The Long-Term Earnings Impact of School Entry Cutoffs: Evidence from a Regression Discontinuity Design

Muhammad Shahryar

Department of Economics

Knox College

October, 2025

Abstract

This paper studies the long-run labor market outcomes (income/earnings) of school entry age policies by employing a sharp Regression Discontinuity Design (RDD) around the September cutoff in the US. Using longitudinal data from the Panel Study of Income Dynamics (PSID), I estimate the causal effect of delayed school entry on adult earnings. The analysis provides results for a standard Ordinary Least Squares (OLS) model, a fully interacted RDD specification model using the full sample, and restricted bandwidth models of ± 3 and ± 6 months around the cutoff boundary for robustness check. The findings show consistent and economically relevant consequences for the younger individuals, born just after the cutoff– those likely to delay their school entry. The full sample RDD specification model reveals a treatment effect of approximately 0.156 log points, which is significant at 5% level. This negative and highly statistically significant coefficient explains that individuals born after the cutoff earn approximately 15.6% less than their peers born just before the cutoff.

1 Introduction

The timing of school entry is an imperative but often ignored determinant of educational and labor market outcomes. In many countries around the world, including United States, institutional rules set by local education authorities say that children must reach a certain age – commonly 5 years old– by a specific cutoff date,September 1st in U.S case, to be eligible to enroll in kindergarten. Resultantly, children born just before the cutoff date start school a full year earlier than those born just after. These policies can sometimes create exogenous variation in school starting age, with little to no personal choice offered near the threshold.

This paper investigates how these cutoff-based differences in school starting age translate into different earnings in adulthood. To answer the research question; *Does delaying school entry, due solely to being born after the cutoff, reduce lifetime earnings potential?* I employ a sharp RDD method using detailed income and demographics data from PSID. This approach compares the individuals born just before and after the cutoff, to have a better shot at understanding the causality.

The theoretical motivation for writing this paper stems from the intersection of human capital framework (Becker 1964) and institutional policies. Heckman and Cunha's model of skill formation (2005, 2007) emphasize a critical role of early childhood investments in explaining adult outcomes. An earlier start may translate into higher earnings accumulation in future by facilitating an early entry into labor market. Alternatively, early entry may also harm children who are cognitively unprepared to start schooling which can lead to very serious behavioral issues, academic difficulties etc. The net effect is in itself a completely different empirical question.

This research largely draws from a rich literature on school entry age effects. Studies such as Angrist and Krueger (1991) and Bedard and Dhuey (2006) have shown that relatively older students in a similar cohort perform better on standardized test scores, are more likely to be in educational system for longer period, and enjoy higher earnings. However, many of these studies in the past rely heavily on reduced-form associations or base level econometric strategies (IVs etc.) with varying assumptions. Through using a sharp RDD method tightly around the cutoff boundary, I intend to improve the internal validity of this research.

As far as methodology is concerned, I estimate both the OLS model and RDD specifications with varying bandwidths around the September cutoff. The use of different bandwidths allowed me to test robustness and sensitivity of my findings – improving the internal validity by manipulation of narrower windows and wider samples. I include key covariates such as gender, parental education and household income to control for potential confounders and improve precision.

Though the narrower bandwidth specifications (± 3 and ± 6 months) give out statistically insignificant results and treatment effects– likely due to a limited sample size or reduced statistical power– the direction of coefficients remain highly consistent with the hypothesis. For instance, in the ± 3 month model, the treated individuals earn approximately **3.7%** less than the ones born just before the cutoff boundary (September 1st). Moreover, in the ± 6 month model the effect increases slightly to **5.3%** , yet still yields an insignificant result.

In summary, this paper provides robust causal evidence that school-entry cutoff policies can play an important role in determining the overall economic consequences for students. The discontinuities induced by rigid birthdate threhsolds can affect the timing of skill acquisition, which in turn translates to differential educational attainment, labor market performance and earnings in long-run.

The findings of this paper highlight the unintended costs of rigid institutional rules and call for a more appropriate approach towards school enrollment policies, ones which consider fair returns in lifetime across the cohorts.

Literature Review

Age-based school admission policies have long been the focus of educational economists, mainly as they relate to academic performance, long term career earnings and fairness in educational opportunity. A growing body of literature has consistently shown that the relative age of a child in a classroom cohort is associated with measurable impacts on (total) educational attainment, academic achievement and monetary returns in long run. This research draws on that body of work and uses Regression Discontinuity Design (RDD) approach to identify causal effects.

A central theme across the many studies on the same topic reveals that the older entrants to the same cohort settings tend to outperform their younger peers in both standardized tests and long-run educational outcomes. For instance, Black, Devereux, and Salvanes (2011) exploit a sharp cut-off in school entry laws of Norway and find that children who start school at an early age are significantly more likely to perform well and stay in the education system longer.

Building on the human capital framework developed by Becker (1964) which conceptualizes that education is an investment for an individual which enhances productivity and earnings later on, subsequent research has shown that early educational interventions greatly influence long term economic returns.For instance, Angrist and Krueger (1991) use quarter-of-birth as an instrument to quantify the effect of compulsory schooling laws on earnings. They highlight the importance of quasi-experimental techniques,and, while their methodology differs from a Sharp RDD, their work emphasizes the value of using natural experiments within institutional settings. Fertig and Kluve (2005) examine similar age-based effects on educational progress by using German data and find that earlier enrollment leads to lower test scores and behavioral difficulties. Similarly, Bedard and Dhuey (2006) find that relatively older students in Grade 4 tend to outperform their younger classmates.

One of the main themes that has unearthed in the existing literature is – school entry age does not just effect short-term test scores but also long term career trajectory and labor market outcomes (Black et al., 2011; Angrist & Krueger, 1991; Bedard & Dhuey, 2006).

As far as empirical methodology concerns, Imbens and Lemieux (2008) provide a foundational guide to RDD approach, specifically for economics. Their work emphasizes greatly on the use of RDD as an identification strategy and widely cited by many papers in applied economics' research. Additionally, applied econometrics methodological concepts such as IVs, RDD, have become a benchmark in identifying key causal relationships in educational economics (Card, 1999; Imbens & Lemieux, 2008)

This research contributes to the existing literature by using a sharp RDD to observational educational, income and demographics data by exploiting cutoffs in age-based school-entry rules of institutions. Unlike some prior papers on test scores and large administrative datasets, this paper will have a central focus on labor market outcomes by using a renowned

dataset of International level. The methodology will somewhat align with Black et al. (2011) and Sandra Black (1999) to ensure internal validity, with alternative bandwidths used to test for robustness of causal inference.

Data Description

The dataset employed in this research is the Panel Study of Income Dynamics (PSID), a nationally representative longitudnal survey conducted in the United States since 1968. It is administered for and updated by the University of Michigan. The PSID collects detailed and thorough information on individuals and households which contains data on education, demographics, income, employment, and family background characteristics.This is actually a panel data which updates the information every two years upon re-surveying of the existing sample and sometimes a new data points (N) are also introduced. It's rich structure makes the PSID very well-suited for studying long term outcomes and causality, especially while using a Regression Discontinuity Design (RDD).

For the purpose of this study, the PSID dataset is used for the years 1999 to 2021, and is utilized for constructing cohorts around the school-entry cutoff dates – targeted at explaining variation in age at school-entry based on birth-month. The dataset provides enough information on respondents' birth month and year, years of education, and labor market outcomes. Additionally, key covariates such as parental education, family income, and gender information are also included to allow for robustness check and internal validity purposes.

The main running variable is *rel_age_months*, which records the number of months relative to September 1st school-entry cutoff. The primary outcome variable is *log_income*, which is the natural log of the labor income measured for the years 2005–2021.

Moving on, the demographic and background characteristics variables include *gender* and date of birth, named *birth_month* and *birth_year*. These variables are used to create exact age and relative cutoff positioning around the September boundary. Additionally, family background is recorded using the early-life household income (*family_income_1999*). This family income variable serves as a proxy for the socioeconomic status prior to entry in labor

market. Parental education is recorded by variables: *father_edu_head, mother_edu_head.*

All in all, these variables give a robust application of the RDD method by constructing birth-based cohorts around a sharp cutoff boundary and allowing for control of crucial covariates that could affect income outcomes in lifetime.

Data Summary and Methodology

Overview of Methodology

This study analyzes whether relative age at school entry causally affects income levels in long run. Using a Septmber 1st cutoff for carrying this out for sample size of U.S, I employ a sharp Regression Discontinuity Design (RDD) where individuals are compared just before and after the cutoff. The discontinuity in the treatment condition is executed using a narrow window just around the cutof Different bandwidths (± 3 and ± 6 months) are used to further validate the results. The treatment indicator is defined as:

$$T_i = 1\{\text{rel_age_months}_i \geq 0\}$$

The indicator function $1\{\cdot\}$ returns a value of 1 if the condition inside the brackets is true, and 0 otherwise.

The running variable `rel_age_months`, measures birth month relative to the September cutoff. The outcome variable is log labor income (`log_income`) at the time individuals turn 45. Other control variables include (e.g., `father_edu_head`, `mother_edu_head`), gender, and baseline socioeconomic status measured by `family_income_1999`.

The baseline OLS model calculates the relationship between labor income (logged) and key individual and parental characteristics:

$$\begin{aligned}\text{log_income}_{it} = \beta_0 &+ \beta_1 \cdot \text{age}_{it} + \beta_2 \cdot \text{gender}_i + \beta_3 \cdot \text{father_edu_wife}_i \\ &+ \beta_4 \cdot \text{mother_edu_wife}_i + \beta_5 \cdot \text{father_edu_head}_i + \beta_6 \cdot \text{mother_edu_head}_i \\ &+ \beta_7 \cdot \text{family_income_1999}_i + \varepsilon_{it}\end{aligned} \tag{1}$$

The main empirical specification estimates a sharp RDD with an added interaction term between treatment condition and running variable.

$$\begin{aligned}\text{log_income}_{it} = \beta_0 &+ \beta_1 \cdot \text{treated}_i + \beta_2 \cdot \text{rel_age_months}_i \\ &+ \beta_3 \cdot (\text{treated}_i \times \text{rel_age_months}_i) + \beta_4 \cdot \text{gender}_i \\ &+ \beta_5 \cdot \text{father_edu_head}_i + \beta_6 \cdot \text{mother_edu_head}_i + \beta_7 \cdot \text{family_income_1999}_i + \varepsilon_{it}\end{aligned} \tag{2}$$

Summary Statistics

I construct two RDD regression models by restricting to bandwidths of ±3 and ±6 months around the September school-entry cutoff. Table 1 summarizes the log income distribution for individuals within each of these windows:

Table 1: Summary statistics for log income across bandwidths and full RDD sample.

Bandwidth	**N**	**Mean Log Income**	**SD Log Income**
±3	12174	1.8433	3.2540
±6	17114	1.8390	3.2519
Full Sample	43426	1.6467	3.1236

These statistics show that mean log income remains stable across bandwidths. Increasing the window size raises the sample size but does not significantly alter the central tendency or variability of log income.

In addition to the bandwidth-specific windows, Table 1 presents an overview of the full PSID dataset used in this study. The dataset contains 43,426 observations across 53 variables. The non-missing observation count (N) varies slightly across variables, typically ranging between 39,000 and 43,000.

Data Visualization

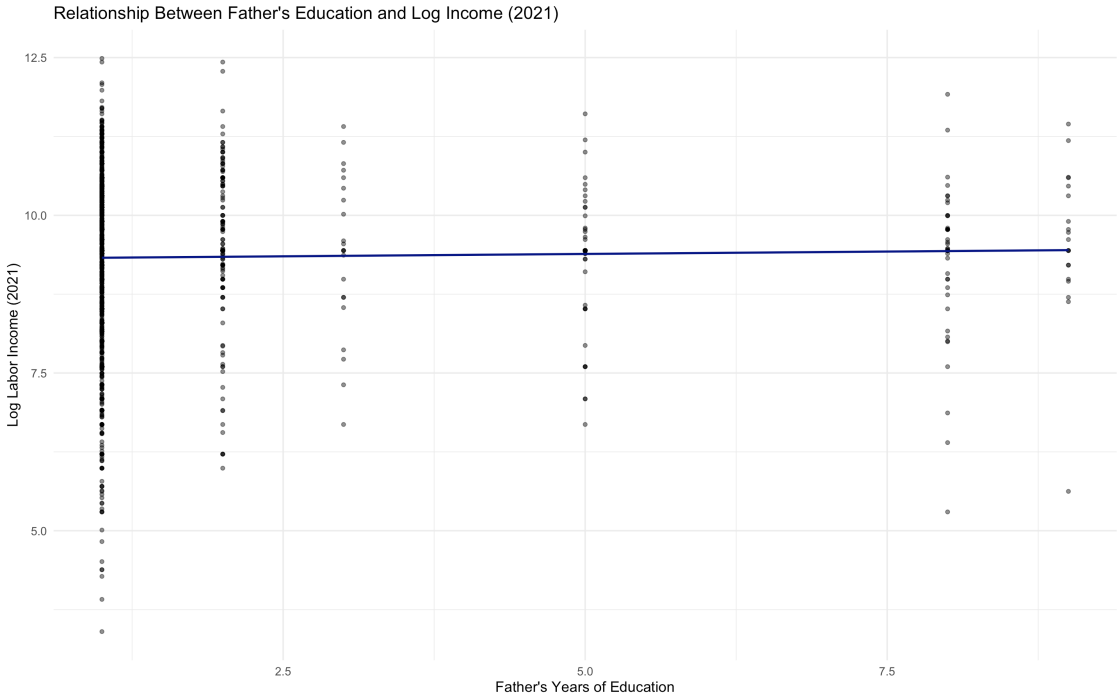


Figure 1: Relationship Between Father's Education and Log Income (2021)

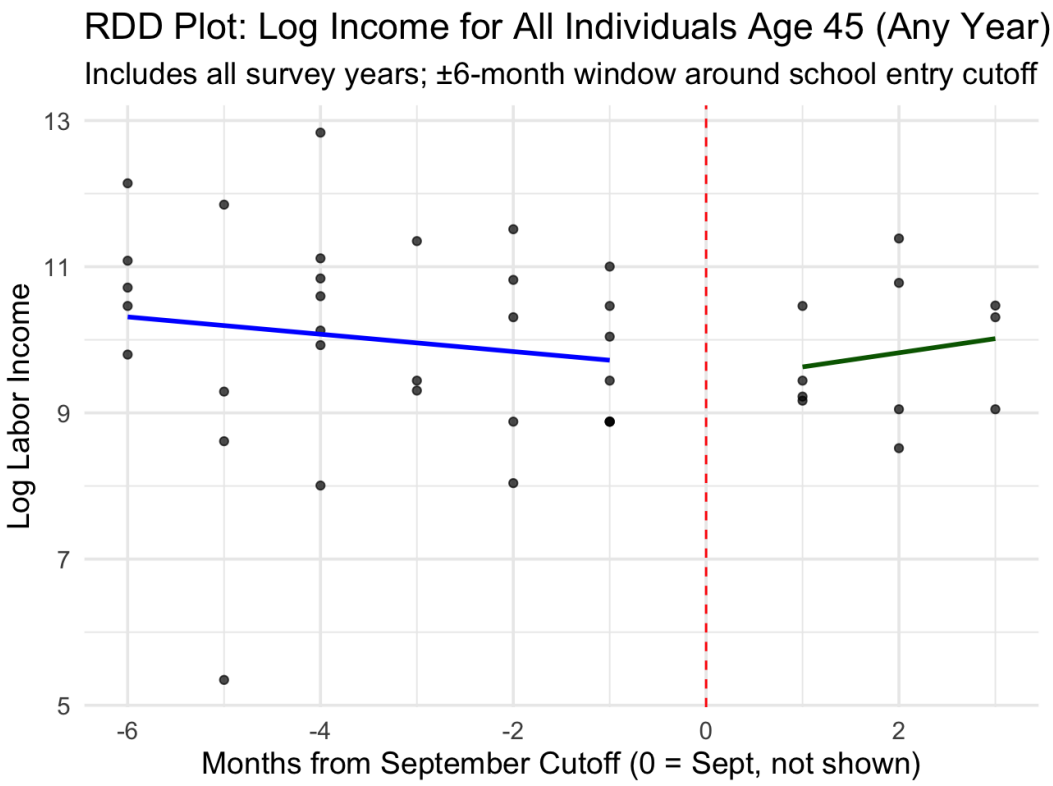


Figure 2: RDD Plot: Log Income for All Individuals Age 45 (Any Year), ±6mo Bandwidth

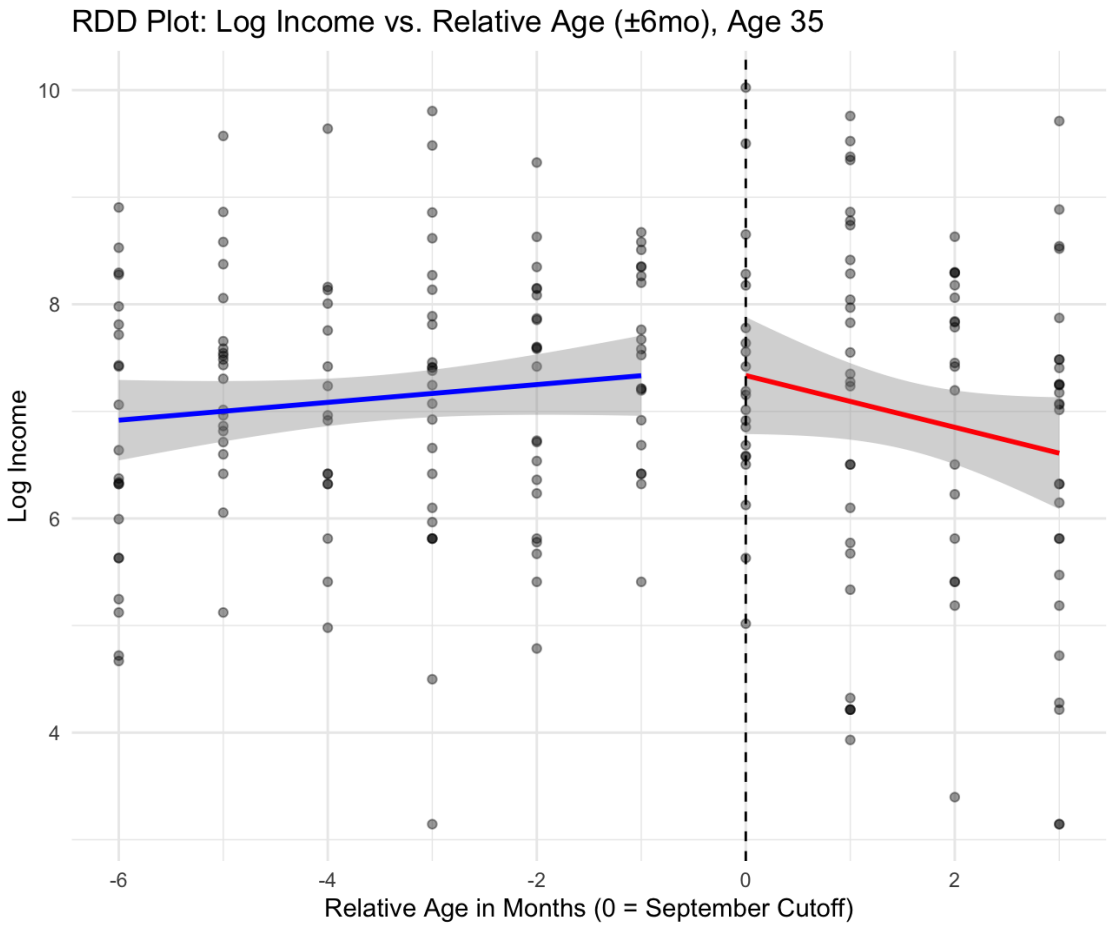


Figure 3: RDD Plot for Age 35: Log Income vs. Relative Age (±6 Months)

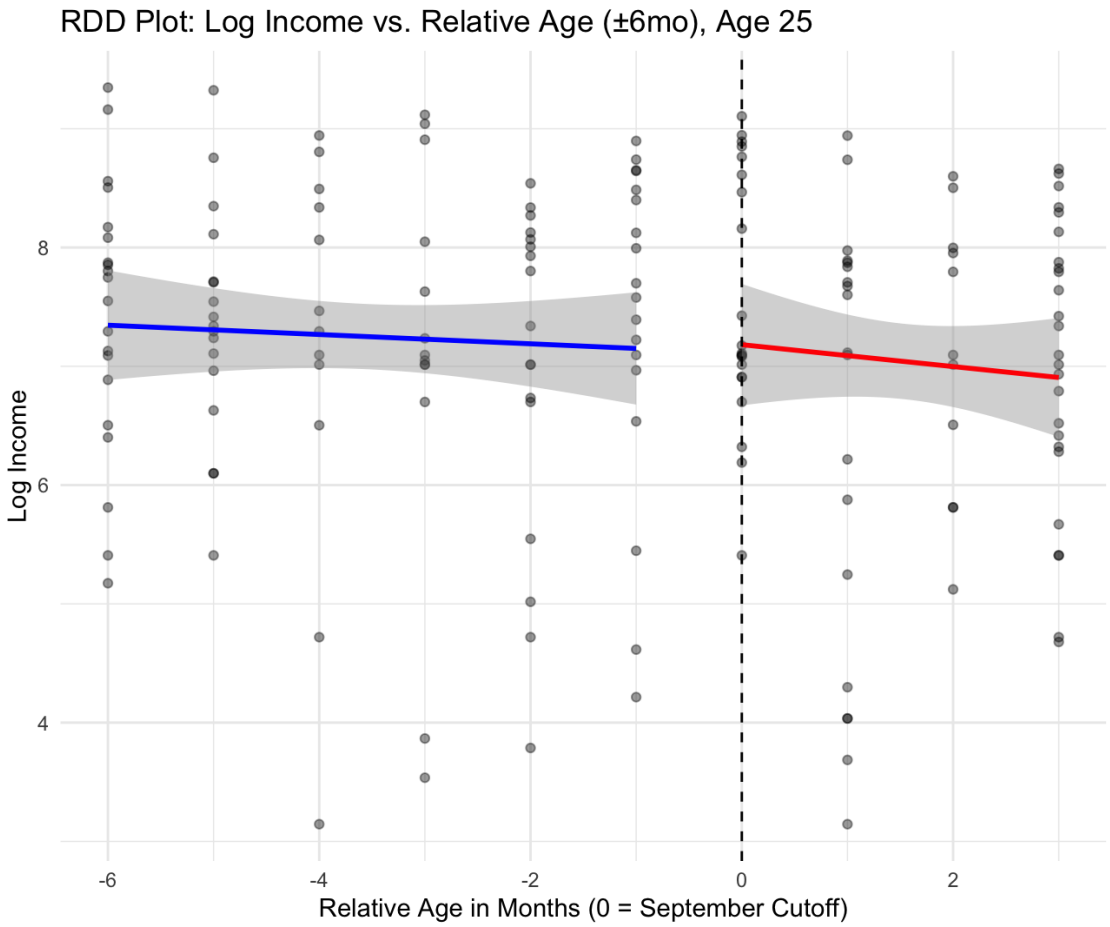


Figure 4: RDD Plot for Age 25: Log Income vs. Relative Age (±6 Months)

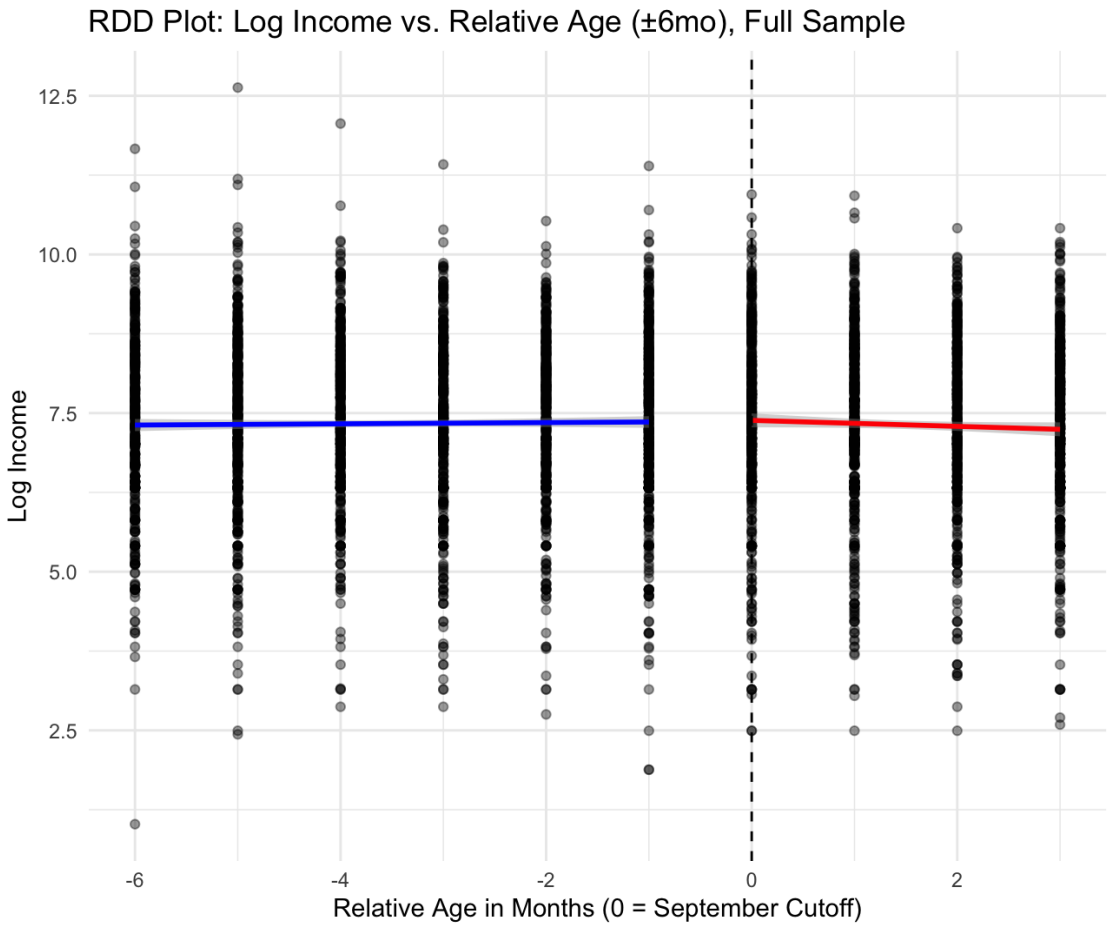


Figure 5: RDD Plot for Full Sample: Log Income vs. Relative Age (±6 Months)

To visually assess the presence of a discontinuity , I plot the log income against the running variable *(rel_age_months)*, keeping them all under a ±6 month bandwidth. These regression discontinuity plots cover the local linear fits on either side of the cutoff, with values deployed at the cut-off boundary (plotted as 0 on x-axis) as well.

Figure 1 displays the relationship between father's year of education and log income for the year 2021 (a starting point of visualizing). The figure shows a near-zero slope of fitted lines, suggesting that father's education is likely uncorrelated with adult earnings for this local window.

Additionally, Figure 2 plots log income for all individuals aged 45, regardless of the PSID survey year, within a ±6-month window around the September school-entry cutoff. A small upward shift at the cutoff provides some evidence that the discontinuity persists for this age group, but we still can't generalize this finding across all cohorts.

Figure 3 shows the results for individuals aged 35, which shows an upward trend and a notable downward trend immediately after the cutoff boundary. Additionally, Figure 4 shows the overall trend for our younger cohort and reveals a flat shift near the threshold. Finally, Figure 5 puts together all eligible individuals and plots their log income in accordnce with their relative age (to the September cutoff) , but there is no clear jump or shift after the boundary. All these plots were plotted by clearing out any zero values for log income of any individuals to make sure the analysis is only done on people who actually had some earnings.

Results and Interpretation

This section presents key findings from the RDD analysis to examine the causal impact of the age of entry into school on adult income, based on September cutoff. The results include findings from a standard Ordinary Least Squares (OLS) model, a fully interacted RDD model using the full sample, and two restricted bandwidth models of ± 3 and ± 6 months. The different bandwidth models are used to test for robustness.

1. Standard OLS Model

The baseline OLS model estimates the relationship between log-income and several covariates including including age, gender, parental education and household income. Even though, it's not a causal specification , it still provides a starting point into understanding the determinants of adult earnings.

In this specification, **age** is positively and significantly associated with log-income ($\beta = 0.00923$, $p = 0.0075$), suggesting increasing returns as the age increases. This is consistent and aligns with typical labor market trends where income rises with age and experience.

The **gender coefficient** for women is estimated at -0.03775, suggesting that females, on average, earn 3. 8% less than men. However, this effect is not statistically significant($p = 0.63$) and thus, should not be overstated. The absence of a robust gender effect here may be due to sample-specific noise, omitted variables, or broader labor force composition within the sample.

All in all, this model explains very little variation in the log-income ($R^2 = 0.034$), which highlights the cons of a simple OLS model while capturing complex economic dynamics. These results call for the use of more robust causal identification strategies, such as the Regression Discontinuity Design (RDD), which is implemented in the following sections.

Table 2: Baseline OLS Regression: Determinants of Log Income

	Estimate	Std. Error
(Intercept)	9.138e+00***	1.377e-01
age	9.231e-03**	3.442e-03
gender_factorFemale	-3.775e-02	7.812e-02
father_edu_wife	8.426e-03	3.968e-02
mother_edu_wife	3.303e-02	4.458e-02
father_edu_head	9.244e-03	3.163e-02
mother_edu_head	9.271e-03	3.757e-02
family_income_1999	3.094e-06***	9.267e-07
Observations	648	
Residual Std. Error	0.9756	
R-squared	0.03426	
Adjusted R-squared	0.02369	
F-statistic	3.243**	

Notes: Standard errors in parentheses. Significance levels: *** p<0.001, ** p<0.01, * p<0.05.

2. Full-Sample Interaction RDD Model

In the main specification, I added an interaction term between treatment condition and the running variable, *rel_age_months*, to analyze the heterogeneity effect as one moves away from the cutoff.

The interaction term is **negative and statistically significant** ($p < 0.01$),implying that the students born just after the cutoff– those who start school a year later– face a persistent earnings penalty. To be more specific, the 'youngest' in a cohort individuals earn approximately 15.6% less than their peers who were born before the cutoff. This is clearly in line with the existing literature that older students within the same cohort tend to outperform their younger classmates in long-run.

This comes as a compelling evidence that delayed school entry, due to rigid institutional cutoffs, translates into long-term economic and monetary disadvantages.

Table 3: Full-Sample Interaction RDD Model

	Estimate	**Std. Error**	**t value**	**Pr(¿|t|)**
(Intercept)	2.396e+00	7.257e-02	33.022	¡ 2e-16 ***
treated	-1.557e-01	6.227e-02	-2.501	0.01240 *
rel_age_months	5.070e-02	6.429e-03	8.753	¡ 2e-16 ***
gender	-2.551e-01	3.137e-02	-8.132	3.43e-16 ***
father_edu_head	2.467e-02	1.216e-02	2.080	0.03757 *
mother_edu_head	4.321e-02	1.499e-02	2.956	0.00312 **
family_income_1999	-9.207e-08	2.461e-07	-0.807	0.41978
treated:rel_age_months	-7.282e-02	7.753e-03	-9.392	¡ 2e-16 ***
Observations	39481			
Residual Std. Error	3.125	(on 39481 df)		
Multiple R-squared	0.004757			
Adjusted R-squared	0.004581			
F-statistic	26.96	(on 7 and 39481 df)		¡ 2.2e-16

Notes: Standard errors in parentheses. Significance levels: *** p<0.001, ** p<0.01, * p<0.05.

3. Bandwidth Robustness: ± 3 and ± 6 Months

In order to check for robustness and avoid any further contamination from broader cohorts, I restrict this estimation to narrower bandwidths around the September cutoff. The ± 3 months model results in a treated coefficient of -0.038, which is statistically insignificant ($p > 0.77$). Similarly, the interaction term is also not significant. However, both coefficients still retain their (negative) signs and direction, supporting the underlying causal effect.

Furthermore, in the ± 6 month window, the treated term is -0.053 and interaction term is positive at 0.0039, both statistically insignificant. These findings reflect weak statistical

power mainly due to sample restrictions and when we have less people to regress upon, the results are automatically less generalizable.

The main reason of using bandwidth specific estimates is that it highlights the importance of robustness checks in causal inference. A narrow window helps control for potential unobserved cohort-level confounding variables. However, none of these bandwidth windows provide any significant result, as far as causal inference is concerned. This suggests that being born just 3 or 6 months apart does not lead to clear or consistent differences in adult earnings.

Table 4: RDD Model: ±3-Month Bandwidth

	Estimate	Std. Error
(Intercept)	2.318e+00***	1.546e-01
treated	-3.773e-02	1.341e-01
rel_age_months	5.558e-03	5.555e-02
gender	-3.560e-01***	5.897e-02
father_edu_head	-2.496e-02	2.338e-02
mother_edu_head	4.432e-02**	2.857e-02
family_income_1999	4.007e-08	4.285e-07
treated:rel_age_months	8.278e-03	6.556e-02
Observations	12164	
Residual Std. Error	3.249	
R-squared	0.003808	
Adjusted R-squared	0.003235	
F-statistic	6.642**	

Notes: Standard errors in parentheses. Significance levels: *** p<0.001, ** p<0.01, * p<0.05.

Table 5: RDD Model: ±6-Month Bandwidth

	Estimate	Std. Error
(Intercept)	2.392e+00***	1.110e-01
treated	-5.259e-02	9.659e-02
rel_age_months	1.864e-02	4.541e-02
gender	-3.686e-01***	4.969e-02
father_edu_head	-1.656e-02*	1.988e-02
mother_edu_head	8.669e-02***	2.416e-02
family_income_1999	1.795e-07	3.795e-07
treated:rel_age_months	3.965e-03	3.947e-02
Observations	17104	
Residual Std. Error	3.246	
R-squared	0.004037	
Adjusted R-squared	0.003629	
F-statistic	9.904***	

Notes: Standard errors in parentheses. Significance levels: *** p<0.001, ** p<0.01, * p<0.05.

Conclusion and Policy Implications

This study provides somewhat limited, yet suggestive evidence that inflexible school-entry cutoffs by institutions, may have adverse economic effects for students in long-run . The most interesting evidence comes from the full-sample interaction model, which reveals a statistically significant and economically meaningful earnings penalty for individuals who start their schooling journey by a year's delay, just because their birthdate fell just after the cutoff. Specifically, these individuals, despite being almost developmentally ready have to wait a full year to begin their education which leads to late exposure to job market.

Essentially, this finding aligns with the existing literature, re-iterating on the evidence that relative age within a cohort plays a crucial role in defining the long-term economic and educational trajectories. For instance, older students in the same 'classroom' tend to perform better academically and financially. However, it is imperative to note that this evidence of (statistical) significance was only observed in the full sample interaction model . The absence of similar significant effects in the narrower ± 3 and ± 6 month bandwidth windows weakens the case for a strong causal relationship between variables, hence, limiting the robustness and generalizability of the results.

Even though, the full sample result is significant and striking, it still does not justify any sweeping policy changes. The lack of consistent statistical significant across all specifications calls for extra controls and caution. It might be possible that the full sample findings only reflect broader cohorts, unobserved heterogeneity, or random variation, rather than a clean local treatment effect.

From a policy standpoint, my findings suggest no immediate or large-scale effect right away. Still, the evidence justifies discussion around flexibility of school-entry criterion.

Potential policy responses could include offering of early assessment tools to determine school readiness of children regardless of birth month. Similarly, pre-K or bridging programs could also be introduced for children who just miss out on cutoffs. In addition, parents could obtain early (tentative) enrollment just to have their kids familiarize with the schooling environment. However, this suggestion is only encouraged to implement for children of borderline age, those very close to 5 years of age.

In summary, this study emphasizes the need for rethinking and redefining the "one-size-fits-all" school-entry policy in favor of approaches that accommodate some variation in child development. Such flexibility in institutional rules can yield collective benefits for both individuals and society.

Acknowledgments and Disclosure

The author thanks faculty at Knox College for helpful comments and guidance during the development of this research. This project received no external financial support. The author declares no financial interests, consulting arrangements, data access restrictions, or other conflicts of interest related to this study.

References

- Angrist, J. D., & Krueger, A. B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics*, 106(4), 979–1014.
- Bedard, K., & Dhuey, E. (2006). The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. *Economics of Education Review*, 25(2), 143–154.
- Black, S. E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *Quarterly Journal of Economics*, 114(2), 577–599.
- Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011). Too Young to Leave the Nest? The Effects of School Starting Age. *Review of Economics and Statistics*, 93(2), 455–467.
- Card, D. (1999). The Causal Effect of Education on Earnings. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 3, pp. 1801–1863). Elsevier.
- Fertig, M., & Kluve, J. (2005). The Effect of Age at School Entry on Educational Attainment in Germany. *IZA Discussion Paper No. 1507*.
- Imbens, G. W., & Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142(2), 615–635.
- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press.
- Cunha, F., & Heckman, J. J. (2007). The Technology of Skill Formation. *American Economic Review*, 97(2), 31–47.
- Cunha, F., Heckman, J. J., Lochner, L, & Masterov, D. V. (2006). Interpreting the Evidence on Life Cycle Skill Formation. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education* (Vol. 1, pp. 697–812). Elsevier.