has allergies (including food, drug, insect, or other)	Binary (1 - yes; 0 - no)
15	ARTHRITIS	A doctor or other health care provider ever told you that this child has Arthritis.	Binary (1 - yes; 0 - no)
16	BLOOD	A doctor or other health care provider ever told you that this child has Blood Disorders (such as Sickle Cell Disease, Thalassemia, or Hemophilia).	Binary (1 - yes; 0 - no)
17	CYSTFIB	A doctor or other health care provider ever told you that this child has Cystic Fibrosis.	Binary (1 - yes; 0 - no)
18	DOWNSYN	A doctor or other health care provider ever told you that this child has Down Syndrome.	Binary (1 - yes; 0 - no)
19	GENETIC	A doctor or other health care provider ever told you that this child has other	Binary (1 - yes; 0 - no)

		genetic or inherited condition.	
20	HEADACHE	A doctor or other health care provider ever told you that this child has frequent or severe headaches, including migraine.	Binary (1 - yes; 0 - no)
21	HEART	A doctor or other health care provider ever told you that this child has Heart Condition.	Binary (1 - yes; 0 - no)
22	K2Q30A	A doctor, other health care provider, or educator ever told you that this child has Learning Disability.	Binary (1 - yes; 0 - no)
23	K2Q31A	A doctor or other health care provider ever told you that this child has Attention Deficit Disorder or Attention-Deficit/Hyperactivity Disorder, that is, ADD or ADHD.	Binary (1 - yes; 0 - no)
24	K2Q32A	A doctor or other health care provider ever told you that this child has Depression.	Binary (1 - yes; 0 - no)
25	K2Q33A	A doctor or other health care provider ever told you that this child has Anxiety Problems.	Binary (1 - yes; 0 - no)
26	K2Q34A	A doctor, other health care provider, or educator ever told you that this child has Behavioral or Conduct Problems.	Binary (1 - yes; 0 - no)
27	K2Q36A	A doctor, other health care provider, or educator ever told you that this child has Developmental Delay.	Binary (1 - yes; 0 - no)
28	K2Q37A	A doctor, other health care provider, or educator ever told you that this child has Speech or other language disorder.	Binary (1 - yes; 0 - no)
29	K2Q38A	A doctor or other health care provider ever told you that this child has Tourette Syndrome.	Binary (1 - yes; 0 - no)
30	K2Q40A	A doctor or other health care provider ever told you that this child has Asthma.	Binary (1 - yes; 0 - no)
31	K2Q41A	A doctor or other health care provider ever told you that this child has Diabetes.	Binary (1 - yes; 0 - no)
32	K2Q42A	A doctor or other health care provider ever told you that this child has Epilepsy or seizure disorder.	Binary (1 - yes; 0 - no)
33	K2Q43B	Does this child have deafness or problems with hearing .	Binary (1 - yes; 0 - no)
34	K2Q60A	A doctor, other health care provider, or educator ever told you that this child has Intellectual Disability (formerly known as Mental Retardation).	Binary (1 - yes; 0 - no)
35	K2Q61A	A doctor or other health care provider ever told you that this child has Cerebral Palsy.	Binary (1 - yes; 0 - no)

36	K2Q35A	A doctor or other health care provider ever told you that this child has Autism or Autism Spectrum Disorder (ASD).	Binary (1 - yes; 0 - no)
37	K2Q35A_1_YEARS	Age of this child when a doctor or other health care provider first told you that they had Autism, ASD, Asperger's Disorder or PDD.	Integer
38	K2Q35C	Autism ASD Severity Description	Categorical (1 - mild; 2 - moderate; 3 - severe)
39	FPR	Family Poverty Ratio	Integer
40	FAMILY_R	Family structure	Binary (1 - both parents; 0 - only one or no parents)
41	HHCOUNT	Number of people living or staying at this address.	Integer
42	HHLANGUAGE	Primary language spoken in the household.	Binary (1 - English; 0 - other)
43	HIGRADE	Highest Level of Education among Reported Adults.	Binary (1 - university; 0 - below university)
44	HOUSE_GEN	Parental Nativity	Binary (1 - all parents born in the US; 0 - any other combination)
45	TOTKIDS_R	Number of Children in Household	Integer
46	TOTFEMALE	Count of Female Children in Household	Integer
47	TOTMALE	Count of Male Children in Household	Integer
48	TOTCSHCN	Count of Children with Special Health Care Needs in Household	Integer
49	MOMAGE	Age of the mother when this child was born.	Integer
50	MENTHEALTH	Mental health of both parents.	Binary (1 - good/excellent; 0 - Fair/poor)
51	PHYSHEALTH	Physical health of both parents.	Binary (1 - good/excellent; 0 - Fair/poor)
52	FIPSST	State FIPS Code	Categorical

Appendix 2: Hyperparameter tuning procedure for each machine learning model.

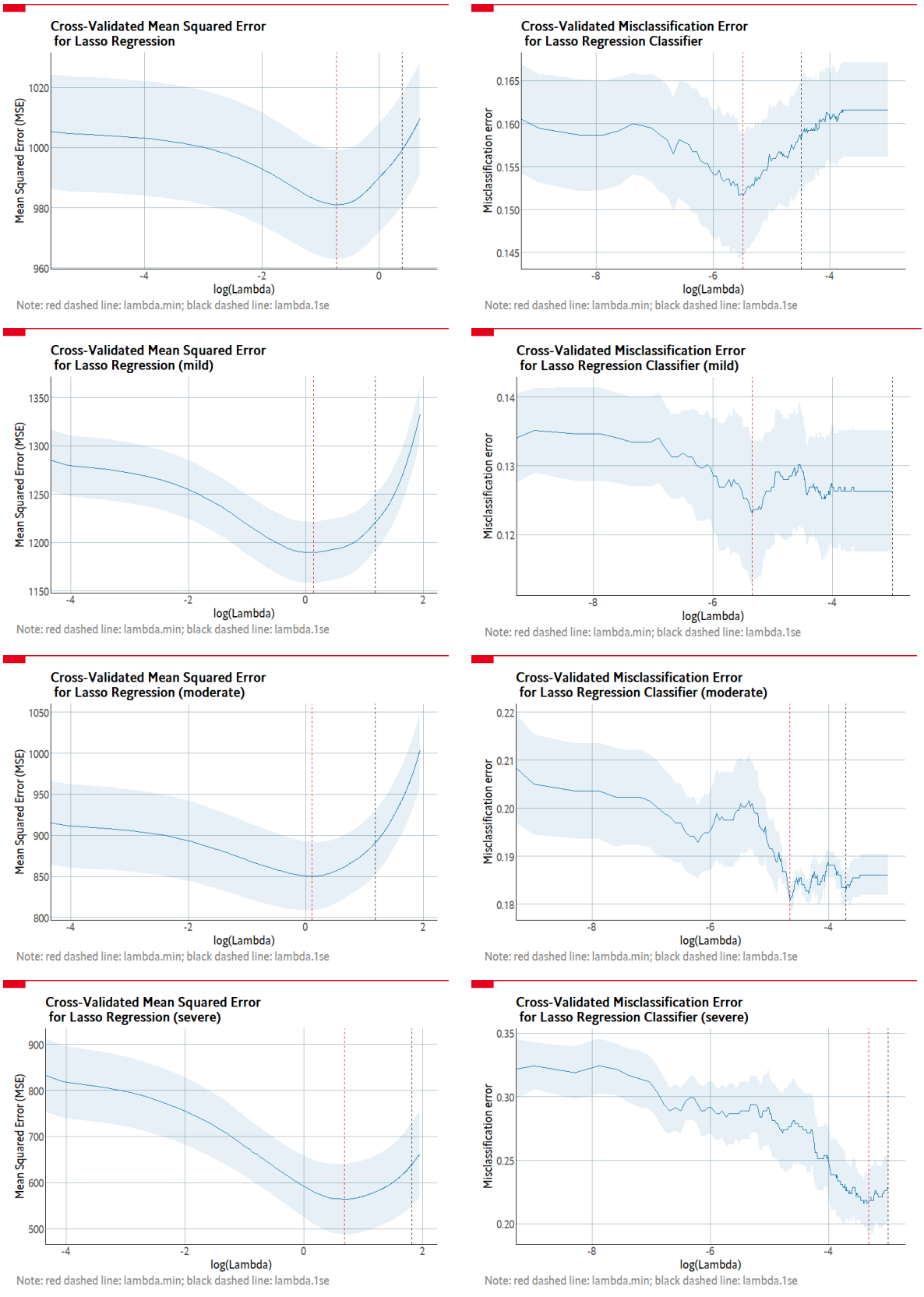

Figure 4

A2.1 LASSO

To ascertain the optimal regularization parameter λ for LASSO models, firstly, both regression and classification variants were trained utilizing the complete dataset. A pair of λ values were procured: one minimizing cross-validation error and another premised on the 1-standard error rule (Figure 4). Then, these values delineated the range for the regularization parameter during the hyperparameter optimization phase for DML estimation. The algorithm scrutinized fifty discrete λ values within the prescribed range, ensuring that the final models attain equilibrium between overfitting and underfitting phenomena.

A2.2 RF

In the context of the RF algorithm, two principal parameters were subjected to optimization: the number of variables examined at each split and the minimum node size. A ±50% range surrounding base values was selected for optimization in both parameters. The base value for the number of variables at each split adheres to the rule of thumb, which is P/3 for regression problems and √P for classification problems. This parameter substantially influences the RF model's diversity, performance, and bias-variance trade-off. The base node size, set at 5, impacts the intricacy and depth of individual decision trees within the RF model.

A2.3 XGBoost

The algorithm is configured to operate for a maximum of 500 iterations, implementing early stopping if no enhancement is observed for 10 consecutive iterations. The learning rate η was optimized within the range of (0.1, 0.5), while the regularization parameter λ was optimized within the range of (0.001, 10). The number of features evaluated at each boosting round was optimized in a manner akin to the RF model, utilizing a ±50% range around the rule-of-thumb values. The maximum depth of individual decision trees, deployed as weak learners during training, was optimized within a range of (1, 10). The tuning process contemplated up to 20 parallel trees for both regression and classification models.

A2.4 Ensemble

The ensemble model incorporates the three above-mentioned models - LASSO, RF, and XGBoost - to amalgamate their predictions using a straightforward arithmetic mean (i.e., they are given equal weight) (Figure 5). Each data point's predictions are computed individually for every model, and subsequently, these predictions are integrated by calculating their arithmetic

mean, yielding a singular prediction for each instance. This process is executed independently for both regression and classification tasks.

The combination of multiple models' predictions offers a substantial enhancement in the ensemble's overall performance by capitalizing on the strengths of each model while simultaneously mitigating their inherent weaknesses. This ensemble model is capable of delivering more precise and robust predictions by attenuating the propensity for overfitting the data, a common pitfall associated with the utilization of singular models. Furthermore, the combination of predictions emanating from a diverse array of models facilitates a more comprehensive representation of the intrinsic data patterns and relationships, ultimately contributing to the refinement of the causal effect estimation.

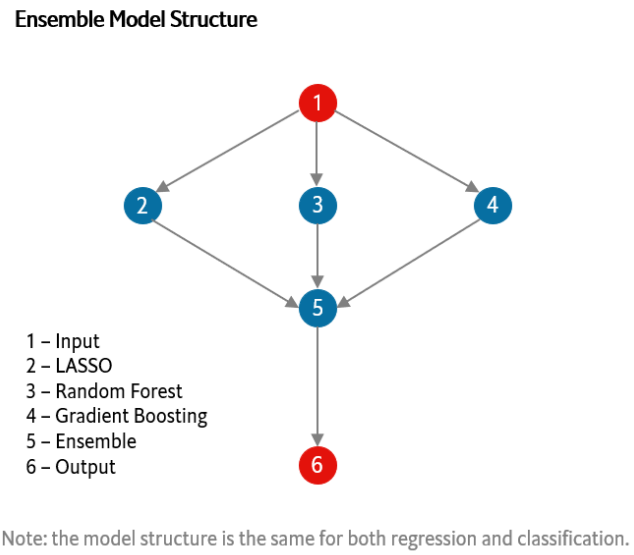

Figure 5