

Jirakom Sirisrisakulchai; Supanika Leurcharusmee

Article

Returns to higher education by field of study in Thailand: Comparative analysis of pre and post-COVID-19 pandemic

Asian Journal of Economics and Banking (AJEB)

Provided in Cooperation with:

Ho Chi Minh University of Banking (HUB), Ho Chi Minh City

Suggested Citation: Jirakom Sirisrisakulchai; Supanika Leurcharusmee (2025) : Returns to higher education by field of study in Thailand: Comparative analysis of pre and post-COVID-19 pandemic, Asian Journal of Economics and Banking (AJEB), ISSN 2633-7991, Emerald, Leeds, Vol. 9, Iss. 1, pp. 2-21,
<https://doi.org/10.1108/AJEB-09-2024-0106>

This Version is available at:

<https://hdl.handle.net/10419/334134>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Returns to higher education by field of study in Thailand: comparative analysis of pre and post-COVID-19 pandemic

Jirakom Siririsakulchai and Supanika Leurcharusmee
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand

Abstract

Purpose – This study estimates returns to higher education across different fields in Thailand for 2019 and 2021, accounting for field selection endogeneity. The comparison offers insights into the impact of the pandemic and other economic shocks on the returns.

Design/methodology/approach – The study applies an econometric causal framework, integrating economic theory with empirical analysis using data from Thailand's socioeconomic surveys in 2019 and 2021. A multinomial treatment effects model with sample selection correction is used to estimate the impact of different fields of study on income, accounting for selection both into higher education in different fields and into employment, addressing potential biases from ability sorting and sample selection.

Findings – The study finds variations in returns to education across fields. In 2019, teaching offered the highest returns on average, followed by healthcare. Social sciences, business and computer-related fields showed moderate returns, while the combined group of science, agriculture, engineering and architecture had non-significant returns, indicating a low weighted average across these diverse fields. In 2021, healthcare exhibited the highest return due to pandemic-driven demand. Across both years, controlling for occupation reduced the estimated returns by approximately 50%, highlighting the role of occupational status in mediating educational returns.

Originality/value – This study uniquely applies an econometric causal framework to analyze returns to higher education by field of study in Thailand. It offers insights for policymakers to align educational programs with labor market demand and emphasizes the importance of data-driven decisions in responding to disruptions.

Keywords Returns to schooling, Field of study, Causal framework, Ability sorting, Multinomial treatment

Paper type Research paper

1. Introduction

Choosing a field of study in higher education is a pivotal decision that significantly impacts an individual's future earnings. As more resources are invested in higher education worldwide, understanding the benefits of different study fields has become increasingly vital. This importance is evident in the substantial growth of global higher education participation. According to UNESCO (2022), the percentage of people enrolled in higher education doubled from 19% in 2000 to about 40% in 2020. This trend is also seen in Thailand, where the proportion of adults with higher education qualifications, including vocational or college degrees, has steadily increased, reaching 20% in 2021 (NSO, 2022).

The COVID-19 pandemic has significantly altered labor markets, increasing demand in sectors like healthcare while reducing it in others such as tourism and hospitality (ADB, 2021). These shifts, coupled with the ongoing Fourth Industrial Revolution, have led to varying returns on education across different fields of study (WEF, 2020). This study compares labor



JEL Classification — C21, J21, J24, J31

© Jirakom Siririsakulchai and Supanika Leurcharusmee. Published in *Asian Journal of Economics and Banking*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

market outcomes from 2019 to 2021, providing insights into how the pandemic and Industry 4.0 have collectively influenced educational returns of higher education across different fields of study in Thailand.

Returns to education have been primarily estimated through two methods: the full-discounting method and the Mincerian earnings function, with the latter being more commonly employed due to its broad applicability (Patrinos and Psacharopoulos, 2018). Global research reveals a consistent average private return of approximately 9% annually per additional year of schooling, with minimal decline over time (Psacharopoulos and Patrinos, 2018). This trend aligns with Tinbergen (1975)'s concept of a "race between education and technology," where skill demand outpaces supply despite increased educational attainment. While general returns to education have marginally decreased, private returns to higher education have risen, exacerbating equity concerns. Goldin and Katz (2009) attribute this to skill-biased technological progress, which has widened income inequality and maintained high premiums for advanced skills. This suggests that educational advancements lag behind the growing demand for higher-level skills, potentially indicating technology's lead in this race. Consequently, analyzing returns by specific fields of study, rather than overall higher education, becomes crucial to identify which skills yield the highest returns in an evolving labor market.

However, estimating returns by field of study presents challenges beyond the typical concerns of ability and selection bias. Ability sorting across fields introduces additional potential biases to the estimation. Research shows that individuals with higher abilities tend to select fields that yield higher returns, complicating the assessment of educational returns (Psacharopoulos and Patrinos, 2018; Arcidiacono, 2004). This ability sorting can lead to overestimation of returns for certain fields, as traditional methods may fail to account for the influence of unobserved characteristics on both educational attainment and labor market outcomes (Eberhard *et al.*, 2017; Lemieux, 2014; Bol and Heisig, 2021).

To address biases stemming from unobserved heterogeneity, namely selection and ability sorting biases, researchers have employed advanced econometric techniques. Structural models, like those used by Arcidiacono (2004), capture individuals' dynamic decisions regarding education and occupation. Quasi-experimental approaches, such as regression discontinuity designs (Canaan and Mouganie, 2018), decomposition method (Bol and Heisig, 2021) and instrumental variable methods (Kirkeboen *et al.*, 2016), have also been utilized to isolate the impact of field-specific education on earnings.

To systematically analyze the causal effect of choosing different fields of study on wages, while clearly communicating the limitations of the model and potential biases in estimating the return to education, it is crucial to adopt a structured approach grounded in the econometric causal framework. This framework provides a rigorous foundation for addressing the complexities inherent in causal inference, ensuring that the analysis is not only methodologically sound but also transparent regarding the assumptions and potential sources of bias. By explicitly framing the analysis within this approach, we offer a clear pathway for other researchers to replicate and critique our findings, while also highlighting the critical factors that could lead to biased estimates, such as unobserved heterogeneity or selection bias. In doing so, we align our research with best practices in econometric analysis, emphasizing the importance of carefully considering model specifications and the robustness of causal interpretations when assessing the returns to education across different fields of study.

To address these potential biases within econometric causal framework, this paper employs a multinomial treatment effects (MTE) model with sample selection correction to analyze the returns to education across various fields of study in Thailand, explicitly addressing both the ability sorting bias and selection into paid employment. The multinomial treatment model allows for the estimation of treatment effects when the treatment variable is multinomial in nature, meaning that individuals can choose among multiple educational fields (Deb and Trivedi, 2006). By comparing educational returns before and after the COVID-19 pandemic, this study aims to provide a comprehensive understanding of how economic shocks influence

student choices and subsequent wage outcomes. This approach enables the analysis of the impact of different fields of study on wage outcomes while controlling for selection bias and unobserved heterogeneity, thereby providing more accurate estimates of the economic returns associated with each field.

This research contributes to the growing body of literature on educational returns, particularly in the context of developing economies like Thailand. By employing a transparent methodology to analyze the returns to higher education across diverse fields of study, while simultaneously accounting for selection bias and field-specific ability sorting, this research aims to provide more precise estimates of the economic returns associated with each academic field. The insights gained can inform policymakers and educational institutions in aligning educational programs with labor market needs and supporting students in making informed decisions about their fields of study, especially in the face of economic disruption and uncertainties.

This paper is organized as follows: [Section 2](#) presents the econometric causal framework, followed by [Section 3](#), which develops the causal model, including the economic modeling of returns to education and field of study decisions. [Section 4](#) discusses the identification of causal parameters, outlining the counterfactual approach, hypothetical models, and empirical strategy. [Section 5](#) details the data sources and sample characteristics. [Section 6](#) presents the results, focusing on the estimated returns to higher education by field of study and the role of occupational status as a mediator. Finally, [Section 7](#) concludes the paper with policy implications and suggestions for future research.

2. Econometric causal framework

The concept of causal inference in econometrics dates back to Ragnar Frisch's seminal lecture in 1930 ([Bjerkholt and Qin, 2011](#)). Frisch conceptualized causality as a thought experiment where economists hypothetically analyze how changes in inputs would influence outputs. This perspective laid the foundation for modern causal analysis in economics, focusing on two fundamental principles: autonomy and directionality ([Heckman and Pinto, 2024](#)).

Autonomy in econometrics refers to the concept explained by [Frisch \(1938\)](#). The term autonomy means autonomy of (causal) relationship and autonomy of a function is defined as functions that exhibit invariance property in the sense that they are invariant to any changes in their arguments ([Heckman and Pinto, 2024](#)). In econometric causal model, this means that each causal relationship in a model is self-contained and not directly affected by changes in unrelated variables. For instance, in a return-to-education model, the relationship between years of schooling and income should remain valid regardless of external economic shifts, provided the model is well-specified. This characteristic of autonomy is essential for isolating the true effect of education on income without contamination from extraneous factors.

For the concept of directionality, [Heckman and Pinto \(2024\)](#) highlighted that causality functions in one direction, flowing from the cause to the effect. In general, if we change the cause, it will affect the outcome, but changing the outcome won't affect the original cause. In the context of education and earnings, this implies that changes in educational attainment influence income, not the other way around. This concept helps avoid issues like circular reasoning or simultaneity bias, ensuring that the estimated return on education reflects a true causal effect rather than just a correlation.

Unlike traditional statistical frameworks, which rely on joint distributions to describe relationships between variables without specifying causal directionality, the econometric causal framework introduces a critical layer of analysis. In statistics, the relationship between two variables can be fully captured by their joint distribution, yet this approach leaves the direction of causality ambiguous. For instance, while correlation or mutual dependence between education and income can be quantified, it does not clarify whether education directly causes higher income or if higher income leads to more education. The econometric causal framework addresses this gap by establishing a structured approach to determining causal

relationships, emphasizing the directionality of influence from cause to effect. This structured framework is particularly valuable in applied econometric analysis, where the goal is often to answer specific causal questions relevant to policy decisions or theoretical inquiries.

The econometric causal framework (Heckman and Pinto, 2024) is built around three core tasks that are essential to causal inference: (1) constructing a causal model, (2) identifying causal parameters, and (3) performing statistical inference. These tasks ensure that the analysis remains aligned with the specific causal questions being investigated by the analyst. We will explain the details of the three key tasks in the following sections to ensure a comprehensive understanding of the causal analysis process in this research.

3. Causal model

The first task, developing a causal model, involves defining a system of structural equations that map the relationships between a set of variables and their power set. Each equation in this system is an invariant mapping from inputs to outputs, where the inputs are understood as the direct causes of the outputs.

Heckman and Pinto (2022) mathematically illustrate how causal models in economics can be constructed, using the Generalized Roy model (Roy, 1951) as an example. They demonstrate how causal relationships can be represented with Directed Acyclic Graphs (DAGs), which clarify the direction of causality. Additionally, they highlight the Local Markov Condition (LMC), which fully characterizes causal models by specifying that a variable is conditionally independent of its non-descendants given its parents in the graph.

The construction of the causal model in this study relies on an economic model for Field of Study Decisions and Wage Outcomes, which will be detailed in the following sections. This framework underpins the analysis by explicitly modeling the selection process into different fields of study and how these choices translate into wage outcomes, allowing for a rigorous examination of the causal relationships involved.

3.1 Economic modeling for the returns to education with endogenous decision on field of study

Individuals make educational decisions, including the level of study and field of study, to maximize their lifetime utility. This modeling framework is based on the work of Altonji *et al.* (2016) in “*The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects*.” However, unlike the original model, which allows for switching fields of study, this paper assumes that individuals select their field of study once. This assumption is made due to limited observations of field-switching in Thailand, where most individuals choose their field of study only once, either in vocational education or at the bachelor’s degree level.

The decision-making process is divided into two distinct periods: the education period and the working period. During the education period (Period 1), individuals face a critical decision regarding whether they should invest in a higher education and which field of study to pursue. The model incorporates $Abil_{ij}$, which is the belief of the individual i regarding his ability in the field of study j . The expected utility during this period is expressed as:

$$U_{ij1} = \alpha_{0j} + \alpha_{1j}Abil_{ij}, \quad (\text{Eq.1})$$

where U_{ij1} represents the utility in the education period for individual i in field j . From this specification, α_{0j} captures the baseline utility associated with field j , and α_{1j} reflects the sensitivity of utility to the individual’s ability $Abil_{ij}$.

In the working period (Period 2 onward), the utility function is represented as:

$$U_{ijt} = NP_j(Abil_{ij}) + \gamma_{1j}Wage_j(Abil_{ij}), \quad (\text{Eq.2})$$

where U_{ijt} denotes the utility for individual i derived from the chosen field j at time $t = 2, \dots, T$ and T is the retirement period. The term $NP_j(Abil_{ij})$ captures the nonpecuniary benefits, which depends on the ability $Abil_i$. The parameter γ_1 reflects the weight given to the expected wage $Wage_j(Abil_{ij})$, which is also contingent on the individual's ability.

Individuals make their educational decisions to maximize the discounted sum of their expected payoffs, expressed as the following expected lifetime utility function:

$$U_i = E \left[\sum_j d_{ij} (U_{ij1} + e_{ij1}) + \sum_{t=2}^T \sum_j \beta^{t-1} d_{ij} (U_{ijt} + e_{ijt}) \right], \quad (\text{Eq.3})$$

where d_{ij} is the dummy variable, which equals 1 if individual i chooses field of study j and 0 otherwise, and e_{ijt} is the idiosyncratic shock.

If an individual does not invest in higher education, then $U_{ijt} = 0$, indicating that their utility is normalized to a baseline of only completing high school or lower education. In this context, U_{ijt} reflects the relative utility compared to not pursuing a specific field of study.

It is crucial to recognize that an individual's ability in a chosen field of study, $Abil_{ij}$, encompasses unobserved characteristics that may influence the decision-making process, including preferences and tastes (Altonji *et al.*, 2016). This unobserved field-specific ability can cause "ability sorting" bias in the returns to education estimation. Specifically, ability sorting can occur when individuals with higher ability in a specific field $Abil_{ij}$ select fields of study that yield higher returns, both during their education and in the labor market. If ability $Abil_{ij}$ influences outcomes in the labor market (i.e. $\gamma_{lj} > 0$), the educational institution's role is to reveal or enhance this ability, which then impacts the individual's productivity and nonpecuniary benefits in the labor market. Conversely, if ability $Abil_{ij}$ matters during the education period (i.e. $\alpha_{lj} > 0$), fields of study with higher potential returns might be avoided by individuals with lower ability due to the academic difficulty, leading to both ability and self-selection bias. Thus, when estimating the treatment effect of a field of study on wage outcomes, it is essential to account for potential ability sorting, which could otherwise bias the estimates.

3.2 Causal model for field of study decisions and wage

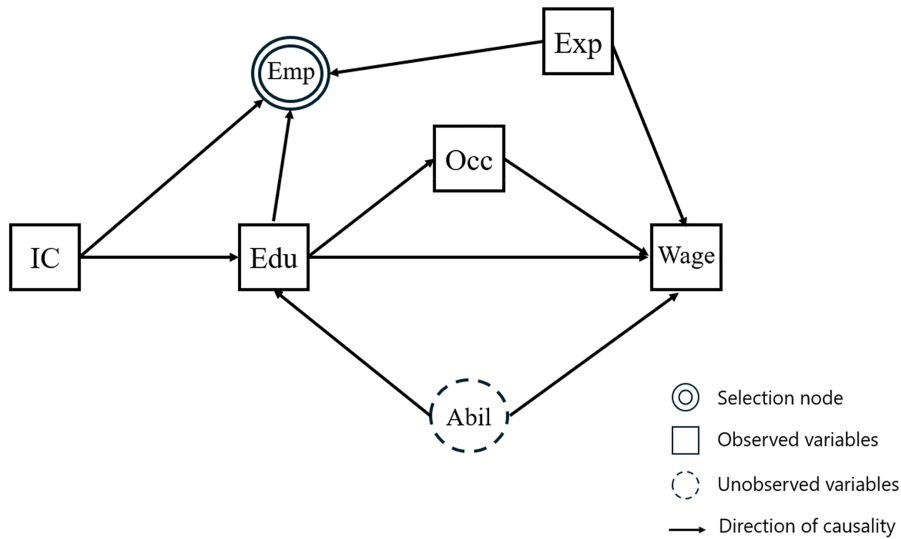
The causal model in this study is based on an economic model (described in Section 3.1) that captures the relationship between field of study decisions and wage outcomes. The model is represented by the Directed Acyclic Graph (DAG) in Figure 1, where the outcome variable is wage (denoted by Wage), the endogenous variable is the level of education segmented by field of study (denoted by Edu), and the confounding factor is ability (denoted by Abil), which influences both educational decisions and labor market outcomes. The exogenous variables include individual characteristics (denoted by IC) such as gender and age, and household background. Additionally, employment status (denoted by Emp) plays a key role as an intermediary variable affecting both occupation and wage outcomes. The structural equations that govern the relationships among these variables are as follows:

$$Wage = f_{Wage}(Edu, Abil, Exp, Occ, \varepsilon_{Wage})$$

$$Occ = f_{Occ}(Edu, \varepsilon_{Occ})$$

$$Exp = f_{Exp}(\varepsilon_{Exp})$$

$$Edu = f_{Edu}(IC, Abil, \varepsilon_{Edu})$$



Source(s): Figure by authors

Figure 1. DAG represents causal model for the returns to education with endogenous decision on field of study

$$Emp = f_{Emp}(IC, Edu, Exp, \varepsilon_{Emp})$$

$$IC = f_{IC}(\varepsilon_{IC})$$

$$Abil = f_{Abil}(\varepsilon_{Abil})$$

In this model, ε_{Wage} , ε_{Occ} , ε_{Exp} , ε_{Edu} , ε_{Emp} , ε_{IC} and ε_{Abil} represent unobserved factors affecting the respective variables. The Local Markov Condition (LMC) fully characterizes the causal model by asserting that each variable is conditionally independent of its non-descendants given its direct parents in the DAG. Specifically:

$$Wage \perp \{Emp, IC\} | \{Edu, Exp, Occ, Abil\}$$

$$Occ \perp \{IC, Exp, Emp, Abil\} | \{Edu\}$$

$$Exp \perp \{Occ, Edu, IC, Abil\}$$

$$Emp \perp \{Wage, Occ, Abil\} | \{IC, Edu, Exp\}$$

$$Edu \perp \{Exp\} | \{IC, Abil\}$$

$$IC \perp \{Exp, Abil\}$$

$$Abil \perp \{Exp, IC\}$$

4. Identification of causal parameters

In this section, we explain Task 2, which involves identifying the causal parameter using a counterfactual approach, and the hypothetical model used for this purpose is detailed in [Sections 4.1–4.3](#). Task 3, focusing on the estimation using the empirical model, is covered in [Section 4.4](#).

4.1 Counterfactual approach

In causal inference, the counterfactual approach is fundamental to estimating causal effects derived from a causal model. The primary objective of a causal model is to estimate the causal effect that answers a specific causal question. A causal effect, or causal parameter, is defined by the counterfactual concept, where the effect is the difference in outcomes resulting from manipulating or intervening on the input variable, while holding the confounding factors constant. This implies that the underlying structural equations remain unchanged even when the input variable is altered.

For instance, consider a scenario where the treatment variable T can take two values, $T = 1$ (treated) and $T = 0$ (untreated). The causal effect is then defined as the difference between the outcome under treatment $Y(1)$ and the outcome without treatment $Y(0)$, $Y(1) - Y(0)$. This difference represents the impact of the treatment on the outcome, assuming that all other factors remain constant. A key point from this definition is that the causal effect is determined entirely by the causal model, without relying on any probability model or statistical assumptions. The causal effect is a theoretical idea that is defined within the causal model and does not depend on the likelihood of the treatment or the way the data is distributed.

However, when we move from theory to empirical estimation, probability models become essential. These models allow us to estimate causal parameters using observed data. For example, to estimate the Average Treatment Effect (ATE), which is the expected difference in outcomes across the population, we would calculate $E[Y(1) - Y(0)]$. Here, the probability model helps us utilize the data to estimate the average causal effect, incorporating statistical assumptions and techniques to ensure that our estimates are accurate and robust.

4.2 Hypothetical models

[Haavelmo \(1943\)](#) introduced the concept of causal manipulation using what is now known as the fix operator ([Heckman and Pinto, 2015](#)). The counterfactual outcome $Y(t)$ can be obtained by fixing the input T in the outcome equation to a specific value $T = t$ within the support of T (denoted as $t \in \text{supp}(T)$). In traditional statistics, when conditioning on a variable, the distribution of all related variables can change, which differs from the effect of the fix operator (or causal operator). The fix operator specifically impacts only the distribution of the descendants of the variable being fixed, leaving other aspects of the model unchanged. Since statistical approaches do not inherently include manipulation using the fix operator, [Heckman and Pinto \(2015\)](#) developed the econometric causal framework, which links the fix operator to statistical conditioning (or conditional distributions). This integration allows for a more precise and controlled analysis of causal effects in empirical research.

[Heckman and Pinto \(2015\)](#) proposed the Hypothetical Model, distinguishing it from the empirical model that generates observable data. The hypothetical model is used to formulate thought experiments involving the manipulation of inputs to determine causality. In this framework, a hypothetical variable \tilde{T} is introduced as an exogenous variable that causes the outcome Y . When \tilde{T} is manipulated, it affects Y without altering other variables in the structural model. The hypothetical model, therefore, translates the causal operation of fixing X into the statistical operation of conditioning on \tilde{T} .

Let M_e represent the empirical causal model used to estimate causal parameters and let M_h represent the causal model formulated using the hypothetical variable \tilde{T} . Heckman and Pinto (2015) introduced criteria to systematically connect the counterfactual and empirical distributions. Specifically, for any disjoint set of variables Y, W in the structural causal model defined by \tilde{T} , and for any values $t, t' \in \text{supp}(T)$, the hypothetical model satisfies:

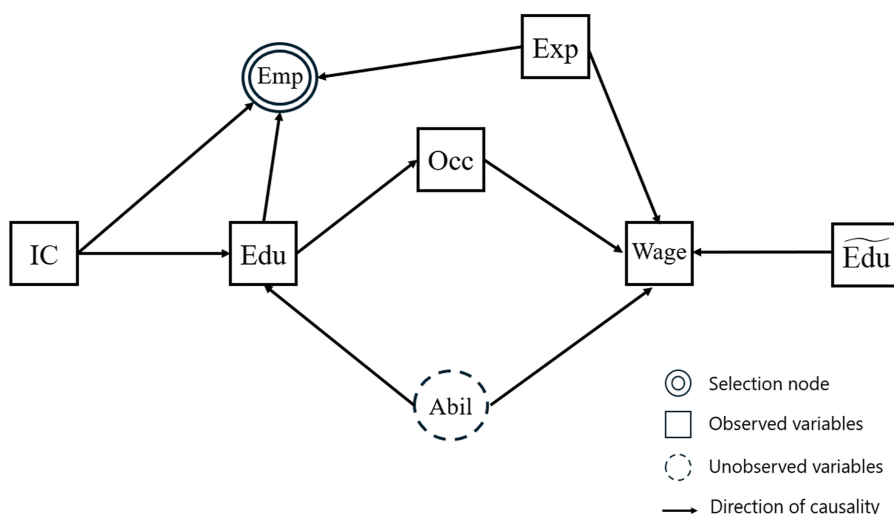
$$Y \perp \tilde{T} \mid (T, W) \Rightarrow P_h(Y|\tilde{T} = t, T = t', W) = P_h(Y|T = t', W) = P_e(Y|T = t', W) \quad (\text{Eq.4})$$

$$Y \perp T \mid (\tilde{T}, W) \Rightarrow P_h(Y|\tilde{T} = t, T = t', W) = P_h(Y|\tilde{T} = t, W) = P_e(Y|T = t, W) \quad (\text{Eq.5})$$

Figure 2 illustrates the hypothetical model corresponding to the causal model depicted in Figure 1. In this hypothetical model, the *Edu*-input in the outcome equation is replaced by the external variable $\tilde{\text{Edu}}$. As a result, in the hypothetical model, there are no arrows pointing from the treatment variable *Edu* to the outcome. This replacement effectively removes the direct influence of the treatment on the outcome, or it fixes the treatment (as fix operator).

Notice that the diagram in Figure 2 specifically illustrates a hypothetical variable setup aimed at estimating the direct effect of education (*Edu*) on wages (*Wage*). The total and indirect effects diagrams are not shown, as they fall outside the scope of this study. For further details on modifying the diagram to identify total, direct, and indirect effects, readers are referred to Heckman and Pinto (2015).

From the LMC of the outcome variable, *Wage* ($\text{Wage} \perp \text{Edu} | (\tilde{\text{Edu}}, \text{Exp}, \text{Occ}, \text{Abil})$), we have that $P_h(\text{Wage} | \tilde{\text{Edu}} = t, \text{Exp}, \text{Occ}, \text{Abil}) = P_e(\text{Wage} | \text{Edu} = t, \text{Exp}, \text{Occ}, \text{Abil})$. However, the variable *Abil* is unobservable, which means that $P_e(\text{Wage} | \text{Edu} = t, \text{Exp}, \text{Occ}, \text{Abil})$ cannot be directly estimated from the data. Therefore, additional assumptions are necessary to enable the estimation of causal parameters from the observed data. These assumptions typically involve either approximating the effects of unobserved variable using observable proxies or



Source(s): Figure by authors

Figure 2. DAG represents hypothetical model for the returns to education with endogenous decision on field of study

employing statistical techniques (such as IV estimation) that allow for the estimation of causal effects in the presence of unobserved confounders.

4.3 Identification of causal parameters

In causal inference, a causal effect (or counterfactual) is identified if it can be expressed in terms of observable data generated by the empirical model (Heckman and Pinto, 2022). Identification, therefore, involves linking hypothetical (counterfactual) outcomes to real-world, observable quantities. In this study, the causal model suggests that variables related to individual characteristics (*IC*), such as parental income, could serve as instrumental variables. However, missing data—particularly non-random missing data, such as when individuals did not live with their parents—on key *IC* variables like parental income could lead to biased estimates if these instruments are used directly (Siririsakulchai and Leucharusmee, 2023).

To address this issue, we adopt a broader identification strategy by employing the control function approach. The control function approach offers a flexible and robust method for dealing with endogeneity by adjusting for unobserved factors that influence both the treatment (educational choices) and the outcome (wages). Heckman and Robb (1985) discussed control function estimators for evaluating the impact of interventions, specifically explaining their applications in both cross-sectional and longitudinal data. An example of a control function approach that relies on parametric assumptions is Heckman's sample selection correction (Heckman, 1979), which corrects for selection bias in the estimation of economic relationships.

Rather than relying solely on potentially incomplete instrumental variables, the control function captures the influence of unobserved factors—such as ability and other latent variables—on both the treatment decision and the outcome. By incorporating the control function into the outcome equation, we adjust for endogeneity bias, ensuring that the causal effect of educational choices on wage outcomes is identified even in the presence of non-random missing data.

This strategy provides a robust alternative to traditional instrumental variable methods, allowing us to estimate the returns to education in a way that is both theoretically sound and empirically feasible given the data limitations. Through this approach, we effectively link the counterfactual outcomes with the observable data, thus achieving identification of the causal effect of education on wages.

From the structural causal model in Figure 1, we redefine the key variables for ease of presentation. Let Y denote the outcome variable, which represents the wage (*Wage*). The variable T represents the level of education (*Edu*), which is the primary endogenous variable of interest in our analysis. The set of exogenous variables, denoted by Z , includes individual characteristics (*IC*), experience (*Exp*), and occupation (*Occ*). To account for unobserved heterogeneity, we introduce V as the set of unobserved variables affecting both the endogenous variable and the outcome, primarily capturing the individual's ability (*Abil*).

We can express the observed outcome in terms of counterfactual outcomes $Y(t) = f_Y(T = t, V, Z, \varepsilon_Y)$ as

$$Y = Y(1) \cdot 1[T = 1] + \dots + Y(J) \cdot 1[T = J], \quad (\text{Eq.6})$$

where $1[\cdot]$ is an indicator function. We introduce the concept of expected utility EU_j^* , which drives the decision-making process and depends on the exogenous variables Z and the unobserved ability factor V . The utility maximization framework allows us to write the outcome equation as:

$$Y = Y(1) \cdot 1 \left[\arg \max_{j=\{1, \dots, J\}} EU_j^* = 1 \right] + \dots + Y(J) \cdot 1 \left[\arg \max_{j=\{1, \dots, J\}} EU_j^* = J \right]. \quad (\text{Eq.7})$$

To facilitate the connection with the empirical model, we introduce the variable d_j as a binary variable representing the observed treatment choice j for $j = 1, 2, \dots, J$. Then, we have

$$Y = Y(1) \cdot 1 \left[\arg \max_{j=\{1, \dots, J\}} Pr(d_j|Z, V) = 1 \right] + \dots + Y(J) \cdot 1 \left[\arg \max_{j=\{1, \dots, J\}} Pr(d_j|Z, V) = J \right], \quad (\text{Eq.8})$$

where $Pr(d_j|Z, V)$ represents the probability of selecting treatment j given the covariates Z and unobserved ability V

Follow Heckman and Pinto (2022), we decompose the counterfactual $Y(j)$ into its mean component μ_j and a deviation from this mean ε_j , such that

$$Y(j) = \mu_j + \varepsilon_j, j = 1, 2, \dots, J. \quad (\text{Eq.9})$$

The unobserved ability V causes $\varepsilon_j, j = 1, 2, \dots, J$. All unobserved variables are statistically independent of the exogenous variables Z . In addition, the unconditional expectation of each ε_j is zero. Thus, we can express the expectation of the outcome Y given Z and the choice of field of study $T \in \{1, 2, \dots, J\}$ as $E_e(Y|Z, X = j) = \mu_j + E_e(\varepsilon_j|Z = z, X = j)$, for $j = 1, 2, \dots, J$.

Given that $d_j = 1$ when the $Pr(d_j = 1|Z, V)$ is maximum, indicates that the individual has optimally selected field j . We can express the expected value of Y given Z and the treatment choice as

$$E_e(Y|Z, T = j) = \mu_j + E_e(\varepsilon_j|Z = z, d_j = 1) \text{ or} \\ E_e(Y|Z, T = j) = \mu_j + E_e\left(\varepsilon_j|Z = z, \max_{j=\{1, \dots, J\}} Pr(d_j = 1|Z, V)\right). \quad (\text{Eq.10})$$

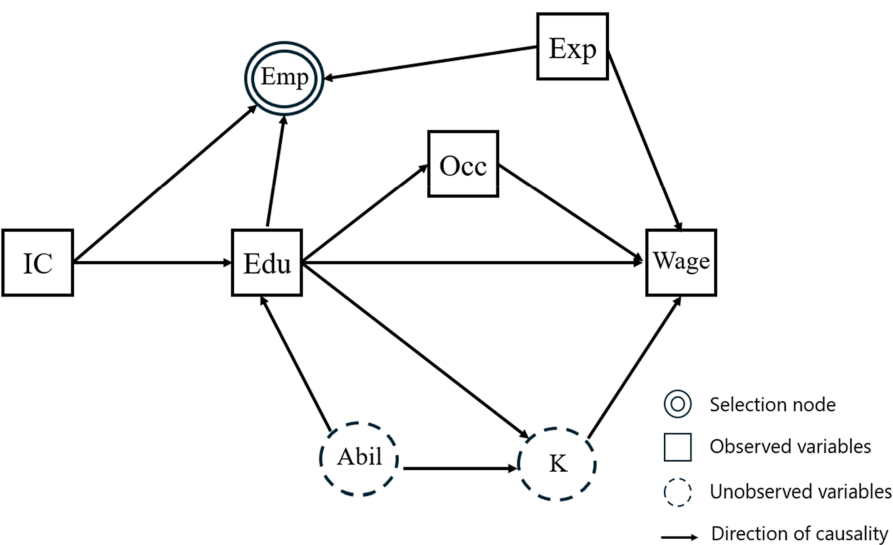
From the independent relationship of Z and $\varepsilon_j, j = 1, 2, \dots, J$, we have that

$$E_e(Y|Z, T = j) = \mu_j + E_e\left(\varepsilon_j \left| \max_{j=\{1, \dots, J\}} Pr(d_j = 1|Z, V) \right.\right). \quad (\text{Eq.11})$$

The expectation of the endogenous error term ε_j can be expressed as a control function of Multivalued Treatment Propensity Scores $K_j(Pr(Z, V))$. This control function adjusts for the endogeneity of the treatment choice, ensuring that the estimates of the returns to education are unbiased and consistent with the underlying causal model. The control function within the causal model is shown in Figure 3.

When constructing the control function, additional assumptions are often required. For example, in the case of a binary treatment $T \in \{0, 1\}$, under the separability assumption of the choice equation, we can express the treatment decision as $T = 1\{Pr(Z) \geq V\}$. This means that if the probability of treatment given the instruments Z is greater than or equal to a certain threshold V , the individual selects $T = 1$. Conversely, if $Pr(Z) < V$, the individual selects $T = 0$. This assumption allows us to model the decision-making process more precisely and construct the appropriate control function to account for potential endogeneity.

Note that, to reduce confusion regarding the notations used for various variables and different tasks in the framework, we outline the approach to defining variable notation as follows. When developing the causal model based on thought experiments and economic theory, we use abbreviations for variables, such as representing education as "Edu." This ensures that the causal diagram is self-explanatory without requiring additional text. For identifying causal effects, we use hypothetical variables by adding a tilde (\sim) to the treatment variable, converting it into a hypothetical variable. This approach allows the hypothetical



Source(s): Figure by authors

Figure 3. DAG represents causal model with control function

model diagram to remain visually intuitive while reflecting the manipulation of the causal graph. When deriving causal probabilities for use in the empirical model, we adopt uppercase English letters commonly used to denote random variables. For instance, Y represents the outcome variable ($Wage$), and T denotes the treatment variable, aligning with standard probability model conventions. Finally, in the empirical model involving sample data, we use lowercase English letters with subscripts to denote observed data for individuals in the sample. For example, y_i represents the wage of individual i . This structured notation ensures consistency and clarity across different sections of the study while maintaining alignment with common econometric and probability modeling practices.

4.4 Empirical model for estimating the treatment effect of field of study

As discussed in the identification strategy in Section 4.3, constructing the control function $K_j(Pr(Z, V))$ requires making parametric assumptions. Specifically, we assume that the latent variable l_j , which proxies for Ability (V) associated with the choice of field of study j . The expected utility for each individual, $i = 1, 2, \dots, N$ can be modeled as $EU_{ij}^* = \alpha_j' z_i + \delta_j l_{ij} + \varepsilon_{ij}$, with independently and identically distributed error terms ε_{ij} , where α_j, δ_j are parameters to be estimated and z_i is the exogenous variable associated with each individual i . We assume that l_{ij} and ε_{ij} are independent of each other. These assumptions, along with the separability assumption, are essential for specifying the probability structure $Pr(d_j|Z, l)$ for multi-treatment group $j = 1, 2, \dots, J$.

Deb and Trivedi (2006) introduced the multinomial treatment effects (MTE) model, which forms the basis for the parametric approach used in this study, with Deb (2009) providing the Stata package for its implementation. This structure allows for flexibility in modeling the heterogeneity of individual choices and the unobserved factors that influence these decisions including field-specific ability. This approach allows for the estimation of treatment effects when the treatment variable (field of study) is multinomial in nature, meaning individuals can choose among multiple educational options. In addition, the inverse Mills ratio is incorporated to the MTE model to correct for selection bias in the observation of income (i.e. selection into paid employment).

The MTE model with sample selection correction consists of two components: a multinomial logit model for the treatment selection (field of study choice) and a regression model for the outcome variable (wage). The selection equation is specified as:

$$\Pr(\mathbf{d}_i | \mathbf{z}_i, \mathbf{l}_i) = g(\mathbf{z}_i' \alpha_1 + \delta_1 l_{i1}, \mathbf{z}_i' \alpha_2 + \delta_2 l_{i2}, \dots, \mathbf{z}_i' \alpha_J + \delta_J l_{iJ}), \quad (\text{Eq.12})$$

where $\mathbf{d}_i = (d_{i1}, d_{i2}, \dots, d_{iJ})$ is a vector of the dummy variable for each field of study j , \mathbf{z}_i is the exogenous observed variable, and $\mathbf{l}_i = (l_{i1}, l_{i2}, \dots, l_{iJ})$ is a vector of unobserved variables affecting the field of study decision, and g is the mixed multinomial logit distribution. For the baseline, $j = 0$ refers to high school or lower education with no field of study.

The outcome equation is given by:

$$E(y_i | \mathbf{d}_i, \mathbf{x}_i, \mathbf{l}_i) = \mathbf{x}_i' \beta + \delta \text{IMR}_i + \sum_{j=1}^J \gamma_j d_{ij} + \sum_{j=1}^J \lambda_j l_{ij}, \quad (\text{Eq.13})$$

where y_i is log of wage, \mathbf{x}_i is the exogenous factors affecting the labor market outcomes, IMR_i is the inverse Mill's ratio for the selection into wage employment, λ_j are the factor-loading parameters capturing the correlation of the treatment and outcome variables through unobserved characteristics and γ_j are the treatment effects of the field of study choice d_j on the wage outcome.

In this model, ability sorting bias arises because individuals with higher abilities in specific fields (A_{ij}) are more likely to select fields of study that yield higher returns. The unobserved variables (l_{ij}) in the model capture these latent abilities, allowing us to account for the influence of ability sorting when estimating the returns to education. By simultaneously estimating the selection and outcome equations, the multinomial treatment model provides consistent estimates of the returns to education for each field of study, thus mitigating the bias introduced by ability sorting.

Notice that, in the DAG shown in Figure 3, the set of variables in \mathbf{z}_i consists of individual characteristics (IC). The set of variables in \mathbf{x}_i includes variables representing education (Edu), occupation (Occ), and experience (Exp). The term $\sum_{j=1}^J \lambda_j l_{ij}$ serves as a control function for the latent variable l_{ij} , which represents ability ($Abil$). Additionally, the inverse Mills ratio (IMR) is included as a control function to address selection bias arising from the decision to work (employment status). This function is modeled as a function of variables representing IC , Edu , and Exp . This clarification ensures a clear understanding of the role of these variables and control functions in the model framework.

5. Data

This study uses Thailand's comprehensive household data from the socioeconomic survey (SES) collected by the national statistical office (NSO) in 2019 (pre-covid) and 2021 (post-covid). The survey composes of 124,874 individuals in 2019 and 130,670 individuals in 2021. As this study estimates returns to education, the scope of the study only covers working aged population whose primary activity is not formal education from 25–60 years old and, thus, there are 64,207 observations in 2019 and 65,885 observations in 2021 [1].

For education and field of study variables, individuals whose highest level of education is high school or lower do not have a specific field of study. In 2019, this baseline group comprised 48,899 individuals, slightly decreasing to 48,376 in 2021.

In contrast, individuals who pursued vocational education, bachelor's degrees, or higher are grouped into specific fields of study. In 2019, 15,308 individuals (23.85% of the sample) pursued higher education in one of these fields, increasing to 17,509 individuals (26.58%) by 2021. These fields of study include teaching, social sciences, business, computer-related

fields, science, healthcare, and other specialized areas. Among these, business and science-related fields had the highest numbers, with notable increases in 2021, reflecting trends in educational preferences.

Regarding employment status, the SES data distinguishes between those in the labor force (employed and unemployed) and those not in the labor force. In this study, the unemployed and non-labor force individuals are grouped as “Not employed or work without pay.” Within the employed group, which includes employers, own account workers, employees, and family contributors without pay, we reclassified contributors in agricultural businesses as “employed with income” by evenly distributing the business’s income among all working members. For non-agricultural businesses, family contributors are classified as “employed with income” only if they report explicit earnings; otherwise, they remain “Not employed or work without pay.”

In 2019, the majority of individuals, 53,337, were classified as “Employed with income,” while 10,870 were “Not employed or work without pay.” By 2021, these numbers slightly increased to 54,828 and 11,057, respectively. This categorization and the reclassification of family contributors provide a clearer understanding of employment trends and income distribution among different educational backgrounds and fields of study. The sample size by employment status and field of study is shown in [Table 1](#).

[Table 2](#) shows the average monthly income by field of study and year. In 2019, healthcare had the highest income at 32,089 THB, which increased to 34,843 THB in 2021. Teaching followed, with income slightly decreasing from 30,501 THB in 2019 to 29,300 THB in 2021. Social science, law, art, and humanities experienced a slight increase in income from 27,702 THB to 28,142 THB. The income in the business and computer-related fields remained stable around 24,000 THB and 20,000 THB, respectively. Science, agriculture, engineering, and architecture also dropped slightly from 24,120 THB to 23,784 THB. High school or lower education had the lowest income, decreasing from 10,009 THB to 9,483 THB. Overall, the total average monthly income remains stable at approximately 13,850 THB.

Table 1. Number of observations by employment status and field of study

Field of study	Year 2019 Not employed or work without pay	Employed with income	Total	Year 2021 Not employed or work without pay	Employed with income	Total
0. High school or lower education (No field of study)	8,596	40,303	48,899	8,378	39,998	48,376
1. Teaching	166	1,655	1,821	153	1,699	1,852
2. Social science, law, art, and humanities	291	1,715	2,006	394	1,956	2,350
3. Business	1,112	4,599	5,711	1,323	5,255	6,578
4. Computer related	132	695	827	139	839	978
5. Science, agriculture, engineer, and architect	406	3,393	3,799	514	3,925	4,439
6. Healthcare	64	645	709	54	753	807
7. Other fields	103	332	435	102	403	505
Total	10,870	53,337	64,207	11,057	54,828	65,885

Source(s): Authors’ calculation using the SES 2019 and SES 2021 datasets for the sample of individuals aged 25–60 years whose primary activity is not attending school, without population weighting

Table 2. Average monthly income by field of study and year

Field of study	Year 2019	Year 2021	Total
0. High school or lower education (No field of study)	10,009	9,483	9,747
1. Teaching	30,501	29,300	29,891
2. Social science, law, art, and humanities	27,702	28,142	27,937
3. Business	24,205	24,026	24,110
4. Computer related	20,402	20,764	20,603
5. Science, agriculture, engineer, and architect	24,120	23,784	23,941
6. Healthcare	32,089	34,843	33,530
7. Other fields	22,823	21,142	21,902
<i>Total</i>	<i>13,871</i>	<i>13,822</i>	<i>13,846</i>

Source(s): Authors' calculation using the SES 2019 and SES 2021 datasets for the sample of individuals aged 25–60 years whose primary activity is not attending school, with population weighting

6. Results

6.1 Returns to higher education by field of study

The primary objective of this analysis is to estimate the returns to higher education by field of study in Thailand, using log monthly income as the outcome variable. Higher education in this context includes vocational education, college, and above, with the baseline category being individuals who have completed high school or lower education without pursuing a specific field of study. By estimating the returns to different fields of study, we aim to provide insights into how educational choices impact income, particularly in the pre- and post-COVID-19 periods.

This study employs two models to estimate returns to higher education by field of study: the Heckman sample-selection model (HM) and the multinomial treatment effects model with sample selection correction (MTE). The HM model corrects for sample selection bias by using the inverse Mills ratio (IMR), ensuring that the estimated returns reflect the true relationship between education and income, even when not all individuals are employed or report income. The MTE goes further by addressing ability sorting, where individuals choose fields based on their abilities, potentially biasing returns if not accounted for. By modeling both field selection and wage outcomes, the MTE controls for unobserved characteristics that influence both. The significance of the IMR in HM and MTE highlights the importance of correcting for both sample selection and ability sorting to avoid biased estimates.

6.1.1 Returns to higher education by field of study in 2019. In 2019, the returns to higher education, as estimated by the Multinomial Treatment Effect (MTE) model illustrated in Table 3, showed significant variation across fields of study. The teaching field had the highest returns, with an MTE estimate of 1.664. This suggests that individuals who pursued higher education in teaching (including vocational, Bachelor's, or higher levels) earned 166.4% more than those with only a high school education or lower. This was followed by the healthcare field, with an MTE estimate of 1.387, and the social sciences, law, arts, and humanities field at 1.227. The computer-related field and business field had returns of 1.178 and 1.127, respectively. While the science, agriculture, engineering, and architecture fields also showed positive returns, these estimates were not statistically significant.

The differences between the MTE and HM estimates indicate the presence of ability sorting bias. In all fields, except for the science, agriculture, engineering, and architecture field, a slight positive difference between MTE and HM suggests that individuals choosing these fields tend to have higher unobserved abilities or a stronger preference for financial returns. In contrast, the negative difference observed in the science, agriculture, engineering, and architecture fields suggests that these fields attract individuals with lower unobserved abilities or a preference for non-monetary benefits, leading to an underestimation of returns in the HM model.

Table 3. Estimation of returns to higher education by field of study in 2019 and 2021 using Heckman sample-selection model (HM) and the multinomial treatment effects model with sample selection correction (MTE)

	(HM2019) ln(income)	(HM2021) ln(income)	(MTE2019) ln(income)	(MTE2021) ln(income)
<i>Field of study</i>				
1. Teaching	1.321*** (0.045)	1.335*** (0.053)	1.664*** (0.205)	1.571*** (0.138)
2. Social science, law, art, and humanities	1.074*** (0.048)	1.163*** (0.049)	1.227*** (0.143)	1.291*** (0.137)
3. Business	1.019*** (0.035)	1.136*** (0.039)	1.127*** (0.328)	1.432*** (0.105)
4. Computer related	0.808*** (0.089)	0.904*** (0.051)	1.178*** (0.198)	1.091*** (0.101)
5. Science, agriculture, engineer, and architect	0.932*** (0.044)	1.016*** (0.042)	0.066 (0.088)	0.112* (0.061)
6. Healthcare	1.159*** (0.075)	1.471*** (0.047)	1.387*** (0.174)	1.943*** (0.089)
7. Other fields	0.963*** (0.074)	1.005*** (0.108)	1.304*** (0.093)	1.158*** (0.163)
Female	0.178*** (0.033)	0.182*** (0.033)	0.039 (0.121)	0.005 (0.113)
Age	0.066*** (0.009)	0.068*** (0.010)	0.118*** (0.014)	0.140*** (0.014)
Age ²	−0.001*** (0.000)	−0.001*** (0.000)	−0.002*** (0.000)	−0.002*** (0.000)
Urban	0.995*** (0.032)	0.927*** (0.031)	1.067*** (0.047)	0.979*** (0.047)
athrho	−2.534*** (0.029)	−2.562*** (0.039)		
lnsigma	1.151*** (0.012)	1.146*** (0.012)		
IMR			−2.839*** (0.475)	−2.792*** (0.491)
Constant	7.396*** (0.196)	7.381*** (0.196)	6.655*** (0.288)	6.247*** (0.279)
Observations	64,207	65,885	53,337	54,828

Note(s): (1) Bootstrapped standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)
(2) Explanatory variables used for the selection (into employment) equation in the HM models and the construction of the Inverse Mill's Ration (IMR) for the MTE models include age, and dummy variables for married, urban, college, and vocational degree
(3) The MTE models were estimated using the `mtreatreg` command in Stata, developed by [Deb \(2009\)](#). Full estimation results are provided in [Appendix A](#)
Source(s): Authors' estimations using the SES 2019 and SES 2021 datasets for the sample of individuals aged 25–60 years whose primary activity is not attending school (with population weighting)

6.1.2 Returns to higher education by field of study in 2021: a comparison. The comparison of the returns to higher education between 2019 and 2021 is shown in [Table 4](#). In 2021, the return to education in the healthcare field increased by 40.1%, becoming the highest among all fields with an MTE estimate of 1.943. This substantial increase likely reflects the heightened demand for healthcare professionals during the COVID-19 pandemic. The teaching field maintained a high return, with an MTE estimate of 1.571, a slight and statistically insignificant decline from 1.664 in 2019. Other fields exhibited some changes, but the changes were not statistically significant. This lack of significance may indicate either no systematic change or heterogeneous effects across subgroups. Overall, returns remained relatively stable across most fields, except for the marked increase observed in healthcare.

Table 4. Returns to higher education in 2019 and 2021 by field of study

Fields of study	Returns to education 2019	Returns to education 2021	Difference	SE	t-stat	p-value
1. Teaching	1.664	1.571	−0.093	0.246	−0.378	0.706
2. Social science, law, art, and humanities	1.227	1.291	0.064	0.199	0.322	0.747
3. Business	1.127	1.432	0.306	0.345	0.887	0.375
4. Computer related	1.178	1.091	−0.088	0.223	−0.395	0.693
5. Science, agriculture, engineer, and architect	0.066	0.112	0.045	0.106	0.424	0.672
6. Healthcare	1.387	1.943	0.556	0.196	2.841	0.004

Note(s): Comparison of MTE2019 and MTE2021 estimates

Source(s): Authors' calculation using the SES 2019 and SES 2021 datasets

In addition to field of study, gender, age, and urban residence also influence income levels. For the female variable, the coefficients in both MTE2019 and MTE2021 are not statistically significant, suggesting no substantial gender wage gap after accounting for selection effects. The coefficients for age and age squared indicate a nonlinear relationship with income. The results suggest that income initially increases with age, reflecting returns to experience, but this growth diminishes and eventually reverses at later stages of the life cycle. The turning point occurs earlier in MTE2019 compared to MTE2021, suggesting that income growth with age was slightly more sustained in 2021. Urban residence consistently shows a substantial and statistically significant wage premium, with coefficients of 1.067 in MTE2019 and 0.979 in MTE2021, though the slight decline suggests a marginal reduction in urban advantages.

6.2 Occupational status as a mediator for the effects of education to income

The net returns to education for each field of study are influenced by differences in occupational statuses. This section estimates returns while controlling for occupation, including (1) employer (business owners with at least one employee), (2) own account workers (agriculture), (3) own account workers (non-agriculture), with (4) public or private employee as the baseline. The comparison between models with and without occupational status control highlights the indirect effect of field of study on wage outcomes through occupational status sorting.

The results from both 2019 and 2021 show a consistent pattern of income across different occupational statuses. Employers and employees, who serve as the base group in the regressions, had the highest average monthly income. In contrast, own-account workers—business owners without employees or freelancers—earned lower incomes, especially in the agricultural sector. Regarding the effect of COVID-19 in 2021, the income of employers and own-account workers outside the agricultural sector dropped more sharply compared to the employee group.

When examining the returns to higher education by field of study while controlling for occupation (Model MTE2019oc in Table 5), the 2019 estimates were approximately 50% lower across most fields compared to the estimates without controlling for occupational status (Model MTE2019 in Table 3). This substantial reduction suggests that a significant portion of the observed returns to education is mediated through occupational sorting, particularly into roles such as employer and salaried employee positions, which are associated with higher earnings. In all fields, the returns declined when controlling for occupation, but this effect was most pronounced in the case of the science, agriculture, engineering, and architecture field. For this group, the estimated return changed from a slight positive (0.066) to a negative value (−0.294), representing a more extreme reduction compared to other fields. This suggests that

Table 5. Estimation of returns to higher education by field of study controlled for occupational status in 2019 and 2021

	(MTE2019oc) ln(income)	(MTE2021oc) ln(income)
<i>Field of study</i>		
1. Teaching	0.826*** (0.314)	0.729*** (0.131)
2. Social science, law, art, and humanities	0.615*** (0.232)	0.717*** (0.129)
3. Business	0.590 (0.363)	0.815*** (0.119)
4. Computer related	0.660** (0.260)	0.599*** (0.107)
5. Science, agriculture, engineer, and architect	−0.294*** (0.073)	−0.196*** (0.059)
6. Healthcare	0.705*** (0.174)	1.186*** (0.081)
7. Other fields	0.736*** (0.089)	0.636*** (0.149)
<i>Occupation</i>		
1. Employer	0.001 (0.117)	−0.016 (0.091)
2. Own account workers (Agri)	−2.722*** (0.048)	−2.735*** (0.048)
3. Own account workers (Non-agri)	−0.187*** (0.033)	−0.306*** (0.032)
Female	0.274** (0.111)	0.277*** (0.102)
Age	0.103*** (0.013)	0.112*** (0.013)
Age ²	−0.001*** (0.000)	−0.001*** (0.000)
Urban	0.480*** (0.041)	0.468*** (0.040)
IMR	−3.355*** (0.429)	−3.440*** (0.445)
Constant	7.562*** (0.268)	7.439*** (0.253)
Observations	53,337	54,828

Note(s): (1) Bootstrapped standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)
(2) Explanatory variables used for the selection (into employment) equation in the HM models and the construction of the Inverse Mill's Ration (IMR) for the MTE models include age, and dummy variables for married, urban, college, and vocational degree
(3) The MTE models were estimated using the `mtreatreg` command in Stata, developed by Deb (2009). Full estimation results are provided in [Appendix A](#)
Source(s): Authors' estimations using the SES 2019 and SES 2021 datasets for a sample of individuals aged 25–60 years whose primary activity is not attending school (with population weighting)

the higher earnings seen among graduates in these fields are largely driven by their entry into higher-paying occupations rather than the intrinsic value of the degree.

Controlling for occupation also alters the estimated coefficients for gender, age, and urban variables, highlighting the role of occupational sorting in wage differences. For the female variable, the coefficient increased from 0.039 to 0.274 when accounting for occupation, indicating that the initially lower earnings for women are largely due to differences in occupational roles. Women are less likely to be employers (1.50% for women vs 3.35% for

men) and more likely to work as own-account workers outside agriculture (27.99 vs 25.58% for men), who earn less income on the average [2]. Once adjusted for occupational sorting, the higher earnings potential of women within similar roles becomes evident. The coefficients for age and age squared demonstrate the nonlinear impact of experience on earnings for both models. Controlling for occupation reduces the age coefficient (from 0.118 to 0.103) and slightly diminishes the absolute value of the age squared coefficient (from -0.002 to -0.001), suggesting that part of the age-related earnings advantage is linked to occupational sorting. The marginal effect of age remains positive until later stages of the life cycle, with the turning point shifting from age 29.5 in MTE2019 to age 51.5 when occupations are controlled. This indicates that occupational sorting amplifies returns for younger workers while mitigating declines for older workers [3]. For urban residence, the coefficient decreases significantly (from 1.067 to 0.480) after adjusting for occupation, indicating that much of the urban wage premium arises from sorting into higher-paying occupations. Regardless, the remaining positive effect highlights advantages related to location, such as better access to high-paying occupations, independent of occupational factors.

The results for 2021 show similar patterns to those observed in 2019 when comparing estimates with and without controlling for occupation. Across both years, controlling for occupation consistently reduces the estimated coefficients for all fields of study, indicating the significant role of occupational sorting in wage differences.

7. Conclusion

This study applies the econometric causal framework developed by Heckman and Pinto (2022) to estimate the returns to higher education by field of study in Thailand, with a focus on comparing the pre- and post-COVID-19 periods. Using data from Thailand's socioeconomic surveys (SES) collected in 2019 and 2021, we analyzed a sample of individuals aged 25–60 who are not engaged in formal education, examining how the pandemic has influenced educational returns across various fields.

Central to our approach is the application of the economic causal framework, linking the returns to education with an endogenous field of study model. This framework accounts for both observed and unobserved abilities affecting individuals' choices of field of study and subsequent labor market outcomes, as illustrated in Figures 1–3. We employ the control function approach to adjust for unobserved factors, using the multinomial treatment effects (MTE) model by Deb (2009) with sample selection correction as the main empirical tool. This method enables accurate estimation of the treatment effects of different fields of study on income while addressing biases from ability sorting and sample selection.

The results reveal significant variation in the returns to higher education by field of study. In 2019, teaching and healthcare had the highest returns (166.4 and 138.7% higher income, respectively), while returns for science, agriculture, engineering, and architecture were minimal or not significant. By 2021, healthcare exhibited the highest return (194.3%), likely driven by increased demand during the COVID-19 pandemic, while teaching maintained strong but slightly lower returns (157.1%). Across both years, controlling for occupation reduced the estimated returns by about 50%, indicating that a substantial part of the observed returns is due to occupational sorting into higher-paying roles. This effect was most pronounced for science, agriculture, engineering, and architecture, where positive returns turned significantly negative, suggesting that the earnings advantage is primarily from access to well-paid jobs rather than field-specific skills. The analysis of demographic factors further highlighted the role of occupational sorting, particularly for gender, age, and urban residence, underscoring the continued importance of experience and access to high-paying industries in explaining wage differences.

The variation in returns across fields of study suggests a need for policies that better align educational offerings with labor market demand. High returns in fields like teaching and healthcare indicate strong labor market value. The low or negative returns in science,

agriculture, engineering, and architecture suggest potential issue and require further study to explore specific causes and develop appropriate strategies. Additionally, efforts should be made to facilitate transitions into higher-paying occupations, ensuring that the benefits of higher education are equitably distributed. This can begin by systematically estimating and reporting returns to various types of education, enabling individuals to make informed decisions.

However, the study has limitations. The reliance on cross-sectional data and the lack of household panel data in Thailand with detailed field of study information are significant constraints. Small sample sizes in certain fields may also limit the precision of the estimates. Nonetheless, the dataset used remains the best available for this type of analysis in Thailand. Future research could build on these findings by considering the matching between educational fields and occupational outcomes, as highlighted by Lemieux (2014), and by incorporating factors like unemployment risk and income variability to better understand wage differentials across fields.

Notes

1. According to Thailand's NSO, the working aged population is 15 years old and above. However, as this study aims to examine the return to education for those who already finished their formal education, the age range is selected to be 25–60 years old.
2. The distribution of male and female workers across occupation was tabulated using the SES 2019 data.
3. The marginal effect of age is calculated as: $\frac{d(\ln(\text{income}))}{d(\text{Age})} = \beta_{\text{Age}} + 2 \cdot \beta_{\text{Age}^2} \cdot \text{Age}$. Unlike other variables with constant effects, the marginal effect of age varies across the life cycle due to the quadratic specification. In MTE2019 (without occupation controls), the marginal effect turns negative at age 29.5 years old. In the model with occupation controls, the turning point shifts 51.5 years old.

References

- Altonji, J.G., Arcidiacono, P. and Maurel, A. (2016), "The analysis of field choice in college and graduate school: Determinants and wage effects", in Hanushek, E., Machin, S. and Woessmann, L. (Eds), *Handbook of the Economics of Education*, Elsevier, Vol. 5, pp. 305-396, doi: [10.1016/b978-0-444-63459-7.00007-5](https://doi.org/10.1016/b978-0-444-63459-7.00007-5).
- Arcidiacono, P. (2004), "Ability sorting and the returns to college major", *Journal of Econometrics*, Vol. 121 Nos. 1-2, pp. 343-375, doi: [10.1016/j.jeconom.2003.10.010](https://doi.org/10.1016/j.jeconom.2003.10.010).
- Asian Development Bank (ADB) (2021), *COVID-19 and Labor Markets in Southeast Asia: Impacts on Indonesia, Malaysia, the Philippines, Thailand, and Viet Nam*, Asian Development Bank, Manila, available at: <https://www.adb.org/publications/covid-19-labor-markets-southeast-asia>
- Bjerkholt, O. and Qin, D. (Eds). (2011), *A Dynamic Approach to Economic Theory: Lectures by Ragnar Frisch at Yale University*, Routledge.
- Bol, T. and Heisig, J.P. (2021), "Explaining wage differentials by field of study among higher education graduates: evidence from a large-scale survey of adult skills", *Social Science Research*, Vol. 99, 102594, doi: [10.1016/j.ssresearch.2021.102594](https://doi.org/10.1016/j.ssresearch.2021.102594).
- Canaan, S. and Mouganie, P. (2018), "Returns to education quality for low-skilled students: evidence from a discontinuity", *Journal of Labor Economics*, Vol. 36 No. 2, pp. 395-436.
- Deb, P. (2009), *MTREATREG: Stata Module to Fits Models with Multinomial Treatments and Continuous, Count and Binary Outcomes Using Maximum Simulated Likelihood*, Statistical Software Components, available at: <https://ideas.repec.org/c/boc/bocode/s457064.html>
- Deb, P. and Trivedi, P.K. (2006), "Maximum simulated likelihood estimation of a negative binomial regression model with multinomial endogenous treatment", *STATA Journal*, Vol. 6 No. 2, pp. 246-255, doi: [10.1177/1536867x0600600206](https://doi.org/10.1177/1536867x0600600206).

- Eberhard, B., Podio, M., Pérez Alonso, A., Radovica, E., Avotina, L., Peiseniece, L., Caamaño Sendon, M., Gonzales Lozano, A. and Solé-Pla, J. (2017), "Smart work: the transformation of the labour market due to the fourth industrial revolution", *International Journal of Business and Economic Sciences Applied Research*, Vol. 10 No. 3, pp. 47-66.
- Goldin, C. and Katz, L.F. (2009), *The Race between Education and Technology*, Harvard University Press.
- Frisch, R. (1938), "Autonomy of economic relations: Statistical versus theoretical relations in economic macrodynamics", Paper given at League of Nations Reprinted in D.F. Hendry and M.S. Morgan (1995) *The Foundations of Econometric Analysis*, Cambridge University Press.
- Haavelmo, T. (1943), "The statistical implications of a system of simultaneous equations", *Econometrica*, Vol. 11 No. 1, pp. 1-12, doi: [10.2307/1905714](https://doi.org/10.2307/1905714).
- Heckman, J.J. (1979), "Sample selection bias as a specification error", *Econometrica*, Vol. 47 No. 1, pp. 153-161, doi: [10.2307/1912352](https://doi.org/10.2307/1912352).
- Heckman, J.J. and Pinto, R. (2015), "Causal analysis after Haavelmo", *Econometric Theory*, Vol. 31 No. 1, pp. 115-151, doi: [10.1017/s026646661400022x](https://doi.org/10.1017/s026646661400022x).
- Heckman, J.J. and Pinto, R. (2022), "The econometric model for causal policy analysis", *Annual review of economics*, Vol. 14 No. 1, pp. 893-923, doi: [10.1146/annurev-economics-051520-015456](https://doi.org/10.1146/annurev-economics-051520-015456).
- Heckman, J.J. and Pinto, R. (2024), "Econometric causality: the central role of thought experiments", *Journal of Econometrics*, Vol. 243 Nos. 1-2, 105719, doi: [10.1016/j.jeconom.2024.105719](https://doi.org/10.1016/j.jeconom.2024.105719).
- Heckman, J.J. and Robb, R. Jr (1985), "Alternative methods for evaluating the impact of interventions: an overview", *Journal of Econometrics*, Vol. 30 Nos. 1-2, pp. 239-267, doi: [10.1016/0304-4076\(85\)90139-3](https://doi.org/10.1016/0304-4076(85)90139-3).
- Kirkeboen, L.J., Leuven, E. and Mogstad, M. (2016), "Field of study, earnings, and self-selection", *Quarterly Journal of Economics*, Vol. 131 No. 3, pp. 1057-1111, doi: [10.1093/qje/qjw019](https://doi.org/10.1093/qje/qjw019).
- Lemieux, T. (2014), "Occupations, fields of study and returns to education", *Canadian Journal of Economics/Revue canadienne d'économique*, Vol. 47 No. 4, pp. 1047-1077, doi: [10.1111/caje.12116](https://doi.org/10.1111/caje.12116).
- National Statistical Office (NSO) (2022), *Statistical Yearbook Thailand 2022*, National Statistical Office, Bangkok.
- Psacharopoulos, G. and Patrinos, H.A. (2018), "Returns to investment in education: a decennial review of the global literature", *Education Economics*, Vol. 26 No. 5, pp. 445-458, doi: [10.1080/09645292.2018.1484426](https://doi.org/10.1080/09645292.2018.1484426).
- Roy, A.D. (1951), "Some thoughts on the distribution of earnings", *Oxford Economic Papers*, Vol. 3 No. 2, pp. 135-146, doi: [10.1093/oxfordjournals.oep.a041827](https://doi.org/10.1093/oxfordjournals.oep.a041827).
- Sirisrisakulchai, J. and Leucharusmee, S. (2023), "Revisiting returns to education in Thailand: structural causal model framework", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 31 No. 2, pp. 223-241, doi: [10.1142/s0218488523400135](https://doi.org/10.1142/s0218488523400135).
- Tinbergen, J. (1975), *Income Distribution: Analysis and Policies*, North- Holland.
- UNESCO (2022), "Higher education global data report", *The 3rd UNESCO World Higher Education Conference*, available at: <https://www.right-to-education.org/resource/higher-education-global-data-report-contribution-world-higher-education-conference>
- World Economic Forum (WEF) (2020), *The Future of Jobs Report 2020*, World Economic Forum, Geneva, available at: <https://www.weforum.org/publications/the-future-of-jobs-report-2020/>

Supplementary material

The supplementary material for this article can be found online.

Corresponding author

Supanika Leucharusmee can be contacted at: supanika.l@cmu.ac.th

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com