Proceedings of the 89th Annual International Meeting of the Psychometric Society, Prague, Czech Republic (2024).

Assessing the marginal effect under partial observability in a selection context

Eduardo Alarcón-Bustamante,[*†‡¶§] Pablo Espinoza,[†] David Torres Irribarra,[§‡¶] and María Inés Godoy[†]

†Departamento de Evaluación, Medición y Registro Educacional (DEMRE), Universidad de Chile, Santiago de Chile, Chile
‡Interdisciplinary Laboratory of Social Statistics, Santiago de Chile, Chile
¶Millennium Nucleus on Intergenerational Mobility: From Modelling to Policy (MOVI), Santiago de Chile, Chile
§Escuela de Psicología, Pontificia Universidad Católica de Chile, Santiago de Chile, Chile
*Corresponding author. Email: eduardoalarcon@uchile.cl

Abstract

The predictive capacity of admission tests depends on their ability to anticipate academic performance, which is directly related to the marginal effect of test scores. However, in selection contexts, this effect is not identifiable since academic performance is only observed for admitted students. In this paper, we first propose a decomposition of the marginal effect based on the law of total probability, distinguishing a within-group effect, which measures how academic performance varies with test scores within each program, and a between-group effect, which captures how test scores explain differences in average predictions across programs and influence admission probabilities. We then propose identification bounds for the marginal effect based on contextual assumptions about the admission system.

Keywords: missing at random, predictive validity, ignorability, academic performance, informative assumptions

1. Introduction

Understanding the predictive capacity of university admission tests is crucial, as it reveals how strongly test scores are related to future academic performance. This relationship is captured by the marginal effect of scores on university performance: if changes in test scores lead to substantial changes in university performance, the test is considered to have a high predictive capacity; conversely, if the relationship is weak, the test has a low ability to differentiate future performance of students (see for instance Geiser & Studley, 2002; Goldhaber et al., 2017).

When the student population can be partitioned into meaningful groups – for instance, by academic program, institution, or demographic characteristics – the predictive capacity of a test is typically assessed in two ways: by comparing correlation or regression coefficients – the marginal effect in linear regression models – across groups (see for instance Ayers & Peters, 1977; Grobelny, 2018; Manzi & Carrasco, 2021). However, these approaches can obscure relevant information, as they do not explicitly account for the structure of groups, such as the selectivity levels of academic programs or the profiles of admitted students. To address this limitation, this study proposes a decomposition of the marginal effect based on the law of total probability. This decomposition distinguishes two components: the within-group effect, which measures how academic performance varies with test scores within each program, and the between-group effect, which captures how test scores explain differences in average predictions across programs and influence the probability of admission to each one.

Due to the lack of observability of the academic performance of non-selected applicants, estimating the marginal effect in this context presents a challenge: while test scores are available for the entire pool of applicants, academic performance is only observed for those who were admitted to the university. From a statistical modeling perspective, the primary consequence of this missing information is that the joint distribution of academic performance and test scores in the applicant population cannot be fully known. As a result, the conditional distribution of academic performance given test scores is not identifiable. Consequently, any parameter of interest related to this distribution will not be identifiable without making any assumption (Koopmans, 1949; Manski, 1993). In particular, since the conditional expectation of academic performance given test scores is not identified, any function of this expectation will also be unidentified. Thus, its derivative with respect to test scores, which defines the marginal effect, cannot be identified.

A common practice in empirical research is to work only with observed data, which is equivalent to assuming that the missing data mechanism follows a missing at random process, also known as the ignorability assumption (Florens & Mouchart, 1982; Hirano & Imbens, 2004; Imbens, 2000; Manski, 2003; Rosenbaum & Rubin, 1983). This assumption implies that the conditional expectation of academic performance given test scores is identical for both selected and non-selected applicants. If this restriction is assumed, the conditional expectation of academic performance given test scores in the whole population of applicants can be point-identified, which in turn allows for the identification of the marginal effect. However, in a selection context, this assumption is unrealistic, as applicants are selected precisely based on the belief that non-selected applicants would have performed worse than those who were admitted (Alarcón-Bustamante et al., 2025; Grassau, 1956). Instead of imposing strong assumptions to recover the point estimate of the marginal effect, this study employs a partial identification approach. This approach does not seek to find a single value for the parameter of interest but rather to bound its possible range based on weaker assumptions and observed data (Manski, 1993; Tamer, 2010). Specifically, we obtain identification bounds for the marginal effect that provide valuable information about its magnitude and direction without assuming a priori that non-selected applicants would behave the same as selected ones.

To illustrate the proposed approach, we use admission data from applicants to three programs within the School of Biology at a university in Chile: Marine Biology, Biology, and Biochemistry. We focus exclusively on students admitted through the regular selection process, in which applicants are selected based on a linear combination of admission test scores, high school GPA, and high school ranking. As a result, two applicants with the same Mathematics test score, for instance, may have different admission outcomes due to differences in their prior academic performance. For example, one applicant may be selected, and therefore we can observe their performance, despite having the same score on this test as another applicant who was not selected, whose performance remains unobserved. The dataset we use includes 289 observations, of which 125 correspond to selected students who completed their first year in at least one program, and 164 to non-selected applicants. In this study, we specifically focus on the Mathematics test score, which ranges from 150 to 850 points, and on academic performance measured through the first-year university GPA, on a 1.0 to 7.0 scale, where 4.0 is the minimum passing grade.

The paper is organized as follows: Section 2 formally presents the identification problem from a modeling perspective. Section 3 introduces the decomposition of the marginal effect under the ignorability assumption, providing a detailed interpretation of both the within-group and between-group effects. Section 4 presents identification bounds for these effects under weaker assumptions than ignorability, allowing for the construction of identification regions based on the context of the selection problem. This section also includes the results of the estimated bounds along with a detailed interpretation. Finally, Section 5 discusses the study's conclusions and provides a broader discussion.

2. Modeling framework and identification issues

Let M be the sample space of all applicants to the School of Biology. For each $m \in M$, we define the following random variables: $X(m)$ the scores of the m^{th} applicant in the Mathematics selection test; $Y(m)$ the GPA of the m^{th} applicant, and $Z(m) = 1$ the m^{th} applicant is selected, and 0 otherwise. By the law of total probability (Kolmogorov, 1950) the conditional expectation $E(Y \mid X)$ can be decomposed as

$$E(Y \mid X) = E(Y \mid X, Z = 1)P(Z = 1 \mid X) + E(Y \mid X, Z = 0)P(Z = 0 \mid X). \tag{1}$$

In (1), $E(Y \mid X, Z = 1)$ is the expected GPA, conditional on the scores for those selected applicants, and $P(Z = 1 \mid X)$ is the proportion of selected applicants given the scores. $E(Y \mid X, Z = 0)$ is the expected GPA, conditional on the scores for those non-selected applicants, which is, in fact, impossible to be estimated because the GPA is not observed for those non-selected applicants. Thus, $E(Y \mid X)$ is said to be non-identified. The marginal effect is given by the derivative of (1), with respect to X, i.e.,

$$\begin{aligned}\frac{dE(Y \mid X)}{dX} = & \frac{dE(Y \mid X, Z = 1)}{dX}P(Z = 1 \mid X) + E(Y \mid X, Z = 1)\frac{dP(Z = 1 \mid X)}{dX} + \\ & \frac{dE(Y \mid X, Z = 0)}{dX}P(Z = 0 \mid X) + E(Y \mid X, Z = 0)\frac{dP(Z = 0 \mid X)}{dX}.\end{aligned} \tag{2}$$

In (2), $\frac{dE(Y|X,Z=0)}{dX}$ represents the effect of test scores on the performance of non-selected applicants. However, this effect cannot be estimated because academic performance is not observed. If selection is assumed to be conditionally independent of academic performance given the test scores, then $E(Y \mid X, Z = 1) = E(Y \mid X, Z = 0)$. This assumption acts as an identification restriction (San Martín & González, 2022), allowing the expected academic performance of selected applicants to serve as a valid proxy for that of non-selected applicants. Thus, under this restriction, $E(Y \mid X)$ becomes point-identified, as it is equal to $E(Y \mid X, Z = 1)$. This allows for the direct estimation of the marginal effect without ambiguity. Now, let us incorporate information about the applicants' selection status. We introduce a new random variable, $G : M \to 0, 1, \ldots, L$, where $G(m) = g$ if applicant m was selected for program g ($g = 1, \ldots, L$), and $G(m) = 0$ if they were not selected. Given the previously defined role of Z, this definition implies that the set of units $m \in M$ for which $Z(m) = 0$ —i.e., those with unobserved outcomes— is equivalent to the set of non-selected applicants, namely those for whom $G(m) = 0$. On the other hand, the set of units with $Z(m) = 1$ —i.e., those with observed outcomes— corresponds to the union of all individuals selected into one of the L programs, that is, those for whom $G(m) = g$ for $g = 1, \ldots, L$. For our data, we set $L = 3$. Without loss of generality, we define $G(m) = 1$, $G(m) = 2$, or $G(m) = 3$ if they are selected for the Marine Biology, Biology, or Biochemistry program, respectively. Thus,

$$E(Y \mid X) = E(Y \mid X, G = 0)P(G = 0 \mid X) + \sum_{g=1}^{3} E(Y \mid X, G = g)P(G = g \mid X). \tag{3}$$

Taking the derivative with respect to X in (3), it is obtained that:

$$\begin{aligned}\frac{dE(Y \mid X)}{dX} = & \frac{dE(Y \mid X, G = 0)}{dX}P(G = 0 \mid X) + \sum_{g=1}^{3} \frac{dE(Y \mid X, G = g)}{dX}P(G = g \mid X) + \\ & E(Y \mid X, G = 0)\frac{dP(G = 0 \mid X)}{dX} + \sum_{g=1}^{3} E(Y \mid X, G = g)\frac{dP(G = g \mid X)}{dX}.\end{aligned} \tag{4}$$

Although we have incorporated information about the applicants, the marginal effect is not identified because it depends on both effect of test scores on university performance and the performance in the non-selected population, namely $G = 0$.

2.1 Partial identifiability of the Marginal Effect

In many cases, the support of Y is inherently bounded by definition. In general, if Y is restricted to the interval $[y_0(X), y_1(X)]$, where $y_0(X)$ and $y_1(X)$ are two functions of X, such that $y_0(X) \leq y_1(X)$, the conditional expectation in those non-selected applicants must also satisfy

$$y_0(X) \leq E(Y \mid X, G = 0) \leq y_1(X).$$

For example, in Chile, GPA Y is bounded between 1.0 and 7.0 (two constant functions of X), implying that $1.0 \leq E(Y \mid X, G = 0) \leq 7.0$. Although the exact value of $E(Y \mid X, G = 0)$ is unknown due to the lack of observed data for non-selected applicants, the bounded nature of Y ensures that its conditional expectation must lie within this interval. This assumption is reasonable in settings where the outcome variable has a known and fixed support, in which case the boundedness of the unobserved conditional expectation becomes a direct implication of the known range of Y. However, this is not universally applicable. For instance, when IRT models are used, sometimes it is of interest to estimate the conditional expectation of the ability parameter given background covariates (e.g., comparing mathematics abilities of students conditional on sex or type of school). If some individuals did not answer the test, their ability remains unobservable and cannot be estimated. In this case, even if we attempt to impose bounds on the non observed conditional expectation, this is not theoretically justifiable, since the latent ability parameter has an unbounded support by definition. In our case, the boundedness is justified both theoretically and empirically. Nevertheless, bounding the outcome variable Y is not sufficient to restrict the derivative of the unobserved conditional expectation, with respect to X (Manski, 1989). In fact, the derivative of the unobserved conditional expectation is not necessarily bounded. In such cases, even small changes in X could induce extremely large variations in the outcome variable, causing the slope of the function to tend toward infinity. This poses a substantive challenge for the partial identification approach, as it implies that the marginal effect may be unbounded. To address this, in this paper we propose a set of contextually grounded assumptions that aim to restrict the plausible range of the derivative, thereby enabling informative bounds on the marginal effect. In practical terms, bounding the marginal effect requires formulating assumptions under which the following inequality holds:

$$D_0(X) \leq \frac{dE(Y \mid X, G = 0)}{dX} \leq D_1(X),$$

where $D_0(X)$ and $D_1(X)$ are two functions of X, satisfying $D_0(X) \leq D_1(X)$. These functions encode plausible lower and upper bounds on the derivative of the unobserved conditional expectation, based on contextual knowledge of how X influences the outcome. These bounds rely jointly on beliefs concerning both $E(Y \mid X, G = 0)$ and its derivative. Taken together, these assumptions enable the computation of partial identification bounds as expressed in (5) as follows:

$$\frac{dE(Y \mid X)}{dX} \in \Bigg[D_0(X)P(G=0 \mid X) + \sum_{g=1}^{3} \frac{dE(Y \mid X, G=g)}{dX} P(G=g \mid X) + \gamma_0(X)\frac{d^+P(G=0 \mid X)}{dX} - \gamma_1(X)\frac{d^-P(G=0 \mid X)}{dX} + \sum_{g=1}^{3} E(Y \mid X, G=g)\frac{dP(G=g \mid X)}{dX},$$
$$D_1(X)P(G=0 \mid X) + \sum_{g=1}^{3} \frac{dE(Y \mid X, G=g)}{dX} P(G=g \mid X) + \gamma_1(X)\frac{d^+P(G=0 \mid X)}{dX} - \gamma_0(X)\frac{d^-P(G=0 \mid X)}{dX} + \sum_{g=1}^{3} E(Y \mid X, G=g)\frac{dP(G=g \mid X)}{dX} \Bigg] \quad (5)$$

(see a proof in Appendix 1). Here, $\frac{d^+P(G=0|X)}{dX}$ and $\frac{d^-P(G=0|X)}{dX}$ are the positive and negative part of $\frac{dP(G=0|X)}{dX}$, respectively. The length $L(X)$, of the interval is given by:

$$L(X) = [D_0(X) - D_1(X)]P(G=0 \mid X) + [\gamma_1(X) - \gamma_0(X)] \left| \frac{dP(G=0 \mid X)}{dX} \right|$$

(see a proof in Appendix 2). These bounds provide a structured way to quantify the lack of information in the marginal effect by integrating information about the support of Y and reasonable assumptions about how $E(Y \mid X, G=0)$ evolves with X. Manski (1989) propose identification bounds for this quantity by assuming that the probability of observing the outcome is an increasing function of X. In Alarcón-Bustamante et al. (2020) identification bounds for the marginal effect were built by using monotonicity assumptions. Alarcón-Bustamante et al. (2023) propose to use the results of Stoye (2007) to define identification bounds for the regression coefficients in a linear regression model to study the predictive validity of selection factors.

In what follows, we analyze the marginal effect of math scores on academic performance under the assumption of ignorability, which not only allows for the point identification of this effect but also provides a clear framework for interpreting the relationship between math scores and academic performance among applicants who were selected into the system. We then explore different scenarios for non-selected applicants to extend the analysis beyond the observed population and consider how the effect of math scores would manifest in the entire pool of applicants. While the true effect in the absence of ignorability cannot be known with certainty, these scenarios will allow us to establish partial identification bounds that delineate a plausible range for the marginal effect without requiring the assumption of conditional independence.

3. The marginal effect under ignorability

Under the ignorability assumption, we have that $E(Y|X) = E(Y|X, Z=1)$. Thus, the marginal effect is given by

$$\frac{dE(Y|X)}{dX} = \frac{dE(Y|X, Z=1)}{dX}.$$

For our data, the GPAs are observed in different programs, namely $G = g \in \{1, 2, 3\}$. Thus,

$$E(Y|X, Z=1) = \sum_{g=1}^{3} E(Y|X, Z=1, G=g)P(G=g|X, Z=1). \quad (6)$$

Taking the derivative with respect to X in Equation (6), and noting that under the ignorability assumption we only have information about the selected applicants, for whom $\sum_{g=1}^{3} P(G = g \mid X, Z = 1) = 1$, we obtain:

$$\begin{aligned}\frac{dE(Y|X, Z = 1)}{dX} &= \sum_{i} \frac{dE(Y|X, Z = 1, G = g)}{dX} P(G = g|X, Z = 1) \\ &+ \sum_{g \neq g'} [E(Y|X, Z = 1, G = g) - E(Y|X, Z = 1, G = g')] \frac{dP(G = g|X, Z = 1)}{dX}.\end{aligned}$$

In Alarcón-Bustamante et al. (2021), this quantity is referred to as the *Global Marginal Effect*. The authors show that its value remains unchanged regardless of the chosen reference group, $G = g'$. Therefore, the global marginal effect possesses an invariance property.

We have decomposed the global marginal effect, isolating the impact of test scores into two distinct components, namely:

- **Within-group effect**

$$W(X) = \sum_{g} \frac{dE(Y|X, Z = 1, G = g)}{dX} P(G = g|X, Z = 1);$$

  it measures how academic performance varies with test scores within each program. This within-group effect is manifested in the average weighted effect trough the score scale, where the weights are given by the proportion of selected applicants in each program.

- **Between-group effect**,

$$B(X) = \sum_{g \neq g'} [E(Y|X, Z = 1, G = g) - E(Y|X, Z = 1, G = g')] \frac{dP(G = g|X, Z = 1)}{dX};$$

  it captures how test scores explain differences in average predictions across groups and how test scores affect the probability of observing a student in one program or another.

3.1 Results under ignorability

In this section, we present the results of this decomposition, highlighting the relative contribution of $W(X)$ and $B(X)$ to the prediction of academic performance, always within the framework of the ignorability assumption. To estimate the conditional expectations involved, we employed a linear model, while the required probabilities were estimated using a multinomial regression. Although we used linear regression for specifying the conditional expectations, which may not fully align with the nature of the data, this choice was made purely for illustrative purposes.

Figure 1 illustrates the relationship between math scores, academic performance, and the distribution of students across programs. The left panel shows that within each program, there is a positive relationship between admission scores and GPA, indicating a within-group effect: on average, higher scores are associated with higher academic performance. The plot in the central panel reveals that the probability of belonging to each program changes with math scores. As scores increase, students are more likely to belong to certain programs and less likely to be in others, suggesting the presence

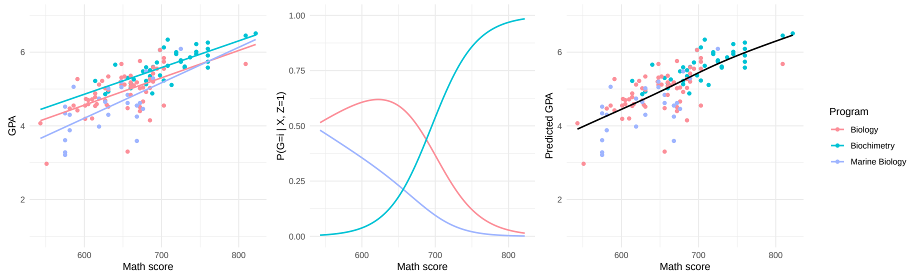

Figure 1. Results under ignorability

of a between-group effect. This redistribution of students across programs contributes to the overall effect of math scores on GPA, as it not only affects performance within each program but also alters the composition of students observed at different score levels. Moreover, since $B(X)$ depends on the differences in predicted values across programs, it is important to note that as scores increase, these differences tend to decrease. This suggests that at the higher end of the scale, the variability in academic performance across programs diminishes, moderating the impact of the between-group effect. In particular, this structure of student allocation across programs directly influences the shape of the overall relationship between test scores and academic performance. As observed in the right panel plot, the result of the law of total probability produces a nonlinear relationship, even though the regressions within each program were modeled as linear. This nonlinearity becomes more evident at the upper end of the scale, where the growth of GPA with respect to math scores slows down. This occurs because students with higher scores are more likely to belong to programs where the slope of the relationship between test scores and performance is lower. As a result, even though the relationship within each program is linear, the combination of these effects at the aggregate level introduces curvature in the predicted academic performance.

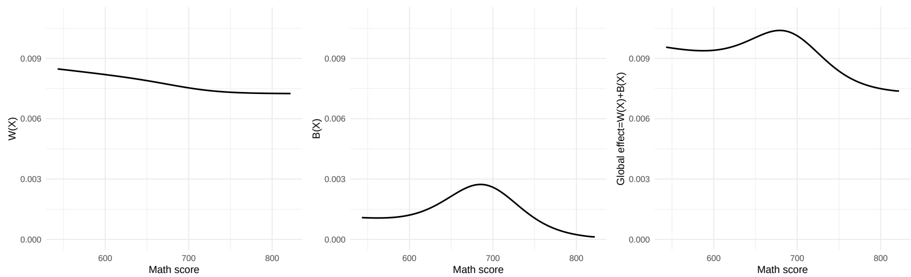

Figure 2. Results under ignorability: $W(X)$ in the left panel; $B(X)$ in the central panel; Global Effect in the right panel.

In Figure 2, the results of computing within and between group effect are shown. The within-group effect, $W(X)$, decreases as X increases, indicating that the relationship between test scores and academic performance within each program is stronger at lower score levels but does not vanish at higher levels. It is important to note that this interpretation is on average, as $W(X)$ is a weighted combination of the within-program effects, reflecting the contribution of each group according to the distribution of students.

The between-group effect, $B(X)$, captures how the distribution of students across programs influences the prediction of academic performance. For lower scores, differences in expected values between programs are more pronounced, contributing to a positive $B(X)$. Additionally, in this region, students are primarily distributed between two programs (Biology and Marine Biology), allowing for greater variability in program assignment and, consequently, a stronger impact of changes in composition. In the middle range, where students are more evenly distributed across all three programs, $B(X)$ reaches its peak, reflecting the highest variability in student distribution and performance differences across programs. However, at the upper end of the scale, the probability of observing students in programs other than biochemistry declines considerably, reducing the contribution of the between-group effect. Furthermore, in this region, the differences in expected values across programs nearly vanish, further attenuating $B(X)$. This implies that at higher score levels, the impact of student redistribution across programs is near to be zero, as the composition becomes more homogeneous and expected performance levels converge.

The combination of both effects, $W(X) + B(X)$, results in a nonlinear relationship where the global impact of math scores is highest in the middle range of the scale and decreases at higher values. This indicates that, although math scores influence academic performance at all levels, their intensity varies depending on the interaction between within-group and between-group effects, with greater sensitivity in the middle range and a progressive decline at the extremes of the scale.

It is important to highlight that the within-group effect, $W(X)$, consistently maintains a higher magnitude across the entire scale, whereas the between-group effect, $B(X)$, reaches its peak in the middle of the distribution but never surpasses $W(X)$. This suggests that, in our data, the relationship between test scores and academic performance is stronger within programs than through changes in the composition of students across programs. While the assignment of students between programs contributes to performance predictions, its impact remains more limited compared to the differences observed within each program. Therefore, math scores are a better predictor of academic performance within each program than through their influence on the distribution of students across programs. This suggests that variability in academic performance is primarily driven by differences within programs, while changes in student composition based on test scores have a more limited effect.

The analysis under the ignorability assumption provided a complete interpretation of the global effect of math scores on academic performance using only information from selected applicants. This outcome follows directly from the assumption that selection is conditionally independent of performance given test scores, allowing the estimation to be conducted solely within the observed group. However, this restriction raises an important question: what assumptions can be made about the effect of X among non-selected applicants in order to gain insights into the predictive capacity of the test for the entire applicant population? Addressing this question requires exploring alternative identification strategies that extend beyond the observed population. In the next section, we examine such alternatives and propose partial identification bounds that characterize the plausible range of the marginal effect. These assumptions are not imposed arbitrarily as abstract mathematical conditions but are instead constructed based on the selection problem and the specific context in which the admissions process operates.

4. Bounding the Within and Between groups effect

Rewriting Equation (4) in an equivalent form, and noting that $\sum_{g=0}^{3} P(G = g \mid X) = 1$, we obtain that the global marginal effect can be written as follows:

$$\begin{aligned}\frac{dE(Y \mid X)}{dX} &= \frac{dE(Y \mid X, G = 0)}{dX} P(G = 0 \mid X) + \sum_{g=1}^{3} \frac{dE(Y \mid X, G = g)}{dX} P(G = g \mid X) + \\ &\sum_{g=1}^{3} \left[E(Y \mid X, G = g) - E(Y \mid X, G = 0)\right] \frac{dP(G = g \mid X)}{dX}.\end{aligned}$$

The decomposition of the marginal effect remains evident in this formulation, where the within-group effect is given by

$$W(X) = \frac{dE(Y \mid X, G = 0)}{dX} P(G = 0 \mid X) + \sum_{g=1}^{3} \frac{dE(Y \mid X, G = g)}{dX} P(G = g \mid X),$$

which is not identified since it depends on the impact of X on academic performance for non-selected applicants. However, if it is assumed that $D_0(X) \leq \frac{dE(Y|X,G=0)}{dX} \leq D_1(X)$, then

$$\begin{aligned} W(X) \quad \in \quad & \Bigg[D_0(X)P(G = 0 \mid X) + \sum_{g=1}^{3} \frac{dE(Y \mid X, G = g)}{dX} P(G = g \mid X); \\ & D_1(X)P(G = 0 \mid X) + \sum_{g=1}^{3} \frac{dE(Y \mid X, G = g)}{dX} P(G = g \mid X) \Bigg]. \end{aligned} \tag{7}$$

The width of this bound is $L_W(X) = [D_0(X) - D_1(X)]P(G = 0 \mid X)$, and quantify how the lack of information about non-selected applicants affects the estimation of $W(X)$. Its magnitude depends on the difference $D_0(X) - D_1(X)$, which captures the uncertainty about how the expected performance of non-selected applicants varies as a function of the mathematics score, and on the probability $P(G = 0 \mid X)$, which indicates the proportion of students with score X who were not selected. If this probability is high or the difference in derivatives is large, the identification bound widens, reflecting a greater impact of missing information on the estimation. Conversely, if both values are small, the bound narrows, allowing for better precision in estimating the within-group marginal effect. Therefore, the severity of the identification problem for $W(X)$ is directly determined by $L_W(X)$, and with it, what an investigator is willing to assume about how the marginal effect would have been for non-selected applicants had they been selected.

Similarly, the between-group effect is given by:

$$B(X) = \sum_{g=1}^{3} \left[E(Y \mid X, G = g) - E(Y \mid X, G = 0)\right] \frac{dP(G = g \mid X)}{dX},$$

where non-selected applicants are used as the reference group, considering the invariance property of the global effect. However, $B(X)$ is also not identified, as it is based on the expected performance of an unobserved population. However, assuming that $\gamma_0(X) \leq E(Y \mid X, G = 0) \leq \gamma_1(X)$ and, after some algebraic manipulation, taking into account that $\sum_{g=0}^{3} P(G = g \mid X) = 1$, we obtain:

$$\begin{aligned} B(X) \quad \in \quad & \left[\sum_{g=1}^{3} \left[E(Y \mid X, G = g) \frac{dP(G = g \mid X)}{dX} - \left(\gamma_1(X) \frac{d^{+}P(G = g \mid X)}{dX} - \gamma_0(X) \frac{d^{-}P(G = g \mid X)}{dX} \right) \right]; \right. \\ & \left. \sum_{g=1}^{3} \left[E(Y \mid X, G = g) \frac{dP(G = g \mid X)}{dX} - \left(\gamma_0(X) \frac{d^{+}P(G = g \mid X)}{dX} - \gamma_1(X) \frac{d^{-}P(G = g \mid X)}{dX} \right) \right] \right]. \end{aligned} \tag{8}$$

It is not difficult to show that the width of the bound is

$$L_B(X) = [\gamma_1(X) - \gamma_0(X)] \left| \frac{dP(G = 0 \mid X)}{dX} \right|.$$

The bound width for $B(X)$ represents the quantification of how the lack of information affects the estimation of the between-group effect due to the absence of data on the performance of non-selected applicants. This expression shows that the bound width depends on two key factors. First, the difference $\gamma_1(X) - \gamma_0(X)$ reflects the range of possible values that the conditional expectation of Y in non-selected applicants could take, introducing more or less variability in the estimation of the between-group effect. Second, the term $\left|\frac{dP(G=0|X)}{dX}\right|$ indicates how the proportion of non-selected applicants varies with respect to X. If this quantity is large, it means that the sample composition changes rapidly with the math score, amplifying the uncertainty around $B(X)$.

Consequently, the severity of the identification problem for $B(X)$ depends on the interaction between these two components: if the interval in which the conditional expectation lies is wide and/or the proportion of non-selected applicants varies sharply with X, the between-group effect will be harder to bound without assuming additional information about the performance of non-selected applicants: a conservative assumption (CA), which provides the widest contextual bounds for both the conditional expectation and its derivative—meaning that any plausible solution under alternative assumptions should lie within these bounds—and an empirical contextual assumption (EA), which is also used to construct contextual bounds for both the conditional expectation and its derivative.

4.1 Assumptions and results

Under the ignorability assumption, it was sufficient to impose a restriction on the conditional expectation, as this allowed us to work exclusively with the selected applicants. However, when considering scenarios for the non-selected applicants, additional restrictions are needed not only on the conditional expectation but also on its derivative. This necessity arises because, in the absence of ignorability, the identification of the marginal effect depends on both the expected performance of non-selected applicants and how it varies with changes in test scores. In what follows, we propose a set of assumptions to bound both $W(X)$ and $B(X)$, which together enable the construction of identification intervals for the global marginal effect.

Assumptions for bounding $W(X)$

To construct bounds for $W(X)$, it is necessary to impose restrictions on the marginal effect among non-selected applicants. Specifically, this involves proposing functions $D_0(X)$ and $D_1(X)$ that represent contextually plausible lower and upper bounds for the derivative of the unobserved conditional expectation. The informativeness of the resulting bounds depends on the nature of these assumptions.

Assumption 1 (CA): A baseline belief is that higher test scores are associated with better academic performance, implying a non-negative marginal effect. This assumption naturally extends to non-selected applicants, as the test is designed to capture academic competencies, and there is no reason to expect a reversal of this relationship among them. Another belief about the effect of test scores is that the maximum possible effect of math scores on university performance is observed within the selected population. Consequently, it follows that the effect of non-selected applicants, had they been selected, would not be expected to exceed the maximum effect observed among those who were admitted. Translating these beliefs into marginal effects terms, it follows that:

$$0 \leq \frac{dE(Y \mid X, G=0)}{dX} \leq \max_x \left\{ \frac{d}{dX} \left[\sum_{g=1}^{3} E(Y \mid X, G=g) P(G=g \mid X) \right] \Bigg|_{X=x} \right\}.$$

These bounds can be interpreted as the most conservative plausible assumption for the marginal effect in the non-selected population, as they rely only on the general assumption of monotonicity and the empirical maximum observed within the selected population.

Assumption 2 (EA): Since the assumptions in this framework are contextual, empirical insights can be used to improve the informativeness of the bounds. In this case, it is plausible that the effect of test scores on academic performance among non-selected applicants—had they been admitted—could have been greater than the lowest effect observed within the selected population. That is, while the effect remains positive, it may have exceeded the minimum identified among admitted students. However, we do not expect it to surpass the effect observed in the most similar group among those admitted. For our data, we define similarity based on the average math score: non-selected applicants averaged 627 points, compared to 632 for MB, 647 for B, and 707 for BC. Given this proximity, we consider the Marine Biology (MB) group the most comparable. Therefore, we can assume that the marginal effect among non-selected applicants would not surpass the maximum effect observed in the MB group ($G = 1$). This belief, can be mathematically translated as follows:

$$\min_x \left\{ \frac{d}{dX} \left[\sum_{g=1}^{3} E(Y \mid X, G = g) P(G = g \mid X) \right] \Bigg|_{X=x} \right\} \leq \frac{dE(Y \mid X, G = 0)}{dX} \leq \max_x \left\{ \frac{d}{dX} \left[E(Y \mid X, G = 1) \right] \Big|_{X=x} \right\}.$$

Both Assumptions 1 and 2 lead to different functions $D_0(X)$ and $D_1(X)$, which in turn determine the width and informativeness of the resulting bounds. These functions reflect the contextual beliefs imposed on the marginal effect among the non-selected applicants. The corresponding bounds are then computed by replacing the corresponding $D_0(X)$ and $D_1(X)$ functions into the identification interval (7) for bounding $W(X)$.

Assumptions for bounding $B(X)$

Bounding $B(X)$ requires imposing restrictions on the expected academic performance of non-selected applicants, conditional on their test scores. Formally, this involves specifying both $y_0(X)$ and $y_1(X)$

Assumption 1 (CA): In the most conservative case, we rely on the known bounds of the GPA scale, namely $y_0(X) = m_0$ and $y_1(X) = m_1$ (in the case of Chile, $m_0 = 1.0$ and $m_1 = 7.0$). This implies that

$$m_0 \leq E(Y \mid X, G = 0) \leq m_1.$$

Assumption 2 (EA): To construct more informative bounds, we incorporate contextual information from the observed GPAs. It is plausible to assume that non-selected applicants—had they been admitted—could have performed better, on average, than the lowest-performing student among those selected. This implies that the expected GPA for non-selected applicants, conditional on their test scores, may exceed the minimum GPA observed in the selected population. At the same time, given that these applicants were not admitted, it is reasonable to assume that their expected performance would not surpass the maximum GPA observed in the most similar group in terms of math scores—namely, the Marine Biology (MB) group. This assumption reflects the belief that, although some non-selected students may have had potential to perform above the weakest admitted students, it is unlikely that their average performance would have exceeded that of the strongest students in the most comparable program.

$$\min\{Y : G \in \{1, 2, 3\}\} \leq E(Y \mid X, G = 0) \leq \max\{Y : G = 1\}$$

These assumptions provide plausible values for the bounding functions $y_0(X)$ and $y_1(X)$. In this setting, both functions are constant with respect to X, reflecting context-specific empirical beliefs that are reasonable to impose given the selection mechanism under study. The corresponding bounds are then computed by replacing the corresponding $y_0(X)$ and $y_1(X)$ functions into the identification interval (8) for bounding $B(X)$.

Results

The results presented in Figure (3) are derived from the previously established identification bounds. These graphs analyze the probabilities of belonging to each group as a function of math scores, as well as the identification bounds for the within-group and between-group marginal effects, along with the global effect.

The upper left panel illustrates how the composition of applicants varies across the score distribution, which is key to understanding the implications of selection. The remaining panels display the identified bounds for $W(X)$, $B(X)$, and their sum $W(X) + B(X)$, which represents the global marginal effect. To highlight the impact of different restrictions, we report two sets of results corresponding to the assumptions previously described. One set is based on Assumption 1, which adopts conservative assumptions (CA), and the other on Assumption 2, which incorporates more contextual and empirical assumptions (EA). For the global marginal effect (GME), the bounds are constructed as the pointwise sum of the corresponding bounds under each assumption—i.e., the bounds under Assumption 1 for $W(X)$ and $B(X)$ are summed to produce the global bounds under Assumption 1, and analogously for Assumption 2.

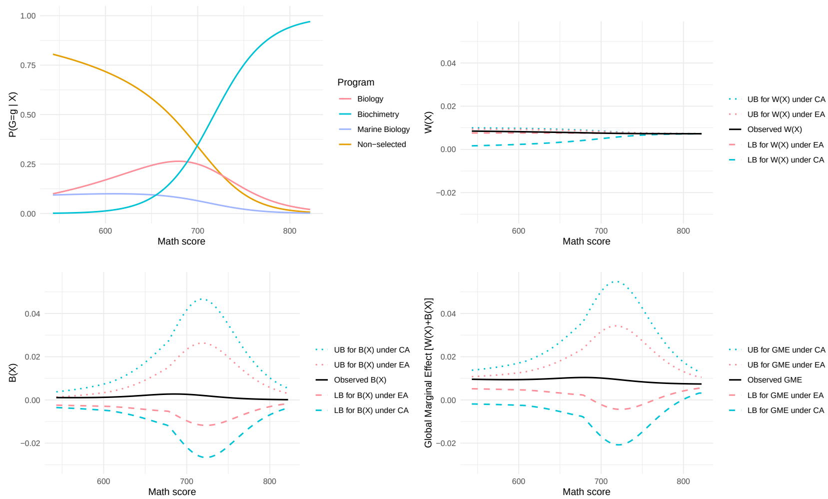

Figure 3. Estimated group membership probabilities and identification bounds $W(X)$, $B(X)$, and global marginal effects under contextual (CA) and empirical-assumption (EA).

In the upper-left panel of Figure (3), the distribution of applicants' group membership is displayed as a function of their math score. The probability of not being selected, $G = 0$, is high for low scores but gradually decreases as the score increases, approaching zero at the upper end of the scale. This implies that for higher scores, almost all applicants are placed in one of the selected programs. It is important to note that the probability of belonging to Biochemistry increases with the score, while the probabilities of being in Biology or Marine Biology are relatively higher at the lower end of the scale. Biochemistry is the program with the highest predicted GPA values, whereas Biology and Marine Biology exhibit lower predicted values. This plays an important role in shaping the behavior of the marginal effect both within and between groups.

In the upper-right panel of the figure, the identification bounds for $W(X)$ are shown. The lower bound reflects the marginal effect under the assumption that math scores have no impact on the academic performance of non-selected applicants, had they been admitted. For low scores, the lower bound for $W(X)$ is close to zero under CA, as in this part of the scale the estimation is strongly influenced by the assumption imposed on the non-selected applicants due to their high proportion in these score levels. In contrast, for high scores, the influence of the non-selected applicants gradually disappears, and $W(X)$ increasingly resembles the marginal effect observed among the selected applicants, as the proportion of non-selected applicants at these levels is nearly zero.

Compared to the conservative assumption (CA), the empirically anchored assumption (EA) produces narrower and more informative bounds for $W(X)$ across the entire score distribution. Interestingly, the bounds under EA closely approximate those obtained under the ignorability assumption. However, these conclusions are now justified through substantive, contextual reasoning. That is, if a policy-maker is willing to assume that non-selected applicants—had they been admitted—could have produced a marginal effect higher than the smallest observed among the selected, but not exceeding the maximum effect observed in the most similar group (in terms of math scores), then *the within-group effect, $W(X)$, is positive and it tends to decrease as X increases, indicating that the relationship between test scores and academic performance within each program is stronger at lower score levels but does not vanish at higher levels.* Although numerically similar to those obtained under ignorability, these conclusions are now grounded in contextual assumptions.

In the lower-left panel of Figure 3, the identification bounds for $B(X)$ are presented. From the data, it is observed that the sum of the derivatives of $P(G = g \mid X)$ for $g \in 1, 2, 3$ is always positive across the entire range of X. This implies that both the lower and upper bounds rely only on the positive part of this sum, resulting in the use of m_1 in the lower bound and m_0 in the upper bound under the conservative assumptions (CA). The use of $m_1 = 7.0$ in the lower bound leads to all differences being negative, as no predicted values reach 7.0 in any of the observed programs. Conversely, the upper bound under CA, based on $m_0 = 1.0$, results in all differences being positive. The peaks in $B(X)$ are found in the central part of the graph, where the highest variability in program assignment occurs, which in turn amplifies the dispersion in the estimates of $B(X)$. When contextual–empirical assumptions (EA) are introduced, the width of the bounds is substantially reduced across the entire range of X, making them more informative. Nevertheless, even under these assumptions, the sign of $B(X)$ cannot be identified, as the bounds continue to include zero. This limitation arises due to the persistent lack of information regarding the performance of non-selected applicants. Incorporating additional assumptions—derived from theoretical or empirical considerations—may allow for further refinement of the bounds and ultimately lead to the identification of the sign of $B(X)$ within specific score regions.

The bottom right panel of Figure presents the identification bounds for the global effect, defined as the sum of the identification bounds for $W(X)$ and $B(X)$. Under conservative assumptions (CA), these bounds indicate that for most of the score scale, the sign of the global effect cannot be identified, as the bounds include zero throughout. Only at the upper end of the scale (above 800 points), where the proportion of non-selected applicants is near zero, does the sign become identifiable. However, when combining contextual–empirical assumptions (EA) for both the conditional expectation and the marginal effect among non-selected applicants, the bounds become substantially narrower and more informative. In particular, under EA, the sign of the global marginal effect can be identified across nearly the entire score range. The only exception occurs in the central region of the scale—approximately between 700 and 750 points—where the high variability in group membership, conditional on test scores, makes it not possible to determine whether the global effect is positive or negative. This demonstrates that, although information about non-selected applicants remains unobserved, adopting contextually grounded and empirically justified assumptions enhances the informativeness of the bounds. As a result, it becomes possible to draw meaningful conclusions about the direction and magnitude of the global marginal effect across most of the score distribution.

5. Conclusion

This paper addressed the estimation of the marginal effect of admission test scores on academic performance in a selection context, where performance was only observed for admitted applicants. We proposed a decomposition of the marginal effect based on the law of total probability, distinguishing a within-group effect, which measures the relationship between scores and performance within each program, and a between-group effect, which captures how scores explain differences in average predictions across programs and affect admission probabilities.

Under the ignorability assumption, we were able to interpret the decomposition of the marginal effect using only information from admitted applicants. This interpretation follows directly from assuming that the conditional expectation of academic performance given the test score is the same for selected and non-selected applicants, allowing us to work exclusively with the observed sample. In the analyzed data, we found that, within the overall effect, the within-group effect had a greater influence than the between-group effect. This suggests that, for this group of selected applicants, mathematics scores do not play a significant role in explaining differences in predicted GPA or in determining whether an applicant is selected into one program or another. However, this relationship could be reversed in other contexts, such as universities with different levels of selectivity for the same program.

Since the ignorability assumption may not be realistic in selection contexts, we proposed identification bounds for the marginal effect based on contextual assumptions about the admission system. The assumptions maintained in this paper reflect possible imperfections in the selection process, allowing for the possibility that non-selected applicants could have performed better than the lowest-performing selected students, though not better than those in the most similar observed group. This dual assumption aligns with the Fallible Selection Assumption (FSA) described by Alarcón-Bustamante et al. (2025), where the system may fail to admit some applicants who would have benefited from admission, but still captures the strongest effect within the selected group.

Under CA, the resulting bounds for the global effect are wide and do not allow sign identification across most of the scale. In contrast, EA assumptions substantially narrow the bounds and enable sign identification in nearly the entire score range, excluding only the region with the highest assignment variability. Nevertheless, although ignorability yields similar numerical conclusions regarding the sign of the effect, it relies on strong and unjustifiable assumptions. The partial identification approach, by contrast, achieves similar interpretability through empirically grounded, transparent restrictions—enhancing both credibility and contextual relevance.

The proposed approach not only enables the interpretation of the marginal effect under different selection scenarios but also expands the partial identification framework to partitioned populations. These results highlight the value of incorporating selection process information into the analysis of the predictive validity of admission test scores. At its core, drawing conclusions from data always requires the articulation of assumptions, as the logic of inference follows the principle: data + assumptions = conclusions. Holding data fixed, different assumptions naturally lead to different conclusions (Manski, 1993, 2003, 2013). Strong assumptions like ignorability may yield precise results, but often lack contextual justification and weaken interpretability. In contrast, the partial identification approach explicitly acknowledges uncertainty and derives conclusions from assumptions that are substantively grounded in the problem at hand. This improves transparency and aligns the inferences with real-world institutional features, making them more interpretable and credible. As articulated in Manski's Law of Decreasing Credibility, *the credibility of inferences decreases with the strength of the assumptions maintained*.

It is important to note that the assumptions imposed are associated with desirable properties of the admission system; however, other types of assumptions can be used when more information is available. For instance, contextual assumptions could be incorporated—if the data corresponded to the pandemic period, one could introduce assumptions about whether predictability increased or decreased, or formulate year-by-year assumptions based on specific contextual factors. Additionally, if information from other years is available, this could be explicitly incorporated into the modeling process by analyzing the intersection of students who were not selected in one year but were in another. This would allow for the incorporation of assumptions regarding whether predictability increases or decreases over time.

Acknowledgement

Eduardo Alarcón-Bustamante was partially funded by the Postdoctoral FONDECYT grant No. 3220422. Also, the author was partially funded by the FONDEF IDeA I+D ID22I10228 project. David Torres Irribarra was partially funded by Agencia Nacional de Investigación y Desarrollo (ANID) with grants FONDECYT 1240929. María Inés Godoy was partially funded by the FONDEF IDeA I+D ID22I10228 project.

References

Alarcón-Bustamante, E., González, J., Torres Irribarra, D., & San Martín, E. (2025). From missing data to informative GPA predictions: Navigating selection process beliefs with the partial identifiability approach. *British Journal of Mathematical and Statistical Psychology*, *78*(2), 647–671. https://doi.org/10.1111/bmsp.12377

Alarcón-Bustamante, E., San Martín, E., & González, J. (2020). Predictive validity under partial observability. In M. Wiberg, M. Molenaar, J. González, U. Böckenholt, & J. Kim (Eds.), *Quantitative psychology. imps 2019* (Vol. 322). Springer, Cham. https://doi.org/10.1007/978-3-030-43469-4_11

Alarcón-Bustamante, E., San Martín, E., & González, J. (2021). On the marginal effect under partitioned populations: Definition and interpretation. In M. Wiberg, D. Molenaar, J. González, U. Böckenholt, & J. Kim (Eds.), *Quantitative psychology. imps 2020* (Vol. 353). Springer, Cham. https://doi.org/10.1007/978-3-030-74772-5_15

Alarcón-Bustamante, E., Varas, I. M., & San Martín, E. a. (2023). On the impact of missing outcomes in linear regression. *Chilean journal of Statistics*, *14*(1), 26–36. https://doi.org/10.32372/chjs.14-01-02

Ayers, J. B., & Peters, M. (1977). Predictive Validity of the test of English as Foreign Language for Asian graduate students in Engineering, Chemistry, or Mathematics. *Educational and Psychological Measurement*, (37), 461–46. https://doi.org/10.1177/001316447703700221

Florens, J., & Mouchart, M. (1982). A note on noncausality. *Econometrica*, *50*(3), 583–592. https://doi.org/10.2307/1912602

Geiser, S., & Studley, R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, *8*(1), 1–26. https://doi.org/10.1207/S15326977EA0801_01

Goldhaber, D., Cowan, J., & Theobald, R. (2017). Evaluating prospective teachers: Testing the predictive validity of the edTPA. *Journal of Teacher Education*, *68*(4), 377–393. https://doi.org/10.1177/0022487117702582

Grassau, E. (1956). Análisis estadístico de las pruebas de bachillerato. *Anales de la Universidad de Chile*, (102), 77–93.

Grobelny, J. (2018). Predictive Validity toward Job Performance of General and Specific Mental Abilities. A Validity Study across Different Occupational Groups. *Business and Management Studies*, *4*(3), 1–12. https://doi.org/10.11114/bms.v4i3.3297

Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. In W. A. Shewhart & S. S. Wilks (Eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-data Perspectives: An Essential Journey with Donald Rubin's Statistical Family* (pp. 73–84). Wiley Series in Probability; Statistics. https://doi.org/10.1002/sim.8835

Imbens, G. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, *87*(3), 706–710. https://doi.org/10.3386/t0237

Kolmogorov, A. N. (1950). *Foundations of the theory of probability*. Chelsea Publishing Co.

Koopmans, T. C. (1949). Identification problems in economic model construction. *Econometrica*, *17*(2), 125–144. https://doi.org/10.2307/1905689

Manski, C. (1989). Anatomy of the selection problem. *The Journal of Human Resources*, *24*(3), 343–360. https://doi.org/10.2307/145818

Manski, C. (1993). Identification problems in the social sciences. *Sociological Methodology*, *23*, 1–56. https://doi.org/10.2307/271005

Manski, C. (2003). *Partial identification of probability distributions*. New York: Springer.

Manski, C. (2013). *Public policy in an uncertain world: Analysis and decisions*. Harvard University Press.

Manzi, J., & Carrasco, D. (2021, September). Validity Evidence of the University Admission Tests in Chile: Prueba de Selección Universitaria (PSU). https://doi.org/10.1007/978-3-030-78390-7_14

Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.2307/2335942

San Martín, E., & González, J. (2022). A Critical View on the NEAT Equating Design: Statistical Modelling and Identifiability Problems. *Journal of Educational and Behavioral Statistics*, *47*(4), 406–437. https://doi.org/10.3102/10769986221090609

Stoye, J. (2007). Bounds on generalized linear predictors with incomplete outcome data. *Reliable Comput*, *13*, 293–302. https://doi.org/10.1007/s11155-006-9030-5

Tamer, E. (2010). Partial identification in econometrics. *Annual Review of Economics*, *2*(1), 167–195. https://doi.org/10.1146/annurev.economics.050708.143401

Supplementary Material

Appendix 1. Derivation of the identification bounds for the marginal effect

Proof. If $D_0(X) \leq \frac{dE(Y|X,G=0)}{dX} \leq D_1(X)$, thus

$$D_0(X)P(G=0 \mid X) + \sum_{g=1}^{3} \frac{dE(Y \mid X, G=g)}{dX} P(G=g \mid X) \leq \frac{dE(Y \mid X, G=0)}{dX} P(G=0 \mid X) + \sum_{g=1}^{3} \frac{dE(Y \mid X, G=g)}{dX} P(G=g \mid X) \leq D_1(X)P(G=0 \mid X) + \sum_{g=1}^{3} \frac{dE(Y \mid X, G=g)}{dX} P(G=g \mid X). \quad (9)$$

Let $\frac{d^+\lambda(X)}{dX}$ and $\frac{d^-\lambda(X)}{dX}$ be the positive and negative part of $\frac{d\lambda(X)}{dX}$, respectively, such that:

$$\frac{d^+\lambda(X)}{dX} = \max\left\{0, \frac{d\lambda(X)}{dX}\right\} = \begin{cases} \frac{d\lambda(X)}{dX} & \text{if } \frac{d\lambda(X)}{dX} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\frac{d^-\lambda(X)}{dX} = \max\left\{0, -\frac{d\lambda(X)}{dX}\right\} = \begin{cases} -\frac{d\lambda(X)}{dX} & \text{if } \frac{d\lambda(X)}{dX} < 0 \\ 0 & \text{otherwise.} \end{cases}$$

Where $\frac{d\lambda(X)}{dX} = \frac{d^+\lambda(X)}{dX} - \frac{d^-\lambda(X)}{dX}$, and $\left|\frac{d\lambda(X)}{dX}\right| = \frac{d^+\lambda(X)}{dX} + \frac{d^-\lambda(X)}{dX}$. Let us consider that $\gamma_0(X) \leq E(Y \mid X, G=0) \leq \gamma_1(X)$. Thus, the positive part of $E(Y \mid X, G=0)\frac{dP(G=0|X)}{dX}$ is bounded as follows:

$$\gamma_0(X)\frac{d^+P(G=0 \mid X)}{dX} \leq E(Y \mid X, G=0)\frac{d^+P(G=0 \mid X)}{dX} \leq \gamma_1(X)\frac{d^+P(G=0 \mid X)}{dX}. \quad (10)$$

Analogously, the negative part is bounded as follows:

$$\gamma_0(X)\frac{d^-P(G=0 \mid X)}{dX} \leq E(Y \mid X, G=0)\frac{d^-P(G=0 \mid X)}{dX} \leq \gamma_1(X)\frac{d^-P(G=0 \mid X)}{dX} \quad (11)$$

By subtracting (11) from (10), it is obtained that

$$\gamma_0(X)\frac{d^+P(G=0 \mid X)}{dX} - \gamma_1(X)\frac{d^-P(G=0 \mid X)}{dX} \leq E(Y \mid X, G=0)\frac{dP(G=0 \mid X)}{dX} \leq \gamma_1(X)\frac{d^+P(G=0 \mid X)}{dX} - \gamma_0(X)\frac{d^-P(G=0 \mid X)}{dX}. \quad (12)$$

Thus, if the term $\sum_{g=1}^{3} E(Y \mid X, G=g)\frac{dP(G=g|X)}{dX}$ is added to (12), and the resultant inequality is added to (9), then the interval in (5) holds. □

Appendix 2. Length of the identification interval for the marginal effect

Proof. Subtracting the lower bound from the upper one in (5), it is obtained that

$$\begin{aligned}
L(X) &= [D_1(X) - D_0(X)]P(G=0 \mid X) + \left[\gamma_1(X)\frac{d^+P(G=0 \mid X)}{dX} - \gamma_0(X)\frac{d^-P(G=0 \mid X)}{dX}\right] - \\
&\quad \left[\gamma_0(X)\frac{d^+P(G=0 \mid X)}{dX} - \gamma_1(X)\frac{d^-P(G=0 \mid X)}{dX}\right] \\
&= [D_1(X) - D_0(X)]P(G=0 \mid X) + [\gamma_1(X) - \gamma_0(X)]\frac{d^+P(G=0 \mid X)}{dX} + \\
&\quad [\gamma_1(X) - \gamma_0(X)]\frac{d^-P(G=0 \mid X)}{dX} \\
&= [D_1(X) - D_0(X)]P(G=0 \mid X) + [\gamma_1(X) - \gamma_0(X)]\left|\frac{dP(G=0 \mid X)}{dX}\right|
\end{aligned}$$

□