Textbook citations increase subsequent citations in economics, except for the most cited articles

by

Danae Bouwer

Dissertation presented in partial fulfilment of the

requirements for the degree of

Master of Philosophy (Financial Technology)

in the

Department of Economics

University of Cape Town

Supervisor: Prof. Co-Pierre Georg

2023

Plagiarism Declaration

COMPULSORY DECLARATION:

1. This dissertation has been submitted to Turnitin (or equivalent similarity and originality checking software) and I confirm that my supervisor has seen my report and any concerns revealed by such have been resolved with my supervisor.
2. I certify that I have received Ethics approval (if applicable) from the Commerce Ethics Committee.
3. This work has not been previously submitted in whole, or in part, for the award of any degree in this or any other university. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works of other people has been attributed, and has been cited and referenced.

Student number	BWRDAN003
Student name	Danae Bouwer
Signature of Student	Signed by candidate
Date:	12/10/2023

Acknowledgements

I am deeply grateful to all the individuals who have been instrumental both in my completing this research and in supporting me through my master's degree journey.

My heartfelt appreciation goes out to my supervisor, Prof. Co-Pierre Georg, for his unwavering guidance, invaluable insights, and, most importantly, his extreme patience throughout the entire research process.

I extend my thanks to my employer, the Financial Innovation Hub, for fostering an inspiring academic environment, providing ample time, and offering the financial and research resources essential for the completion of this dissertation.

I am sincerely indebted to my colleagues and peers at the Financial Innovation Hub who generously granted me the time and flexibility to work on this dissertation. Your enduring patience with my moments of stress and complaints over the past year and more is truly appreciated.

To my family and friends, I owe a debt of gratitude for their unwavering support and encouragement throughout this journey. My parents, grandparents, and sister – your love and belief in me have sustained me through the challenges and triumphs. To my partner, Thomas, your shoulder to lean on and to cry on has been a source of comfort and strength. My friends, your understanding, patience, and acceptance of my countless excuses and absences from gatherings have not gone unnoticed (I might need to come up with new excuses if I wish to skip an event now!).

Last, and most importantly, I owe a heartfelt thank you to my remarkable brain, the true goddess of genius. Without your boundless determination, strength, and intelligence, I would not have navigated these years of uphill battles.

Danae Bouwer

Abstract

This dissertation explores the intricate dynamics of information attribution and dissemination within academia, focusing on how textbook citation events influence subsequent citation patterns of academic literature. Rooted in the science of science, this study employs a comprehensive dataset of collaboration, affiliation, and citation metrics. The primary objective has been to uncover the nuanced mechanisms underlying the provenance of information by investigating how the attribution of ideas changes following a "notable event" – the citation of academic literature in well-known microeconomics textbooks.

Defining two distinct effects associated with a "notable event" that could influence the subsequent citation counts of academic literature, this study examined the cessation effect, linked with decreasing citations, and the signalling effect, tied to increasing citations. Constructing a curated citation dataset of 610 observations and 18 variables, articles were categorised into treatment and control groups. The treatment group includes articles cited in microeconomics textbooks, while the control group, meticulously matched to ensure comparability, comprises uncited articles. A multiple linear regression model contrasts the citation trajectories of cited and uncited articles after the textbook citation event, simultaneously exploring the influence of various control variables on citation count.

My findings shed light on the significant impact of textbook citation events on subsequent citation counts, ultimately enhancing the attribution of ideas. Notably, cited articles consistently achieve higher post-event citations, highlighting the reinforcement of attribution and confirming the presence of a signalling effect. However, for highly cited articles, a theoretical reduction in post-event citations was observed, suggesting a potential weakening of attribution and indicating the presence of a cessation effect.

In addition to these central findings, the analysis of control variables revealed compelling insights. A positive correlation was identified between journal ranking, article length, and author citedness with citation counts, while a negative relationship surfaced between title length and citations. These secondary findings contributed to a deeper understanding of citation dynamics within academia.

Drawing practical implications from these findings, this study affirms the robustness and enduring nature of academic referencing practices, elucidating that citing sources in textbooks significantly bolsters the preservation of information attribution. Furthermore, by drawing parallels between academic citation and data lineage, the research underscores the broader implications for data provenance, emphasising the importance of documenting data origins, sources, and methods thoroughly to ensure data reliability and traceability.

Overall, this research offers a comprehensive analysis of the impact of textbook citation events on information attribution and dissemination within the field of economics. The study serves as a valuable contribution to the science of science, enriching the understanding of citation patterns and the complex social dynamics that shape the provenance of information in academia and beyond.

Keywords: textbook citations, science of science, citation analysis, attribution, dissemination, signalling effect, cessation effect, data provenance.

Table of Contents

Table of Figures

List of Tables

Chapter 1: Introduction

This dissertation is rooted in the emerging field of the *science of science*, employing a comprehensive dataset of collaboration, affiliation, and citation metrics. The primary objective is to investigate citation patterns of academic literature referenced in microeconomics textbooks. Employing a methodology commonly known as the *textbook approach*, I analyse how a notable event which, I argue, is the citing of academic literature in textbooks, influences the attribution and dissemination of ideas, offering insights about the provenance of information within academia. I have investigated the following two research questions:

1. Does citing academic literature in a textbook affect the subsequent attribution thereof?
2. If subsequent attribution is affected, is it strengthened or weakened, and what mechanisms are possibly responsible for this change?

Citing sources in academia is a mandatory social standard when researchers write and publish papers (Hunter, 2006). This practice emerged in the mid-19th century and has invariably been applied in the field of science ever since (De Solla Price, 1986). As a result, academic referencing provides a reliable, consistent and standardised record of the origin and history of information, which is crucial to understanding the attribution and dissemination of scientific knowledge (Hyland, 1999). Consequently, the field of science, or academia, serves as an interesting laboratory for the provenance of information. By employing citation analysis, a method used to study academic references, one can trace the ownership of ideas to specific academic research articles (attribution) and expose how these ideas spread (dissemination). The spread of ideas gradually manifests as discernible citation patterns, and in this dissertation, I propose that this happens after a *notable event*. The literature extensively investigates the impact of certain notable events on subsequent citation patterns. Aizenman and Kletzer (2011) focus on the premature death of prominent economists throughout history as a notable event, contending that it exerts a significant negative effect on citation patterns. They emphasise the pivotal role of academic networking among economists as a key factor contributing to this influence. Alternatively, McMahan and McFarland (2021) identify citations in review articles as a notable event, revealing a noteworthy negative effect on subsequent citation patterns. Their study suggests that researchers tend to cite review articles instead of the original papers, leading

to discernible shifts in citation behaviours. Further expanding on the inquiry, Card and DellaVigna (2013) delve into the effect of citations in prominent economics journals, unearthing evidence of a significant positive impact. They attribute this effect to the perceived superiority of articles cited within these prestigious journals, thus highlighting the significance of academic prestige in shaping citation dynamics. Additionally, Bjork, Offer and Söderberg (2013) investigate the notable event of winning the Nobel Prize and its influence on subsequent citation patterns for the authors awarded this prestigious honour. Their research uncovers a pronounced positive effect, indicating that this event acts as a signal of the importance and relevance of specific authors and their ideas in the field, resulting in increased citation patterns. In Yuret's (2023) recent study, a notable event is defined as a citation in a widely used microeconomics textbook, and the findings reveal a compelling link to increased citation patterns for the literature cited within this influential text. The author attributes this pronounced pattern to the heightened attention and visibility that the textbook confers upon the cited works (Yuret, 2023).

Like Yuret (2023), I also define a notable event as a citation in well-known microeconomics textbooks. I chose textbooks for several reasons. First, textbooks summarise major concepts and theories produced by research in the literature and play a pivotal role in the certification of knowledge (Rotfeld, 2000). For example, Thomas Kuhn discusses the importance of textbooks in science and academic disciplines and asserts that textbooks are the foremost means of education for encapsulating the knowledge in a particular field (Kuhn, 1970). Similarly, in his work, De Solla Price (1986) posits that extensive volumes of scholarly research eventually undergo consolidation into textbooks, effectively presenting the essential information amassed in a particular field. Second, there is a scarcity of research focusing specifically on examining how textbooks affect subsequent citation patterns. This highlights a gap in current knowledge regarding the impact of textbooks on citation practices in academic literature. Third, while the scientific community acknowledges textbooks as the culmination of where the current corpus of scientific research is consolidated and presented as factual information, the fate of this knowledge, after it undergoes verification within a textbook, remains uncertain.

A unique opportunity thus arises from studying the changes in citation patterns of research articles that embody the most significant ideas acknowledged and endorsed by the scientific community. This allows me to determine whether the attribution of ownership of these ideas remains intact or becomes diluted following the citation event in a textbook. This offers

captivating insights into the provenance of information within the realm of science because, as suggested, academia has a unique social convention to attribute to the origins of ideas. By using citation patterns to understand attribution, it is possible to estimate how strong or binding this social convention really is. Additionally, one can use the newfound understanding of attribution to learn about the potential value of creating and maintaining provenance in other types of information, specifically data. Data provenance refers to the complete history of the origin, custody, and transformations throughout the life cycle of data (Gupta, 2009). Studying how information ownership is established, preserved, and eventually lost in academia can offer valuable insight into best practices for creating and maintaining provenance in data. For example, one can learn about the importance of documenting the sources and methods used to generate data and the need to maintain this information over time to ensure that the data remain reliable and useful.

Following textbook citation analysis or, more formally, the *textbook approach*, I collected a treatment group of research articles cited in nine well-known microeconomic textbooks. At the same time, I created a control group of comparable research articles not cited in any of the nine textbooks. I used the regression model below and ran various multivariate regressions to compare the count of citations before and after the notable citation event of the treatment group with that of the control group.

$$\log(\widehat{Citations12YearsAfter}) = \beta_0 + \delta_0 CitedInBook + \beta_1 \log(Citations10YearsBefore - Avg) + \delta_1 CitedInBook\# \log(Citation10YearsBefore - Avg) + control\ variables + fixed\ effects + u \quad (1)$$

I regressed three variables of interest, namely, *"Citations 10 Years Before"*, *"Cited In Book"* and *"Cited In Book#Citations 10 Years Before – Avg"* on *"Citations 12 Years After"* while controlling for author, article and journal level variables. *"Citations 10 Years Before"* represents pre-event citations and measures aggregate citation count ten years before the research articles are cited in the microeconomics textbook. *"Cited In Book"* is a dummy that measures whether or not the article is cited in the microeconomics textbook, and *"Cited In Book#Citations 10 Years Before – Avg"* is an interaction term between the dummy and the pre-event citation variables. *"Citations 12 Years After"* represents post-event citations and measures aggregate citation count 12 years after research articles are cited in the microeconomics textbook. This comparative analysis determines whether a notable citation event affects subsequent citation count and, if it does, whether a) the treatment group is cited

less in the years following the notable event compared to the control group, or b) the treatment group is cited more in the years following the notable event compared to the control group. As I asserted previously, a change in attribution can be inferred from the shift in citation patterns. In this dissertation, this pattern becomes evident when examining the disparity in citation counts between the treatment and control groups. Consequently, the fewer citations observed in the treatment group imply a loss of attribution, indicating that the provenance of information within academia diminishes when research is cited in textbooks. Conversely, the higher citations observed in the treatment group imply no loss of attribution, thereby demonstrating that provenance remains intact. Building on this premise, the subsequent exploration delves into two potential explanations for this change in attribution, specifically referred to as the *cessation effect* and the *signalling effect*.

The cessation effect can be observed when citation counts decrease after a notable event. It describes a point at which the citing of research articles becomes obsolete. Several reasons could explain this mechanism. For instance, when an article is cited in a well-known economics textbook, this is most likely because it supports a common and prominent economic theory (Rothman, 1971). Thus, it is possible that some of the academic literature included in the textbook is already relatively well known or *common knowledge*. If this is the case, individuals might deem it unnecessary to reference the pioneering article and instead refer to the theory itself. A second explanation for this mechanism is that the textbook may promote an article's findings to the status of common knowledge, once again leading individuals to reference the theory itself rather than the pioneering article (Aizenman & Kletzer, 2011). A third explanation for a perceived cessation effect is that researchers may reference the popular textbook instead of the original research article after it is included in the textbook (McMahan & McFarland, 2021).

The signalling effect, on the other hand, can be observed when citation counts increase after the notable event. This mechanism predicts prominent and relevant research articles and describes a point at which they "take off". Several reasons could explain this mechanism, too. For instance, authors of well-known economics textbooks tend to cover seminal economics theories and topics, necessitating the inclusion of the most relevant, pioneering and influential research in their field (Korom, 2018). Owing to the nature of this research, it garners the most attention, leading to growing citations. Thus, citations increase over time owing to the nature of the academic literature included in textbooks rather than the act of including said papers in

a textbook. A second explanation of this mechanism could be the exposure that economics textbooks bring to academic literature. Well-known economics textbooks have a broad readership and strong reputation. Thus, the citing of academic literature in such a textbook could attract additional attention and spark interest resulting in an increase in citations and mentions (Yuret, 2023).

To demonstrate the practical application of these concepts, I employed an example from my dataset. Robert Aumann, a microeconomics theorist, has authored a groundbreaking research paper on utilising general financial claims and the fundamental equilibrium concept (Aumann et al., 1983). David Mark Kreps, an economist, has authored a microeconomics theory textbook (Kreps, 1990). Within the chapter addressing pure exchange and general equilibrium, Kreps reinforces his arguments by referencing Aumann's paper. In this dissertation, I undertook a test to examine whether the inclusion of Aumann's paper in Kreps' textbook correlates with the subsequent citations of Aumann's work and, if so, what insights this observed citation pattern may yield regarding the provenance of information. If a discernible effect emerges, specifically a decrease in subsequent citations (as evidenced by lower post-event citations in the treatment group), it indicates a weakening of attribution and the dominance of a cessation effect. This suggests that the provenance of information is compromised, resulting in the assimilation of knowledge into the realm of common knowledge when a research article is cited in a well-known textbook. Alternatively, if a notable effect arises, specifically an increase in subsequent citations (as indicated by higher post-event citations in the treatment set), it underscores strengthened attribution and the prevalence of a signalling effect. This implies that the provenance of information remains intact, and it cannot be contended that a textbook citation renders the information common knowledge.

I performed two hypothesis tests to address my research questions. The first hypothesis test determines the significance of the difference in slopes (*"log(Citations 12 Yrs After)"*|*"log(Citations 10 Yrs Before)"*) between the treatment group and the control group. The second hypothesis test investigates whether the average *"log(Citations 12 Yrs After)"* is equal for the treatment and control groups when they have the same *"log(Citations 10 Yrs Before)"*.

I first found a notable difference in slopes, indicating that academic literature cited in textbooks tends to have lower post-event citations compared to academic literature not cited in textbooks, but only at exceptionally high levels of pre-event citations. This finding suggests the presence

of a cessation effect among highly influential or ground-breaking research articles, providing some evidence, albeit limited, of weakened attribution in these cases.

Second, I found a highly significant positive difference in the average post-event citations between the treatment and control groups. This suggests that including academic literature in textbooks influences subsequent citations of the cited articles. Moreover, the positive difference implies that, at average levels of pre-event citations, post-event citations for academic literature cited in textbooks are comparatively higher than those not cited in textbooks. These findings indicate that attribution is strengthened for most research articles cited in textbooks, highlighting the presence of a signalling effect.

Figure 1.1 illuminates the central and most notable outcome of this dissertation, showcasing the dominant signalling effect among the cited articles. The visualisation depicts that both cited and uncited articles follow a comparable citation trajectory for years leading up to the textbook citation event. However, after reaching time T (representing the textbook citation event), a clear divergence occurs. The blue line, representing the cited articles, surpasses the red line, symbolising the articles that are not cited. This observation highlights a notable increase in attribution for cited articles following the textbook citation event.

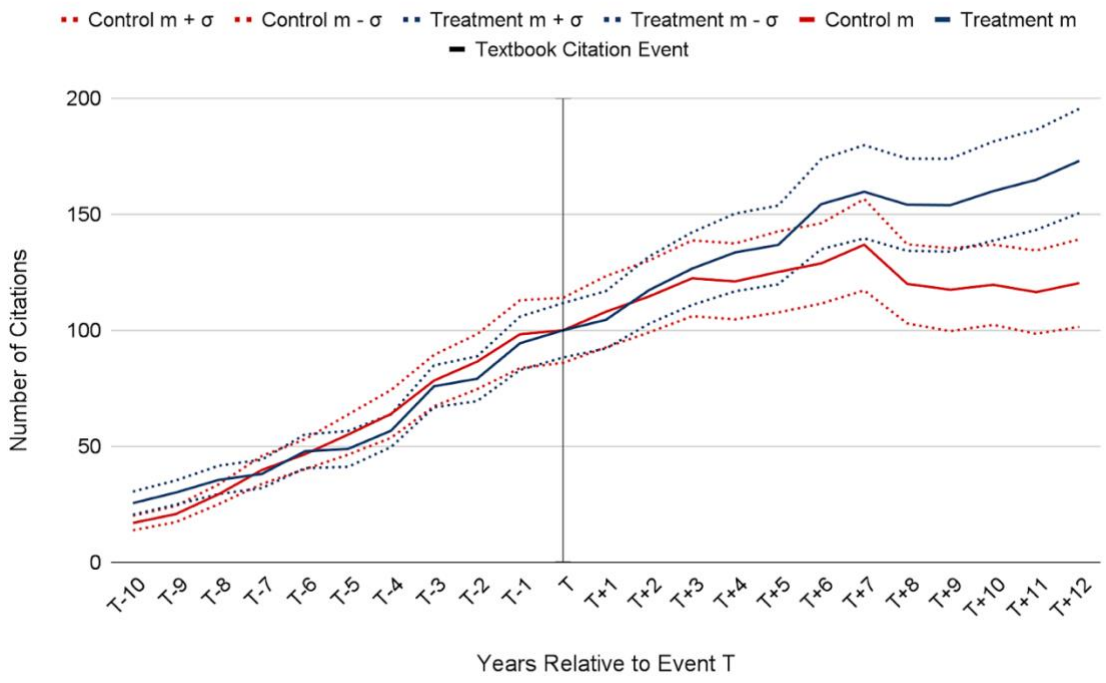


Figure 1.1: Comparison of normalised annual average citations before and after the textbook citation event (Time T)

Note: This figure shows the mean and error ranges for cited and uncited articles for the ten years before and 12 years after the textbook citation event (time T). The blue lines (solid and dotted) represent the treatment group mean and mean plus/minus standard deviation, respectively. The red lines (solid and dotted) represent the control group mean and mean plus/minus standard deviation, respectively. Before Time T, both groups exhibit a similar citation trajectory. However, at T, a clear divergence occurs, with cited articles experiencing significantly higher citation trends compared to uncited ones.

The subsequent sections of this dissertation are organised as follows: Chapter 2 presents the comprehensive literature review, delving into relevant studies and scholarly works covering topics such as the science of science, citation analysis, the cessation effect, the signalling effect and the textbook approach. Chapter 3 outlines the methodology employed, detailing the approach and procedures utilised. This section also outlines the limitations of this research. Chapter 4 presents the results obtained from the study and engages in a thorough discussion of their implications. Chapter 5 concludes this dissertation by summarising the key findings, reflecting on the research process, and offering insights for future research directions.

Chapter 2: Literature Review

In this chapter, I provide further elaboration and contextualise the topics explored in the introduction of this dissertation by connecting them with the most relevant literature. I also review the literature on the science of science, citation analysis, cessation effect, signalling effect, and the textbook approach.

2.1 The science of science

In academia, the practice of academic referencing is a unique social convention that plays a crucial role in the establishment of intellectual property and the diffusion and expansion of scientific knowledge. This convention can be traced back to the mid-19th century, around 1850, marking the emergence of a familiar pattern of explicit references to previous work in scientific periodicals (De Solla Price, 1986). Academic referencing serves three key purposes. First, it acknowledges and gives credit to previous contributions. Second, it establishes priority claims. Third, by referencing earlier studies, researchers establish a connection between their research and the existing body of knowledge, highlighting the interconnectedness and cumulative nature of scientific progress (Hyland, 1999).

Owing to the generous trail of reference data that the practice of academic referencing has left, a new multidisciplinary field of study called the *science of science* has emerged. This field seeks to understand the scientific process itself, including the methods used to conduct scientific research, the creation and diffusion of new scientific knowledge, and the social and economic factors that influence scientific progress. Researchers in this field use data-driven approaches, such as citation analysis, funding analysis, and collaboration networks to study patterns and trends in scientific research (Fortunato et al., 2018). Newman (2001) explores social factors driving scientific progress, specifically the structure of collaboration networks. His findings uncover “small worlds” within these networks, where randomly selected scientists are often connected by short paths through intermediate acquaintances. Roshani et al. (2021) employ funding analysis to investigate the influence of economic factors on scientific advancement. By studying the relationship between research funding and citation-based performance, they predict the dissemination of new scientific discoveries. Their findings highlight that funded research papers garner more citations than unfunded ones, indicating the significant role of funding in citation trajectories. Park, Leahey and Funk (2023) employ

citation analysis on a vast dataset of scientific papers and patents to explore the emergence and spread of new knowledge. Their findings reveal a diminishing trend of disruptive innovation in science and technology, indicating that recent papers and patents contribute less to pushing the boundaries of knowledge.

Among the aforementioned methods used in the science of science, citation analysis holds particular relevance to this dissertation. Citation analysis is a research method used to examine the frequency and patterns of citations within research documents and, amongst other things, it seeks to understand how scientific knowledge is created and spread. Numerous factors affecting citation trends are explored extensively in research papers. Tahamtan, Safipour Afshar and Ahamdzadeh (2016) provide a comprehensive review of the literature on factors predicting citation performance in research publications. They identify 28 factors and show that the literature groups these into three general categories: journal-related factors, like the journal impact factor and prestige; paper-related factors, like early citation performance, length of the paper and characteristics of the title; and author-related factors, like author gender, number of co-authors and author affiliation. Their study reveals that early citation performance alone is insufficient to define the quality and impact of research papers, as citation impact relies on a combination of factors. Additionally, the researchers suggest that certain factors among the recognised 28 have greater effectiveness in attracting citations, including the journal impact factor and the number of authors.

Building upon these insights, I have utilised key explanatory variables, specifically early citation performance and the inclusion of an article in a textbook, within a citation analysis framework to investigate their impact on future citation performance. Furthermore, I leveraged this citation analysis framework to understand the underlying mechanism influencing citation performance. Through this analysis, I identified two potential effects: a negative cessation effect and a positive signalling effect, both contributing to the direction of citation performance. Alongside these key variables, I included various journal, author, and paper-related factors, ensuring the inclusion of the journal impact factor and the number of authors. By including these additional factors as control variables, I can account for their influence on the citation performance of research publications within my dataset while also delving deeper into their relationship with citation outcomes.

2.2 The cessation effect

A *cessation effect* occurs when the citation count of academic research decreases after a notable event. There are no studies that test directly for a cessation effect. Still, one related concept is the idea of "intellectual lineages" in economics, which refers to the way in which the ideas and theories of influential economists are passed down through generations of scholars. This concept suggests that, while the original pioneer authors may become less frequently cited over time, their ideas may continue to be cited and built upon by subsequent generations of scholars who are part of the same intellectual lineage. This concept is alluded to by Breit and Hudson (1997). In their study, the authors highlight that the ideas put forth by senior economists can gradually become part of the collective knowledge of the field. As these ideas are incorporated and utilised by other researchers, they might no longer be explicitly cited, leading to a decline in citations over time. Anderson, Levy and Tollison (1989), in their study, explore the intellectual counterpart to the efficient-market hypothesis, which suggests that the efficient transmission of knowledge leads to the inclusion of all classical findings in contemporary texts. The researchers examine a list of 262 economists compiled by Stigler and Friedland for the 1985 Calendar of Great Economists (covering individuals who have passed away since 1600) and analyse the citations to their work between 1982 and 1984. They investigate the decay of an economist's work over time, and their findings indicate that citations decay more rapidly in economics compared to other social sciences. The study supports the efficient-market hypothesis in the literature on the history of economics, suggesting that, over time, an economist's work becomes common knowledge and does not require citation. While the authors offer evidence that citations decline as knowledge is gradually absorbed into the scientific community, they do not mention explicitly an event that triggers this decay. However, one creative interpretation could be that the death of a prominent economist serves as such an event.

One article by Aizenman and Kletzer (2011) explores this idea. The authors collected a sample of research papers written by well-known economists who died before the age 65, from 1975 to 1997, to determine the impact of premature death on citation patterns. They found evidence of a term they call the "citation death tax", which suggests that articles written by prominent authors who have died prematurely receive fewer citations in the years after the author's death than the most closely matching papers written by other authors. These authors establish that an event, such as the premature death of an economist, can cause a citation decline and attribute

this observed effect to the impact of academic networking and strategic citation behaviour in economics.

Another related concept is the "half-life of knowledge", which was initially proposed by Fritz Machlup (1962). This term describes the duration it takes for half of the knowledge in a specific field to become outdated or surpassed (Machlup, 1962). This suggests that time or age could initiate a citation decline. According to De Solla Price (1986), the age of a paper significantly influences the number of citations it receives, with an apparent decline observed once the paper reaches a certain age. This idea relates to the cessation effect to some extent, since the concept of the "turning of age" can be interpreted as an event that initiates citation decline. Various researchers have attempted to quantify this age threshold and the magnitude of the decline. For example, Arao, Santos and Guedes (2017) studied the half-life and obsolescence of knowledge in the literature science field by analysing the references in theses and dissertations of master's and doctoral degree students. From their sample, they calculated a half-life of 14 to15 years. In a study by Galiani and Gálvez (2017), the researchers examine citation evolution in the years following a paper's publication date. They found that citations follow a life-cycle pattern characterised by a growth period, peak, and subsequent decline. In economics, the life cycle of research is longer compared to other fields, and the decline in citations is also more gradual.

Certain citation events, such as citations in review articles, represent another related concept. McMahan and McFarland (2021) conducted a comprehensive analysis using a vast research article database to investigate the impact of academic review articles on the publications that they cite. Their study reveals a significant decline in future citations for research papers cited in formal review articles. The authors attribute this decline to the phenomenon that the review article is cited instead of the specific articles referenced within the review. This research paper is an illustrative instance of another investigated event in the literature that can contribute to the decline in citations, providing compelling evidence of the cessation effect.

The concepts, theories, and methodologies depicted in research papers inevitably become outdated. This notion aligns with the concept of the half-life of knowledge, where research papers gradually lose relevance as time passes. In addition to the natural life cycle and ageing process of research that eventually renders it obsolete, the emergence of fresh ideas that either debunk or significantly enhance previous notions can disrupt the citation patterns of research papers that advocate for those outdated ideas. In his work, Kuhn (1970) proposes that new ideas are initially disregarded but that they accumulate over time, eventually leading to a paradigm

shift where a new perspective replaces an older one. Consequently, as new ideas enter the realm of research, the significance of older ideas gradually diminishes, resulting in a decline in citations for studies representing outdated concepts. Simultaneously, there is a natural increase in citations for papers that embrace these novel ideas. Leahey, Lee and Funk (2023) determine that this phenomenon is particularly prominent in the study of new methods, instruments, and techniques documented in the literature. Thus, a cessation effect occurs when a new idea supersedes an old one, while simultaneously, this occurrence serves as evidence of a signalling effect (discussed in the subsequent section) for research that supports these fresh ideas.

2.3 The signalling effect

The *signalling effect* refers to a scenario where the citation count of academic research increases following a notable event. Although there is limited research specifically addressing the concept of a signalling effect in post-event citations within economics textbooks, certain studies offer evidence suggesting that significant recognition events can act as signals, garnering attention and citations for particular researchers and their contributions.

One such study by Card and DellaVigna (2013) aims to analyse how articles published in prestigious economics journals affect citation patterns and how attention is distributed among researchers and their work within the field. The authors highlight the substantial influence that publications in top journals can have on the academic success of economists. To investigate this influence, they compiled Google Scholar citations for articles published in the top five economics journals between 1970 and 2012, revealing noteworthy trends. Their analysis demonstrates that papers published in these esteemed journals receive a considerable number of citations, showcasing how being published in a highly rated journal can act as a signal for the subsequent achievements and recognition of the authors.

In their study, Bjork, Offer and Söderberg (2013) employ time-series data to examine the citation trajectories of Nobel Prize winners in economics. Their analysis reveals that, for most prize winners, citation patterns follow a bell-shaped curve, indicative of the typical innovation cycle observed in economic knowledge. However, they identify various alternative trajectories, one of which is the “staying power” trajectory, where citations reach a peak around the time of the prize and remain high after that. Additionally, the study demonstrates a brief period of heightened citation activity known as the “prize citation premium”, characterised by a sharp increase in citation counts immediately following the year of the award. In this context, the

Nobel Prize in Economics serves as a signal to the economics community about the importance and relevance of certain ideas and research streams, leading to increased citations and attention for the laureates and their work.

The work by Merton (1988) can be seen as an illustrative example of the signalling effect. In this study, Merton explores the idea that recognition and prestige in science are often based on cumulative advantage, where early success and recognition lead to further success and recognition. This can lead to a "Matthew effect", where the most successful and recognised researchers receive the most attention and credit, even if their work is not fundamentally better or more innovative than their peers. Here, early success serves as a signal to later success.

2.4 The textbook approach

By selecting well-known microeconomic textbooks as representatives of notable events, I postulate that the act of referencing academic research within these textbooks may have a potential impact on the subsequent citation count of the cited research. This, in turn, holds significant implications for the attribution of information within the academic realm. Consequently, the foundation of this dissertation is rooted in textbook citation analysis, commonly referred to as the *textbook approach*.

Textbooks are widely acknowledged as essential vessels that encapsulate the knowledge of a particular field (Kuhn, 1970). In his book, De Solla Price (1986) discusses the exponential growth of science. Still, at each stage along the way, papers can be compressed into review articles and eventually into textbooks (De Solla Price, 1986), suggesting that textbooks serve as comprehensive resources that consolidate and present fundamental concepts, theories, and practices within a given discipline. Korom (2018) explains that textbook authors are tasked not only to include well-established fundamentals of the discipline in their textbooks but also take account of recent work in a rapidly moving set of research problems over which there is substantial disagreement. Thus, by deciding to include specific research findings in their textbooks, authors are essentially promoting what they believe are the most important and pressing topics in economics research at the time. These points bring us to the assertion that textbooks play a pivotal role in the certification of knowledge. It was Norman W. Storer who first popularised the concept of certification of knowledge: "Only when a contribution is 'certified', that is, acceptable to other scientists under the canons of proof they share, will it be welcomed" (Storer 1966:119). This statement suggests that, in the scientific community, a

contribution or finding is considered valuable and accepted only when it meets certain standards of proof and is deemed acceptable by other scientists. According to Rothman (1971), the concept of certification entails the presence of three distinct stages in accepting a new idea: "a trial period during which professional evaluation occurs; a transitional stage when members of the discipline become convinced of the merits of the new knowledge; and a final phase when a number of members of the discipline accept the contribution as a valid extension of knowledge" (Rothman 1971:126). At stage three, the subject becomes a legitimate topic for inclusion in textbooks, and omitting it may be considered a serious oversight (Rothman, 1971). The notion that textbooks serve as a certification of knowledge has undoubtedly gained recognition within the scientific community. Nonetheless, the post-inclusion fate of these certified ideas in textbooks remains shrouded in ambiguity, thereby constituting the fundamental objective of this dissertation.

Given the acknowledged significance of textbooks as resources that certify, consolidate and present fundamental ideas within a given discipline, several research papers have adopted a research perspective primarily focusing on examining textbook citations, termed the *textbook approach*. One study by Liner (2002) focuses on ranking economics journals through textbook citations. Liner (2002) collected citations from six graduate-level textbooks in the fields of microeconomics, macroeconomics, and econometrics. By tallying the references to each journal, he calculated the journal rank based on the frequency of citations received. The top nine journals identified in their study align closely with those found in comparable studies. In a similar vein to my study, Liner (2002) also gathered citations from economics textbooks. However, the extent of the analysis differs from the approach employed in my study, since he did not conduct a comprehensive citation analysis of the references. Additionally, our research objectives diverge significantly because Liner (2002) aimed to rank economics journals while I focus on the influence of textbooks on post-event citation trends and their corresponding meaning for the dissemination and provenance of information.

In their study, Breit and Hudson (1997) seek to discern the differentiation between economists' reputations and influence. Their approach involves examining a sample of 19 introductory economics textbooks, comparing the percentage of textbooks in which an author is mentioned to the total number of scientific references to the author's work. The findings reveal substantial disparities between these two metrics. As a result, the authors emphasise the importance of establishing a clear differentiation between reputation, which is best evaluated through journal

citations, and influence, which can be better assessed by considering textbook citations. Their study is slightly more related to this dissertation as they employ textbook citation analysis. However, our research questions are dissimilar.

Korom (2018) also uses textbook citation analysis in a comprehensive study of 30 widely used introductory textbooks in the disciplines of economics, psychology, and sociology. The study examines two primary questions regarding eminence in economics textbooks: the factors contributing to attaining eminence and the durability of textbook eminence over time. The study finds that "certified eminence" in textbooks is primarily associated with the number of pages referencing a scholar. Regarding the durability of textbook eminence, the study finds that eminence tends to be short-lived overall. Many leading scholars of their time have a marginal presence in contemporary textbooks, while a select few, such as John M. Keynes, maintain enduring prominence. Korom's (2018) study shares similarities and differences with this dissertation. Both studies employ citation analysis to examine the influence of citations in academic literature. However, while Korom focuses on the authors referenced in textbooks, my research compares cited articles with uncited ones. Despite diverging research questions, there is a notable similarity: Korom (2018) attempts to identify the expiration histories of scholarly recognition, exploring how long star scholars remain heavily cited in textbooks; comparably, one of the aims of my study is to identify a notable point at which scholarly recognition starts to diminish.

The research article authored by Yuret (2023) demonstrates a striking resemblance and relevance to the objectives of this dissertation. Specifically, Yuret's (2023) study explores the citation performance of references found within the widely used microeconomics textbook by Mas-Colell, Whinston and Green (1995), which is among the nine textbooks employed in this dissertation. This singular edition, published in 1995, has maintained its prominence over 25 years, particularly in introductory doctoral economics courses. Exploiting the distinct characteristics of this textbook, Yuret (2023) seizes a unique opportunity to investigate whether the textbook's success is reflected in the performance of its referenced works. The study compares the citation performance of references in Mas-Colell et al. (1995) after its publication year to that of a carefully selected comparison group of publications. To ensure an appropriate comparison, Yuret (2023) ensures that the comparison publications match the same year and journal and possess a comparable number of early citations (pre-1995). The study's findings reveal that the references in Mas-Colell et al. (1995) exhibit superior citation performance

compared to the publications in the comparison group after 1995, which evidences a signalling effect. Yuret's (2023) research shares several key similarities with this dissertation yet exhibits notable differences. Both studies employ a citation analysis approach, comparing the references within a textbook to a comparison group. The objectives align closely, as we both aim to examine the citation performance of these references following the textbook's publication. However, in this dissertation, I extend the analysis to include a broader selection of microeconomics textbooks, while Yuret (2023) focuses solely on one textbook. I also incorporate additional factors that may influence the citation performance of the references in my dataset, whereas Yuret's (2023) study does not consider these factors. Furthermore, our analytical methods differ: I employ a linear regression analysis on the entire dataset, encompassing both textbook references and their comparison publications, while Yuret's (2023) study adopts a like-for-like comparison, matching each reference to its most closely related comparison publication.

The literature examined in this review shares certain similarities with the present dissertation. However, there appears to be a limited body of research that specifically employs the textbook approach or similar methodologies to investigate the provenance of information. The works by Aizenman and Kletzer (2011), McMahan and McFarland (2021), and Yuret (2023) are particularly relevant in this context, as they employ similar methodologies and yield comparable findings. Aizenman and Kletzer (2011) explore the citation patterns of research papers authored by prominent economists who experienced premature deaths. McMahan and McFarland (2021) investigate the impact of citations in formal review articles on future citation counts. Both studies provide evidence of a cessation effect. Notably, these research papers, along with the present dissertation, share a common thread: they all examine the effect of significant events on subsequent citation counts. However, it is important to note that these events differ considerably. Therefore, it can be stated broadly that noteworthy events can influence subsequent citations and, consequently, the attribution of information. Yet, it remains unclear whether a citation in a well-known textbook can elicit such an effect. This dissertation enhances the existing literature by addressing this lack of clarity directly. Yuret (2023) explores the citation patterns of microeconomics textbooks and identifies evidence of a signalling effect. However, the author's analysis is based on a single textbook and does not account for additional factors that may impact citation performance. In contrast, this dissertation improves upon the existing literature by examining multiple textbooks and by incorporating additional control variables into the analysis.

Chapter 3: Methodology

In this chapter, I present a comprehensive overview of my methodology. I explain the process I use to construct a citation dataset in detail. The dataset consists of 610 observations and 18 variables, which are further categorised into a treatment and control group. Furthermore, I provide a detailed description of these variables, explaining their anticipated economic relationship with the dependent variable. Additionally, I articulate two hypotheses, outline the multivariate linear model employed, and discuss the analysis conducted to address the research questions. Finally, I state the limitations of this dissertation.

3.1 Data collection

The data collection process encompasses multiple steps, and I compiled the final dataset by integrating data inputs from various providers. I created a dataset that includes a sampled list of articles along with their corresponding metadata. I utilised this dataset in a multivariate regression analysis, comparing a treatment group consisting of articles cited in microeconomics textbooks with a control group comprising articles that are not cited.

In this dissertation, I intentionally differentiate between *articles* and *papers* and consistently refer to them as such. During the initial data collection stage, papers cover various publication types, such as articles, books, notes, and more. Later, articles specifically represent publication types limited to articles alone, forming a subset of the previously defined papers. In the final data collection stage, I exclude all publication types except articles from the sample. Thus, the final sample consists exclusively of articles.

This study utilises notable data sources, including MIT OpenCourseWare, Scopus, Gender.io, and the CentER for Research in Economics at Tilburg University. Furthermore, I retrieved the electronic versions of the microeconomic textbooks from various sources to digitise the references. A detailed description of each data collection step is provided in the subsequent sections. Descriptions of all variables mentioned in the subsequent sections are included in **Table B.1** of **Appendix B**.

3.1.1 Building the treatment dataset

The treatment group consists of a sample of research articles cited in microeconomics textbooks and their corresponding metadata.

3.1.1.1 Collecting and processing data from microeconomics textbooks

The first stage of the process involves the collection of a sample of microeconomics textbooks. I started by browsing MIT OpenCourseWare to construct a list of popular microeconomics textbooks.[1] Using this list, I collected a sample of nine microeconomics textbooks from various online sources. These nine textbooks form the basis of all the other data I collected because they contain the references I used to form the treatment group of articles cited. The titles of the nine textbooks, along with their authors and publication dates, are provided in **Table 3.1**.

Table 3.1: Sample of microeconomics textbooks used to compile the treatment group of cited articles

Textbook Title	Author	Publication Date
Microeconomics behaviour, institutions, and evolution	Samuel Bowles	2006
A course in microeconomic theory	David M. Kreps	1990
Development microeconomics	Pranab Bardhan, Christopher Udry	1999
Evolutionary microeconomics	Jacques Lesourne, André Orléan, Bernard Walliser, P. Bourgine, E. Fauchart, J.-F. Laslier, L. Marengo, F. Moreau, G. Umbhauer	2006
Methodology for a new microeconomics	Lawrence A. Boland	1987
Microeconomic theory	Andreu Mas-Colell, Michael D Whinston, Jerry R Green	1995
Microeconomics	Jeffrey M Perloff	2006
Microeconomics analysis	Hal R Varian	1992
Microeconomic foundations I.	David M Kreps	2013

[1] The Massachusetts Institute of Technology provides content from thousands of courses through MIT OpenCourseWare, which is free and open access (Massachusetts Institute of Technology, n.d.).

Note: This table shows the title, authors and publication dates of the sample of nine microeconomics textbooks used in this study. The references within each of these textbooks are collected and then consolidated to form the treatment group.

Next, I scanned each textbook for references and created text strings of these references. Each reference contains information such as the paper's title, the authors and the publication date. Because this data is unsorted, is formatted as a comma-delimited text string, and spans multiple rows in Excel, I wrote a small Python program to extract each reference. The Python program takes the text string of reference data from each textbook as input, processes the data by locating the "next reference" separator that is inherent to that textbook, and produces an Excel file containing unique rows of comma-delimited reference data, ultimately ensuring that each reference starts on a new row in Excel. From there, it is possible to use Excel functions to separate the reference text strings into individual columns. This data includes the *"First Author's Surname*", *"Paper Title",* and *"Publication Date"*. I added the *"Textbook Title"* and *"Book Year"* as columns to keep track of the references relating to each textbook. The resulting cleaned data, i.e. the *"First Author's Surname"*, *"Paper Title*" and *"Publication Date"*, serve as input for the next stage of the data collection process.

3.1.1.2 Matching papers to the Scopus database and retrieving related metadata

The next data collection stage involves retrieving the references and their corresponding metadata from a database called Scopus. It is necessary to match the references listed in the nine microeconomics textbooks to an existing database and to find additional information about these references that can be used as the dependent, independent and control variables in the analysis.

Scopus is an abstract and citation database that hosts many papers from various publishers across several disciplines. Additionally, it stores data about the papers themselves, along with data about the authors who write them. Scopus provides API functionality, which enables its users to retrieve diverse datasets. Specifically, the APIs provide programmatic access to journals and books, citation data, abstracts, research metrics and more (Elsevier developer portal, n.d.). To retrieve the data provided via the APIs, a user must access Scopus and a Scopus API key. **Appendix A** details the process of creating a Scopus account through your institution (the University of Cape Town, in this instance), creating an API key, and additional details about relevant APIs and views used for collecting data in this analysis —as mentioned in the following sections.

I used statistical analysis and programming languages, R and Python, to retrieve the data from Scopus. I categorised the metadata as journal-level data, paper-level data and author-level data.

First, I used an R package called RScopus to retrieve most of the data. RScopus is a Scopus database API interface developed to make Scopus data retrieval simple and easy for users familiar with R (Muschelli, 2022). I wrote a script that utilises the ***Scopus Search API*** and collects the following variables: *"Article Scopus ID"*, *"Article Title"*, *"Publication Name"*, *"Publication ISSN"*, *"Publication Issue Number"*, *"Publication Page Range", "Publication Cover Date"*, *"Source Type"*, *"Document Type"*, *"Number of Authors"*, *"Author Full Name"*, *"Author First Name"*, *"Author Last Name"*, *"Author Scopus ID"*, *"Author Affiliation ID"* and *"Author Affiliation Name"*.

Next, I wrote three Python scripts, each using the Python Requests library and Scopus APIs to retrieve additional data. The first script uses the ***Author Retrieval API*** to retrieve author-level information. The variables I retrieved with this script include *"Co-author Count", "Document Count", "Author Citation Count",* and *"H-Index"*. The second script retrieves the yearly citation counts from 1950 to 2020 using the ***Citations Metadata View*** of the ***Citation Overview API***. The third script collects journal-level data such as the *"Journal Citation Count"*, *"SNIP Score",* and *"SJR Score'.* To collect this data, I used the Scopus ***Serial Title API***.

3.1.2 Building the control dataset

The control group consists of a sample of articles and their corresponding metadata that are not cited in microeconomics textbooks.

3.1.2.1 Collecting the reference data and corresponding metadata to construct the control group

The control group is defined as a citation dataset that can be compared to the treatment group. In other words, it shares a similar distribution of yearly citation counts but differs from the treatment group in that the titles included in the control group are not cited in microeconomics textbooks.

Forming the control group involves using the ***Scopus Search API*** to retrieve a sample of titles related to each research paper in the treatment group. The goal is to gather a corresponding group of papers that are published in the same journals, issues, and years as the papers in the treatment group but are not cited in microeconomics textbooks. I retrieved this data from

Scopus using RScopus. The script utilises the input variables *"Publication ISSN"*, *"Publication Issue Number"*, and "*Publication Cover Date"*, obtained from the first R script, to collect a list of control group paper titles and Scopus IDs. For each treatment group paper, I collected a sample of 15 control group papers. I curated the collected paper titles to ensure uniqueness between the two datasets. To match the treatment group data, I employed existing R and Python scripts to collect the remaining metadata. Finally, I sampled the control group to reduce its size to 1,000 unique Scopus IDs for better comparability.

3.1.3 Combined citations dataset

The combined citations dataset consists of the combined treatment and control group articles.

3.1.3.1 Collecting additional variables for the combined dataset

The full citation dataset includes the treatment group and control group. I added the variable *"Cited in Book"* to distinguish the two types of entries, where 1 indicates a treatment group entry, and 0 indicates a control group entry.

In this next phase, I collected supplementary data to create three new variables: *"Author Gender", "First Author Gender",* and *"Affiliation Ranking"*. I obtained these variables from two data sources, namely Gender.io and The CentER for Research in Economics and Business at Tilburg University.

To create the *"Author Gender'* variable, I used Gender.io, a website that provides a database of common names and their corresponding genders. I employed a Python script to retrieve the genders (male or female) of the authors in the dataset by making API requests to Gender.io. I rotated IP addresses and collected data over several days to surpass the limit of 1,000 requests per day. If the gender information was unavailable, I marked the entry as undefined. Additionally, I defined the *"First Author Gender"* variable by assigning the gender of the first author to all authors of the respective paper.

For obtaining affiliation rankings and creating the *"Affiliation Ranking"* variable, I utilised the university rankings database of The CentER for Research in Economics and Business, which is based on publications in leading economics journals (Tilburg University, n.d.). The database covers rankings from 2004 onwards, and I standardised the rankings using the reference year 2019. I copied the ranking tables into an Excel workbook, where I used index-match formulas to retrieve the relevant university ranking for each author's affiliation. I manually identified

and corrected the matches for entries with discrepancies between the Scopus and University Ranking database naming conventions. Affiliations not listed in the rankings database were labelled as unranked.

3.1.3.2 Calculated variables

To prepare the dataset for its final form, I performed additional calculations and cleaning steps. First, I included the variable *"Article Publication Year"*, which corresponds to the *"Publication Cover Date"* year. Second, I added the variable *"Number of Pages'*, calculated as the difference between the last and first page numbers in the *"Publication Page Range"*, plus 1. Third, I created the variable *"Title Length"* by counting the number of letters in each paper's title.

Furthermore, I established the variables *"Citations 10 Yrs Before*" and *"Citations 12 Yrs After"* by summing up the yearly citation counts. To calculate these variables, I extracted the citation count per year for the ten years prior and, separately, for the 12 years following the *"Book Year"* (the year the book was published). This is accomplished using an index-match formula in Excel. This process generated 22 variables labelled T-10 to T+12, with T representing the book's publication year. I calculated *"Citations 10 Yrs Before"* by aggregating T-10 to T-1 and calculated *"Citations 12 Yrs After"* by aggregating T+1 to T+12.

3.1.3.3 Distribution comparison

To ensure an appropriate comparison between the control and treatment groups, it is important to confirm that the *"Citations 10 Yrs Before"* and *"Citations 12 Yrs After*" variables share the same distribution. I achieved this by taking the logarithm of the *"Citations 10 Yrs Before"* and *"Citations 12 Yrs After"* variables for both the control and treatment groups while excluding outliers. Specifically, I removed values below six and above 2000.

Next, I plotted the transformed variables on histograms and observed that the control and treatment sets exhibit similar distributions. **Figures 3.1** and **3.2** illustrate the histograms for both variables. These findings validate the suitability of the control and treatment groups for conducting a meaningful comparison.

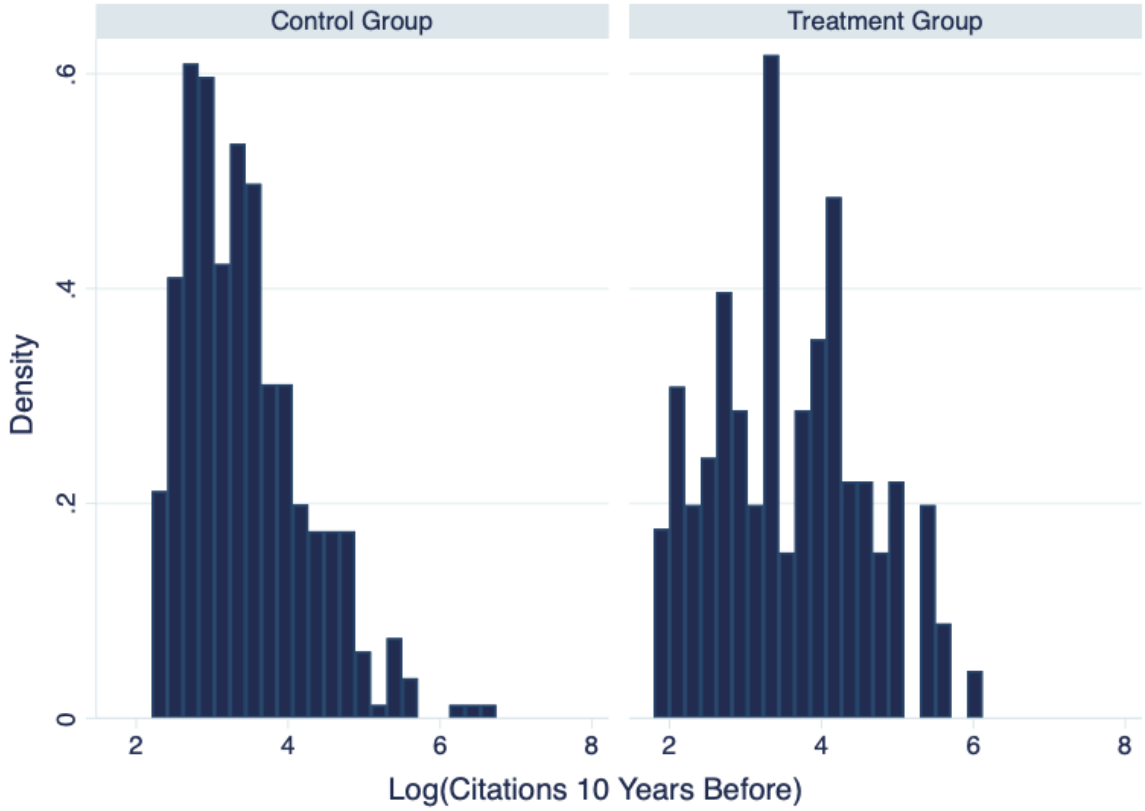


Figure 3.1: Histograms showing the distribution of "*log(Citations 10 Yrs Before)*" for the treatment group (cited) and the control group (not cited)

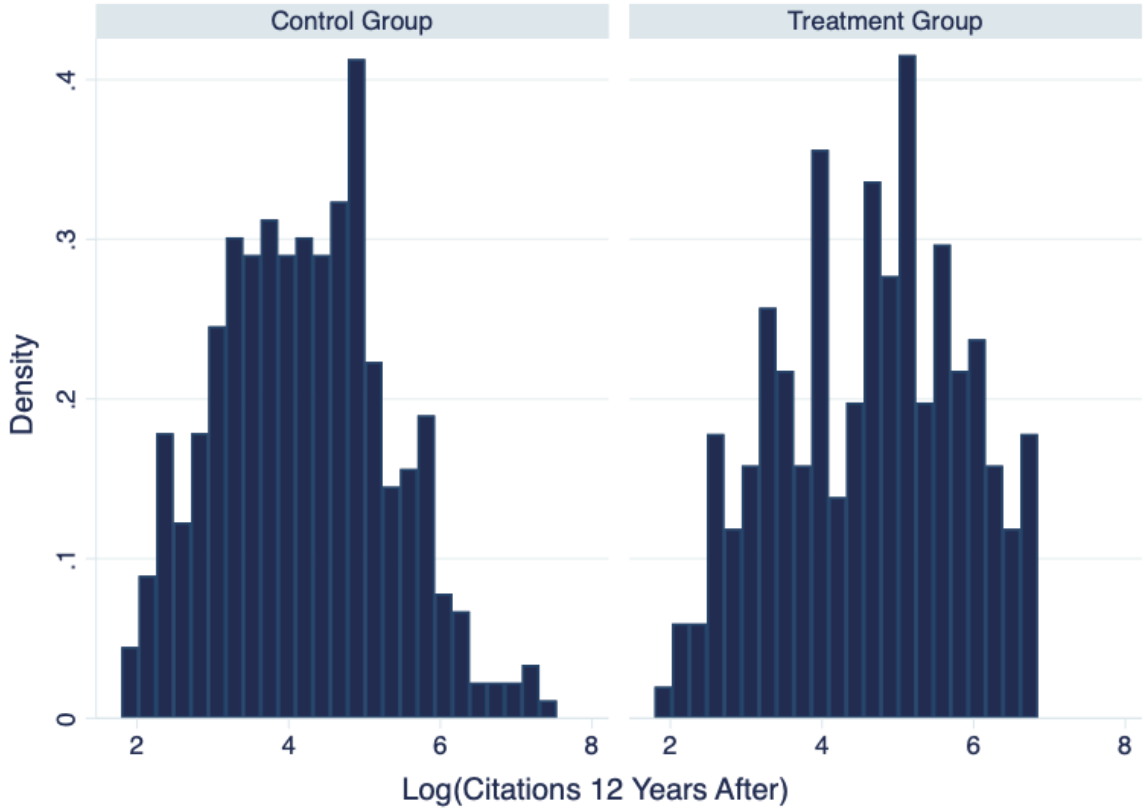


Figure 3.2: Histograms showing the distribution of "*log(Citations 12 Yrs After)*" for the treatment group (cited) and the control group (not cited)

The data collection process produced a citation dataset comprising 1926 observations and 33 variables. The subsequent step entails cleaning and curating the dataset to prepare it for analysis.

3.2 Data cleaning and preparation

I used STATA, a statistical software package developed for data manipulation, visualisation, statistics, and automated reporting, to clean, curate and inspect the dataset to prepare it for analysis.

3.2.1 Removing undesired observations and creating dummy variables

First, I dropped duplicates. The *"Article ID"* and the year the microeconomics textbook was published (i.e., the *"Book Year"* variable) define a unique entry. Hence, if a paper appeared more than once and was sourced from different textbooks published in different years, it was not dropped from the dataset. However, when a paper appeared more than once and was sourced from different textbooks published in the same year, it was dropped from the dataset.

Additionally, if a paper appeared more than once and was sourced from the same textbook, it was dropped from the dataset.

Second, I dropped all observations classified as Editorials, Letters, Notes, Reviews and Short Surveys as defined by the *"Document Type"* variable. Similarly, I dropped all observations classified as a Book or Book Series, as defined by the *"Source Type"* variable. Consequently, only observations belonging to a Journal and classified as an Article remained in the dataset.

Third, I dropped all the missing data. This includes articles where the *"First Author Gender*" variable or the *"Affiliation Ranking"* variable is undefined.

Finally, I created a dummy variable called *"Male First Author"* from the categorical variable "*First Author Gender"*, where 1 denotes male and 0 denotes female.

3.2.2 Aggregating author-level observations

It is necessary to collapse the author-defined observations belonging to each article into one unique record (based on *"Article ID")* because the dataset contains variables that are defined at the author level (*"Co-author Count"*, *'Document Count"*, *"Author Citation Count"* and *"H-Index"*). I did this to ensure that the dataset is defined at the article level. To achieve this, I dropped any non-numeric variables and then calculated an average value for each of the remaining numeric variables (aggregated by *"Article ID"* and *"Book Year")* while also taking care to remove incomplete observations. Incomplete observations are any record that had missing author-level information that was dropped in the previous stage. The non-numeric variables dropped include *"Textbook Title", "Article Title"*, *"Publication Name"*, *"Publication ISSN"*, *"Publication Issue Number"*, *'Publication Page Range"*, *"Publication Cover Date"*, *"Source Type"*, *"Document Type"*, *"Author Full Name"*, *"Author First Name'*, *'Author Last Name"*, *"Author Scopus ID"*, *"Author Affiliation ID"*, *"Author Affiliation Name"*, *"Author Gender"* and *"First Author Gender"*.

3.2.3 Specifying functional form

I took the log of the variables *"Citations 10 Yrs Before"* and *"Citations 12 Yrs After",* as it is common practice in statistics to take the log of large integer values. Furthermore, these variables are heavily skewed, so taking the log thus normalises them.

3.2.4 Creating categorical variables to capture fixed effects

I added fixed effects to the regression model to avoid omitted variable bias. In STATA, fixed effects can be captured by generating categorical variables that are derived from the variables in the analysis. As such, I controlled for any individual-specific attributes related to years (or decades) and journals. In other words, I controlled for the average differences across time and journals in any observable or unobservable predictors (Wai, 2017). I created variables *"Year Fixed Effects"*, *"Decade Fixed Effects"*, *"Journal Fixed Effects",* and an additional dummy variable, *"Top 5 Journal"*, for this purpose.

3.2.5 Removing highly correlated variables

The correlation matrix (see **Appendix B, Table B.2**) shows that the author-level variables *"Author Citation Count"*, *"Document Count",* and *"H-Index"* are highly correlated. Owing to their correlation and because these variables all capture similar characteristics of the article's author, I dropped *"Author Citation Count*" and *"Document Count"* to avoid multicollinearity. Furthermore, the matrix shows that the author-level variable *"Co-author Count"* is also highly correlated with the abovementioned variables. Because of this high correlation, and because *"Co-Author Count"* and the article-level variable *"Number of Authors"* capture similar article characteristics, I dropped *"Co-Author Count"*" also to avoid multicollinearity.

Furthermore, the correlation matrix shows that the journal-level variables "*SNIP Score"* and *"SJR Score"* are highly correlated. I dropped "*SNIP Score"* owing to its similarity and correlation with *"SJR Score"*.

The final dataset has 610 observations and 18 variables.

3.3 Variable description

The variables in this dissertation are classified either as the dependent variable (*"Citations 12 Years After"*), main independent variables (*"Cited In Book"* and *"Citations 10 Years Before"*), control variables or fixed effects variables. I used the main independent variables to answer the two defined research questions. Additionally, I analysed the control variables and their relationship to the dependent variable to produce several related results.

I collected data related to journals, articles, and authors. Each variable included in the analysis corresponds to one of these three factors. Additionally, identification variables were added to the collected data to form the final dataset.

3.3.1 Identification variables

The variable *"Article Scopus ID*" is the unique identifier for each article. The variable *"Book Year"* defines the year in which each microeconomics textbook was published. Since an article may appear in more than one microeconomics textbook, published in different years, combining these variables serves as a unique identifier for each observation in the citations database.

3.3.2 Journal-related variables

*"SJR Scor*e*"* is a continuous variable that measures the scientific influence of a scholarly journal. SCImago developed the SJR (SCImago Journal Rank) Score which takes into account the number of citations received as well as the prestige of a particular journal. The score is based on the weighted average number of citations obtained over a chosen year for each article published in the journal over the preceding three years. The citation weighting depends on the subject field and the prestige of the citing journal (Guerrero-Bote & Moya-Anegón, 2012). In this study, the *"SJR Score"* represents 2020 calculated values. *"Journal Citation Count"* is a discrete variable and indicates the total number of citations that a journal has received during a given period – 2017-2020.

3.3.3 Article-level variables

"Citations 12 Yrs After" is a discrete variable that measures the total number of article citations since the reference event (where the reference event is defined as the year the article appeared in one of the sampled microeconomics textbooks, i.e. the *"Book Year"*). This is the dependent variable.

"Cited in Book" is a binary variable, reflecting whether an article is included in one of the sampled microeconomics textbooks. A value of 1 signifies that the article is included in a textbook, making it part of the treatment group. Conversely, a value of 0, the base group, indicates that an article is not included in a textbook, making it part of the control group. This is one of the main independent variables.

"Citations 10 Yrs Before" is a discrete variable that measures the total number of article citations in the ten years leading up to the reference event. This is one of the main independent variables.

"Article Publication Year" is a categorical variable that defines the year in which an article was published in a journal.

I utilised several article-level variables as control variables in this study. These variables consist of *"Number of Authors"*, *"Number of Pages"*, and *"Title Length"*. All three variables are discrete. *"Number of Authors"* represents the total count of authors associated with an article. *"Number of Pages"* quantifies the total page count of an article. Lastly, *"Title Length"* gauges the character count in the article's title, serving as a proxy for title conciseness and clarity.

3.3.4 Author-level variables (defined at the article level)

I used the following author-level variables as control variables: *"H-Index"*, *"Male First Author",* and *"Affiliation Ranking*". The *"H-Index"* is a continuous variable, and I used it to measure both the productivity and citation impact of an author's publication. An author's h-index is equal to the count of an author's publications for which that author was cited by other authors the same number of times or more. For example, an author who has published articles A, B, C, D and E, which have been cited 20, 19, 5, 3 and 2 times, respectively, has an h-index of 3. This is because the author has at most three articles that have been cited at least three times (Bornmann & Daniel, 2007). *"Affiliation Ranking"* is a discrete variable and is equal to the average 2019 Tilburg University Ranking achieved by the universities with which the article's authors are associated. I used this variable to measure the rank and prestige of the author's association with their university. In this study, all authors are affiliated with universities; hence, the ranking is standardised. *"Male First Author"* is a binary variable that I use to measure the degree of male dominance in an article. A 1 denotes that the first named author of the article is male. A 0, the base group, denotes that the first named author is female.

3.3.5 Journal and time fixed effects

The *"Top 5 Journal"* variable is derived from the *"Journal Citation Count"* variable and indicates whether an article was published in one of the top five most highly cited journals within the dataset. It takes a value of 1 if the article was published in a top five journal and 0

otherwise. On the other hand, the *"Journal Fixed Effects"* variable, also derived from *"Journal Citation Count"*, assigns a unique ID based on the value of the citation count. This variable captures both observed and unobserved variations in the dependent variable resulting from differences in impact, quality, visibility, rank, or prestige of the specific journal where the article was published.

Similarly, the *"Decade Fixed Effects"* and *"Year Fixed Effects"* variables are derived from *"Article Publication Year"*. They aim to account for observed and unobserved variations in the dependent variable associated with factors specific to the time of article publication. The *"Decade Fixed Effects"* variable is assigned a unique ID based on the decade in which the article was published, while the *"Year Fixed Effects"* variable is assigned a unique ID based on the year of publication.

Summary statistics for the above-mentioned variables (excluding the fixed effects variables) can be found in **Tables B.3–B.8** and **Figures B.1–B.5** of **Appendix B**.

3.4 Expected economic relationship of the control variables

3.4.1 Journal impact, quality, visibility, rank and prestige

The literature indicates that research articles published in prestigious journals receive more citations compared to those published in lower-ranked journals. This trend can be attributed to the enhanced visibility, perceived reliability and trustworthiness associated with high-impact journals. Consequently, when scholars are faced with two similar articles differing only in the journal they were published in, they are more likely to cite the article from the higher-ranked journal (Van der Pol et al., 2015). Moreover, researchers consider high-impact journals as an indicator of article quality, as authors must compete for publication in these journals owing to limited space, ensuring that only the best studies are featured. Callaham, Wears and Weber (2002) demonstrate this by showing that a weak paper published in a highly ranked journal would receive more citations than a strong paper published in a lower-ranking journal. Therefore, I anticipated a positive correlation between the dependent variable and variables that measure journal quality, rank, and prestige.

3.4.2 The number of authors/degree of author collaboration

Empirical evidence consistently demonstrates that co-authored work receives more citations, and researchers have proposed several explanations for this phenomenon. One such

explanation is the Matthew effect, a term coined by Robert K. Merton in 1968. The Matthew effect describes a cumulative process in which renowned authors with well-established communication networks receive greater recognition and popularity for their articles. Consequently, highly cited authors tend to receive even more citations, suggesting that their prominence plays a significant role in citation counts (Merton, 1968). Additionally, the positive correlation between co-authored work and citation counts can be attributed to the impact of citation networks. Each co-author typically possesses their own network of scientific contacts. Collaborating with more co-authors expands these networks, increasing the dissemination of knowledge through presentations at conferences, seminars, and workshops. This broader knowledge diffusion leads to greater recognition, attention, and, ultimately, to more citations (Bosquet & Combes, 2013). Another contributing factor is self-citations. When authors cite their own previously co-authored work, it naturally leads to more citations, particularly when multiple authors cite the same work (Glänzel & Thijs, 2004). Finally, the productivity of co-authors tends to increase as they share their knowledge, resulting in higher quality work that surpasses what individuals could achieve on their own within the same timeframe. The uniqueness and prominence of such collaborative efforts garner appreciation from the scientific community, leading to more citations (Adams et. al, 2005).

Based on these explanations, I expected a positive correlation between the dependent variable and the variables that quantify the number of co-authors participating in the research.

3.4.3 Length of a research article

The literature widely agrees that longer articles tend to receive more citations. Several arguments provide possible explanations for this pattern. First, it has been suggested that longer articles enjoy greater visibility within the journals where they are published, thus attracting more attention from readers (Leimu & Koricheva, 2005). Second, another explanation posits that longer articles offer more content to cite. Regardless of the paper's quality or relevance, the increased length raises the chances of including sections that are relevant to other researchers, thereby increasing the likelihood of citation (Falagas et al., 2013). Third, longer articles are often associated with higher quality. Academic research aims to convey novel and original ideas on pertinent topics, requiring comprehensive justifications and extensive descriptions for effective expression. Consequently, longer publications typically present original ideas and demand sophisticated analyses with detailed explanations. This implies that

these papers make substantial contributions to knowledge, thus labelling them as high-impact publications deserving of citations (Haslam et al., 2008).

Supporting this trend, Card and DellaVigna (2013) found that articles published in the top five economics journals display an average threefold increase in length compared to in the 1970s. This change is attributed to intensified competition for limited journal space. Authors strive to enhance the quality of their papers to improve the likelihood of publication, resulting in longer articles.

Simultaneously, research has also demonstrated that shorter articles tend to receive more citations. Some journals restrict article length owing to limited publishing space, which may incentivise authors to keep their papers concise (Fox, Paine & Sauterey, 2016). However, this explanation raises the possibility of multicollinearity under two conditions: 1) a positive correlation between journal quality and citation count, and 2) the presence of page length restrictions. In this scenario, the negative correlation observed is not attributed to page length but rather to the higher quality of the journal in which the paper is published. Furthermore, many journals, particularly open-access ones, do not impose such restrictions on page length.

An explanation for the observed negative correlation that is harder to refute involves scientists' increasing awareness of the time they spend on professional reading. Given the rising volume of academic articles being published, scientists prefer shorter papers as they allow them to engage with a wider range of published works, thus gaining exposure to more papers (Lozano & Salmerón, 2005).

Given the existing divide in the literature regarding this topic, I anticipated that both a positive and negative relationship with the dependent variable would be equally plausible outcomes in this study.

3.4.4 Title of a research article

The literature on article title characteristics indicates that articles with shorter and more concise titles tend to receive more citations compared to articles with longer titles. This trend is attributed to the enhanced clarity of the message and reader appeal associated with shorter titles. Consequently, shorter titles are more likely to be downloaded, read, and utilised, increasing their chances of being cited (Habibzadeh & Yadollahie, 2010). Based on these

findings, I anticipated a negative correlation between the variables measuring the characteristics of the article title and my dependent variable.

3.4.5 Author citedness

Various studies on author citedness demonstrate that highly cited authors have a greater likelihood of receiving a larger number of citations when they publish new work. These highly cited authors typically hold prominent positions in their respective fields of study and enjoy recognition within the scientific community, which contributes to the increased citation count they receive (Jiang, He & Ni, 2013). Additionally, the Matthew effect can be applied in this context. Well-known authors with established reputations have access to larger citation networks through which they can disseminate and promote their new work. Naturally, their citation counts are expected to be higher compared to authors without a similar network (Merton, 1968). Based on these observations, I anticipated a positive correlation between the variables that measure author citedness and the dependent variable.

3.4.6 Gender of the author (gender bias)

The impact of gender on citation counts has received extensive research attention. Numerous studies consistently demonstrate that publications dominated by male authors tend to receive more citations compared to those not dominated by male authors. One paper by Maliniak, Powers and Walter (2013) attributes this phenomenon to the preference of male authors for citing other male authors more frequently than female authors, while women tend to cite themselves less than they cite men. Other studies indicate that male authors have higher rates of self-citation than their female counterparts, resulting in a disproportionate allocation of citations to male authors (Roberts, Stewart & Nielsen, 2016).

Another potential explanation is the underrepresentation of women in their respective fields of study (Dion, Sumner & Mitchell, 2018). For example, prestigious economic journals often publish fewer women than men, leading to a lower citation count for women (Hengel & Moon, 2020). The Matthew effect may also contribute to the gender citation bias, where certain research subfields are predominantly composed of male researchers, resulting in higher citation rates for male-authored papers (Reece-Evans, 2010). The Matilda effect, which refers to the devaluation and misattribution of women's contributions, has also been proposed as a factor influencing gender bias in citations (Rossiter, 1993).

Observations indicate that males tend to publish more books and academic articles than females, increasing the likelihood of their work being cited (Dion, Sumner & Mitchell, 2018). Additionally, factors such as recognition and amplification of academic achievements have been identified as contributors to an author's success and citation count. An analysis of gender disparities on Twitter by Zhu et al. (2019) reveals that women in academic medicine are less likely to receive amplification, potentially explaining their lower citation rates. Furthermore, a study by Hengel and Moon (2020) suggests that women often produce higher quality work but publish fewer papers, contributing to lower citation counts.

Contrary to the notion of males being referenced more frequently, Hengel and Moon (2020) refute that claim by controlling for outliers, such as older highly cited works predominantly authored by men, typically associated with winning Nobel Prizes. The study concludes that female writers are frequently mentioned after accounting for these outliers.

Despite some contrasting findings, most research supports the notion that male authors tend to receive more citations. Therefore, in this study, I expected that articles dominated by male authors would exhibit higher citation counts.

3.4.7 Author affiliation

According to Lou and He (2015), an article receives more citations when its author is affiliated with a highly ranked or well-known institution. They propose three possible explanations for this phenomenon. First, readers may perceive papers from highly ranked institutions as being of better quality. Second, highly ranked institutions are inherently superior and likely to produce higher quality papers. Finally, researchers tend to reference works from highly ranked institutions because this reflects positively on their own work when evaluated for publication. The Matthew effect, mentioned earlier, could also account for this correlation. Authors affiliated with top-ranking universities are often involved in influential and well-connected research networks, resulting in increased exposure, recognition, and traction for their published papers (Merton, 1968). In this study, I expected that articles written by authors affiliated with highly ranked institutions would have more citations.

3.5 Multivariate regression analysis

In my analysis, I conducted a multivariate regression using STATA to establish a statistical relationship between the main explanatory variables, *"Cited in Book"* and *"Citations 10 Years*

Before", and the dependent variable *"Citations 12 Yrs After"*. I added different pairs of control variables and fixed effects sequentially.

3.5.1 Model and hypotheses

I constructed the multiple regression model as follows:

$$\log(\widehat{Citations12YearsAfter}) = \beta_0 + \delta_0 CitedInBook + \beta_1 \log(Citations10YearsBefore - Avg) + \delta_1 CitedInBook\# \log(Citation10YearsBefore - Avg) + control\ variables + fixed\ effects + u \quad (1)$$

First, this multivariate regression analysis aims to answer the two research questions of this dissertation:

1. Does citing academic literature in a textbook affect the subsequent citation count (attribution) thereof, and
2. If subsequent attribution is affected, is it strengthened or weakened, and what mechanisms are possibly responsible for this change, i.e., a *signalling* or a *cessation effect*?

I formed two hypotheses to test this:

Hypothesis 1: The rate of change, i.e., the slope of *"log(Citations 12 Yrs After)"* with respect to *"log(Citations 10 Yrs Before)",* is the same for articles cited and articles not cited in microeconomics textbooks.

$H_0: \delta_1 = 0$

Hypothesis 2: The average *"log(Citations 12 Yrs After)"* is identical for articles cited and articles not cited in microeconomics textbooks that have the same "*log(Citations 10 Yrs Before)"*.

$H_0: \delta_0 = 0, \delta_1 = 0$

Second, this multivariate regression analysis aims to address several related questions. It investigates the relationship between the control variables and the post-event citations and determines which group of control variables – journal-related, article-related, or author-related

– has the most significant impact on post-event citations. I did this by looking at the signs and significance of the various control variables in this study.

3.5.1.1 Description of the regression model

Equation 1 shows the base model, containing the independent variables *"Cited In Book"* and *"Citations 10 Yrs Before"*, where *"Citations 10 Yrs Before"* is reparameterised and taken as a log. I included an interaction term between *"Cited In Book"* and *"Citations 10 Yrs Before"* and different combinations of fixed effects in this base model. The fixed effects include *"Decade Fixed Effects", "Year Fixed Effects"*, *"Journal Fixed Effects", and "Top 5 Journal"*. In addition, the equation shows more restricted models, where I included the control variables: *"SJR Score"*, *"Number of Authors"*, *"Number of Pages"*, *"Title Length"*, *"H-Index"*, *"Male First Author",* and *"Author Affiliation"*.

I included an interaction term to the regression model, as it was necessary to test the statistical relationship between *"Cited in Book"* and *"Citations 12 Yrs After"* (post-event citations) for a specified level of *"Citations 10 Yrs Before"* (pre-event citations). Put differently, when comparing the treatment set and control set post-event citations, it is pertinent to do so for equal levels of pre-event citations. The coefficient δ_1 captures the interaction effect (Wooldridge, 2012).

Furthermore, to avoid inaccurate estimations of the coefficient of *"Cited in Book"*, I parameterised *"Citations 10 Yrs Before"* by subtracting its mean value. Reparameterisation is often required when adding an interaction term. In this study, multicollinearity arises when both *"Cited in Book"* and *"Cited in Book#log(Citations 10 Yrs Before)"* are included in the regression model. A useful way to think about the multicollinearity is as follows: δ_0 measures the difference in *"log(Citations 12 Yrs After)"* between articles cited and not cited when *"log(Citations 10 Yrs Before)"* = 0, but because there are no instances of *"log(Citations 10 Yrs Before)"* = 0 in the dataset, it becomes difficult to estimate δ_0 at low levels of *"log(Citations 10 Yrs Before)"*. It would be better to estimate the differential at the average *"log(Citations 10 Yrs Before)"* (Wooldridge, 2012).[2]

[2] See **Table B.3** in **Appendix B** for the average value of *"log(Citations 10 Yrs Before)"*.

Additionally, I included the control variables and fixed effects to control for any observed and unobserved variation in the dependent variable, *"Citations 12 Yrs After"*.

3.5.2 Analysis process

I used STATA to run 20 regressions, adding varying combinations of control variables and fixed effects, first starting at the base model and then building towards the most restricted model, which contains all the control variables.[3]

3.6 Limitations

To address the limitations of my dissertation, I considered several factors. The limitations emerge during the data collection phase of this study. Initially, I collected a sample of nine microeconomic textbooks, which is relatively large compared to similar studies utilising the textbook approach. However, missing data significantly reduces the dataset size, potentially affecting the accuracy of the results in representing the population of citations referenced in textbooks.

Another limitation pertains to the gender of the authors in the dataset. I obtained gender information from a database called gender.io, which unfortunately fails to recognise non-US names. Consequently, the sample of articles may exclude research produced by authors from countries with non-US names, potentially biasing the results towards research produced in the US.

Furthermore, including the author affiliation ranking variable introduced some issues. First, the rankings are standardised, meaning that the dataset only included research articles by authors affiliated with ranked research institutions, primarily universities. This introduced a bias towards university-affiliated authors. Additionally, I used university rankings obtained for the year 2019 rather than for the year of publication. This limitation is particularly evident in the Tilburg University ranking dataset, as the university ranking system was initiated in 2004, and most articles in this dataset were published before this year. Consequently, the affiliation ranking in this sample does not accurately represent the ranking at the time of publication but rather provides an estimation.

[3] I present a detailed explanation of each regression in Chapter 4 of this dissertation.

In future studies, addressing these limitations can be achieved by employing a larger sample of textbooks or more carefully curating the control variables so as not to drop too much missing data, excluding variables that measure author gender or ensuring accurate identification of authors with non-US names, and avoiding variables that measure author affiliation or obtaining the affiliation ranking specifically for the year in which the article is published.

Chapter 4: Results and Discussion

In this section, I report on the results of the 20 regressions, state the primary and related findings and discuss these in detail.

4.1 Regression results

I ran 20 regressions. Each regression has a combination of three components. The first component comprises all the variables of interest. These include the interaction term defined as *"Cited In Book#log(Citations 10 Yrs Before - Avg)"* and the main independent variables defined as *"Cited In Book"* and *"log(Citations 10 Yrs Before - Avg)".* The second component comprises the fixed effects. The fixed effects are defined as *"Top 5 Journal"*, *"Year Fixed Effects"*, *"Decade Fixed Effects",* and *"Journal Fixed Effects".* The third component comprises the control variables. There are three types of control variables: journal-related, author-related, and article-related.

The 20 regressions are split into five tables. I included four regressions with various combinations of journal- and year-fixed effects in each of the five regression tables. In the explanations that follow, I refer, in general, to these as **regressions one**, **two**, **three** and **four** (these are not bound to one specific table, but rather to the respective four columns of all five regression tables and are intended to distinguish the effects of the varying combinations of fixed effects). The first four regressions, regressions (4.1.1) to (4.1.4), contain only the first and second components (variables of interest and fixed effects). These regressions form the base model. The remaining 16 regressions have all three components (variables of interest, fixed effects and control variables). A journal control variable was added to regressions (4.2.1) to (4.2.4). The article control variables were added to regressions (4.3.1) to (4.3.4). Author control variables were added to regressions (4.4.1) to (4.4.4). Finally, all control variables were added to regressions (4.5.1) to (4.5.4) and form the most restricted model.

The regression results are presented in regression tables (4.1) through (4.5). In the explanations that follow, I refer to these as **regression tables (4.1)**, (**4.2)**, (**4.3)**, (**4.4)** and (**4.5)** and refer to the regression specific to a regression table as **regressions (4.1.1)** through **(4.5.4)**. The regression coefficients of each variable are presented first, and standard errors are shown in the brackets below the coefficients. The asterisk shows the level of significance. Tick marks indicate the specific combination of fixed effects included in the model. The adjusted R-

squared, number of observations, number of fixed effects, residual degrees of freedom, and residual sum of squares of the full model and reduced model are shown at the bottom section of each table.

Table 4.1: Regression results for the base model with various fixed effect combinations (columns 4.1.1 to 4.1.4)

Base model				
Dependent variable: "log(Citations 12 Yrs After")				
Independent variables	(4.1.1)	(4.1.2)	(4.1.3)	(4.1.4)
Cited In Book#log(Citations 10 Yrs Before - Avg)	-0.1041 (0.0703)	-0.1145 (0.0782)	-0.0853 (0.0734)	-0.1155* (.0697)
Cited In Book	0.4800*** (0.0670)	0.4019*** (0.0772)	0.4853*** (0.0731)	0.4576*** (0.0666)
log(Citations 10 Yrs Before - Avg)	1.0561*** (0.0471)	0.9667*** (0.0541)	0.9764*** (0.0503)	1.0618*** (0.0467)
Year Fixed Effects	✓			✓
Decade Fixed Effects			✓	
Top 5 Journal				✓
Journal Fixed Effects		✓	✓	
Constant	3.1167*** (0.7058)	5.9621*** (0.7673)	3.1140*** (0.7133)	3.1394*** (0.6989)
Adj R-squared	0.6444	0.5801	0.6373	0.6513
No. observations	610	610	610	610
No. of fixed effects	42	78	85	42
Residual df	564	527	521	563
Residual SS Full Model	278.4427	307.2068	262.3042	272.5417
Residual SS Reduced Model	304.8768	323.7449	284.8175	296.7415

*Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1*

Note: Table 4.1 presents the results of four base model regressions. Each model regresses the dependent variable, "*log(Citations 12 Yrs After)*", on the interaction term and independent variables: "*Cited In Book#log(Citations 10 Yrs Before - Avg)*", "*Cited In Book*", and "*log(Citations 10 Yrs Before - Avg)*". The variations across these models are in their fixed effects: regression (4.1.1) includes a "*Year Fixed Effect*", regression (4.1.2) incorporates a "*Journal Fixed Effect"*, regression (4.1.3) integrates both "*Decade*" and "*Journal Fixed Effects*", and regression

(4.1.4) combines a "*Year Fixed Effect*" with a "*Top 5 Journal*" fixed effect. The table provides regression coefficients, standard errors for each variable, and significance levels denoted as *p < 0.05, **p < 0.01, and ***p < 0.001. Additionally, it reports the adjusted R-squared, number of observations, number of fixed effects, residual degrees of freedom, and the residual sum of squares for both full and reduced models. Notably, Regression (4.1.4) displays the highest adjusted R-squared and has the most significant coefficients.

Table 4.2: Regression results for the base model plus a journal control variable (*SJR Score*) with various fixed effects (columns 4.2.1 to 4.2.4)

Base model + Journal control variable				
Dependent variable: "log(Citations 12 Yrs After)"				
Independent variables	(4.2.1)	(4.2.2)	(4.2.3)	(4.2.4)
Cited In Book#log(Citations 10 Yrs Before - Avg)	-0.1071 (0.0696)	-0.1112 (0.0782)	-0.0837 (0.0735)	-0.1165* (0.0691)
Cited In Book	0.5057*** (0.0667)	0.3993*** (0.0772)	0.4830*** (0.0733)	0.4831*** (0.0666)
log(Citations 10 Yrs Before - Avg)	1.0321*** (0.0471)	0.9660*** (0.0541)	0.9759*** (0.0504)	1.0401*** (0.0468)
SJR Score	0.0110*** (0.0031)	0.2340 (0.1960)	0.1178 (0.1841)	0.0096*** (0.0031)
Year Fixed Effects	✓			✓
Decade Fixed Effects			✓	
Top 5 Journal				✓
Journal Fixed Effects		✓	✓	
Constant	3.0883*** (0.6984)	5.9053*** (0.7685)	3.0759*** (0.7162)	3.1116*** (0.6937)
Adj R-squared	0.6518	0.5804	0.6369	0.6566
No. observations	610	610	610	610
No. of fixed effects	42	78	85	42
Residual df	563	526	520	562
Residual SS Full Model	272.1455	306.3763	262.0978	267.9343
Residual SS Reduced Model	301.0952	322.6426	284.3242	294.3823

*Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1*

Note: Table 4.2 presents the results of four base model regressions, including a journal control variable ("*SJR Score*"). Each model regresses the dependent variable, "*log(Citations 12 Yrs After)*", on the interaction term and

independent variables: "*Cited In Book#log(Citations 10 Yrs Before - Avg)*", "*Cited In Book*", and "*log(Citations 10 Yrs Before - Avg)*". The variations across these models are in their fixed effects: regression (4.2.1) includes a "*Year Fixed Effect*", regression (4.2.2) incorporates a "*Journal Fixed Effect"*, regression (4.2.3) integrates both "*Decade*" and "*Journal Fixed Effects*", and regression (4.2.4) combines a "*Year Fixed Effect*" with a "*Top 5 Journal*" fixed effect. The table provides regression coefficients, standard errors for each variable, and significance levels denoted as *p < 0.05, **p < 0.01, and ***p < 0.001. Additionally, it reports the adjusted R-squared, number of observations, number of fixed effects, residual degrees of freedom, and the residual sum of squares for both full and reduced models. Regression (4.2.4) exhibits the highest adjusted R-squared and the most significant coefficients.

Table 4.3: Regression results for the base model plus article control variables ("*Number of Authors", "Number of Pages"* and "*Title Length"*) with various fixed effects (columns 4.3.1 to 4.3.4)

Base model + Article control variables				
Dependent variable: "log(Citations 12 Yrs After)"				
Independent variables	(4.3.1)	(4.3.2)	(4.3.3)	(4.3.4)
Cited In Book#log(Citations 10 Years Before - Avg)	-0.1087 (0.0697)	-0.0742 (0.0778)	-0.0826 (0.0740)	-0.1162* (0.0695)
Cited In Book	0.4449*** (0.0666)	0.4169*** (0.0760)	0.4843*** (0.0730)	0.4375*** (0.0665)
log(Citations 10 Years Before - Avg)	1.0455*** (0.0476)	0.9190*** (0.0542)	0.9603*** (0.0514)	1.0497*** (0.0474)
Number of Authors	-0.0458 (0.0322)	0.0940** (0.0395)	0.0261 (0.0386)	-0.0284 (0.0330)
Number of Pages	0.0072*** (0.0026)	0.0151*** (0.0041)	0.0055 (0.0041)	0.0056** (0.0027)
Title Length	-0.0211** (0.0081)	-0.0118 (0.0094)	-0.0154* (0.0090)	-0.0198** (0.0081)
Year Fixed Effects	✓			✓
Decade Fixed Effects			✓	
Top 5 Journal				✓
Journal Fixed Effects		✓	✓	
Constant	3.2524*** (0.6989)	5.8049*** (0.7601)	3.1272*** (0.7159)	3.2511*** (0.6965)
Adj R-squared	0.6543	0.5942	0.6387	0.6566
No observations	610	610	610	610
No of fixed effects	42	78	85	42

Base model + Article control variables				
Dependent variable: "log(Citations 12 Yrs After)"				
Independent variables	(4.3.1)	(4.3.2)	(4.3.3)	(4.3.4)
Residual df	561	524	518	560
Residual SS Full Model	269.2358	295.2048	259.8121	266.9486
Residual SS Reduced Model	291.7767	312.3799	282.1616	288.9178

*Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1*

Note: Table 4.3 presents the results of four base model regressions, including article control variables ("*Number of Authors*", "*Number of Pages*" and "*Title Length*"). Each model regresses the dependent variable, "*log(Citations 12 Yrs After)*", on the interaction term and independent variables: "*Cited In Book#log(Citations 10 Yrs Before - Avg)*", "*Cited In Book*", and "*log(Citations 10 Yrs Before - Avg)*". The variations across these models are in their fixed effects: regression (4.3.1) includes a "*Year Fixed Effect*", regression (4.3.2) incorporates a "*Journal Fixed Effect"*, regression (4.3.3) integrates both "*Decade*" and "*Journal Fixed Effects*", and regression (4.3.4) combines a "*Year Fixed Effect*" with a "*Top 5 Journal*" fixed effect. The table provides regression coefficients, standard errors for each variable, and significance levels denoted as *p < 0.05, **p < 0.01, and ***p < 0.001. Additionally, it reports the adjusted R-squared, number of observations, number of fixed effects, residual degrees of freedom, and the residual sum of squares for both full and reduced models. Regression (4.3.4) has the highest adjusted R-squared and the most significant coefficients.

Table 4.4: Regression results for the base model plus author control variables ("H-Index", "Male First Author" and "Author Affiliation") with various fixed effects (columns 4.4.1 to 4.4.4)

Base model + Author control variables				
Dependent variable: "log(Citations 12 Yrs After)"				
Independent variables	(4.4.1)	(4.4.2)	(4.4.3)	(4.4.4)
Cited In Book#log(Citations 10 Years Before - Avg)	-0.1054 (0.0701)	-0.1170 (0.0775)	-0.0903 (0.0731)	-0.1161* (0.0695)
Cited In Book	0.4691*** (0.0670)	0.3817*** (0.0769)	0.4693*** (0.0731)	0.4488*** (0.0666)
log(Citations 10 Years Before - Avg)	1.0396*** (0.0479)	0.9370*** (0.0543)	0.9547*** (0.0507)	1.0445*** (0.0475)
H-Index	0.0031 (0.0020)	0.0077*** (0.0023)	0.0053** (0.0022)	0.0036* (0.0020)
Male First Author	0.1096 (0.0887)	0.0928 (0.1031)	0.1264 (0.0964)	0.0884 (0.0881)
Author Affiliation	-0.0002 (0.0002)	-2.93e-05 (0.0002)	0.0001 (0.0002)	-7.67e-05 (0.0002)

Year Fixed Effects	✓			✓
Decade Fixed Effects			✓	
Top 5 Journal				✓
Journal Fixed Effects		✓	✓	
Constant	3.0778*** (0.7046)	5.5545*** (0.7740)	3.0399*** (0.7102)	3.0892*** (0.6983)
Adj R-squared	0.6463	0.5878	0.6413	0.6527
No observations	610	610	610	610
No of fixed effects	42	78	85	42
Residual df	561	524	518	560
Residual SS Full Model	275.4625	299.8600	257.9671	270.0229
Residual SS Reduced Model	300.6795	314.7916	278.8916	293.2968

*Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1*

Note: Table 4.4 presents the results of four base model regressions, including author control variables ("*H-Index*", "*Male First Author*" and "*Author Affiliation*"). Each model regresses the dependent variable, "*log(Citations 12 Yrs After)*", on the interaction term and independent variables: "*Cited In Book#log(Citations 10 Yrs Before - Avg)*", "*Cited In Book*", and "*log(Citations 10 Yrs Before - Avg)*". The variations across these models are in their fixed effects: regression (4.4.1) includes a "*Year Fixed Effect*", regression (4.4.2) incorporates a "*Journal Fixed Effect"*, regression (4.4.3) integrates both "*Decade*" and "*Journal Fixed Effects*", and regression (4.4.4) combines a "*Year Fixed Effect*" with a "*Top 5 Journal*" fixed effect. The table provides regression coefficients, standard errors for each variable, and significance levels denoted as *p < 0.05, **p < 0.01, and ***p < 0.001. Additionally, it reports the adjusted R-squared, number of observations, number of fixed effects, residual degrees of freedom, and the residual sum of squares for both full and reduced models. Regression (4.4.4) exhibits the highest adjusted R-squared and the most significant coefficients.

Table 4.5: Regression of the most restricted model (including all control variables) with various fixed effect combinations (columns (4.5.1) to (4.5.4)), explaining "*log(Citations 12 Yrs After)"*

Base model + Author control variables				
Dependent variable:" log(Citations 12 Yrs After)"				
Independent variables	(4.5.1)	(4.5.2)	(4.5.3)	(4.5.4)
Cited In Book#log(Citations 10 Years Before - Avg)	-0.1112 (0.0695)	-0.0748 (0.0773)	-0.0861 (0.0738)	-0.1179* (0.0691)

Base model + Author control variables				
Dependent variable:" log(Citations 12 Yrs After)"				
Independent variables	(4.5.1)	(4.5.2)	(4.5.3)	(4.5.4)
Cited In Book	0.4615*** (0.0670)	0.3967*** (0.0758)	0.4681*** (0.0732)	0.4532*** (0.0669)
log(Citations 10 Years Before - Avg)	1.0156*** (0.0484)	0.8927*** (0.0544)	0.9395*** (0.0519)	1.0200*** (0.0483)
SJR Score	0.0080** (0.0032)	0.2060 (0.1923)	0.0939 (0.1840)	0.0075** (0.0032)
Number of Authors	-0.0457 (0.0320)	0.0941** (0.0394)	0.0261 (0.0386)	-0.0304 (0.0329)
Number of Pages	0.0060** (0.0027)	0.0142*** (0.0041)	0.00521 (0.0041)	0.0047* (0.0028)
Title Length	-0.0171** (0.0083)	-0.0108 (0.0094)	-0.0145 (0.0090)	-0.0165** (0.0083)
H-Index	0.0036* (0.0019)	0.0071*** (0.0022)	0.0053** (0.0022)	0.0038* (0.0019)
Male First Author	0.0638 (0.0880)	0.0638 (0.1021)	0.1036 (0.0970)	0.0558 (0.0878)
Author Affiliation	1.95e-05 (0.0002)	6.04e-5 (0.0002)	0.0001 (0.0002)	6.61e-05 (0.0002)
Year Fixed Effects	✓			✓
Decade Fixed Effects			✓	
Top 5 Journal				✓
Journal Fixed Effects		✓	✓	
Constant	3.1622*** (0.6951)	5.3818*** (0.7681)	3.0186*** (0.7163)	3.1610*** (0.6932)
Adj R-squared	0.6587	0.6006	0.6417	0.6606
No observations	610	610	610	610
No of fixed effects	42	78	85	42
Residual df	557	520	514	556
Residual SS Full Model	263.9201	288.2987	255.6473	262.0044
Residual SS Reduced Model	287.5836	303.7194	276.3065	284.9606

*Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1*

Note: Table 4.5 presents the regression results of four base model regressions, including all control variables ("*Number of Authors*", "*Number of Pages*", "*Title Length*", "*H-Index*", "*Male First Author*", "*Author Affiliation*"

and "*SJR Score*"). Each model regresses the dependent variable, "*log(Citations 12 Yrs After)*", on the interaction term and independent variables: "*Cited In Book#log(Citations 10 Yrs Before - Avg)*", "*Cited In Book*", and "*log(Citations 10 Yrs Before - Avg)*". The variations across these models are in their fixed effects: regression (4.5.1) includes a "*Year Fixed Effect*", regression (4.5.2) incorporates a "*Journal Fixed Effect"*, regression (4.5.3) integrates both "*Decade*" and "*Journal Fixed Effects*", and regression (4.5.4) combines a "*Year Fixed Effect*" with a "*Top 5 Journal*" fixed effect. The table provides regression coefficients, standard errors for each variable, and significance levels denoted as *p < 0.05, **p < 0.01, and ***p < 0.001. Additionally, it reports the adjusted R-squared, number of observations, number of fixed effects, residual degrees of freedom, and the residual sum of squares for both full and reduced models. Regression (4.5.4) exhibits the highest adjusted R-squared and the most significant coefficients among all regressions in this thesis.

4.2 Discussion

4.2.1 General trends and commentary on the different regressions

4.2.1.1 Multicollinearity

It is important to note that regressions two and three, which include *"Journal Fixed Effects"* and *"Top 5 Journal"* variables, exhibit high average variable inflation factors (VIF) when *"SJR Score"* is included in the regressions (shown in regression tables (4.2) and (4.5)). This high VIF is a result of multicollinearity between *"SJR Score"* and certain *"Top 5 Journal"* and *"Journal Fixed Effects*" dummy variables. It is crucial to recognise that multicollinearity only affects the collinear variables themselves, and it does not impact the coefficients of the variables of interest or compromise the effectiveness of the control variables as controls. However, it is worth noting that the coefficients of collinear control variables may be less stable owing to increased standard errors (Allison, 2012). To obtain more accurate coefficient estimates, relying on regressions one and four for *"SJR Score"* is advisable. The average VIFs of all regressions are displayed in **Table C.1** of **Appendix C**.

4.2.1.2 Overfitting

I included different combinations of *"Journal Fixed Effects"*, *"Top 5 Journal"*, *"Year Fixed Effects"*, and *"Decade Fixed Effects"* in each of the five regression tables and indicated the number of fixed effects incorporated at the bottom of each table. The inclusion of *"Journal Fixed Effects"* in regressions two and three leads to a substantial number of parameters, as it has 78 fixed effects. According to the one in ten-rule in statistics, it is recommended to have at least 10 observations per parameter to avoid overfitting and to ensure the detection of reasonably sized effects with sufficient power (Harrel, 2022). Since regressions two and three

fall short of this criterion, I considered the possibility of overfitting, implying decreased accuracy (and a higher probability of failing to reject the null hypothesis) when testing the specified hypotheses.

4.2.1.3 Adjusted R-squared

The adjusted R-squared of the regressions ranges between 0.5801 and 0.6606. Regression two (which includes *"Journal Fixed Effects"*) has the lowest adjusted R-squared in all five regression tables. Regression three (which includes *"Journal Fixed Effects"*, while *"Decade Fixed Effects"*) has the second lowest. Lower adjusted R-squared values are expected for these regressions since they have more parameters. However, the adjusted R-squared values of regression three are not much lower than those of regressions one (which includes *"Year Fixed Effects"*) and four (which includes *"Year Fixed Effects"* and *"Top 5 Journal"*). In contrast, regression two has a much lower value, and I ascribe this to the fact that undefined time-specific attributes have a noticeable impact on post-event citations. Regressions one and four have the highest adjusted R-squared in all five regression tables. Regression (4.5.4) has the highest adjusted R-squared.

Regression (4.5.4) does not have a high VIF, meets the one-in-ten rule criterion, and has the highest adjusted R-squared values. For these reasons, I used regression (4.5.4) when numerically interpreting the results in the following sections.

4.2.2 Primary findings: variables of interest

The variables of interest in this study include the interaction term *"Cited In Book#log(Citations 10 Yrs Before - Avg)"*, as well as the two main independent variables, *"Cited In Book"* and *"Citations 10 Yrs Before"*. These variables play a crucial role in addressing the two research questions of this study. The first question aims to assess the impact of a textbook citation event on subsequent citation counts. Building on this, the second question delves into understanding the specific mechanisms underlying this effect and how it influences subsequent citation counts. To address these questions, two hypothesis tests are formulated based on the coefficients of the variables of interest. The first hypothesis test is denoted as H_0: $\delta_1 = 0$, while the second hypothesis test is represented as H_0: $\delta_0 = 0, \delta_1 = 0$.

I restate Equation 1, as discussed in the methodology section, below:

$$\log(\widehat{Citations12YearsAfter}) = \beta_0 + \delta_0 CitedInBook + \beta_1 \log(Citations10YearsBefore - Avg) + \delta_1 CitedInBook\# \log(Citation10YearsBefore - Avg) + control\ variables + fixed\ effects + u \quad (1)$$

4.2.2.1 Theoretical interpretation of the coefficients of the variables of interest

When a model includes an interaction term, relevant variable coefficients should be interpreted together to derive meaning from them. Generally, the coefficients of the variables of interest are interpreted as follows:

The coefficient of *"Cited In Book"* (δ_0) is used to indicate the y-intercept of *"Cited In Book"* = 0 (articles not cited). The sum of the constant and the coefficient of *"Cited In Book"* ($\beta_0 + \delta_0$) is used to indicate the y-intercept of *'Cited In Book"* = 1 (articles cited). Therefore, the coefficient of *"Cited In Book"* measures the intercept difference between articles cited when pre-event citations are equal to zero. In this study, however, the coefficient measures the difference at the average level of pre-event citations (3.48), as *"log(Citations 10 Years Before)",* parameterised to improve the estimation accuracy of the coefficient of *"Cited In Book*"

The coefficient of *"log(Citations 10 Yrs Before - Avg)"* (β_1) represents the partial effect of *"log(Citations 10 Yrs Before - Avg)"* on *"log(Citations 12 Yrs After)"* and shows the approximate percentage change in *"Citations 12 Yrs After"* for *"Cited In Book"* = 0 when *"Citations 10 Yrs Before – Avg"* increases by 1%. The sum of the coefficients of *"log(Citations 10 Yrs Before - Avg)"* and the interaction term ($\beta_1 + \delta_1$) represents the approximate percentage change in *"Citations 12 Yrs After"* for *"Cited In Book"* = 1 when *"Citations 10 Yrs Before – Avg"* increases by 1%. The coefficient of the interaction term thus measures the difference in the rate of change (slopes) of post-event citations with respect to pre-event citations between articles cited and not cited.

Figure 4.1 represents the theoretical relationship (with application to the results of this study) of the variables of interest described above, where the two lines represent the two separate groups (*"Cited In Book"* = 0 and *"Cited In Book"* = 1), each with their intercept and slope.

Figure 4.1: The difference in post-event citations between articles cited and not cited at various levels of pre-event citations *("log(Citations 12 Yrs After)"/"log(Citations 10 Yrs Before - Avg)"*)

Note: The blue line represents the slope of the treatment group (cited) and is smaller than the red line, representing the slope of the control group (not cited). Point **a** (treatment group y-intercept) lies above point **b** (control group y-intercept). Point **c** represents the intersection of the two slopes. The treatment group consistently surpasses the control group for most levels of pre-event citations, indicating higher post-event citations. However, a point of intersection exists, suggesting that the control group exhibits more post-event citations at very high levels of pre-event citations.

4.2.1.2 Practical interpretation of the coefficients of the variables of interest

Tables 4.1 to **4.5** show the regression results of the variables of interest. I interpret the regression results as follows:

The interaction term coefficient (δ_1) is negative in all the regressions. This coefficient is significant at the 10% significance level in some regressions. The variable *"Cited In Book"* (δ_0), has a large positive coefficient and is significant at the 1% significance level in all the regressions. Similarly, the coefficient of *"log(Citations 10 Yrs Before - Avg)"* (β_1) has a large positive coefficient and is significant at the 1% significance level in all the regressions. The coefficient of the constant (β_0) is positive and significant at the 1% significance level. These results point to three key discussion points.

1. **Articles cited in textbooks have a smaller slope than articles not cited.**

Referring to the theory and **Figure 4.1**, I assert that the slope of *"Cited In Book"*= 0 is β_1, and the slope of *"Cited In Book"* = 1 is $\beta_1 + \delta_1$. The results indicate a negative value for δ_1, implying that the slope of *"Cited In Book"* = 0 is greater than the slope of *"Cited In Book"* = 1. The angled lines in **Figure 4.1** portray the difference in slopes between the two groups. These findings suggest that *"Cited In Book"* = 0 will have a higher return to *"log(Citations 10 Yrs Before)*" – a 1% increase in *"Citations 10 Yrs Before"* will lead to a larger percentage increase in *"Citations 12 Yrs After"* for articles not cited. When applying this to regression (4.5.4), the rate of post-event citations relative to pre-event citations is estimated at 102% for articles that are not cited. Conversely, for cited articles, the rate is calculated as 102% - 11.79% = 90.21%. Consequently, there is a difference of 11.79% less for cited articles.

2. **Articles cited in textbooks have a greater y-intercept than articles not cited.**

In the theoretical framework, I showed that the y-intercept of *"Cited In Book"* = 0 is β_0, while the y-intercept of *"Cited In Book"* = 1 is $\beta_0 + \delta_0$. The results indicate that the y-intercept of *"Cited In Book"* = 1 exceeds that of *"Cited In Book"* = 0, owing to the positive value of δ_0. The disparity in *"log(Citations 12 Yrs After)"* between *"Cited In Book"* = 0 and *"Cited In Book"* = 1 is δ_0, observed when *"log(Citations 10 Yrs Before)"* is 3.48 (representing the average).[4] This is observed as the distance between points **a** and **b** in **Figure 4.1**. Consequently, it becomes evident that, at average levels of pre-event citations, articles that are cited exhibit higher post-event citations. In the context of regression (4.5.4), articles that are not cited possess a y-intercept of 3.16, while articles that are cited have a y-intercept of 3.16 + 0.45 = 3.61, resulting in a difference of 0.45 in favour of articles that are cited.

3. **A threshold of pre-event citations exist at which post-event citations are lower for articles that are cited. However, it is noteworthy that this threshold exceeds the maximum value of pre-event citations observed in this study.**

Figure 4.1 illustrates point **c**, where the lines representing *"Cited In Book"* = 1 and *"Cited In Book"* = 0 intersect. This point signifies the pre-event citation level at which post-event citations are equivalent for both cited and uncited articles. By utilising **Equation 1**, this level can be computed as - δ_0/δ_1. When the values of *"Citations 10 Yrs After – Avg"* exceed this

[4] **Table B.3** of **Appendix B** shows the mean value for pre-event citations and other variable summary statistics.

threshold, post-event citations tend to be lower for cited articles. This implies that, at a certain pre-event citation level (above the calculated average of 3.48), articles that are cited exhibit fewer post-event citations compared to those that are not cited. Utilising this information in regression (4.5.4), the aforementioned point is determined as -0.4532/-0.1179, resulting in a value of 3.84. For pre-event citation levels (adjusted for the average) exceeding this threshold, cited articles tend to exhibit lower post-event citations. However, within this sample, the maximum value of pre-event citations (adjusted for the average) is 3.25, which falls below the calculated slope interception point.[5] Consequently, post-event citations are only lower for articles that are cited when pre-event citations reach very high levels. In this specific sample, such high values are not present.

The preceding discussion relied solely on the signs and magnitudes of the relevant coefficients observed in the specific sample of articles of this study. Consequently, it became imperative to conduct tests in order to ascertain whether there is sufficient evidence against the two null hypotheses proposed, thereby substantiating the validity of the discussion for the wider population of articles.

Hypothesis 1: The rate of change, i.e., the slope of *"log(Citations 12 Yrs After)"* with respect to *"log(Citations 10 Yrs Before)",* is the same for articles cited and articles not cited in microeconomics textbooks.

$H_0: \delta_1 = 0$

I used a t-test to test whether the coefficient of the interaction term (δ_1) is significantly different from zero. For some regressions, especially those that include *"Year Fixed Effects"* and the *"Top 5 Journal"* dummy, the interaction term coefficient is moderately significant – at the 10% significance level. This suggests that there is some evidence against the hypothesis that the rate of change of post-event citations with respect to pre-event citations is the same for articles cited and articles not cited. I reject the null hypothesis at the 10% significance level for these regressions.

In relation to regression (4.5.4), the analysis reveals an economically large and statistically significant difference in slope percentages between articles that are cited and those that are not cited, with a magnitude of -11.79%. This finding provides economic significance and supports

[5] **Table B.3** of **Appendix B** shows the maximum value of pre-event citations adjusted for the average, along with other variable summary statistics.

evidence against the null hypothesis, indicating that the slopes of cited and uncited articles are indeed distinct. Specifically, the slope of cited articles demonstrates a smaller value, indicating a lower return to pre-event citations than those not cited.

Hypothesis 2: The average *"log(Citations 12 Yrs After)"* is identical for articles cited and articles not cited in microeconomics textbooks that have the same "*log(Citations 10 Yrs Before)"*.

H_0: $\delta_0 = 0, \delta_1 = 0$

I used an F-test to test whether the coefficient of *"Cited In Book"* (δ_0) and the interaction term (δ_1) are significantly different from zero. I calculated the F-statistic by comparing the full regression model to a reduced model, which excludes the variables *"Cited In Book"* and the interaction term – simulating a situation where $\delta_0 = 0$ and $\delta_1 = 0$. The results I used to calculate the F-statistics, including the SSE of the full model and the reduced model, along with the residual degrees of freedom, are reported in the last three lines of **Tables 4.1** to **4.5**. The computed F-statistics for the various regressions range from 13.05 to 29.94.[6] This is a very large value for an F random variable with numerator *df* = 2 and denominator of range *df* = 514 to *df* = 564, producing a p-value < 0.001. The small p-value suggests sufficient evidence against the null hypothesis that average post-event citations are the same for articles cited and those not cited with the same pre-event citations. I reject the null hypothesis at the 1% significance level for all the regressions.

In relation to regression (4.5.4), the difference in post-event citations between articles cited and not cited at the average level of pre-event citations is 45.32%. This difference is economically large and has an F statistic of 24.36, which is highly significant for an F random variable with numerator *df* = 2 and denominator *df* = 556. There is sufficient evidence to suggest that the average number of post-event citations differs for articles cited and not cited with the same number of pre-event citations. Moreover, this evidence supports the notion that post-event citations are higher for cited articles.

This difference is clearly shown in **Figure 4.2**, which compares the normalised annual average citations between cited articles (treatment group) and those not cited (control group). Following

[6] Computations and calculated F-statistics are presented in **Tables C.2** to **C.6** of **Appendix C**.

the textbook citation event at time T, cited articles demonstrate a significantly greater number of citations compared to those not cited.

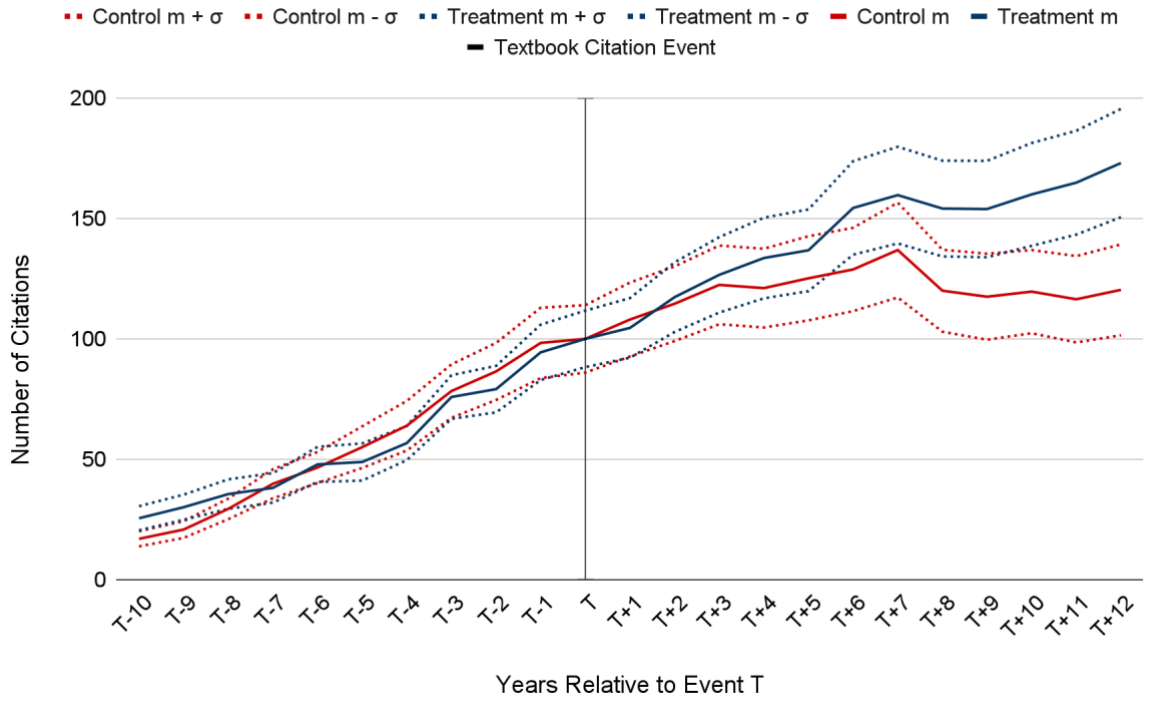


Figure 4.2: Comparison of normalised annual average citations before and after the textbook citation event (Time T), showing error ranges for cited and uncited articles.

Note: This figure illustrates the mean citation trajectories and associated error ranges of both the treatment and control groups, spanning from ten years before the textbook citation event (time T) to 12 years after the event. The blue solid line denotes the mean of the treatment group (cited articles), with blue dotted lines representing one standard deviation above and below this mean. Similarly, the red solid line represents the mean of the control group (uncited articles), and the red dotted lines denote its standard deviation range. Up until time T, the citation trajectories of both groups align closely. However, starting at T, a pronounced divergence is observed. The treatment group (cited articles) exhibits a significantly higher citation trend compared to the control group.

4.2.1.3 Discussion of the findings pertaining to the variables of interest

The findings of this study make a significant contribution to our understanding of the impact of citing academic literature in a textbook on subsequent attribution of the cited literature, as well as on the underlying mechanisms at play. Effectively, this provides answers to the two stated research questions:

1. **Does citing academic literature in a textbook affect the subsequent attribution of said academic literature?**

First, this thesis aims to investigate the impact of citing literature in well-known textbooks on subsequent attribution of the cited literature. Based on the results obtained, it is evident that a significant disparity exists in post-event citation counts between academic literature cited in textbooks and literature that is not cited. This is supported by the highly significant F-statistic, which allows for the rejection of H_0: $\delta_0 = 0, \delta_1 = 0$, as stated in the second hypothesis. Thus, the findings indicate that citing literature in a textbook does indeed influence the post-event citation count of the cited literature. Considering post-event citation counts as a proxy for attribution, it can be stated that textbook citation events exert a significant effect on the subsequent attribution of cited literature.

2. **If subsequent attribution is affected, is it strengthened or weakened, and what mechanisms are possibly responsible for this change?**

Having established the significant impact of the textbook citation event on attribution, the second objective of this thesis is to determine whether this effect strengthens attribution, aligning with a signalling effect, or weakens attribution, aligning with a cessation effect. From the preceding discussion and hypotheses tests, I summarise the key findings as follows:

Articles that are cited have higher levels of post-event citations compared to articles that are not cited, particularly at lower levels of pre-event citations. This assertion is supported by the significant positive coefficient of the variable *"Cited In Book"*, indicating that articles cited have 45.32% more post-event citations compared to articles not cited for average levels of pre-event citations. The rejection of H_0: $\delta_0 = 0, \delta_1 = 0$, as stated in the second hypothesis, serves as further support for this statement. As the number of pre-event citations increases, the disparity between the two groups diminishes. Eventually, there comes a point where post-event citations for cited articles become lower than those for uncited articles. This assertion is supported by the significant coefficient of the interaction term, which leads to the rejection of H_0: $\delta_1 = 0$ as stated in the first hypothesis. Nevertheless, in this study, even though the gap narrows, cited articles consistently exhibit higher post-event citations. This is due to the fact that the maximum value of pre-event citations, calculated as 3.25, exceeds the point of intersection of the slopes, calculated as 3.84 (represented by point **c** in **Figure 4.1**).

Based on this summary, it is evident that when academic articles are cited in well-known textbooks, a general trend emerges: their subsequent citations tend to increase. These findings offer compelling evidence of strengthened attribution, which can be attributed to the underlying

signalling effect. This implies that literature incorporated into textbooks have amplified acknowledgement and appreciation.

It is worth noting that, within the broader population of research articles, there is a possibility of a different pattern emerging. Particularly, articles with exceptionally high levels of pre-event citations, often associated with groundbreaking and highly influential research, may exhibit some evidence of weakened attribution once they are referenced in well-known textbooks. This highlights the presence of a cessation effect. Thus, highly cited literature included in textbooks may see a decrease in acknowledgement and appreciation, potentially marking a point at which information becomes common knowledge. While the significance of this effect is not pronounced in this study, it is important to acknowledge its potential existence. Future studies could focus specifically on highly cited research articles to investigate this phenomenon further.

This study aimed to explore the concept of data provenance within scientific research. The results of this study have two primary applications: first, to examine the extent to which the social convention of academic referencing serves as a robust, binding, and enduring mechanism; and second, to apply the new-found knowledge of provenance in academia to other types of information, particularly data.

Based on the compelling evidence that confirms strengthened attribution, I can confidently assert that provenance remains intact. There is little evidence to suggest that the social convention of academic referencing is weak or lacks durability. However, it is important to consider the moderately significant results that confirm weakened attribution for articles exhibiting exceptionally high levels of pre-event citations. These findings shed light on a plausible scenario where academia may encounter challenges in preserving the complete provenance of information. It implies that the social convention of academic referencing may have a vulnerability or a diminishing effect when highly cited papers are incorporated into well-known textbooks.

As mentioned before, this study establishes the effectiveness of the social convention of academic referencing in preserving the provenance of information in academia. This provides valuable insights into the establishment and maintenance of provenance for other forms of information, particularly data. With its emphasis on referencing sources and acknowledging data ownership, academic referencing can serve as a model for establishing data provenance. By ensuring that metadata include information about the origin of the data, the owner/producer,

and the sources and methods employed in their generation, we can attribute ownership and establish data provenance, thereby mitigating the risk of the loss of data provenance.

This study investigated the mechanisms that strengthen or weaken attribution, with a specific focus on the *cessation* and *signalling* effects. In the introduction, I presented a detailed explanation of these mechanisms and proposed potential reasons for their occurrence. As previously discussed, this dissertation strongly supports the presence of a signalling effect. However, the precise cause of this effect remains uncertain. It is unclear whether the signalling effect arises solely from the objective superiority of the articles included in textbooks or whether it stems from their popularisation through textbook citation. Similarly, the exact factor responsible for the perceived cessation effect among highly cited research remains uncertain. For example, it is unclear whether this effect occurs because highly cited research included in textbooks is already widely recognised as common knowledge*;* whether the inclusion in textbooks promotes these highly cited research findings to the status of common knowledge, or is due to individuals that tend to cite the well-known textbook instead of the highly cited research after its inclusion. Future studies could delve deeper into investigating the origins of these effects.

4.2.2 Related results: control variables

Control variables enhance the internal validity of a study by mitigating the influence of confounding and extraneous variables. This allows for the establishment of a robust statistical relationship between the dependent variable and the variables of interest. In line with earlier literature discussed in Chapter 3, this study controlled for various factors related to the journal, article, and author, which have been found to impact article citation counts.

In this section, I analyse the coefficients of the control variables, considering their signs, magnitude, and significance. Additionally, I informally test the secondary "hypotheses" outlined in Chapter 3 and offer potential explanations for any observed disagreements with prior research findings.

4.2.2.1 Journal control variables: findings and discussion

The journal control variable, *"SJR Score"*, is included in regression tables (4.1) and (4.5) and is used as a proxy to measure journal impact.

SJR score

The variable *"SJR Score"* exhibits a positive coefficient across all regression analyses. This indicates that articles published in journals with higher SJR scores receive a higher number of post-event citations compared to those published in journals with lower SJR scores. This finding aligns with existing literature that suggests a positive relationship between journal ranking and citation counts.

Although the sign of the coefficient remains consistent across all regressions, there is variation in the size of the coefficients. Regressions one and four exhibit smaller coefficients, while regressions two and three have relatively larger coefficients. Notably, regressions two and three include *"Journal Fixed Effect"*" which is highly collinear with *"SJR Score"*. This high collinearity leads to inflated variance inflation factors (VIFs) and introduces potential instability in the coefficient estimates of *"SJR Score"* in these regressions. Furthermore, regressions two and three have a lower number of degrees of freedom owing to the inclusion of *"Journal Fixed Effect"*, which results in decreased precision.

Additionally, *"SJR Score"* is statistically significant at the 5% or 1% significance level in regressions one and four but does not reach significance in regressions two and three. The lack of significance in the latter regressions can be attributed to the multicollinearity issue and the reduced degrees of freedom (Hanck, 2015). Consequently, I conclude that the coefficient estimates in regressions one and four provide a more reliable indication of the true coefficient size. Holding all other variables constant, an increase of one unit in *"SJR Score"* is associated with an average increase of less than 1% in *"Citations 12 Yrs After"*, based on the coefficients obtained from regressions one and four.

Applying the aforementioned analysis to regression (4.5.4), it is observed that a one-unit increase in an article's SJR score corresponds to a 0.75% increase in *"Citations 12 Yrs After"*. This coefficient is statistically significant at the 5% level, indicating that there is enough evidence to support the proposition that articles published in journals with higher SJR scores tend to receive more post-event citations.

4.2.2.2 Article control variables: findings and discussion

In this study, I incorporate several article control variables, namely *"Number of Authors*" *"Number of Pages"*, and *"Title Length"*. These article control variables are included in regression tables (4.3) and (4.5).

Number of authors

The variable *"Number of Authors*" exhibits a negative coefficient in regressions one and four of each regression table, while it shows a positive coefficient in regressions two and three of each regression table. Furthermore, the size of the coefficient is consistently small and lacks statistical significance in most of the regressions. Therefore, no significant relationship exists between the degree of scientific collaboration, represented by the number of authors, and citation counts.

Number of pages

In all regressions, the variable *"Number of Pages*" exhibits a positive coefficient. This suggests that articles with more pages tend to receive more post-event citations. This finding aligns with previous literature that highlight a positive relationship between article length and citations. According to this literature, longer articles tend to receive more citations owing to their increased visibility, comprehensive content, and perceived higher quality.

The coefficients of *"Number of Pages"* are all smaller than 2%, suggesting that they are not economically large. Furthermore, tables three and five show that these coefficients are either significant at the 1%, 5% or 10% significance level for all relevant regressions. I affirm that, while holding all other variables fixed, a one-unit change in *"Number of Pages"* increases *"Citations 12 Yrs After"* by less than 2% on average.

Applying the above discussion to regression (4.5.4), a one-page increase in the *"Number of Pages"* variable corresponds to a 0.47% increase in *"Citations 12 Yrs After"*. This coefficient is statistically significant at the 10% significance level, providing some evidence to support the notion that articles with a greater number of pages tend to have higher post-event citations.

Title length

The variable *"Title Length"* exhibits a negative coefficient in all regressions, indicating that articles with shorter titles have more post-event citations than those with longer titles. This finding aligns with prior research on article-title characteristics, which suggests that articles with more concise titles tend to receive higher citation counts.

Similar to the previous control variables, the coefficients for *"Title Length"* are not economically large. Moreover, they are significant at the 5% significance level in regressions one and four but less significant in regressions two and three. The lower significance in these

regressions is attributed to the higher standard errors resulting from the increased number of parameters. By utilising the coefficients from regressions one and four, it can be concluded that, while keeping all other variables constant, a one-letter increase in *"Title Length"* leads to an average decrease of less than 3% in *'Citations 12 Yrs After"*.

Applying the above discussion to regression (4.5.4), a one-letter increase in *"Title Length"* results in a 1.65% decrease in *"Citations 12 Yrs After"*. This coefficient is significant at the 5% significance level, indicating that there is sufficient evidence to support the notion that articles with longer titles have lower citation counts.

4.2.2.3 Author control variables: findings and discussion

The author control variables are *"H-Index"*, *"Male First Author",* and *"Author Affiliation*". The author control variables are included in regression tables (4.4) and (4.5).

H-index

"H-Index" exhibits a positive coefficient in all regressions, indicating a positive relationship between the average h-index of authors and the number of post-event citations for their articles. This finding aligns with the existing literature, which suggests that higher author citedness is associated with higher article citedness.

The significance of *"H-Index"* varies across regressions, with most showing significance at the 1%, 5%, or 10% level. Despite its significance, the effect size is small, as a one-unit increase in *"H-Index"* leads to less than a 1% average increase in *"Citations 12 Yrs After*".

Applying this interpretation to regression (4.5.4), a one-unit increase in *"H-Index"* corresponds to a 0.38% increase in *"Citations 12 Yrs After"*. The coefficient is significant at the 10% significance level, providing some evidence that articles with higher average h-indices tend to receive more citations.

Male first author

'Male First Author" exhibits a positive regression coefficient in all regressions, indicating that articles with a male first author tend to have more citations in this study. This finding aligns with the existing literature, which demonstrates that publications by men generally receive more citations.

The expected difference in post-citations between articles with a male first author and those with a female first author ranges from 5.58% to 12.64% in the relevant regressions, indicating a substantial difference. However, none of these coefficients is significant. Consequently, I cannot confidently assert that the observed relationship between post-event citations and *"Male First Author"* accurately reflects the entire population of articles.

Applying these interpretations to regression (4.5.4), the analysis reveals that, on average, articles with male first authors receive approximately 5.58% more post-event citations than those with female first authors. However, this result is not statistically significant, providing no evidence to support the presence of a difference in post-event citations between articles with male and with female first authors.

Author affiliation

The variable *"Author Affiliation"* exhibits a negligible coefficient (ranging from zero to three decimal places) in all regressions, and none of the coefficients is statistically significant. These findings indicate that *"Author Affiliation"* contributes very little to the variation in post-event citations and, as a result, no definitive interpretation can be drawn from the sign and significance of this control variable.

4.2.3 Related results: a comparison between journal, article, and author control variables

The literature clearly distinguishes between journal-, article-, and author-level effects, with studies often focusing on specific variables within each category to determine their significance. However, in this study, these distinct effects are utilised as control variables to minimise error-term variation. By running separate regressions for the journal, article, and author variables, I could assess the importance of each variable category in explaining the variation in *"Citations 12 Yrs After"*.

Regression tables (4.2), (4.3), and (4.4) incorporate the three variable categories individually. To compare these non-nested models, I utilised the adjusted R-squared. The results revealed that regression table (4.3), which includes article variables, exhibits the highest adjusted R-squared. Journal variables rank second, while author variables demonstrate the lowest adjusted R-squared.

Therefore, I have demonstrated in this study that article-level variables exert the greatest influence on determining the future citation performance of academic research.

Chapter 5: Conclusion

The primary objective of this research has been to explore the intricate dynamics of information attribution and dissemination within the realm of science, particularly investigating the influence of textbook citation events on subsequent citation counts of academic literature. By doing so, the aim was to deepen our understanding of the complex mechanisms underlying the provenance of information within academia and to uncover the practical implications that emerge from these insights. Building upon existing literature, including studies by Aizenman and Kletzer (2011), McMahan and McFarland (2021), and Yuret (2023), this research provides clarity on the nature of textbook citation events and their influence on citation counts. Furthermore, it enhances the field by analysing a larger sample of textbooks and considering several additional factors that have been shown to impact citation patterns, thereby strengthening the empirical foundation for future research.

The central questions for this research are as follows:

1. Does citing academic literature in a textbook affect the subsequent attribution thereof?
2. If subsequent attribution is affected, is it strengthened or weakened, and what mechanisms are possibly responsible for this change?

To address these questions, I constructed a comprehensive citation dataset consisting of 610 observations and 18 variables, which were divided into two distinct groups: a treatment group (comprising cited articles) and a control group (comprising articles not cited). The treatment group of articles was sourced from nine prominent microeconomics textbooks, while the control group of articles was obtained from the Scopus citation database by matching articles based on similar characteristics to those in the treatment group, including journal, issue, and year of publication. I developed scripts in Python and R to retrieve the citation data and control variables from various databases, including Scopus, Gender.io, and the CentER for Research in Economics at Tilburg University, and cleaned the data using STATA. Utilising the curated citations dataset, I employed a multiple linear regression model that incorporates an interaction term (*"Cited In Book#Log(Citations 10 Yrs Before)"*), variables of interest "*Cited In Book"* and "*Log(Citations 10 Yrs Before)"*, and a combination of various control variables and fixed effects to observe the difference in post-event-citation counts between the control and treatment groups of articles.

The regression results revealed significant findings. The interaction term has a negative coefficient, indicating a difference in slopes between the two groups, with the control group having a larger slope. Furthermore, the variables of interest display significant positive coefficients, indicating a difference in the y-intercept between the two groups, with the treatment group having a larger intercept. Hypothesis testing yielded significant t-statistics and F-statistics, resulting in the rejection of both the null hypotheses. These findings demonstrate that cited articles have notably higher levels of post-event citations compared to uncited articles, particularly at lower levels of pre-event citations. However, as the number of pre-event citations increases, the disparity between the two groups diminishes. Eventually, a point of slope interception is reached, where post-event citations of cited articles are lower than those of uncited articles for levels of pre-event citations beyond this point. Nonetheless, in this study, even as the gap narrows, pre-event citations do not ever exceed this point of intersection. Therefore, cited articles consistently exhibit higher post-event citations throughout the observed range of pre-event citations.

Moreover, the regression results uncover significant findings related to the control variables. The analysis reveals a highly significant positive association between journal ranking, as measured by the *"SJR Score"* variable, and the citation count of published literature. Similarly, the *"Number of Pages"* variable demonstrates a moderately significant positive relationship with citation count, indicating that longer papers tend to receive more citations. The study also identifies a moderately significant positive relationship between author citedness, measured by the *"H-Index'* variable, and article citation count. On the other hand, the variable "*Title Length*" indicates a highly significant negative relationship, with shorter titles being associated with higher citation counts.

The findings of this dissertation have led to several key insights and discussions. First, it provides compelling evidence that textbook citation events significantly impact the subsequent citation counts of academic literature. This suggests that citing academic literature in textbooks has a noteworthy effect on subsequent attribution and demonstrates that these events are crucial in shaping the attribution dynamics within the academic community. Moreover, textbook citation events predominantly contribute to increased citation counts for cited literature, indicating a strengthening of attribution and implying that literature incorporated into textbooks have amplified acknowledgement and appreciation. The identified mechanism responsible for this strengthened attribution is the signalling effect. It is likely that citing academic literature in textbooks signals the importance and quality of the cited works to

readers, leading to heightened recognition and subsequent citation. However, the precise underlying reason for this mechanism remains uncertain and future studies could delve deeper into the exact factors contributing to this effect.

Second, although this study did not include articles with exceptionally high levels of pre-event citations in its dataset, statistical evidence supports the notion that such articles experience a decrease in citation count after being included in well-known textbooks. This phenomenon indicates weakened attribution and the presence of a cessation effect among highly influential and groundbreaking research articles. Consequently, highly cited literature included in textbooks may receive reduced acknowledgement and appreciation, potentially reaching a point where the information becomes common knowledge. While the magnitude of this effect may not be substantial in the context of this study, it is noteworthy to recognise its potential existence. Future studies could focus specifically on highly cited research articles to investigate this phenomenon further. Additionally, since the exact cause of the observed cessation effect remains uncertain, it would be valuable for future studies to delve into contributing factors.

Third, the findings regarding the control variables enhance our understanding of citation patterns in academia and underscore the influence of certain journal-, article-, and author-related variables on citation counts. However, it is important to note that the inclusion of control variables posed certain limitations, including the exclusion of a substantial portion of observations owing to missing data, the omission of research by authors with non-US names, and the exclusion of non-university-affiliated research. Future studies should address these limitations to enhance the robustness and generalisability of their findings.

Finally, the findings have important practical implications for the provenance of information. First, the study confirms that citing research in textbooks strengthens the attribution of information, ensuring that its provenance is preserved. This highlights the robustness and enduring nature of the social convention of academic referencing. Second, the insights gained from understanding provenance in academia can be applied to other forms of information, particularly data. By ensuring that metadata includes information about the origin, ownership, and the sources and methods employed in the generation of the data, it is possible to establish data provenance and mitigate the risk of the loss of data provenance.

Reference list

Adams, J.D., Black, G.C., Clemmons, J.R. & Stephan, P.E. (2005). Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981–1999. *Research Policy*, 34(3):259–285. doi:https://doi.org/10.1016/j.respol.2005.01.014.

Aizenman, J. & Kletzer, K. (2011). The life cycle of scholars and papers in economics – the 'citation death tax'. *Applied Economics*, 43(27):4135–4148. doi:https://doi.org/10.1080/00036846.2010.485930.

Allison, P. (2012). *When can you safely ignore multicollinearity?* [online] Statistical Horizons. Available at: https://statisticalhorizons.com/multicollinearity

Anderson, G.M., Levy, D.M. & Tollison, R.D. (1989). The half-life of dead economists. *The Canadian Journal of Economics*, 22(1):174. doi:https://doi.org/10.2307/135467

Arao, L.H., da Costa Santos, M.J.V. & Guedes, V.L.S. (2017). The half-life and obsolescence of the literature science area: a contribution to the understanding the chronology of citations in academic activity. *Qualitative and Quantitative Methods in Libraries*, 4(3):603–610.

Aumann, R.J., Katznelson, Y., Radner, R., Rosenthal, R.W. & Weiss, B. (1983). Approximate purification of mixed strategies. *Mathematics of Operations Research*, 8(3):327–341.

Bjork, S., Offer, A. & Söderberg, G. (2013). Time series citation data: the Nobel Prize in economics. *Scientometrics*, 98(1):185–196. doi:https://doi.org/10.1007/s11192-013-0989-5.

Bornmann, L. & Daniel, H.-D. (2007). What do we know about the *h* index? *Journal of the American Society for Information Science and Technology*, 58(9):1381–1385. doi:https://doi.org/10.1002/asi.20609

Bosquet, C. & Combes, P.-P. (2013). Are academics who publish more also more cited? Individual determinants of publication and citation records. *Scientometrics*, 97(3):831–857. doi:https://doi.org/10.1007/s11192-013-0996-6

Breit, W. & Huston, J.H. (1997). Reputation versus influence: The evidence from textbook references. *Eastern Economic Journal*, 23(4):451–456.

Callaham, M., Wears, R.L. & Weber, E. (2002). Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *JAMA*, 287(21):2847. doi:https://doi.org/10.1001/jama.287.21.2847.

Card, D. & DellaVigna, S. (2013). Nine Facts about Top Journals in Economics. *Journal of Economic Literature*, 51(1):144–161. doi:https://doi.org/10.1257/jel.51.1.144.

De Solla Price, D.J. (1986). *Little science, big science– and beyond*. New York: Columbia University Press.

Elsevier developer portal. (n.d.). [online] Available at: https://dev.elsevier.com/

Dion, M.L., Sumner, J.L. & Mitchell, S.M. (2018). Gendered citation patterns across political science and social science methodology fields. *Political Analysis*, 26(3):312–327. doi:https://doi.org/10.1017/pan.2018.12

Tilburg University economics ranking - Tilburg University. (n.d.). [online] Available at: https://econtop.uvt.nl/

Falagas, M.E., Zarkali, A., Karageorgopoulos, D.E., Bardakas, V. & Mavros, M.N. (2013). The impact of article length on the number of future citations: A bibliometric analysis of general medicine journals. *PLoS ONE*, 8(2):e49476 doi:https://doi.org/10.1371/journal.pone.0049476

Fortunato, S., Bergstrom, C.T., Börner, K., Evans, J.A., Helbing, D., Milojević, S., Petersen, A.M., Radicchi, F. et al. (2018). Science of science. *Science*, 359(6379). doi:https://doi.org/10.1126/science.aao0185

Fox, C.W., Paine, C.E.T. & Sauterey, B. (2016). Citations increase with manuscript length, author number, and references cited in ecology journals. *Ecology and Evolution*, 6(21):7717–7726. doi:https://doi.org/10.1002/ece3.2505

Galiani, S. & Gálvez, R.H. (2017). The life cycle of scholarly articles across fields of research. *SSRN Electronic Journal*. doi:https://doi.org/10.3386/w23447

Glänzel, W. & Thijs, B. (2004). Does co-authorship inflate the share of self-citations? *Scientometrics*, 61(3):395–404. doi:https://doi.org/10.1023/b:scie.0000045117.13348.b1

Guerrero-Bote, V.P. & Moya-Anegón, F. (2012). A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *Journal of Informetrics*, 6(4):674–688. doi:https://doi.org/10.1016/j.joi.2012.07.001

Gupta, A. (2009). Data provenance. *Encyclopedia of database systems*, [online] 608–608. doi:https://doi.org/10.1007/978-0-387-39940-9_1305

Habibzadeh, F. & Yadollahie, M. (2010). Are shorter article titles more attractive for citations? Cross-sectional study of 22 scientific journals. *Croatian Medical Journal*, 51(2):165–170. doi:https://doi.org/10.3325/cmj.2010.51.165.

Hanck, C. (2015). Significant predictors become non-significant in multiple logistic regression. Available at: https://stats.stackexchange.com/questions/27257/significant-predictors-become-non-significant-in-multiple-logistic-regression

Harrell, F. (2022). Regression modeling strategies. [online] Available at: https://hbiostat.org/doc/rms.pdf

Haslam, N., Ban, L., Kaufmann, L., Loughnan, S., Peters, K., Whelan, J. & Wilson, S. (2008). What makes an article influential? Predicting impact in social and personality psychology. *Scientometrics*, 76(1):169–185. doi:https://doi.org/10.1007/s11192-007-1892-8

Hengel, E. & Moon, E. (2020). *Gender and quality at top economics journals*. University of Liverpool Repository.

Hunter, J. (2006). *The importance of citation*. [online] Available at: https://web.grinnell.edu/Dean/Tutorial/EUS/IC.pdf [Accessed 1 Aug. 2023].

Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics*, 20(3)341–367. doi:https://doi.org/10.1093/applin/20.3.341

Jiang, J., He, D. & Ni, C. (2013). The correlations between article citation and references' impact measures: What can we learn? *Proceedings of the American Society for Information Science and Technology*, 50(1):1–4. doi:https://doi.org/10.1002/meet.14505001162

Korom, P. (2018). Does scientific eminence endure? Making sense of the most cited economists, psychologists and sociologists in textbooks (1970–2010). *Scientometrics*, 116(2):909–939. doi:https://doi.org/10.1007/s11192-018-2781-z

Kreps, D.M. (1990). *A course in microeconomic theory*. Princeton, N.J.: Princeton University Press.

Kuhn, T.S. (1970). *The structure of scientific revolutions*. Chicago: The University of Chicago Press.

Leahey, E., Lee, J. & Funk, R.J. (2023). What types of novelty are most disruptive? *American Sociological Review*, 88(3):562–597. doi:https://doi.org/10.1177/00031224231168074

Leimu, R. & Koricheva, J. (2005). What determines the citation frequency of ecological papers? *Trends in Ecology & Evolution*, 20(1):28–32. doi:https://doi.org/10.1016/j.tree.2004.10.010

Lou, W. & He, J. (2015). Does author affiliation reputation affect uncitedness? *Proceedings of the Association for Information Science and Technology*, 52(1):1–4. doi:https://doi.org/10.1002/pra2.2015.1450520100103.

Lozano, S. & Salmerón, J.L. (2005). Data envelopment analysis of OR/MS journals. *Scientometrics*, 64(2):133–150. doi:https://doi.org/10.1007/s11192-005-0245-8

Machlup, F. (1962). *The production and distribution of knowledge in the United States*. Princeton, N.J.: Princeton University Press.

Maliniak, D., Powers, R. & Walter, B.F. (2013). The gender citation gap in international relations. *International Organization*, 67(4), pp.889–922. doi:https://doi.org/10.1017/s0020818313000209.

Massachusetts Institute of Technology. (n.d.). *The Massachusetts Institute of Technology (MIT)*. [online] Available at: https://web.mit.edu/

McMahan, P. & McFarland, D.A. (2021). Creative destruction: The structural consequences of scientific curation. *American Sociological Review*, 86(2):341–376. doi:https://doi.org/10.1177/0003122421996323.

Merton, R.K. (1968). The Matthew effect in science. *Science*, 159(3810):56–63.

Merton, R.K. (1988). The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property. *Isis*, 79(4):606–623.

Muschelli, J. (2019). *Package 'rscopus'*. [online] Available at: https://cran.r-project.org/web/packages/rscopus/rscopus.pdf

Newman, M.E.J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409. doi:https://doi.org/10.1073/pnas.98.2.404

Park, M., Leahey, E. & Funk, R.J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, 613(7942):138–144. doi:https://doi.org/10.1038/s41586-022-05543-x.

Reece-Evans, L. (2010). Gender and citation in two LIS e-journals: A bibliometric analysis of LIBRES and information research. *Library and Information Science Research E-Journal*, 20(1). doi:https://doi.org/10.32655/libres.2010.1.2

Roberts, M., Stewart, B. & Nielsen, R. (2016). Matching methods for high-dimensional data with applications to text. Unpublished manuscript.

Roshani, S., Bagherylooieh, M.-R., Mosleh, M. & Coccia, M. (2021). What is the relationship between research funding and citation-based performance? A comparative analysis between critical disciplines. *Scientometrics*, 126(9):7859–7874. doi:https://doi.org/10.1007/s11192-021-04077-9.

Rossiter, M.W. (1993). The Matthew Matilda effect in science. *Social Studies of Science*, 23(2):325–341.

Rotfeld, H.J. (2000). Book review: The textbook effect: Conventional wisdom, myth, and error in marketing. *Journal of Marketing*, 64(2):122–126. doi:https://doi.org/10.1509/jmkg.64.2.122.18003.

Rothman, R.A. (1971). Textbooks and the certification of knowledge. *The American Sociologist*, 6(2):125–127.

Storer, N.W. (1966). Science as a social system. *The social system of science.* New York: Holt, Rinehart and Winston.

Tahamtan, I., Safipour Afshar, A. & Ahamdzadeh, K. (2016). Factors affecting number of citations: a comprehensive review of the literature. *Scientometrics*, 107(3), pp.1195–1225. doi:https://doi.org/10.1007/s11192-016-1889-2

Van der Pol, C.B., McInnes, M.D.F., Petrcich, W., Tunis, A.S. & Hanna, R. (2015). Is quality and completeness of reporting of systematic reviews and meta-analyses published in high impact radiology journals associated with citation rates? *PLoS One*, 10(3). doi:https://doi.org/10.1371/journal.pone.0119892

Wai, S. (2017). *Fixed effects in Stata*. [online] www.youtube.com. Available at: https://www.youtube.com/watch?v=H95BHswbT3w [Accessed 13 Oct. 2022].

Wooldridge, J.M. (2012). *Introductory econometrics: A modern approach*. Mason, OH: South-Western, Cengage Learning.

Yuret, T. (2023). The citation performance of the references in the standard graduate-level microeconomics textbook: Mas-Collel et al. (1995). *Scientometrics*, 128:1473–1484. doi:https://doi.org/10.1007/s11192-023-04650-4.

Zhu, J.M., Pelullo, A.P., Hassan, S., Siderowf, L., Merchant, R.M. & Werner, R.M. (2019). Gender differences in Twitter use and influence among health policy and health services researchers. *JAMA Internal Medicine*, 179(12):1726–1729. doi:https://doi.org/10.1001/jamainternmed.2019.4027.

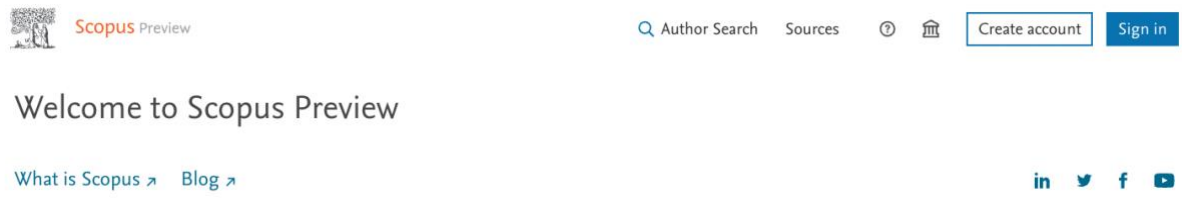
Appendix A

This section details Scopus account creation and API key creation and explains the four Scopus APIs I used in this study.

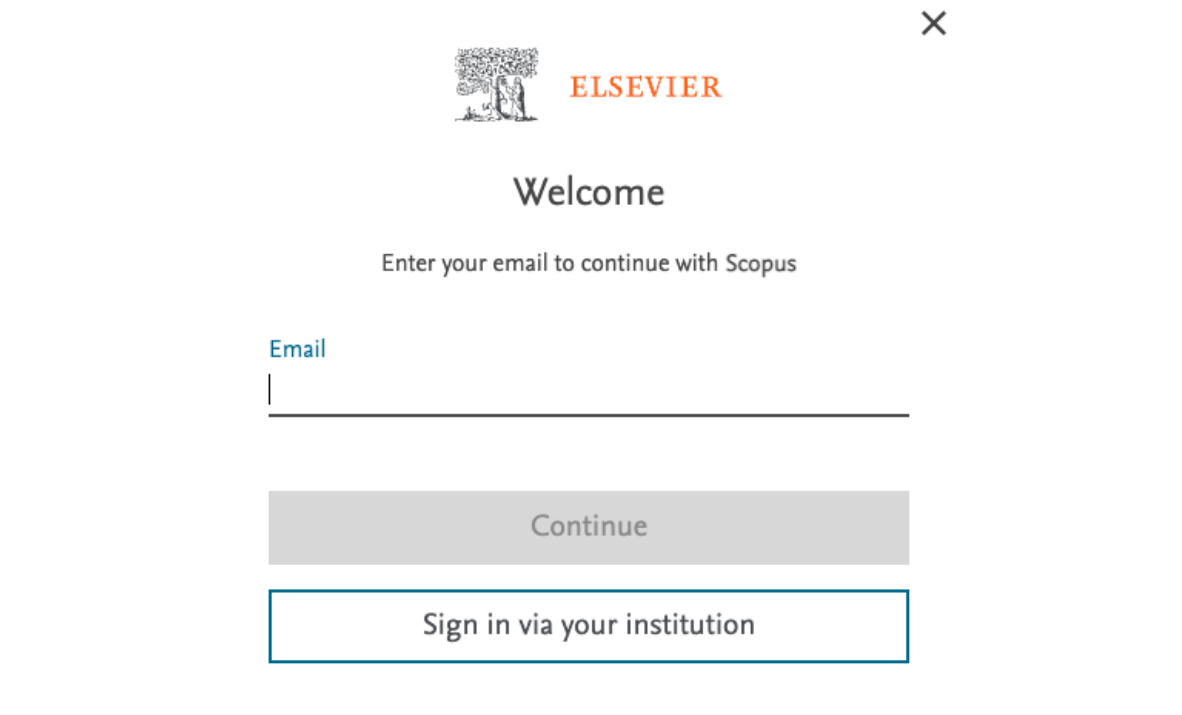
A1 Scopus access guide

A Scopus user must create a Scopus account and log in via an institution to access Scopus API keys and article metadata. This section explains the process of creating a Scopus account via an institution.

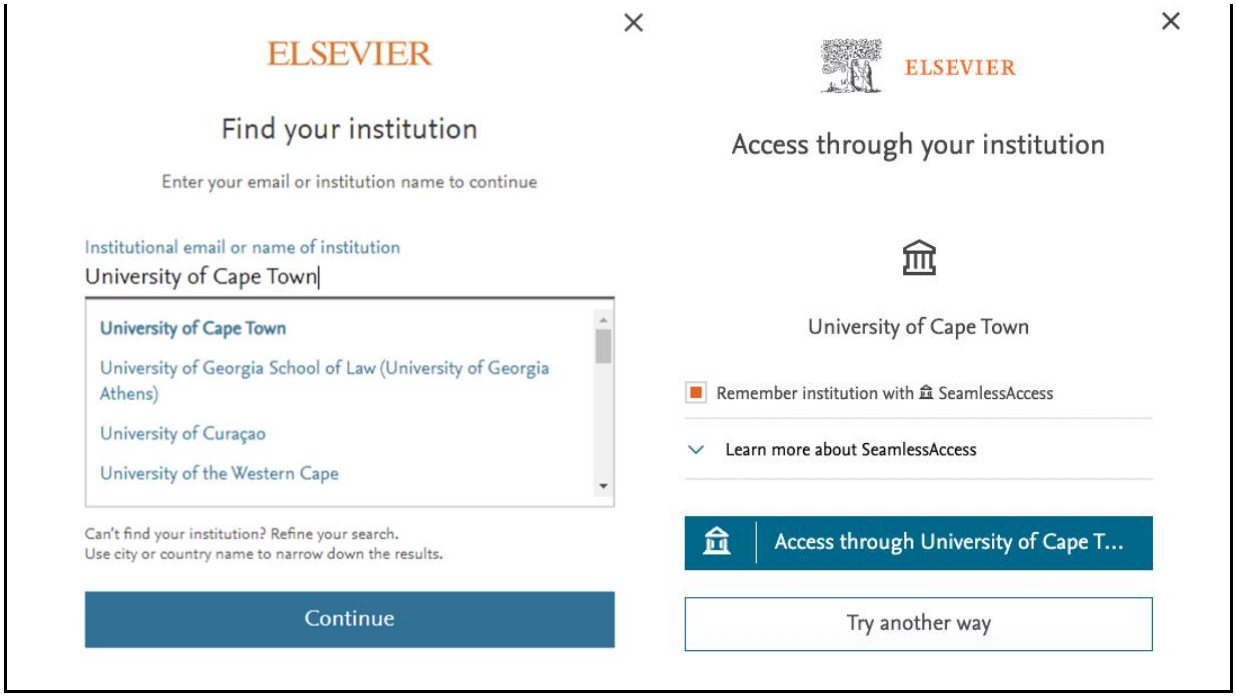
Step 1: Go to the Scopus website and select "Create account".



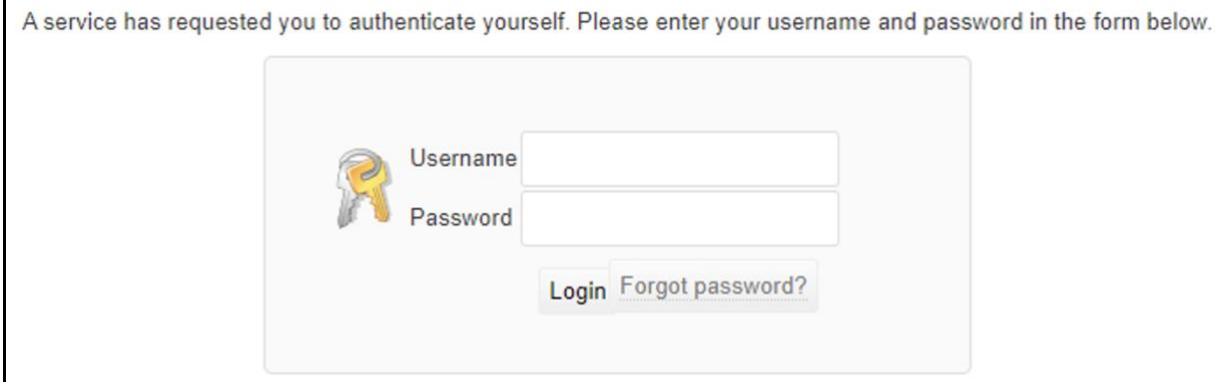
Step 2: Select "Sign in via your institution".



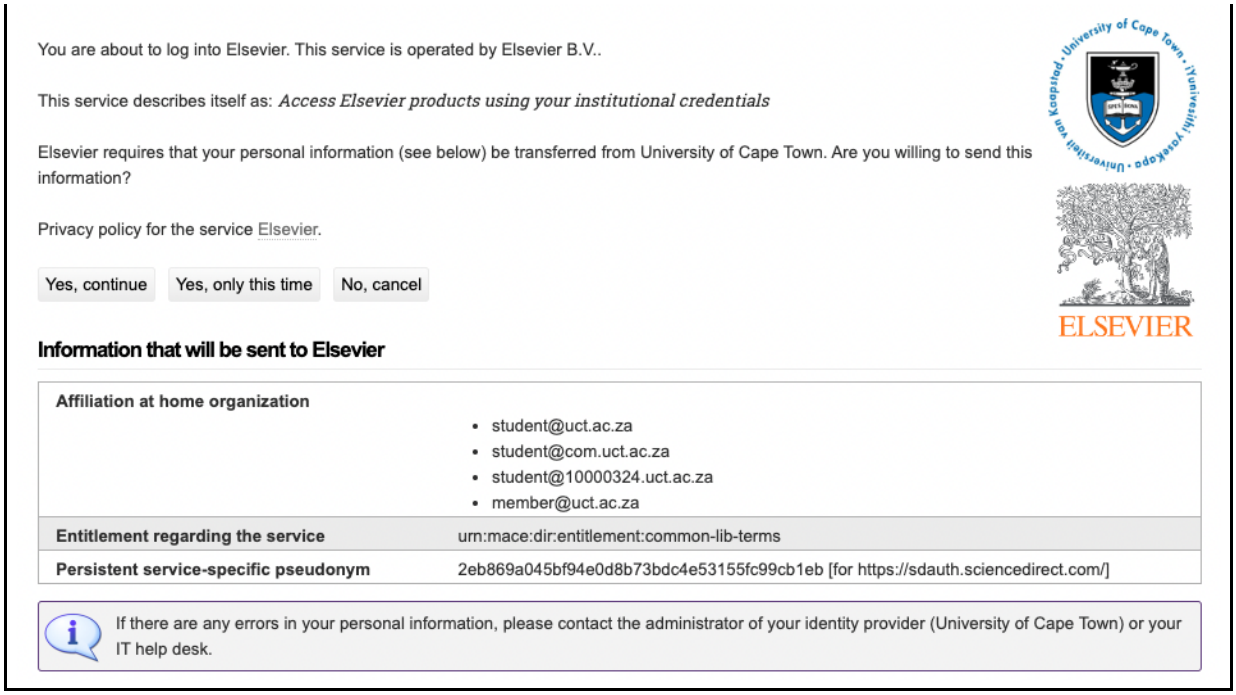
Step 3: Search for your institution, continue and select "Access through [relevant institution]".



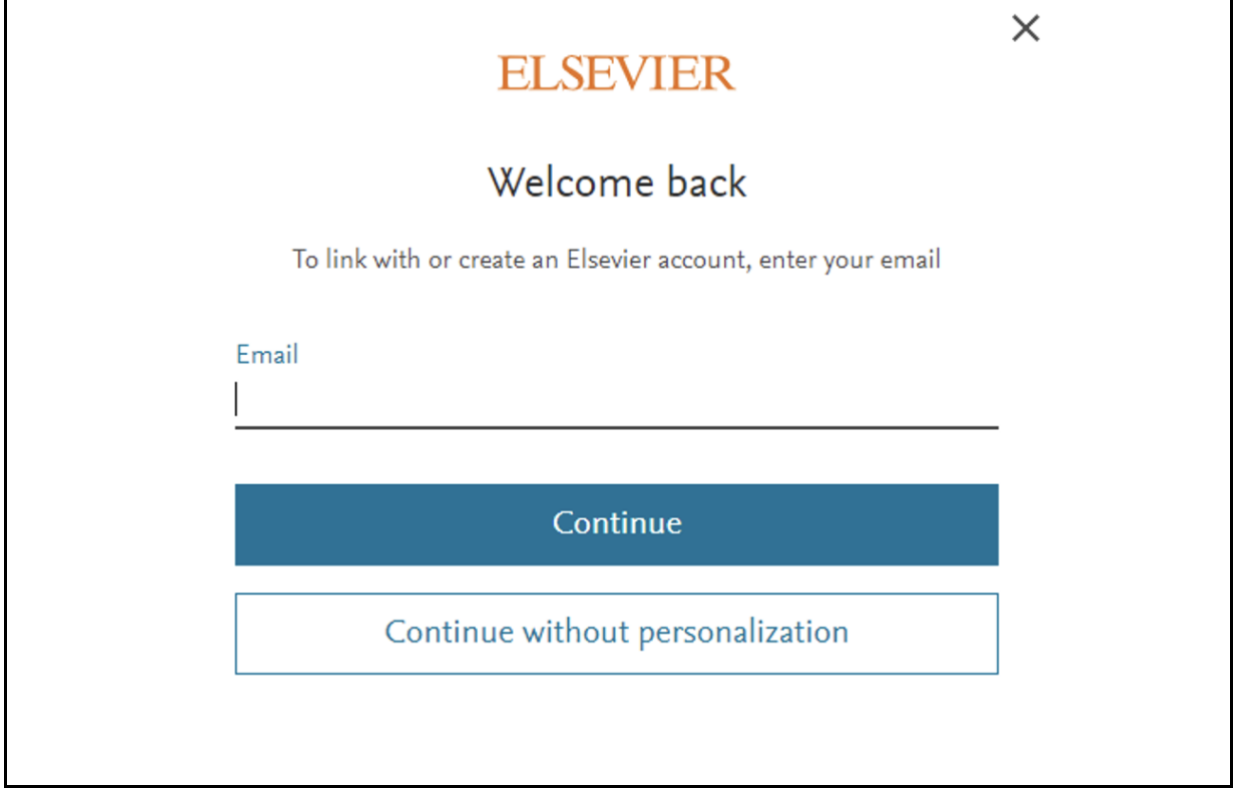
Step 4: Enter your institution login credentials and select "login".



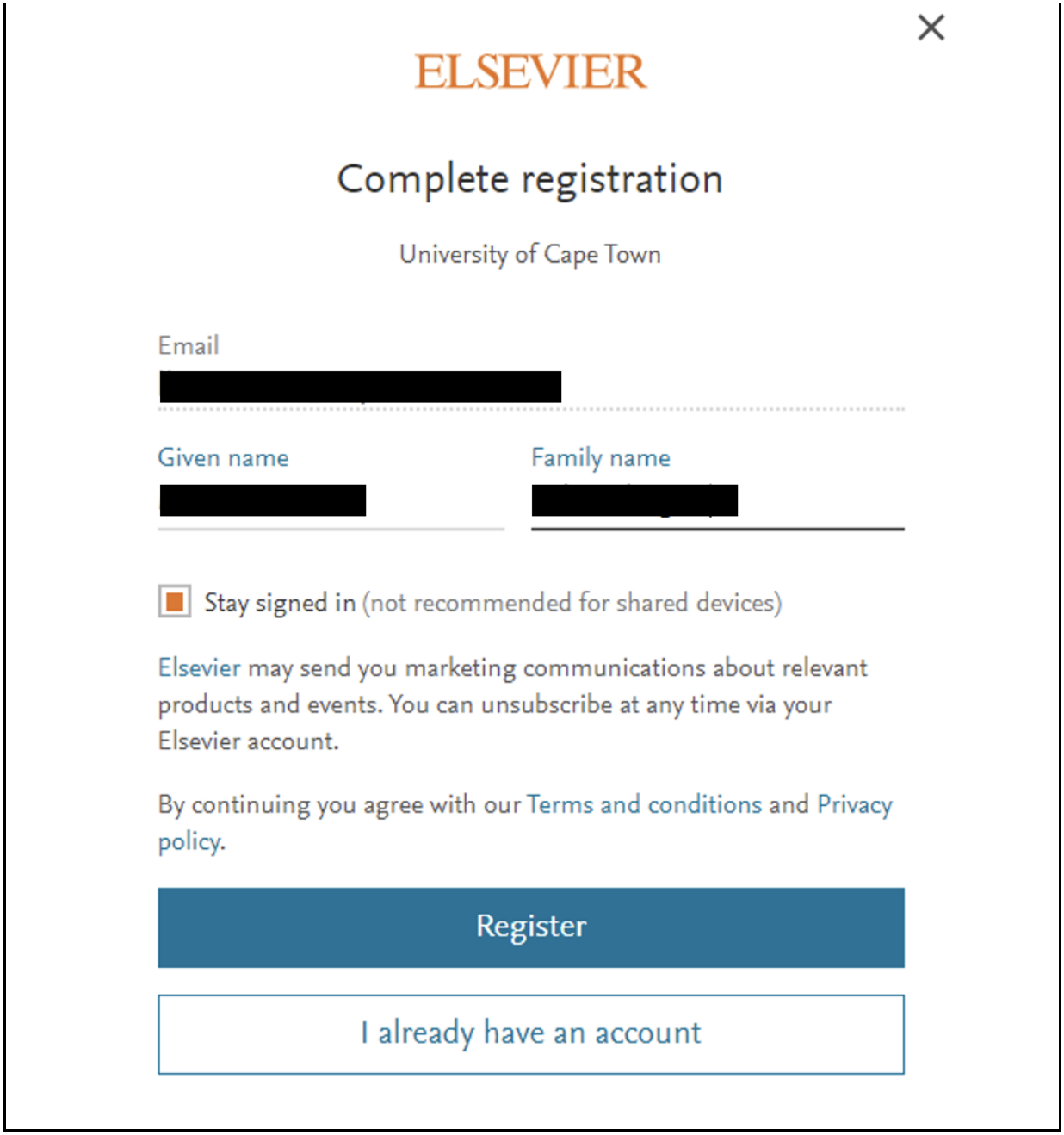
Step 5: Select "Yes, only this time".



Step 6: Enter your email address and select “Continue”.



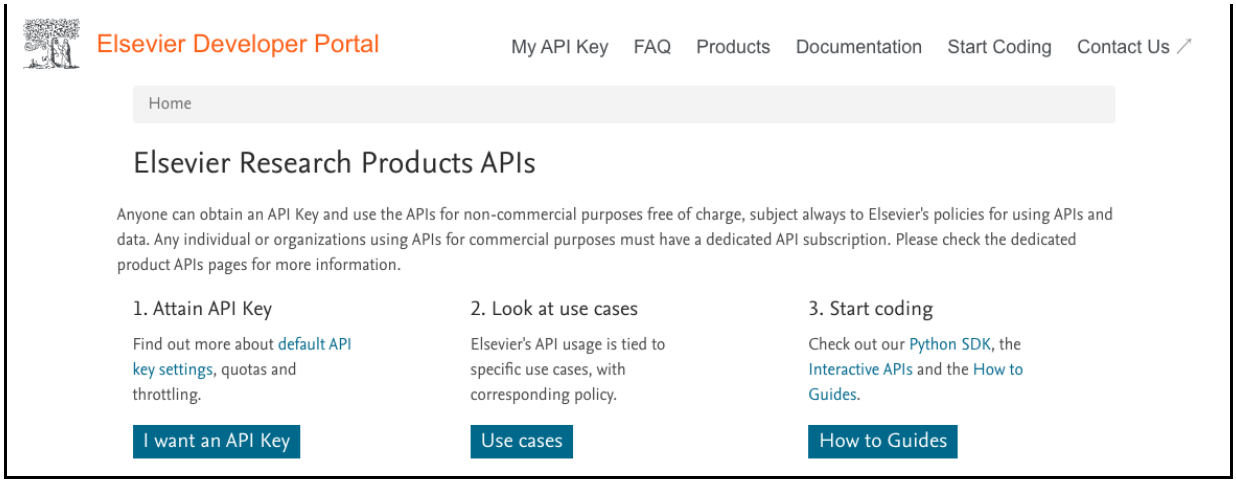
Step 7: Enter your Name and Surname and select “Register”.



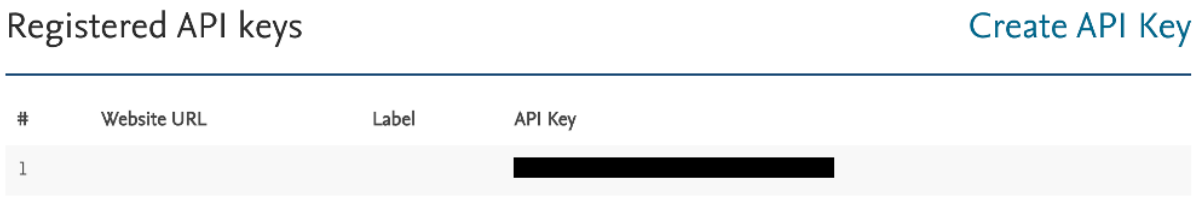
A2 Scopus API key

Scopus provides its users with an API key that allows bulk download functionality of a wide variety of article metadata hosted in their database. To access the bulk download functionality, a user requires an API key. This section explains how to create a Scopus API key.

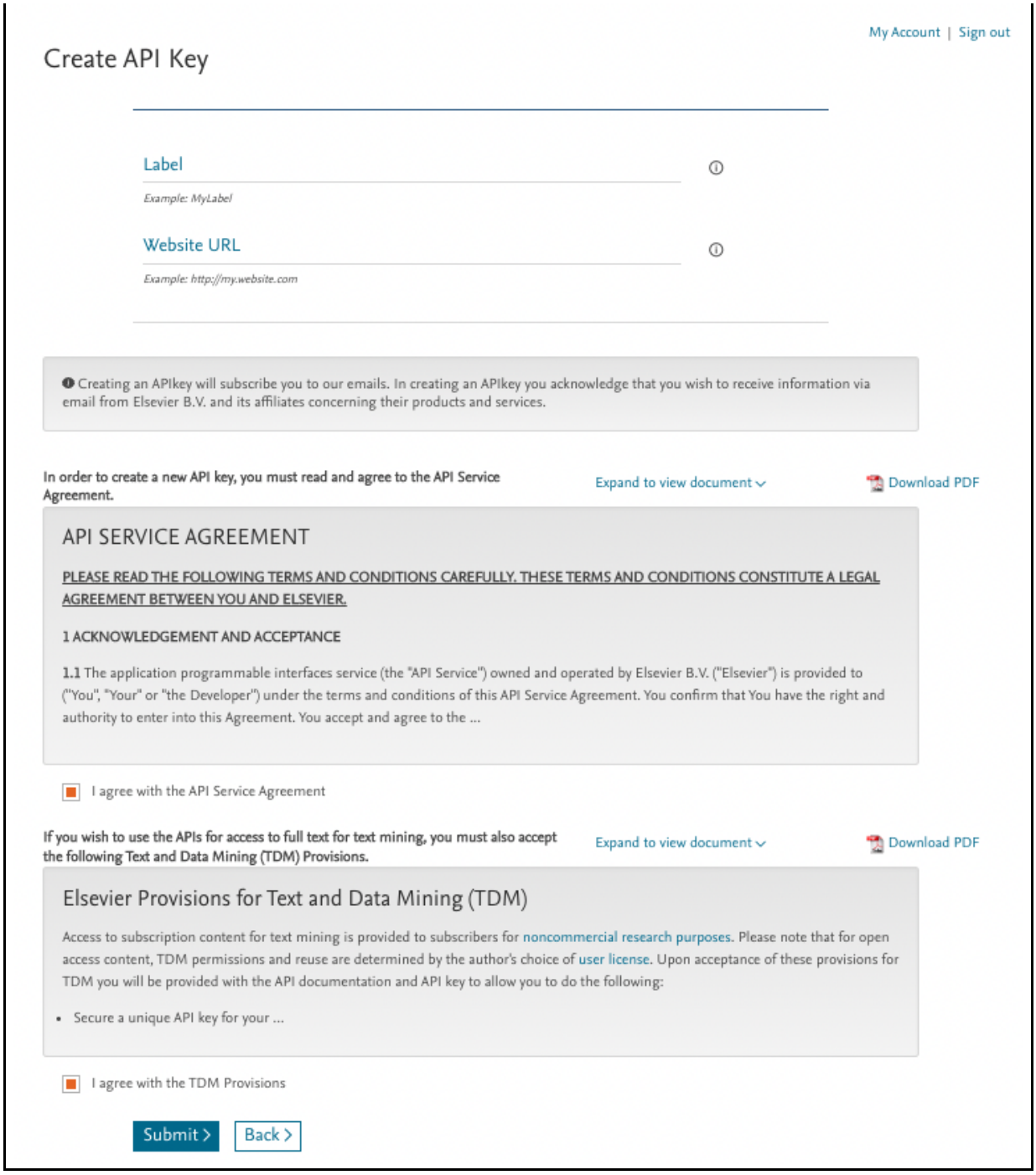
Step 1: Navigate to the Elsevier Developer Portal and select "I want an API Key".



Step 2: Select "Create API Key"



Step 3: Accept the terms and conditions and select “Submit”.



Step 4: View and access the API key under the API key list, as shown in Step 2.

A3 Scopus APIs

Scopus provides a variety of APIs tailored to specific data needs. Each API can bulk download metadata based on a given set of parameters. A user must have a Scopus API key to use the provided APIs. Each API key has a specified number of weekly requests and limited data access. To increase the number of permitted requests or to gain access to specific information,

it is possible to contact Scopus Support. This section briefly discusses each API that I used to retrieve data in this study.

Scopus Search API

The Scopus Search API can search the Scopus database based on a given set of parameters, for example, article title, author name, publication date, etc. The Scopus Search Views represent the relevant metadata that can be retrieved using the Scopus Search API. The API mimics the advanced search functionality found on the Scopus website.

Author Retrieval API

The Author Retrieval API can retrieve information about authors in the Scopus database given their author ID, Electronic Identifier or ORCID. The Scopus Author Retrieval Views chart lists the metadata that is available to access.

Citations Overview API

The Citations Overview API can retrieve citation metadata such as yearly citation counts and summaries and takes the Scopus ID as input. The Citation Overview Views chart on Scopus provides details on specific data that can be retrieved using this API.

Serial Title API

The Scopus Serial Title API can retrieve publication metadata by using the publication ISSN as input. The Serial Title Views chart on Scopus shows the metadata that can be retrieved using this API.

Appendix B

B1 Variable description

Table B.1: Description of variables used in the dissertation and their inclusion status in the final dataset

	Description	Included in the final dataset
Textbook Title	The microeconomics textbook in which an article was cited.	No
Book Year	The microeconomics textbook publication year.	Yes
Article Scopus ID	The unique Scopus identifier of the research article.	Yes
Article Title	The name of the research article.	No
Publication Name	The name of the periodical in which the research article appears.	No
Publication ISSN	The unique identifier of the periodical.	No
Publication Issue Number	The specific edition or release of the periodical.	No
Publication Page Range	The starting page and ending page number of the research article in the periodical.	No
Publication Cover Date	The specific month and year that the periodical was released.	No
Source Type	The specific type of periodical or publication, i.e., Book, Book Series, Journal, etc.	No
Document Type	The specific type of document, i.e., Articles, Editorials, Reviews, etc.	No
Number of Authors	The total count of authors of a research article.	Yes
Author Full Name	The name and surname of the author of a research article.	No
Author First Name	The first name of the author of a research article.	No
Author Last Name	The surname of the author of a research article.	No
Author Scopus ID	The unique Scopus identifier of the author of a research article.	No
Author Affiliation ID	The unique Scopus identifier of an author's affiliated institution.	No
Author Affiliation Name	The name of the author's affiliated institution.	No

	Description	Included in the final dataset
Co-author Count	The total number of prior co-authors of the author of a research article.	No
Document Count	The total number of documents written by the author of a research article.	No
Author Citation Count	The total number of citations received by the author of a research article.	No
H-Index	A bibliometric measure that quantifies the impact of an author's work based on both the number of their publications and the number of citations those publications have received.	Yes
Yearly Citation Count (T-10 to T+12)	The total number of citations received by a research article on an annual basis from ten years before the textbook citation event (T-10) to twelve years after the textbook citation event (T+12).	No
Journal Citation Count	The total number of times that all articles published in a specific journal have been cited in other scholarly publications.	Yes
SNIP Score	A bibliometric indicator that measures the average impact of citations received by a journal's publications, considering the subject field's citation potential and publication frequency.	No
SJR Score	A bibliometric indicator that assesses the prestige and influence of scholarly journals based on the citations received by their articles, taking into account the citation weight of the citing sources.	Yes
Cited In Book	An indicator distinguishing whether an article belongs to the treatment group (cited) or the control group (not cited).	Yes
Author Gender	The gender (male or female) of the author of a research article.	No
First Author Gender	The gender (male or female) of the primary or first-named author of a research article.	No
Affiliation Ranking	The Tilburg University Ranking of the research article's authors' affiliated university.	Yes
Article Publication Year	The year in which the article was published in a journal.	Yes
Number of Pages	The page count of an article.	Yes
Title Length	The character counts in the article's title.	Yes
Citations 10 Yrs Before	The total number of article citations in the ten years leading up to the textbook citation event.	Yes
Citations 12 Yrs After	The total number of article citations since the textbook citation event.	Yes

	Description	Included in the final dataset
Male First Author	An indicator distinguishing whether the primary or first named author of an article is male or female.	Yes
Top 5 Journal	Fixed effect indicating whether an article was published in one of the top five most highly cited journals within the dataset.	Yes
Journal Fixed Effects	A unique ID assigned to each value of 'Journal Citation Count' to accommodate both observed and unobserved variation in the dependent variable.	Yes
Decade Fixed Effects	A unique identifier generated based on the decade in which the article was published.	Yes
Year Fixed Effects	A unique identifier generated based on the year in which the article was published.	Yes

Note: The variables with an inclusion status "Yes" are included in the final dataset and used in the regression analysis. The variables with an inclusion status "No" are not included in the regression analysis but serve various purposes, such as data collection inputs, calculation of new variables, data categorisation, or plotting graphs and tables for summary statistics and visualisation of the data. Some variables are not used and are dropped from the dataset due to their correlation with the included variables.

B2 Correlation matrix

Table B.2: Correlation matrix of variables (excluding fixed effects)

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*	*(11)*	*(12)*	*(13)*	*(14)*	*(15)*	*(16)*	*(17)*	*(18*
(1) Number of Authors	1.00																	
(2) Co-Author Count	0.29	1.00																
(3) Document Count	0.13	0.65	1.00															
(4) Author Citation Count	0.09	0.49	0.63	1.00														
(5) H-Index	0.13	0.61	0.85	0.82	1.00													
(6) Journal Citation Count	0.46	0.44	0.21	0.12	0.17	1.00												
(7) SJR Score	0.02	-0.11	-0.03	0.17	0.08	-0.01	1.00											
(8) SNIP Score	0.07	-0.08	-0.01	0.16	0.09	0.09	0.91	1.00										
(9) Affiliation Ranking	0.10	-0.01	-0.03	-0.16	-0.14	0.01	-0.18	-0.12	1.00									
(10) Number of Pages	-0.10	-0.15	-0.05	0.03	-0.00	-0.27	0.22	0.15	-0.15	1.00								
(11) Title Length	0.14	0.11	0.05	-0.05	0.02	0.11	-0.18	-0.17	0.13	-0.05	1.00							
(12) Citations 10 Yrs Before	0.10	0.04	0.13	0.33	0.25	0.06	0.21	0.24	-0.09	0.19	-0.12	1.00						
(13) Citations 12 Yrs After	0.08	0.03	0.14	0.28	0.21	0.05	0.17	0.19	-0.06	0.12	-0.09	0.72	1.00					
(14) Cited In Book	-0.17	-0.11	-0.03	0.01	-0.03	-0.14	-0.14	-0.18	-0.07	0.12	-0.07	0.18	0.08	1.00				

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*	*(11)*	*(12)*	*(13)*	*(14)*	*(15)*	*(16)*	*(17)*	*(18*
(15) Book Year	0.18	0.09	0.03	-0.03	0.03	0.08	0.13	0.20	0.11	-0.00	0.11	0.06	0.09	-0.40	1.00			
(16) Article Publish Date	0.30	0.18	0.08	0.07	0.14	0.17	0.12	0.22	0.06	0.07	0.11	0.10	-0.09	-0.38	0.58	1.00		
(17) Male First Author	-0.08	-0.02	0.07	0.09	0.09	-0.10	0.03	-0.01	-0.07	0.07	-0.15	0.04	0.02	0.10	-0.14	-0.13	1.00	
(18) Top 5 Journal	0.36	0.35	0.17	0.07	0.12	0.81	-0.09	0.01	0.16	-0.26	0.11	-0.03	-0.03	-0.13	0.06	0.15	-0.09	1.00

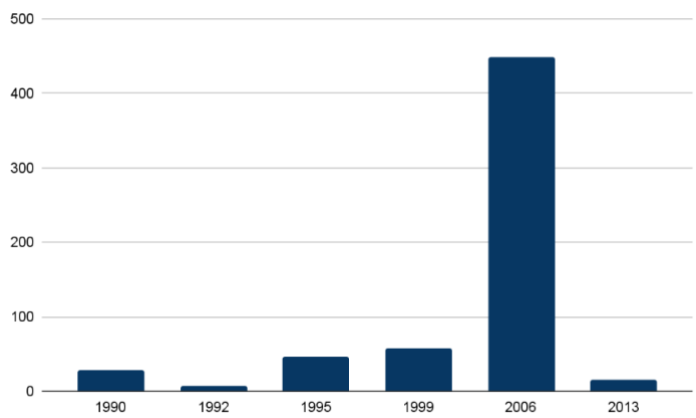
Note: Highly correlated author-level variables include *"Author Citation Count"*, *"Document Count*" *"H-Index"*, and *"Co-Author Count"*. Highly correlated journal-level variables include *"SNIP Score"* and *"SJR Score"*.

B3 Summary statistics

Table B.3: The mean, standard deviation, minimum, and maximum of the numeric variables

	Mean	Standard Deviation	Min	Max
Journal Citation Count	11920.04	44869.31	64	271357
SJR Score	10.08	10.40	0.25	34.57
log(Citations 12 Yrs After)	4.34	1.19	1.79	7.31
log(Citations 10 Yrs Before)	3.48	0.87	1.79	6.74
log(Citations 10 Yrs Before - Avg)	0.00	0.87	-1.69	3.25
Number Of Authors	1.75	1.00	1	8
Number of Pages	19.60	11.38	1	115
Title Length	8.71	3.76	1	24
H-Index	27.12	15.81	1	103.50
Affiliation Ranking	101.71	151.63	1	904

Table B.4: Frequency table of the categorical variable *"Book Year"* showing the number of articles per year of textbook publication

Book Year	Frequency	Percentage	Histogram
1987	5	0.82	
1990	29	4.75	
1992	7	1.15	
1995	47	7.70	
1999	58	9.51	
2006	448	73.44	
2013	16	2.62	

Total	610	100.00	**Figure B.1: Histogram of articles per book publication year**

Table B.5: Frequency table of the categorical variable *"Cited In Book"* showing the number of articles cited and not cited in textbooks

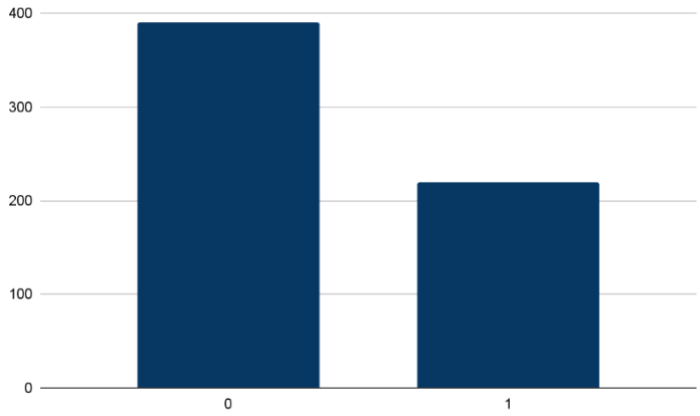
Cited In Book	Frequency	Percentage	Histogram
0	390	63.93	
1	220	36.07	
Total	610	100.00	**Figure B.2: Histogram of articles cited and not cited**

Table B.6: Frequency table of the categorical variable *"Male First Author"* showing the number of articles with male first authors and female first authors

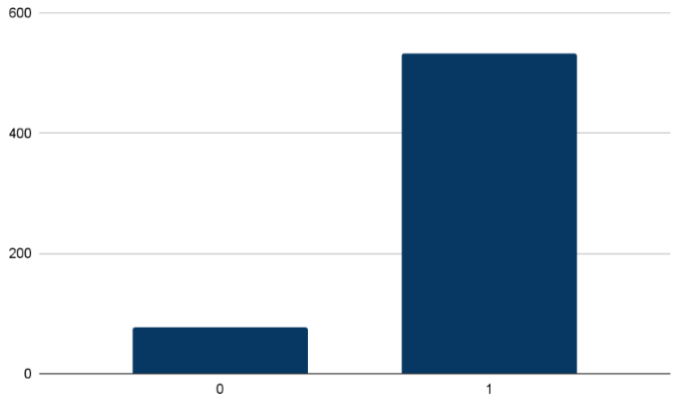
Male First Author	Frequency	Percentage	Histogram
0	78	12.79	
1	532	87.21	
Total	610	100.00	**Figure B.3: Histogram of articles with male first author and female first author**

Table B.7: Frequency table of the fixed effect variable *"Top 5 Journal"* showing the number of articles published in a top 5 journal and published in lower-ranked journals

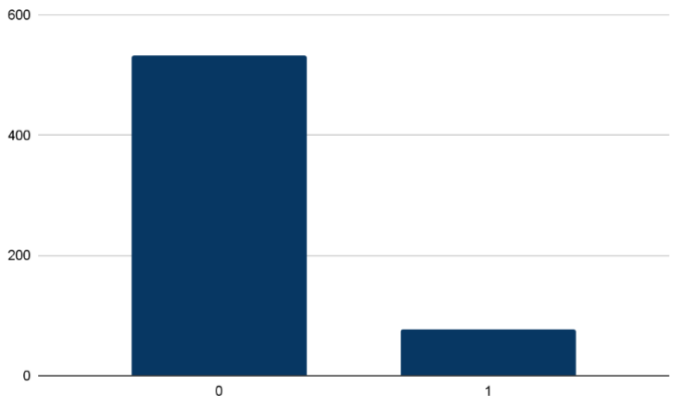
Top 5 Journal	Frequency	Percentage	Histogram
0	532	87.21	
1	78	12.79	
Total	610	100.00	

Figure B.4: Histogram of articles published in a top 5 journal and published in lower-ranked journals

Table B.8: Frequency table of the categorical variable *"Article Publication Year"* showing the number of articles per year of article publication

Article Publication Year	Frequency	Percentage	Article Publication Year	Frequency	Percentage	Article Publication Year	Frequency	Percentage
1935	1	0.16	*1979*	14	2.30	*1994*	24	3.93
1950	1	0.16	*1980*	5	0.82	*1995*	25	4.10
1956	2	0.33	*1981*	8	1.31	*1996*	70	7.87
1967	1	0.16	*1982*	11	1.80	*1997*	48	7.87
1968	1	0.16	*1983*	6	0.98	*1998*	34	5.57
1969	1	0.16	*1984*	3	0.49	*1999*	20	3.28
1970	3	0.49	*1985*	6	0.98	*2000*	28	4.59
1971	5	0.82	*1986*	21	3.44	*2001*	23	3.77
1972	5	0.82	*1987*	4	0.66	*2002*	34	5.57
1973	10	1.64	*1988*	23	3.77	*2003*	7	1.15

Article Publication Year	Frequency	Percentage	Article Publication Year	Frequency	Percentage	Article Publication Year	Frequency	Percentage
1974	17	2.79	*1989*	19	3.11	*2004*	1	0.16
1975	7	1.15	*1990*	18	2.95	*2005*	4	0.66
1976	12	1.97	*1991*	18	2.95	*2006*	3	0.49
1977	9	1.48	*1992*	17	2.79			
1978	9	1.48	*1993*	32	5.25			
						Total	610	100.00

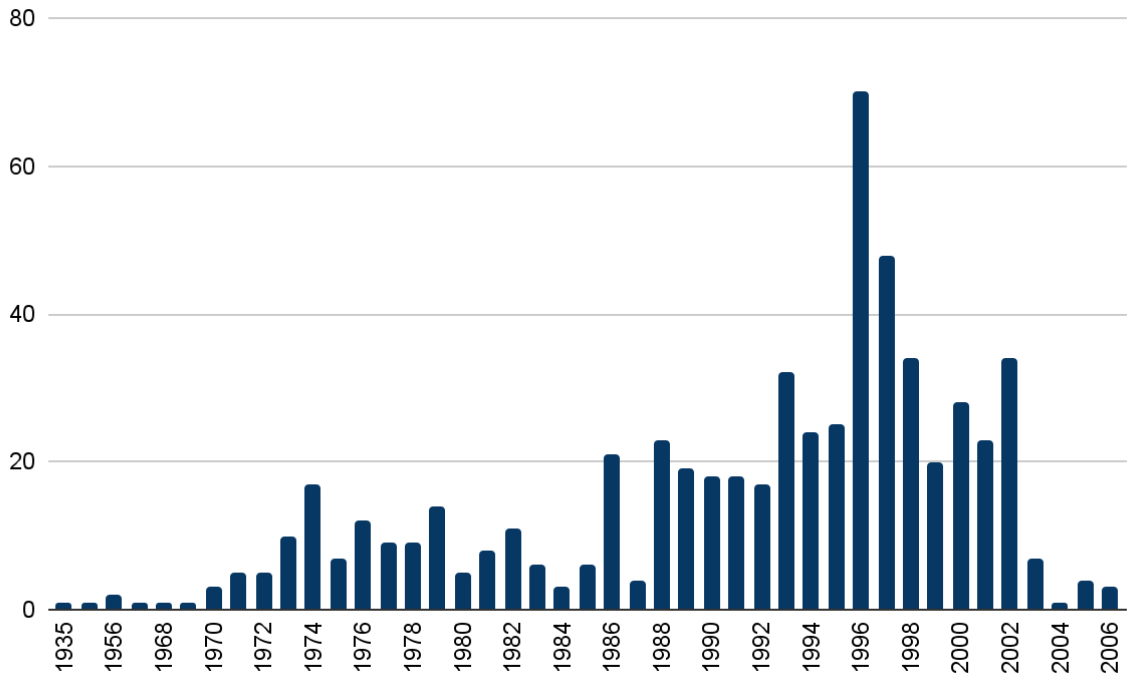


Figure B.5: Histogram of articles per article publication year

Appendix C

C1 Regression average VIFs

Table C.1: The average Variable Inflation Factors (VIF) for all the regressions

Regression	VIF	Regression	VIF
(1.1)	14.00	*(3.3)*	16.45
(1.2)	8.11	*(3.3)*	13.04
(1.3)	16.78	*(4.1)*	13.38
(1.4)	13.73	*(4.2)*	7.91
(2.1)	13.75	*(4.3)*	16.44
(2.2)	150.59	*(4.4)*	13.14
(2.3)	152.34	*(5.1)*	12.53
(2.4)	13.48	*(5.2)*	142.44
(3.1)	13.27	*(5.3)*	145.11
(3.2)	8.00	*(5.4)*	12.33

C2 F-statistic calculation

Table C.2: F-statistic calculation for regressions one to four of the base model

Table 1 - Base model				
	(1.1)	*(1.2)*	*(1.3)*	*(1.4)*
Residual df	564	527	521	563
Residual SS Full Model	278.4427	307.2068	262.3042	272.5417
Residual SS Reduced Model	304.8768	323.7449	284.8175	296.7415
N = (RSSr - RSSf)/2	13.2171	8.2690	11.2567	12.0999
D = (RSSf)/Residual df	0.4937	0.5829	0.5035	0.4841

Table 1 - Base model				
	(1.1)	*(1.2)*	*(1.3)*	*(1.4)*
F = N/D	26.7718	14.1852	22.3584	24.9952

Table C.3: F-statistic calculation for regressions one to four of the base model plus a journal control variable

Table 2 - Base + Journal control variable				
	(1.1)	*(1.2)*	*(1.3)*	*(1.4)*
Residual df	563	526	520	562
Residual SS Full Model	272.1455	306.3763	262.0978	267.9343
Residual SS Reduced Model	301.0952	322.6426	284.3242	294.3823
N = (RSSr - RSSf)/2	14.4749	8.1332	11.1132	13.2240
D = (RSSf)/Residual df	0.4834	0.5825	0.5040	0.4768
F-Statistic = N/D	29.9448	13.9633	22.0485	27.7377

Table C.4: F-statistic calculation for regressions one to four of the base model plus article control variables

Table 3 - Base + Article control variables				
	(1.1)	*(1.2)*	*(1.3)*	*(1.4)*
Residual df	561	524	518	560
Residual SS Full Model	269.2358	295.2048	259.8121	266.9486
Residual SS Reduced Model	291.7767	312.3799	282.1616	288.9178
N = (RSSr - RSSf)/2	11.2705	8.5876	11.1748	10.9846
D = (RSSf)/Residual df	0.4799	0.5634	0.5016	0.4767
F-Statistic = N/D	23.4840	15.2432	22.2796	23.0433

Table C.5: F-statistic calculation for regressions one to four of the base model plus author control variables

Table 4 - Base + Author control variables				
	(1.1)	*(1.2)*	*(1.3)*	*(1.4)*
Residual df	561	524	518	560
Residual SS Full Model	275.4625	299.8600	257.9671	270.0229
Residual SS Reduced Model	300.6795	314.7916	278.8916	293.2968
N = (RSSr - RSSf)/2	12.6085	7.4658	10.4623	11.6370
D = (RSSf)/Residual df	0.4910	0.5723	0.4980	0.4822
F-Statistic = N/D	25.6782	13.0464	21.0083	24.1338

Table C.6: F-statistic calculation for regressions one to four of the most restricted model (base model plus all control variables)

Table 5 - Restricted model: Base + all control variables				
	(1.1)	*(1.2)*	*(1.3)*	*(1.4)*
Residual df	557	520	514	556
Residual SS Full Model	263.9201	288.2987	255.6473	262.0044
Residual SS Reduced Model	287.5836	303.7194	276.3065	284.9606
N = (RSSr - RSSf)/2	11.8318	7.7104	10.3296	11.4781
D = (RSSf)/Residual df	0.4738	0.5544	0.4974	0.4712
F-Statistic = N/D	24.9708	13.9070	20.7685	24.3577

Appendix D: Editing Certificate

Editing Certificate

Ricky Woods Academic Editing Services
Cell: +27 (0)83 3126310
Email: rickywoods604@gmail.com

To Whom It May Concern

University of Cape Town

Editing of a Master's Dissertation

I, Marietjie Alfreda Woods, hereby certify that I have completed the editing and correction of the dissertation: **Textbook citations increase subsequent citations in economics, except for the most cited articles** by **Danae Bouwer.** I believe that the dissertation meets with the grammatical and linguistic requirements for a document of this nature.

Name of Editor: Marietjie Alfreda Woods

Qualifications: BA (Hons) (Wits); Copy-editing and Proofreading (UCT); Editing Principles and Practice (UP); Accredited Text Editor (English) (PEG)

MA (Ricky) Woods

7 August 2023

Signed by candidate