IONA

Volume VIII

IONA

JOURNAL OF ECONOMICS

VOLUME VIII
2023

Vancouver School of Economics

University of British Columbia
Vancouver, BC

Cover Design: Natalia Ramirez
Logo Design: Sean Durfee

All published content featured in the *IONA Journal of Economics* is available online at ionajournal.ca.

Please refer to the excluded appendices and references lists in our digital version.

Printed in Canada.

First Printing, 2023.

Contents

III. REFERENCES & APPENDICES

A. FEATURE PAPER

B. PAPERS

i. Authors

ii. IONA Journal Team

Faculty Foreword

Dear Reader,

It is my pleasure to provide a foreword for the eighth volume of the IONA Journal of Economics, the premier undergraduate research forum for the Vancouver School of Economics – and in my opinion, one of the finest examples of undergraduate-run and managed research journals in any discipline in the world.

I have been involved in supervising undergraduate research, and served as a reviewer for this journal, since my start here at UBC. I firmly believe (and research shows!) that practical, hands-on experience with research is one of the most impactful learning experiences possible. Retrospective reviews from alumni and faculty of their own undergraduate education highlight their first research projects as being formative for their careers and academic interests.

I hope that by reading this volume, you will see why. Undergraduate research has a unique position in academia, as being the closest to what is interesting, current, and relevant to young people today. This is an aspect of scholarship which is often overlooked, in our preoccupation with citation counts and top journal publications. I often tell my students that they have a unique position in that they can identify aspects of the economic world that the older generations overlook or do not recognize – and undergraduate research gives them a chance to explore these ideas in a profound and deep way.

To pick only a few examples from this volume: Mea Srisan's work using the homophily of K-Pop idol's body shapes to capture the returns to lookism in Korea is something we faculty might not have struck upon. Ting Guo's deep and analytical insight into the product review ecosystem on Amazon and its relationship to video gaming is another example of a current and interesting topic. These papers deal with the environment of the daily life of undergraduates, and pose a unique perspective on what is economically interesting – often very different from faculty members, with tenure and childrearing on their mind.

I will not belabour the point, at the editors' request. Instead, I will address my comments to those of you reading who might be students yourselves. Undergraduate research is one of the best things you can do for your academic and professional careers – and it does not require a fancy title or RA-ship to get started. It also does not require any special levels of academic brilliance or mathematical prodigiousness. What it does require is determination, hard-work, a willingness to handle set-backs, and a creative mind that thinks deeply and analytically about economic problems and ideas. None of this is beyond the scope of an undergraduate student here at UBC, especially as you move into your final year of study.

The economist James Viner once said that "economics is what economists do," and while I think that is true, there is a deeper meaning to this statement. These young scholars have produced work which highlights the breadth of economics, as a discipline spanning a domain rich in content, but also the methodology of what economists do. For any economist, from the most mainstream to the most heterodox, it is our methodology that defines us – and where our deepest disagreements arise. This volume makes it clear that the emerging generation of economists will continue in this proud tradition; focusing on understanding the economic world around us through rigorous methodologies and a pragmatic focus on models.

I have advised and supervised hundreds of young researchers, and one of the privileges of working here at UBC is the intellectual strength of our undergraduate cohorts. The best of our undergraduate students, as showcased here, is on par with any cohort from any program in the world – and their potential is unlimited.

So, dear reader, if you are not a student, use this as an opportunity to get a sneak peek at what the next generation of economic thinkers and leaders look like – where their interests lie, where their values point them, and what shape their thinking takes on. While we, the faculty, may do our best to try to shape that thought, it will ultimately grow and flourish in ways that we cannot predict today... but maybe we can get a hint from the outstanding scholarship evidenced here, in the IONA Journal's eight volume.

If you are a student, use this as an inspiration; a template to frame what you might or could do. Realize that, although this work is outstanding, the secret is simply hard work, creativity, and a little perspective – and we, your faculty mentors, are here to help point you in the right direction, when you need that little extra push – so don't hesitate to reach out and get started. Now! Go! You can do it!

Sincerely,

Dr. Jonathan L. Graves
Assistant Professor of Teaching and Majors Advisor
Vancouver School of Economics,
The University of British Columbia

Student Foreword

Dear Reader,

On behalf of the Vancouver School of Economics Undergraduate Society (VSEUS), I congratulate the IONA Journal team this year on another successful launch. The IONA journal's capacity to showcase novel and innovative research ideas, met with the team's dedication to excellence, ensures that important economic problems are considered and the voices of our community are heard. Valentina and Colby, I hope you find fulfillment in knowing that you have done an excellent job managing this year's team. Together, VSEUS hopes that our shared mandate for enhancing student academic experiences at the Vancouver School of Economics continues to grow.

Readers, I challenge you to absorb, reflect on, and perhaps challenge some of the empirical methods presented in this edition of the IONA journal. I hope you learn something new from the papers by our student community and join me in celebrating another successful year of the IONA journal!

Sincerely,

Erin Song
President
Vancouver School of Economics Undergraduate Society

Letter from Editor-in-Chief

Dear Reader,

The papers you are about to read span a diverse range of topics, from self-rated health to body shape and career outcomes, from consumer reviews online to value stocks and earnings surprises. Despite their uniqueness, these papers share a common trait: they showcase the high standard UBC undergraduate students can meet with enough effort and determination. Mea Srisan's paper, in particular, serves as a clear example of this in exploring how deviation from ideal body image in South Korea affects one's labour market prospects. Investigating an interesting question in a novel context, all while writing with concision and clarity, Mea's work is excellent; consequently, we have decided to feature it at the start of this year's volume.

As you peruse this year's journal and each of the four papers within, I urge you to reflect on the steps which led to their inclusion in Volume VIII. These papers were chosen after months of deliberation among our Editorial Board, a team composed of 22 undergraduate economics students. Additionally, they underwent rigorous scrutiny from our Faculty Review Committee composed of professors whom we have listed at the end of this volume. In total, we received 45 submissions and are delighted to present four excellent papers in this year's journal.

If you have already completed your undergraduate career, no matter how long ago, we hope that you reflect on the intense effort, ingenuity, and passion involved in writing these papers. If you are presently an undergraduate student, we hope that these papers inspire you to be ambitious in your own undergraduate thesis. UBC's economics degree is unique in that it requires all students to write a senior thesis. While this may seem daunting or even frustrating, I cannot understate the weight of this opportunity. Whether you are considering graduate school, work opportunities, or just want to earn a high grade, I want to be clear that writing a strong paper is an achievable, worthwhile goal.

This year, the IONA journal introduced comprehensive econometrics training staggered by knowledge level with the aim of exposing editors to statistics on a deeper level and causal inference more broadly. I couldn't help but feel immense joy during this process. Seeing the genuine desire to learn outside the classroom shine through the eyes of my fellow team members during these sessions invigorated me. To any undergraduates reading, I want to stress that the IONA journal does not merely showcase exceptional undergraduate research; joining the team also provides you with an opportunity to advance your knowledge and gain early exposure to academic research. I encourage you in particular to read this volume of the IONA journal, itself a legacy of eight years of hard work, with deep interest; consider contributing your own skills to future editions of the journal, whether through submission of a paper or joining the team as an editor.

Lastly, I want to thank all of the professors at the Vancouver School of Economics. We are blessed to have highly intelligent, inspiring and caring faculty at this school. They have been monumental in instilling in me the value of this journal's presence and conducting research as a career more generally. I hope that their excellence is reflected in the excellence of the IONA Journal perennially.

With gratitude,

Colby Chambers
Editor-in-Chief
IONA Journal of Economics Volume VIII

Letter from the Director of Business Operations

Dear Reader,

Within the pages of this journal, you will find more than just research papers; you will find the collective effort and dedication of editors, authors, and faculty members who have embraced the spirit of intellectual curiosity and academic excellence. It is with great pleasure that I welcome you to explore the inspiring contributions presented in this volume.

The success of this journal is a testament to the hard work and commitment of our diligent editors who have meticulously curated and refined the submissions, ensuring the highest standards are upheld.

To the authors, the driving force behind the intellectual depth of this journal, I extend my warmest congratulations. Your research, analysis, and innovative insights have added vibrancy and diversity to the field of economics. Your pursuit of knowledge and your willingness to delve into uncharted territories exemplify the true spirit of scholarly inquiry.

A special acknowledgment goes to our esteemed faculty members, whose guidance and expertise have played a pivotal role in shaping the future of economics through the mentorship of our students. Your role in fostering a community of scholars who are driven by intellectual curiosity is paramount.

In the words of the economist Edmund Burke, "The first and simplest emotion which we discover in the human mind, is curiosity." It is this very curiosity that fuels our quest for knowledge, drives us to question established theories, and inspires us to explore new horizons. Our commitment to nurturing this curiosity is not only essential for the growth of individual scholars but also for the advancement of the field as a whole.

As you delve into the research papers showcased in this volume, I encourage you to embrace the spirit of curiosity and wonder that embodies our Economics community. Let these works spark your own curiosity, challenging you to ponder new ideas, question prevailing notions, and seek solutions to the pressing economic challenges of our time.

Once again, I extend my heartfelt congratulations to all contributors for their exceptional work, dedication, and unwavering commitment to the pursuit of knowledge. As we celebrate the achievements showcased in this volume, I urge everyone, regardless of their economic level or expertise, to allow their eyes to get lost in the pages of the journal, to immerse themselves in the captivating world of economics.

Within these pages lie the seeds of inspiration, waiting to ignite the spark of curiosity within each of us. Let us all embrace the joy of learning, engage in meaningful discussions, and together, shape the future of our ever-evolving field.

With heartfelt thanks,

Valentina Ramirez
Director of Business Operations
IONA Journal of Economics Volume VIII

Contributors

EDITOR-IN-CHIEF
Colby Chambers

DIRECTOR OF BUSINESS OPERATIONS
Valentina Ramirez

SENIOR EDITORS
Bi Yu Li
Chanya Chawla
Chris Haun
Jasmine Arora
Manan Aggarwal
Megumi Bernardo
Pius Lau

JUNIOR EDITORS
Andie Bartolome
Anna Chernesky *
Casandra Lim
Himanaya Bajaj
Ishaan Sahai
Janesh Tulsan
Joshua Evangelista
Julia Hedda Lima Pastana
Leona Yu
Luigi Vicencio
Noah Monash
Pranav Vinod
Priyanka Srinivas
Shreya Iyer
Teona Mariuta

AUTHORS
Mea Srisan
Bi Yu Li
Derek Situ
Ting (Dongting) Guo

* OUTSTANDING EDITOR

FACULTY REVIEW COMMITTEE

Dr. Angela Redish	*Vancouver School of Economics*
Dr. Jonathan Graves	*Vancouver School of Economics*
Dr.Scott Orr	*UBC Sauder School of Business*
Dr. Kevin Milligan	*Vancouver School of Economics*

SPECIAL THANKS

Dr. Jonathan Graves	*Vancouver School of Economics*

SPONSORS

Vancouver School of Economics
Vancouver School of Economics Undergraduate Society

I

FEATURE PAPER

Body Shape and Career Outcomes in Entry-level Jobs: The Study of South Korean Young Adults

Mea Srisan

ECON 490: Seminar in Applied Economics

ABSTRACT

This paper examines the relationship between body shape and job career outcomes in entry-level jobs by focusing on South Korean young adults. Using panel survey data from the Korean Education and Employment Panel from 2009 to 2011, this paper finds that undesirable body shapes have negative associations with career outcomes in entry-level jobs for both men and women. For women, having a less desirable body shape results in lower probability of employment, lower wages, and lower probability of getting paid bonuses while men with undesirable body shapes experience lower chance of employment and lower chance of having permanent job status. Alternative analyses are performed to illustrate the robustness of the results. The results provide supporting evidence to existing literature documenting appearance-based bias in career settings.

* Vancouver School of Economics, 6000 Iona Dr, Vancouver, BC V6T 1L4. I would like to thank my thesis supervisor, Professor Jonathan Graves, for his continued support and guidance throughout the process. His encouragement and motivation allowed me to push myself and learn to apply concepts to solve real-world problems through this project.

INTRODUCTION

Most of us accept the fact that there is a return to being physically attractive. The presence of appearance-based discrimination has been well documented in many areas such as citizens' voting for political candidates (Clifford & Walster, 1973), teachers' judgments of students (Efrain & Patterson, 1974), and the selection of actors in the US film industry (Smith et al., 1999). Not surprisingly, physical appearance also plays a role in the career success of an individual. Although people often believe that factors such as educational background, cognitive abilities, and technical skills are the primary determinants of success in work-related contexts, research consistently demonstrates that physical appearance also significantly influences workplace dynamics. Hamermesh and Biddle (1994) examined the US and Canadian household surveys and found that workers who are considered to be above-average in physical attractiveness earn about 10% to 15% more than workers who are below-average in beauty when working the same jobs. According to Mobius and Rosenblat (2006), the size of this so-called beauty premium is almost as wide as the race and gender wage gap in the US labor market. Recently, Pfeifer (2012) observed German General Social Survey data and showed that more attractive candidates are on average more likely to be hired and earn on average higher wages than less attractive candidates.

Body shape is a salient attribute of physical appearance, and many studies have documented harmful weight-based stereotypes and discrimination. Levine and Schweitzer (2015) found that obesity is often associated with negative traits, such as a lack of self-esteem and physical unattractiveness. These negative stereotypes give way to stigma and discrimination against overweight individuals in multiple domains of living, especially in work settings (Brownell, 2001). Flint et al. (2016) reported that overweight candidates are deemed less qualified during the recruiting process when compared to normal-weight candidates. According to research by Carr and Friedman (2005), individuals with a body mass index (BMI) of above 35 were 84% more likely to have experienced job-related discrimination when compared to individuals with average weight. Similarly, Ball et al. (2002) and Schulte et al.

(2007) found evidence that overweight workers receive lower starting wages, work longer hours, and are more likely to have jobs associated with lower socioeconomic status than normal-weight individuals. Moreover, many studies have shown that the influence of body shapes on career outcomes differs across genders. Mason (2012) investigated data from the 1997–2008 National Longitudinal Survey of American Youth and found that obese women (BMI > 30) earned 25% less than non-obese women while such weight penalty was much smaller among men. More recently, Li et al. (2021) and Li et al. (2022) examined the Chinese labor market and found that, apart from receiving a higher degree of the slim premium than their male counterparts, women with lower body weight are more likely to have long-term employment contracts and better career opportunities than women with higher body weight.

Despite the lack of agreement in the research, there may be evidence suggesting that the role of physical appearance on career success varies across different experience levels. One study supporting this claim was conducted by Fuller and Raman (2017). According to the study, 37% of employers ranked experience as the most important qualification in a potential candidate. Since work experience provides valuable hands-on skills that cannot be compensated for by physical attractiveness, the presence of weight bias in entry-level jobs that require minimal professional experience should be more significant when compared to jobs that require candidates with more sophisticated experience levels. As most of the studies on the topic are based on career industries that require more experience or training such as lawyers (Biddle & Hamermesh, 1998) or management executives (Frieze et al., 1991), the significance of body images in entry-level jobs has not been sufficiently clarified.

Therefore, my research builds on the existing discussion on body shape and its effect on career-related outcomes by focusing on entry-level jobs. Specifically, my research identifies and quantifies the relationship between body shape and job market outcomes in the early adulthood of South Koreans. Since early adulthood is a period in which individuals are generally new to the workforce and have only accumulated enough professional experience to work in entry-level jobs, they may have not acquired enough skills to offset the stigma stemming from weight discrimination. In addition, I select South Korean young adults as the population of interest because South Korea is known for its strict beauty standards and physical appearance has a large influence on the South Korean job market (Park et al., 2019). As pointed out in a Los Angeles Times article by Stiles (2017), many companies in South Korea routinely ask applicants to submit their photos, leading many job applicants to undergo plastic surgeries to enhance their physical appearance.

To explore the relationship between ideal body shape and job market outcomes of Korean young adults, I obtained data from the Korean Education and Employment Panel (KEEP), a nationally representative dataset comprising 2,000 middle school seniors (9th graders) and 4,000 high school seniors (12th graders). The panel study surveyed the same cohort of middle school seniors and high school seniors every year for 10 years starting from 2004. For each survey respondent, I have data on their education records, employment, household details, demographic information, and personal characteristics. I extract data on the high school student cohort for three years from the 2009, 2010, and 2011 surveys. To measure the respondents' body shape, I utilize weight and height data of K-pop idols released by their companies and compute a dissimilarity index to measure the dissimilarity between each respondent and a group of individuals who are likely to have body types that are considered ideal in South Korean culture. In choosing the dependent variables of this study, I consider not only the pecuniary aspects of job outcomes but also the non-pecuniary aspects of career outcomes. In the KEEP data, job market outcomes are measured by 1) employment status, 2) permanent job status, 3) monthly wage, and 4) bonus provision. Ordinary least squares regression is used to estimate the effect of body shape on monthly wages while logistic regression with marginal effects is used for employment status, permanent job status, and bonus provision.

My results provide important insight on beauty privilege; they show that undesirable body shape negatively affects career outcomes in entry-level jobs for both men and women. For women, having a less desirable body shape results in lower probability of employment, lower wages, and lower probability of getting paid bonuses while men with less desirable body shape experience lower chance of employment and lower chance of having permanent job status. I also perform robustness checks by estimating alternative specifications using a dissimilarity index computed from BMI rather than weights and heights. My primary results are largely consistent with these checks. In Section V, I provide a rationale for why inconsistent results from the robustness analysis may not be a concern in my study.

This paper's main contribution to the existing literature is three-fold. First, it provides supporting evidence to a wide range of studies that have examined the presence of appearance-based bias and discrimination in work settings. In addition, while most papers tend to focus merely on wage premiums or penalties, my paper further explores the effect of body shape on other qualitative aspects of a job. Second, my findings bolster the existing evidence of a larger weight penalty in terms of wages among women. Third, my paper uses the weights and height of K-pop idols as benchmarks for ideal bodies and thus deals with the subjectivity and cultural inconsistency issues that other researchers have faced in trying to find an accurate measure of body shape.

The remainder of the paper is organized as follows. Section II presents the data. Section III describes the empirical methodology. Section IV presents the baseline results and robustness analysis. Section V discusses the results and Section VI offers conclusions.

DATA

A. Student KEEP Data

The student panel data for this study are obtained from the Korean Education and Employment Panel (KEEP), a nationally representative dataset comprising 2,000 middle school seniors (9th graders) and 4,000 high school seniors (12th graders) from high schools. The KEEP panel study surveyed the same cohort of middle school seniors and high school seniors every year for 10 years starting from 2004. In each survey, respondents were asked questions regarding education records, employment, household details, professional experience, and personal characteristics including their weight and height. I focus on the high school student cohort for three years from the 2009, 2010, and 2011 surveys which are the most recent years that the English version of the survey data was available. While I have data prior to 2009, I do not include the earlier years because most respondents were still full-time students. The middle school student cohort is excluded from this study because they were still attending college from 2009 to 2011. The response rates of the annual survey during the period from 2009 to 2011 for this study were 76.2%, 76.5%, and 75.8%, respectively. After dropping observations with no weight or height information (135 observations), I exclude observations whose BMI is under 10 (one observation) or over 40 (eight observations) as those are the lower and upper limits for human survival (Lee et al., 2019). The following exclusion criteria are also applied to a total of 8072 person-year observations: monthly wages over the ninety-ninth percent level (33 observations), and invalid logarithm of monthly wages[1] (five observations). The final sample consists of 8034 person-year observations (4847 males and 3187 females). The summary statistics for the KEEP dataset are provided in Tables 1 and 2.

The employment rate is 51.1% for men and 60.7% for women. The higher employment rate among women could be the result of the mandatory national military service for Korean men which takes around 21 to 36 months. Typically, most Korean men complete military service in their early 20s; they will first complete one or two years of university studies before starting their military service. Since men take longer to finish college, they generally start working later than women of the same age, resulting in a higher proportion of full-time students among men and higher employment rate among women. Most employed men and women in the data are regularly paid workers (92.7% for men and 97.1% for women). The proportions of self-employed workers and family workers is higher for men. The average monthly wages earned by men and women are 1520000 KRW (approximately 1576 CAD) and 1330000 (approximately 1379 CAD) respectively. The majority of the employed respondents have permanent jobs (82.4% for men and 85% for women). Approximately 62.8% of male and 54.9% of female regularly paid workers receive bonuses. Approximately 82.6% and 80.6% of men and women in the

1 This is caused by observations with 0 monthly wages (log of 0 is undefined).

Table 1: Descriptive statistics of the qualitative variables of KEEP respondents

N = 8072	Male (N = 4847) Freq (%)	Female (N = 3187) Freq (%)
Attending School Full-time	2495 (51.1)	624 (19.6)
Employed	1708 (35.3)	1933 (60.7)
Regularly Paid Worker	1583 (92.7)	1877 (97.1)
Self-Employed	98 (5.8)	44 (2.3)
Family Worker	27 (1.6)	12 (0.6)
Job Market Outcomes		
Permanent Position [a,c]	1407 (82.4)	1643 (85.0)
Receiving Bonus[b,c]	992 (62.8)	1030 (54.9)
Education Level		
High School or Lower	844 (17.4)	618 (19.4)
College or higher	3995 (82.6)	2565 (80.6)
Proxy for Family Income		
Took Loan for College	399 (8.2)	662 (20.8)
Work Experience/Qualifications		
Job Training [c]	307 (6.3)	216 (6.8)
Certificate [c]	509 (10.5)	371 (11.6)
Internship [c]	370 (7.6)	1056 (33.1)
Smoking Behavior		
Currently Smoking	2285 (47.2)	235 (7.4)
Does Not Smoke	2559 (52.8)	2952 (92.6)
Drinking Behavior		
Frequent Drinker	476 (9.8)	166 (5.2)
Occasional Drinker	3899 (80.5)	2540 (79.7)
Does Not Drink	466 (9.6)	479 (15.0)

[a] Calculated only for the employed sample.
[b] Calculated only for paid workers
[c] Binary indicator

Table 2: Descriptive statistics of the quantitative variables for KEEP respondents

N = 8072	Male (N = 4847)							Female (N = 3187)						
	min	*max*	*mean*	*sd*	*p50*	*p25*	*p75*	*min*	*max*	*mean*	*sd*	*p50*	*p25*	*p75*
Weight (in kg)	40	125	71	11	70	63	77	36	100	53	7	52	48	56
Height (in cm)	154	198	175	5	175	172	179	140	179	162	5	162	158	165
BMI	13	39	23	3	23	21	25	15	35	20	2	20	18	21
Dissimilarity Index	0.06	10	2	1	2	1	2	0.03	14	2	1	1	0.7	2
BMI Dissimilarity Index[a]	0.001	13	1	1	0.5	0.2	1	0.003	16	2	2	0.7	0.3	2
Monthly Wage [b] (in 10,000 Korean Won)	4	400	152	54	150	120	180	12	400	133	40	130	110	150

[a] Only used for robustness checks. More details about this in the later section
[b] Calculated only for the employed sample.

data have completed higher education.

Job Market Outcomes.— There are several characteristics in a job profile that make a job appealing to young adults entering the labor market for the first time. Some determinants of the desirability of a job include monetary compensation, the socioeconomic status of the job, job security, permanent status and labor union presence, and job contents (Kim & Han, 2015). Even though remuneration is generally the key aspect that constitutes a desirable job, non-monetary aspects of a job, such as job security and future prospects, also heavily define a job's appeal (Brown, 1980; Jencks et al., 1988). Since workers value more than just wages, to comprehensively address the correlation between ideal body shape and career-related outcomes, both pecuniary and non-pecuniary outcomes must be taken into account.

In the KEEP data, job market outcomes are measured by 1) employment status, 2) permanent job status, 3) monthly wage, and 4) bonus provision. Monthly wages were measured in units of 10000 Korean Won (KRW). Respondents who attended school full-time are excluded from the analysis as it is likely that they were not actively looking for jobs. Respondents who helped their family at least 18 hours per week on an unpaid basis are considered employed. A permanent position is defined as either self-employment, a family worker, or a regularly paid job with a contract of one year or more. When analyzing the effect of body shape on bonus provision, I restrict the sample to only regularly paid workers. Each dependent variable except for monthly wage is represented by a binary variable.

Body Shape and Problems with Measuring It.— The key independent variable of this study is body shape. Since body shape cannot be measured directly, determining each respondent's body shape requires the use of a proxy variable. One plausible candidate is BMI. As Swami (2006), Fisher, and Voracek (2006) suggest, BMI is one of the significant determinants of physical attractiveness since body weight is inversely related to physical attractiveness. However, both studies also note that although BMI is a salient indicator of female physical attractiveness, BMI plays a much more minor role in determining male physical attractiveness because the negative correlation between attractiveness and body weight is much more pronounced among women than men. More importantly, the image of an ideal body varies across cultures; whether or not one's body shape is considered ideal depends on where they are. While some countries place a greater emphasis on curves, others prefer thinner frames. Due to cultural differences in ideal body shapes, using BMI to determine whether one's body shape is preferred or not is problematic (Swami, 2006).

B. K-Pop Idol Data

To address the cultural inconsistency in ideal body shape, I use data on the weights and heights of K-pop idols who are more likely to uphold the South Korean body standards as a benchmark for ideal bodies. The K-pop idol data used in this study were obtained from an open-access data source Rlist.io. The dataset is comprised of personal information of K-pop idols released by their companies. For each idol observation, I have information on their weight, height, and birth date. After dropping 986 observations with no weight and height information, the final K-pop idol sample consists of 1306 observations (811 males and 495 females). I summarize the weights, heights, and BMIs of the K-pop idol dataset in Table 3. By looking at the summary statistics, we can see that the majority of the K-pop idols are taller and lighter than the KEEP survey respondents. This suggests that South Koreans prefer slimmer and taller figures.

Table 3: Descriptive statistics of the quantitative variables for the K-pop idol data

N = 1306	Male (N = 811)							Female (N = 495)						
	min	*max*	*mean*	*sd*	*p50*	*p25*	*p75*	*min*	*max*	*mean*	*sd*	*p50*	*p25*	*p75*
Weight (in kg)	47	90	62	5	61	58	65	34	59	47	3	46	45	48
Height (in cm)	165	192	178	4	178	175	181	141	180	165	5	165	162	168
BMI	15	28	19	1	19	19	20	14	21	17	1	17	17	18

C. Dissimilarity Index

To measure how ideal each KEEP respondent's body is, I compute a dissimilarity index to measure the dissimilarity between each respondent and a representative group of K-pop idols. To compute the dissimilarity index, I first scale the weight and height of all idols in the dataset before grouping them into an appropriate number of clusters based on their weights and heights using a K-means clustering algorithm. Data scaling and clustering are done separately for each gender. Next, I scale the weights and heights of the KEEP respondents relative to the idol data. After scaling the data, I compute the Euclidean distances between each respondent's scaled weight-height datapoint and each center of the K-pop idol. The dissimilarity index of each respondent is defined as the minimum of all the distances between each respondent and all centers of K-pop clusters. Under this framework, having a smaller dissimilarity index corresponds to having a more ideal body shape. The summary statistics of the dissimilarity index are shown in Table 2. Figures 1-2 show the histograms of the dissimilarity index for men and women. Figures 3-4 show the scatter plots of weights and heights of K-pop idols for male and female idols as well as the visualizations of K-Means clustering.

EMPIRICAL METHODOLOGY

To estimate the impact of body shape on job market outcomes among young adults, I run logistic regression models with marginal effects to analyze the effect on binary dependent variables. For the continuous dependent variable, the analysis is done through an ordinary least squares regression. The key identifying assumption underlying this framework is that conditional on my controls, any omitted variables which are correlated with dissimilarity do not influence my employment outcomes. The potential threat to this is of course an omitted variable bias as there are numerous factors that contribute to career success. For example, individuals with a smaller dissimilarity index may make distinct lifestyle choices or exhibit attitudes and work ethics that are positively linked to career outcomes. They might possess higher self-confidence, enabling them to confidently request raises or excel at work, compared to someone equally qualified in the same job but with a larger index. The following regression specification is used for the baseline analysis:

$$Y_{it} = \beta_0 + \beta_1 Homophily_{it} + \beta_2 PerChar_{it} + \alpha_t + \beta_3 JobChar_{it} + \beta_4 Exp_{it} + \beta_5 Health_{it} + \epsilon_{it}$$

where the subscripts *i* and *t* denote individual *i* and year *t*, respectively. Y is either the logarithm of monthly wages or one of three dummy dependent variables: employment status, permanent job status, or bonus provision. Dissimilarity denotes the dissimilarity index of how different each respondent is to the representative group of K-pop idols. *PerChar* represents the vector of the respondent's personal characteristics including education level and a dummy indicator of whether they took out a student loan for college, which is a proxy for family income. α_i denotes year fixed effects. *JobChar* denotes the vector of the respondent's job characteristics including the job industry, location, and type of employment[2]. *Exp* represents a vector of the respondent's professional experience and qualifications including obtainment of certificates, job training, or internship experiences. Health represents a vector of the respondent's health characteristics including drinking and smoking behavior. ε_{it} is the error term. β_1 represents the estimated impact of having a socially preferred body shape on career outcomes. All the regression analyses in this study are run separately for male and female respondents. All estimators are computed using heteroskedasticity-robust standard errors.

RESULTS

A. Baseline Results

Tables 4-7 present the baseline results of each outcome variable of interest. Column 1 of each table displays the results of parsimonious specifications which include only the controls for personal characteristics and year-fixed effects. I extend the parsimonious specifications by incrementally adding controls for job characteristics (column 2), professional experience (column 3), and health characteristics (column 4).

The regression results shown in Table 4 confirm that there is a statistically significant association between body shape and the probability of employment for both genders. A unit increase in the dissimilarity index is associated with an approximately 1.5% decrease in the probability of employment

2 The control for type of employment is only added when analyzing the effect of body shape on monthly wages

Table 4: Regression Estimates on Employment Status

Dependent Variable: Employment Status

	Probability of Employment					
	(1)		(2)		(3)	
	Men	Women	Men	Women	Men	Women
Dissimilarity Index	-0.0159 *	-0.0122 *	-0.0155 *	-0.0128 *	-0.0154 *	-0.0127 *
	(0.00702)	(0.00569)	(0.00703)	(0.00567)	(0.00699)	(0.00563)
Controls						
Personal Characteristic Controls						
Education Level	Yes		Yes		Yes	
Proxy for Family Income	Yes		Yes		Yes	
Year	Yes		Yes		Yes	
Profesional Experience Controls						
Certificates	No		Yes		Yes	
Job Training	No		Yes		Yes	
Internship Experience	No		Yes		Yes	
Health Characteristic Controls						
Drinking Habit	No		No		No	
Smoking Habit	No		No		No	
Sample Size	2342	2558	2341	2557	2340	2557

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

that both men and women with less desirable body shapes are less likely to be employed. Since every coefficient from Table 4 falls within the same range, the influence of body shape on the probability of employment does not seem to differ significantly between men and women.

Table 5: Regression Estimates on Log of Monthly Wages

Dependent Variable: Log of Monthly Wage (in 10,000 Korean Won)

	Log Monthly Wage							
	(1)		(2)		(3)		(4)	
	Men	Women	Men	Women	Men	Women	Men	Women
Dissimilarity Index	-0.0053	-0.0136 *	-0.0064	-0.0114 *	-0.0066	-0.0121 *	-0.0071	-0.0116 *
	(0.00939)	(0.00548)	(0.00892)	(0.00524)	(0.00890)	(0.00516)	(0.00887)	(0.00526)
Controls								
Personal Characteristic Controls								
Education Level	Yes		Yes		Yes		Yes	
Proxy for Family Income	Yes		Yes		Yes		Yes	
Year	Yes		Yes		Yes		Yes	
Job Charactistic Controls								
Job Industry	No		Yes		Yes		Yes	
Work Location	No		Yes		Yes		Yes	
Type of Employment	No		Yes		Yes		Yes	
Profesional Experience Controls								
Certificates	No		No		Yes		No	
Job Training	No		No		Yes		No	
Internship Experience	No		No		Yes		No	
Health Characteristic Controls								
Drinking Habit	No		No		No		Yes	
Smoking Habit	No		No		No		Yes	
Sample Size	1541	1751	1534	1745	1534	1744	1534	1745

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 5 shows the relationship between the dissimilarity index and monthly wages. For men, there is no statistically significant relationship between the dissimilarity index and monthly wages. For women, a unit increase in the dissimilarity index is associated with an approximately 1.1% decrease in monthly wages which means that there is a negative relationship between having a less desirable body shape and monthly wages. The estimated effect for women does not change in magnitude considerably after adding more controls in columns 2, 3, and 4.

Table 6: Regression Estimates on Job Permanent Status

Dependent Variable: Job Permanent Status	Probability of Having Permanent Job Status							
	(1)		(2)		(3)		(4)	
	Men	Women	Men	Women	Men	Women	Men	Women
Dissimilarity Index	-0.0226 **	-0.0097	-0.0150 *	-0.0034	-0.0153 *	-0.0038	-0.0148 *	-0.0033
	(0.00767)	(0.00665)	(0.00701)	(0.00636)	(0.00701)	(0.00635)	(0.00704)	(0.00634)
Controls								
Personal Characteristic Controls								
Education Level	Yes		Yes		Yes		Yes	
Proxy for Family Income	Yes		Yes		Yes		Yes	
Year	Yes		Yes		Yes		Yes	
Job Charactistic Controls								
Job Industry	No		Yes		Yes		Yes	
Work Location	No		Yes		Yes		Yes	
Profesional Experience Controls								
Certificates	No		No		Yes		No	
Job Training	No		No		Yes		No	
Internship Experience	No		No		Yes		No	
Health Characteristic Controls								
Drinking Habit	No		No		No		Yes	
Smoking Habit	No		No		No		Yes	
Sample Size	2342	2558	1683	1918	1683	1917	1683	1918

*** p < 0.001; ** p < 0.01; * p < 0.05.

The estimates from Table 6 illustrate interesting results. Unlike other job market outcomes, the regression results demonstrate a statistically significant association between the dissimilarity index and the probability of having permanent job status only for men, but not for women. The regression results after adding controls for job characteristics, professional experience, and health characteristics demonstrate that a unit increase in the dissimilarity index for men reduces the probability of having a permanent job by 1.5%.

Table 7 shows the relationship between the dissimilarity index and the probability of receiving a bonus. Similar to the results we observe for monthly wages, there is a statistically significant negative correlation between the dissimilarity index and the probability of receiving a bonus for women. A unitincrease in the dissimilarity index corresponds to an approximately 2.5% decrease in the probability of receiving a bonus for women. Adding controls for job characteristics, professional experience, and health characteristics in columns 2, 3, and 4 does not alter the estimated effects appreciably.

Table 7: Regression Estimates on Bonus Provision

Dependent Variable: Bonus Provision								
	Probability of Receiving Bonus							
	(1)		(2)		(3)		(4)	
	Men	Women	Men	Women	Men	Women	Men	Women
Dissimilarity Index	-0.0135	-0.0251 ***	-0.0143	-0.0253 **	-0.0163	-0.0258 **	-0.0145	-0.0239 **
	(0.00988)	(0.00829)	(0.00945)	(0.00805)	(0.00943)	(0.00800)	(0.00944)	(0.00804)
Controls								
Personal Characteristic Controls								
Education Level	Yes		Yes		Yes		Yes	
Proxy for Family Income	Yes		Yes		Yes		Yes	
Year	Yes		Yes		Yes		Yes	
Job Charactistic Controls								
Job Industry	No		Yes		Yes		Yes	
Work Location	No		Yes		Yes		Yes	
Profesional Experience Controls								
Certificates	No		No		Yes		No	
Job Training	No		No		Yes		No	
Internship Experience	No		No		Yes		No	
Health Characteristic Controls								
Drinking Habit	No		No		No		Yes	
Smoking Habit	No		No		No		Yes	
Sample Size	2181	2461	1533	1823	533	1822	1533	1823

*** p < 0.001; ** p < 0.01; * p < 0.05.

Overall, the baseline results show that women with less desirable bodies suffer from lower chance of employment, lower monthly wages, and lower chance of getting paid bonuses while men with less ideal bodies are subject to lower chance of employment and lower chance of having permanent jobs. Additionally, we can see that the estimated effects of body shape on all four job market outcomes fall within the approximate range of 1% to 2.5%. Although these numbers seem small, it is noteworthy to keep in mind that the results from the regressions are marginal effects and the effects can still be significant when comparing individuals at the opposite ends of the body shape spectrum. For illustration, consider the female respondent with the smallest dissimilarity index (0.03) and the female respondent with the largest dissimilarity index (14). Suppose that these two women are identical in other characteristics. Then the regression results from Table 5 tell us that the difference between their monthly wages should be as large as 15.37%. Therefore, even though the results from the regression analyses are small, they still provide evidence for the large appearance-based bias in the workplace observed in previous studies, given that the identifying assumption holds.

B. Robustness Checks

I conduct a variety of robustness checks to examine the validity of my results. Rather than computing the dissimilarity index based on weight and height, I compute the dissimilarity index based on the BMIs of the KEEP respondents and K-pop idols. Subsequently, I replace the original dissimilarity index with the new dissimilarity index computed using BMIs and re-estimate the baseline regression specifications using this new independent variable. Since BMI is closely related to weight and height[3], the estimated effects from the alternative specifications should be consistent with the results from the baseline analysis. The process of computing the dissimilarity index is similar to before. I first start by scaling the BMIs of K-pop idols before clustering the idols by their BMIs using the K-means algorithm. Then, I scale the BMIs of the KEEP survey respondents relative to the idol data. Subsequently, I compute the absolute differences between each respondent's scaled BMI and the centroid (the average BMI) of each idol cluster. The dissimilarity index is defined as the smallest absolute difference between each respondent's BMI and all idol cluster centroids. The descriptive statistics of the BMI dissimilarity index are shown in Table 2.

The results from the regressions ran using the dissimilarity index from BMI are presented in Tables 8-11. Looking at the results in Table 8, we see that there is no significant association between BMI dissimilarity and the probability of employment for women which is unlike what we observe from the baseline results in Table 4. The estimated effect for men, on the other hand, demonstrates results that are consistent with the baseline specification; there is a significant negative correlation for men which means that men with less desirable BMIs are more likely to be unemployed. The estimated effect for men is also about the same magnitude obtained from the main specification shown.

3 Body Mass Index (BMI) = weight (kg) / [height (m)]^2

Table 8: Alternative Regression Estimates on Employment Status

Dependent Variable: Employment Status						
	Probability of Employment					
	(1)		(2)		(3)	
	Men	Women	Men	Women	Men	Women
BMI Dissimilarity Index	-0.0141 **	-0.0068	-0.0137 *	-0.0072	-0.0150 **	-0.0070
	(0.00541)	(0.00406)	(0.00542)	(0.00406)	(0.00538)	(0.00402)
Controls						
Personal Characteristic Controls						
Education Level	Yes		Yes		Yes	
Proxy for Family Income	Yes		Yes		Yes	
Year	Yes		Yes		Yes	
Professional Experience Controls						
Certificates	No		No		Yes	
Job Training	No		No		Yes	
Internship Experience	No		No		Yes	
Health Characteristic Controls						
Drinking Habit	No		No		No	
Smoking Habit	No		No		No	
Sample Size	2342	2558	2341	2557	2340	2557

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

The estimated effects of BMI dissimilarity on monthly wages shown in Table 9 are consistent with the baseline results in Table 5. There is a statistically significant association between monthly wages and the BMI dissimilarity index for women, although the magnitude of the effect is smaller than the results observed earlier from the baseline specification.

Table 9: Alternative Regression Estimates on Log of Monthly Wages

Dependent Variable: Log of Monthly Wage (in 10,000 Korean Won)								
	Log Monthly Wage							
	(1)		(2)		(3)		(4)	
	Men	Women	Men	Women	Men	Women	Men	Women
BMI Dissimilarity Index	-0.0036	-0.0108 **	-0.0035	-0.0085 *	-0.0036	-0.0089 *	-0.0051	-0.0086 *
	(0.00796)	(0.00378)	(0.00755)	(0.00366)	(0.00756)	(0.00361)	(0.00754)	(0.00367)
Controls								
Personal Characteristic Controls								
Education Level	Yes		Yes		Yes		Yes	
Proxy for Family Income	Yes		Yes		Yes		Yes	
Year	Yes		Yes		Yes		Yes	
Job Charactistic Control								
Job Industry	No		Yes		Yes		Yes	
Work Location	No		Yes		Yes		Yes	
Type of Employment	No		Yes		Yes		Yes	
Professional Experience Controls								
Certificates	No		No		Yes		No	
Job Training	No		No		Yes		No	
Internship Experience	No		No		Yes		No	
Health Characteristic Controls								
Drinking Habit	No		No		No		Yes	
Smoking Habit	No		No		No		Yes	
Sample Size	1541	1751	1534	1745	1534	1744	1534	1745

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 10: Alternative Regression Estimates on Job Permanent Status

Dependent Variable: Job Permanent Status								
	Probability of Having Permanent Job Status							
	(1)		(2)		(3)		(4)	
	Men	Women	Men	Women	Men	Women	Men	Women
BMI Dissimilarity Index	-0.0191 **	-0.0067	-0.0100	-0.0052	-0.0102	-0.0055	-0.0101	-0.0051
	(0.00606)	(0.00467)	(0.00547)	(0.00429)	(0.00547)	(0.00429)	(0.00547)	(0.00425)
Controls								
Personal Characteristic Controls								
Education Level	Yes		Yes		Yes		Yes	
Proxy for Family Income	Yes		Yes		Yes		Yes	
Year	Yes		Yes		Yes		Yes	
Job Charactistic Control								
Job Industry	No		Yes		Yes		Yes	
Work Location	No		Yes		Yes		Yes	
Professional Experience Controls								
Certificates	No		No		Yes		No	
Job Training	No		No		Yes		No	
Internship Experience	No		No		Yes		No	
Health Characteristic Controls								
Drinking Habit	No		No		No		Yes	
Smoking Habit	No		No		No		Yes	
Sample Size	2342	2558	1683	1918	1683	1917	1683	1918

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Comparing the results of the alternative regression on permanent job status from Table 10 with the baseline results in Table 6, we observe different results for men. The coefficients from Table 10 show no statistically significant association between the BMI dissimilarity index and the probability of having permanent job status for both genders while Table 6 illustrates that there is a significant association for men. Lastly, we examine the estimated effects of BMI dissimilarity on the probability of receiving a bonus in Table 11. The estimated results in Table 11 provide similar results to the baseline specification in Table 7. There is a statistically significant negative correlation between the BMI dissimilarity index and the probability of receiving a bonus for women, but not for men which suggests that women with more desirable BMIs are more likely to get a bonus.

Table 11: Alternative Regression Estimates on Bonus Provision

Dependent Variable: Bonus Provision								
	Probability of Receiving Bonus							
	(1)		(2)		(3)		(4)	
	Men	Women	Men	Women	Men	Women	Men	Women
BMI Dissimilarity Index	-0.0120	-0.0158 **	-0.0123	-0.0156 **	-0.0126	-0.0159 **	-0.0135	-0.0145 **
	(0.00790)	(0.00588)	(0.00770)	(0.00564)	(0.00766)	(0.00561)	(0.00772)	(0.00560)
Controls								
Personal Characteristic Controls								
Education Level	Yes		Yes		Yes		Yes	
Proxy for Family Income	Yes		Yes		Yes		Yes	
Year	Yes		Yes		Yes		Yes	
Job Charactistic Control								
Job Industry	No		Yes		Yes		Yes	
Work Location	No		Yes		Yes		Yes	
Professional Experience Controls								
Certificates	No		No		Yes		No	
Job Training	No		No		Yes		No	
Internship Experience	No		No		Yes		No	
Health Characteristic Controls								
Drinking Habit	No		No		No		Yes	
Smoking Habit	No		No		No		Yes	
Sample Size	2181	2461	1533	1823	533	1822	1533	1823

*** p < 0.001; ** p < 0.01; * p < 0.05.

DISCUSSION

The results from the alternative specifications with BMI illustrate the validity of the estimated effects of body shape on monthly wages and the probability of receiving a bonus for both genders. As for the effect on employment status, the robustness check shows consistent results for men but not for women, and vice versa for permanent job status. Although the inconsistent results cast doubt on the validity of the relationships between body shape and those two job outcomes shown by the baseline results, it may be the case that the dissimilarity index from BMI is less precise than the dissimilarity index from weight and height in determining how ideal a body shape is. According to Song and Baek (2021), BMI is "too sparse to characterize detailed body shapes" and thus the dissimilarity index computed from BMI may have failed to capture the association between body shape and employment status and bonus provision. Hence, despite the inconsistent results from the alternative specification, there might still be significant correlations between body shape and the two career outcomes.

My findings indicate that undesirable body shapes have negative associations with career outcomes in entry-level jobs for both genders. For women, having a less ideal body shape is associated with lower probability of employment, lower wages, and lower chance of getting paid bonuses while men with less ideal body shapes are less likely to be employed and have permanent jobs. These results are consistent with previous studies documenting the wage penalty from being overweight or failing to meet beauty standards. My results also provide supporting evidence for previous studies indicating that women suffer more pronounced weight penalties in terms of wages than men.

My findings further illustrate that penalties from failing to meet body standards apply not only to the monetary aspect but also to the non-monetary aspects of career outcomes. For instance, I have shown that men with less desirable bodies are less likely to have permanent jobs. Therefore, in the discussion of appearance-based bias in work settings, it is essential that aspects of a job other than remuneration are also taken into consideration.

Another key contribution of my paper lies in the use of the dissimilarity index. While BMI has been a popular measurement of physical appearance in studies on the association between physical attractiveness and labor market outcomes, many studies have reported that BMI fails to characterize detailed body shapes (Fisher & Voracek, 2006; Song and Baek, 2021; Swami 2006). Moreover, BMI is a flawed measure of physical appearance because body standards vary across cultures. My paper deals with these two issues by using the weights and heights of a group of individuals whose body shapes are considered ideal as benchmarks and measuring the dissimilarity between South Koreans and those benchmarks.

It is also important to acknowledge the limitations of my study. First, the data were obtained through written surveys. According to D. H. Lee et al. (2011), South Korean people usually over-report their heights and under-report their weights; thus, this self-reported data could bias the relationship between dissimilarity and labor market outcomes. Second, although I have added a variety of controls to mitigate omitted variable bias, there might be other unobserved personal characteristics that can influence job market performance. For instance, body image can impact self-perception, which in turn can influence interview performance, job performance, and ultimately job outcomes, irrespective of the employer's level of bias towards weight. Additionally, there is potential reverse causality in my regression specification. For instance, it may be the case that a change in monthly wage also influences body shape. Such reverse causality can cause an endogeneity issue for my estimates. Future research should develop a systematic approach to address the issues of measurement error and reverse causality as well as a more complete set of control variables to minimize omitted variable bias.

CONCLUSION

In this paper, I utilize the panel survey data from the Korean Education and Employment Panel to investigate the relationship between body shape and career outcomes in entry-level jobs. I employ the weight and height data of K-pop idols released by their companies and compute a dissimilarity index to measure how desirable each respondent's body shape is. Using these data, I show that individuals with less desirable body shapes suffer job market penalties and that such penalties are different for men and women. My paper complements previous literature by providing robust results to the existing work on appearance-based bias and goes beyond it by using benchmarking as a method of measuring body shape. Future research can expand on this study by investigating the long-term effects of appearance-based bias on career outcomes beyond entry-level jobs. Additionally, exploring the intersectionality of body shape with other identity factors, such as gender, race, and age, can provide a more comprehensive understanding of how multiple dimensions of identity influence career opportunities. Furthermore, examining the effectiveness of organizational interventions and policies in mitigating appearance-based biases can inform strategies for creating inclusive work environments.

II

PAPERS

Short-Term Value Stock Outperformance Following Earnings Surprises

Bi Yu Li

ECON 490: Seminar in Applied Economics

ABSTRACT

This paper investigates the short-term capital returns of value stocks following quarterly earnings surprises. Using an event-study framework, the study examines whether value stocks, characterized by lower price-earnings (PE) and price-book (PB) ratios than the market, outperform average and growth stocks. Analyzing the second quarter of the 2022 earnings season, the research finds that a portfolio of 32 value stocks achieved significantly higher cumulative average abnormal returns compared to 32 average stocks. These results support the hypothesis that value stocks outperform their counterparts in the short-term following earnings events. The study suggests the potential development of a short-term value trading strategy based on the likelihood of analyst error and earnings surprises. These findings contribute to the literature on value investing and demonstrate the superior capital returns of value stocks, both in the short and long term.

INTRODUCTION

Academic and professional research have demonstrated extensively that value stocks outperform growth and average stocks in the long run. This paper seeks to answer: do value stocks provide superior short-term capital returns in response to quarterly earnings surprises? Here, we define value stocks as those with lower price-earnings (PE) and lower price-book (PB) ratios than the market, or S&P 500; average stocks as those with PE and PB ratios approximately equal to the market; growth stocks as those with PE and PB ratios higher than the market; and earnings surprises as any earnings event wherein the actual earnings differ from the analysts' consensus earnings estimates.

This paper explores whether there is evidence in support of short-term value stock outperformance following an earnings season. If such evidence is found, a potential short-term trading strategy may be derived wherein an investor longs a portfolio of value stocks before their earnings events and experiences abnormal gains in a period of 15 trading days following the earnings events. This paper uses the second quarter of 2022 earnings season, with event windows starting and ending in July and August.

Using an event-study framework, this paper finds that a portfolio of 32 value stocks achieved 44.75% higher cumulative average abnormal returns over the event window compared to a portfolio of 32 average stocks following the mentioned earnings season. All cumulative average abnormal returns are statistically significant at five percent for both groups. These results support the hypothesis that value stocks do outperform their counterparts in the short-term following earnings events. Further research into this topic over a wider range of earnings seasons is needed.

The subsequent sections of this paper are organized as follows. First, a literature review explores the background of value investing and its long-term outperformance against growth investing. Also reviewed is the existing research on the long-term effects of earnings surprises and value stock outperformance, which serves as the inspiration for this paper. Second, research design, data, and methods are summarized. Third, the results section presents the research findings, discusses limitations, and identifies avenues for further research into this and similar topics. Finally, the conclusion summarizes the results.

BACKGROUND

Although value investing traces its origins to the 19th century, many consider Benjamin Graham, a mentor to Warren Buffett, to be the Father of Value Investing. Dreman (1998) notes that among Graham's favourite investment tools is the low price-to-book value method. Later research investigates several value investing metrics, among which include the price-to-earnings (PE) ratio, price-to-cash flow (PCF) ratio, and dividend yield methods.

Formal statistical research into value investing began in the mid-20th century. Cited by many as one of the earlier works, Nicholson (1960) studies the relationship between PE ratio (or PE multiple) and returns using panel data of 100 trust investment-quality stocks between 1939 and 1959. The author observes that the 20 lowest multiples showed more capital appreciation in all periods than the 20 highest multiples, with similar results for the bottom and top 40 multiples. Furthermore, stocks that outperformed the market averages are predominantly in the low multiple groups. Nicholson also demonstrates that a portfolio rebalanced every decade in the observation period to always include the current 20 lowest multiples would have an appreciation of 14.7 times. The same portfolio and rebalancing with the 20 highest multiples would have only 4.5 times appreciation. Given this evidence, Nicholson posits that low PE stocks show greater long-term returns, on average, compared to growth stocks in both capital gains and dividends.

In a subsequent study, Nicholson (1968) continues his exploration of PE ratios and returns. The author compares stocks with ratios above 20 to those with ratios at or below 10. The results indicate that five-year appreciation averaged 90% for the low PEs but only 32% for the high PEs. Similar results were found over longer periods of six to seven years, and over shorter periods of one to four years. These findings reinforce the low PE value investing principle.

McWilliams (1966) performs a similar analysis of PE ratios with applications to portfolio management. Analyzing returns of 390 exchange-listed stocks from 1953 through 1964, the author concludes that better performance can be obtained from portfolios comprised of low PE stocks relative to those comprised of high PE stocks. Additionally, high PE portfolios consistently underperformed the Dow Jones Industrial Average. Although good-performing stocks can be found in all PE deciles, McWilliams cautions against investing in high PE stocks as there exists significant uncertainty surrounding whether such companies can continue growing at expected rates.

Basu (1983) arrives at similar conclusions while also controlling for firm size. The author investigates earnings yield - that is, the inverse of the PE ratio - and its relationship with the returns of NYSE-listed stocks. Examining a period between 1963 and 1980, the data indicates that the stocks with higher earnings yield and therefore lower PE ratios earned, on average, higher risk-adjusted returns than their low earnings yield counterparts. The earnings yield effect is significant even after controlling for differences in firm size.

In their 1992 paper, Fama and French acknowledged the works of Stattman (1980) and Rosenberg et al. (1985). Both works find a positive and significant relationship between average returns of American stocks and their book equity-to-market equity values, an inverse of the PB ratio. Fama and French (1992) also cite Chan et al. (1992), who discovered a similar positive and significant correlation between Japanese stock returns and their book equity-to-market equity ratios. Fama and French (1992) themselves find that the relationship between earnings yield and abnormal returns is positive, but an even stronger relationship exists between book-price ratio and abnormal returns.

In their 1993 work, Fama and French posit that the following variables, when used alone, have explanatory power on average returns: firm size, earnings yield, and book-to-market equity. Their results indicate that, on average, smaller firms and value firms tend to have better performance. In other words, according to the results of Fama and French (1993), one can expect superior returns from firms with low PE ratios (the inverse of earnings yield) and low PB ratios (the inverse of book-to-market equity).

Dreman and Berry (1995) examined earnings announcements and their effects on returns of stocks with different PE ratios. The authors define earnings surprises as a difference in absolute value between analysts' forecasted earnings and actual earnings announced. They compared stocks in the 80th percentile of highest PE ratios with stocks in the 20th percentile of highest PE-ratios, and found an asymmetry of returns in response to earnings surprises. Between 1973 and 1993, the low-PE ratio stocks earned market-adjusted annualized returns of 7.05% and 5.22% for holding periods of one quarter and one year following earnings surprises. Meanwhile, the high-PE ratio stocks achieved market-adjusted annualized returns of -5.69% and -4.57% for holding periods of one quarter and one year following the earnings surprises. Above and below-market returns continued for 19 quarters following earnings surprises for low- and high-PE ratio stocks. The authors conclude that earnings surprises, whether negative or positive, favour the long-term returns of value stocks, on average. Additionally, the effects are strongest in the surprise quarter, but persist throughout a one-year holding period and beyond.

Dreman and Berry (1995) also examine the separate effects of positive and negative earnings surprises. For positive surprises, wherein the reported earnings are higher than expected, the low-PE stocks earned market-adjusted annualized returns of 20.05% and 9.39% for holding periods of one quarter and one year following earnings surprises. Meanwhile, the high-PE stocks earned market-adjusted annualized returns of 9.39% and 0.32% for holding periods of one quarter and one year following earnings surprises. For negative surprises, wherein the reported earnings are lower than expected, the low-PE stocks earned market-adjusted annualized returns of -4.17% and 0.74% for holding periods of one quarter and one year following earnings surprises. Meanwhile, the high-PE stocks earned market-adjusted annualized returns of -18.49% and -9.54 % for holding periods of one quarter and one year following earnings surprises. The authors demonstrate that both positive and negative earnings surprises favour the low-

PE stocks.

Finally, Dreman and Berry (1995) find that the net impacts of earnings surprises are minimal on stocks in the middle 60 percent of PE ratios. For all earnings surprises (positive and negative), the market-adjusted annualized returns for holding periods of one quarter and one year following the events are -0.57% and -0.36%. This is the rationale for using stocks of average PE and PB ratios as the control group in my paper.

Dreman and Lufkin (1997) incorporate four measures of value into their research. The ratios include PE, PCF, PB, and PD (price-dividend; the inverse of dividend yield) ratios, collectively referred to as the price ratios. The authors report that diversified portfolios of stocks with low value ratios have returns significantly above the market by approximately 300 basis points annually. Conversely, diversified portfolios of stocks with high value ratios have returns below the market by roughly the same amount. Tying all four measures together, the authors demonstrate that low price ratio stocks do outperform both the market and their high ratio counterparts.

Dreman (1998) uses findings from his prior works, along with insights and research from other scholars and industry professionals, to create general observations surrounding value investing. Note that "out-of-favour stocks", "lowest value ratio stocks", "worst stocks", and "value stocks" are used interchangeably here. Note that "in-favour stocks", "highest value ratio stocks", "best stocks", "favourites", and "growth stocks" are used interchangeably here. Also note that by "analyst error", Dreman refers to earnings surprise. Dreman's observations are as follows:

1. Take advantage of the high rate of analyst forecast error by simply investing in out-of-favour stocks.
2. Positive and negative surprises affect "best" and "worst" stocks in a diametrically opposite manner.
3. [Earnings] Surprises, as a group, improve the performance of out-of-favour stocks, while impairing the performance of favourites.
4. Positive surprises result in major appreciation for out-of-favour stocks, while having minimal impact on favourites.
5. Negative surprises result in major drops in the price of

favourites, while having virtually no impact on out-of-favour stocks.
6. The effect of an earnings surprise continues for an extended period of time.

The works cited thus far have demonstrated, extensively, that value stocks outperform both the market and growth stocks in the long run. Moreover, price ratios seem to be better predictors of future return than the Capital Asset Pricing Model. Indeed, Fama and French (1993) find little relation between beta and average returns of U.S. stocks. According to the CAPM, higher beta stocks are supposed to provide higher returns, but this is not the case. Dreman (1998), citing Basu (1977), quotes "low PE stocks provided superior returns, and were also somewhat less risky". In other words, compared to their high PE counterparts, low PE portfolios exhibit higher returns but are not associated with higher risk.

Collectively, these studies advocate for long-term investments in value stocks, with periodic portfolio rebalancing to always include the stocks with lowest price ratios. Even strategies as simple as choosing the lowest PE Dow Jones stocks and rebalancing every five years provides higher returns than the Dow Jones itself (Dreman, 1998). However, the strategies explored by these researchers are applicable only to longer-term investors.

How might a short-term trader benefit from value strategies? This topic has yet to be explored as extensively. Like the works cited previously, this paper endeavours to investigate the relative performance of value stocks. Building upon Dreman's works on returns and earnings surprises, this paper will focus exclusively on whether a disparity in performance between value and average stocks can be identified in the short run: only 15 trading days after earnings surprises.

DESIGN

This paper investigates the cumulative abnormal average returns over a 20-trading day event window for treatment and control groups.

The treatment group includes a sample of 32 firms reporting earnings in the period of July to August 2022 with

PB ratios between zero and two, and PE ratios between zero and 10. Firms comprising the treatment group are randomly selected from over 300 eligible firms during this period. The control group includes a sample of 32 firms reporting earnings in the period of July to August 2022 with PB ratios between three and five, and PE ratios between 17 and 23. Firms matching the control group criteria are much more scarce, resulting in the control sample of 32 firms being approximately equal to the universe of eligible control stocks in this period, when matched on a by-day basis with the treatment group.

These ratios are chosen such that the treatment group, composed of value stocks, has PB and PE ratios at 50% of market average at most and zero at minimum. The control group, composed of average stocks, have PB and PE ratios close to those of the market average. The market averages, based on the S&P 500, are approximately four and 20 for PB and PE ratios, respectively (NASDAQ, 2022).

A firm must be listed on either the New York Stock Exchange or NASDAQ exchange to be included. A firm must also have experienced a surprise in earnings - that is, a difference between actual reported earnings and analysts' consensus forecasted earnings. Both positive and negative earnings surprises are included.

To limit the effects of broad-market movements on returns of one group but not the other, the chosen firms for both groups are matched on a by-day basis, with one exception. For example, if the treatment group includes five firms reporting on a given day, the control group would also include five firms reporting on that same day, and so on.

Day zero is defined as the earnings reporting date for any given firm. The estimation period is defined as day -45 to day -5. The event period is defined as day -4 to day 15.

Daily returns data for all firms are obtained from the Compustat Security Daily Close Price accessed via the Wharton Database Research Services (WRDS). Daily returns are then calculated, for any particular day, as the day's close price divided by the previous day's close price, minus one. For the estimation of a three-factor Beta, daily data is obtained from the Fama French 5 Factor section in the WRDS.

The Fama French (1993) three-factor regression to estimate the Beta of any given firm i, using the daily returns and three-factors data over the estimation period, takes the following form:

$$r_{it} - r_{ft} = a + \beta_i\left[r_{mt} - r_{ft}\right] + s_i SMB + h_i HML + e_t$$

Where:

r_{it} is stock i's return on day t,

r_{ft} is the risk-free rate on day t,

r_{mt} is the market's return on day t,

s_i is the control coefficient for firm size (SMB),

h_i is the control coefficient for firm book-to-market value (HML), and

e_t is the error term.

Using Beta from the previous step, the normal (expected) daily return for any given firm *i* is estimated by the Capital Asset Pricing Model:

$$E[r_i] = r_f + \beta_i(r_m - r_f)$$

Note that the risk-free and market return rates use mean values for both variables over the estimation window.

Abnormal returns for any given firm *i* on any given day *t*, during the event window, are estimated by subtracting that day's return from the firm's normal daily return:

$$ar_{it} = r_{it} - E[r_i]$$

Over the event period, average abnormal returns for any given day *t* is the mean of all abnormal returns for that day. Cumulative average abnormal returns for any given day *k* is the cumulative sum of average abnormal returns on that day plus all prior days.

$$aar_t = \frac{1}{n}\sum_{i=1}^{n} ar_{it}$$

$$caar_k = \sum_{t=1}^{k} aar_t$$

One potential threat to this research design is the presence of firm-specific shocks that may distort results. However, given that the firms of both samples vary in size, industry, and other characteristics, the threat of firm-specific shocks is minimal. Another potential threat is that the magnitude of earnings surprises is not considered, such that a firm is eligible to be included in its respective group so long as it experienced an earnings surprise of any size. The results for both groups are shown below (Table 1, Table 2)

Table 1
Regression results for each treatment stock during the estimation period. Betas are then used in the Capital Asset Pricing Model to determine normal returns for each stock.

Stock	β
AFL	0.928* (0.064)
ARW	1.144* (0.121)
VZ	0.199 (0.104)
CLF	1.869 (0.209)
DCO	0.810* (0.189)
ALLY	1.717* (0.129)
RYI	2.018* (0.296)
IP	0.862* (0.096)
VSH	0.997* (0.132)
LCII	0.895* (0.156)
VHI	1.857* (0.290)

ASC	1.215* (0.309)
SYF	1.606* (0.212)
CRC	1.181* (0.236)
UNVR	1.399* (0.168)
ESTE	1.809* (0.433)
TTE	1.162* (0.141)
NEXA	1.847* (0.293)
SI	2.799* (0.465)
DFH	0.620 (0.398)
LAD	1.294* (0.208)
MT	1.620* (0.183)
NLY	1.018* (0.167)
GPI	1.136545* (0.252272)
WCC	1.806* (0.222)
UBS	1.101* (0.153)
WAL	1.270* (0.129)
TX	1.228* (0.196)
IPI	1.968* (0.329)
SB	1.070* (0.294)

CIVI	1.656* (0.274)
ASX	1.090* (0.254)

Table 2
Regression results for each control stock during the estimation period. Betas are then used in the Capital Asset Pricing Model to determine normal returns for each stock.

Stock	β
BCE	0.678 (0.109)
CR	0.988 (0.143)
CFR	0.973 (0.107)
CMI	1.064 (0.111)
DOV	0.914 (0.085)
FMC	1.164 (0.117)
FRT	0.897 (0.126)
GD	0.823 (0.110)
HAL	1.700 (0.208)
ITT	1.193 (0.119)
EME	1.027 (0.103)
NSC	0.850 (0.084)
SXT	0.679 (0.119)

WMB	0.923 (0.171)
SFBS	0.528 (0.124)
CHGG	1.029 (0.323)
ENLC	1.843 (0.206)
GIB	0.672 (0.113)
HLI	0.947 (0.135)
SITE	1.228 (0.161)
KAI	0.977 (0.1509)
GOLF	0.758 (0.162)
ALV	1.511 (0.209)
FIX	1.053 (0.188)
RBA	0.523 (0.276)
GRMN	0.864 (0.095)
ELV	0.612 (0.144)
LDOS	0.813 (0.105)
AWK	0.783 (0.162)
BKI	0.366 (0.121)
GNRC	1.806 (0.250)
WIT	1.019 (0.154)

RESULTS

The treatment group achieved higher cumulative average abnormal returns over a 20-trading day event window than the control group. By the final event period day, the treatment group had cumulative average abnormal returns of 0.0909 versus 0.0628 for control group. In other words, the treatment group achieved 44.75% higher cumulative average abnormal returns compared to the control group (Figure 1). Hypothesis tests at 5% significance indicate that cumulative average abnormal returns for every given date are all statistically significant for both groups (Table 3, Table 4).

Figure 1
The Treatment Group (Red) achieves higher cumulative average abnormal returns (CAAR) than the Control Group (Blue)

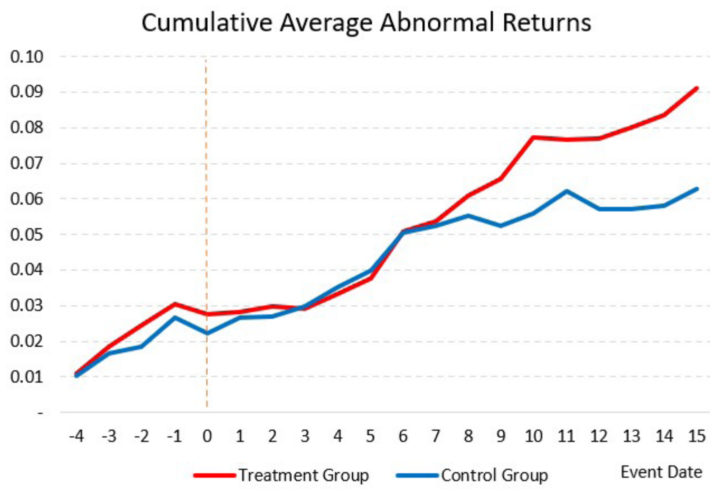


Table 3
Statistics for the treatment group – statistically significant cumulative average abnormal returns (caar) for any given day are denoted by an asterisk (*) and are found by dividing the caar by its standard error (se)

Time	aar	caar
-4	0.011	0.011* (0.004)
-3	0.008	0.0184* (0.006)
-2	0.006	0.024* (0.007)
-1	0.006	0.030* (0.008)
0	-0.003	0.0276* (0.009)
1	0.001	0.028* (0.0103)
2	0.002	0.030* (0.011)
3	-0.001	0.029* (0.012)
4	0.004	0.033* (0.013)
5	0.005	0.038* (0.013)
6	0.013	0.051* (0.014)
7	0.003	0.054* (0.015)
8	0.007	0.061* (0.015)
9	0.005	0.066* (0.016)
10	0.011	0.077* (0.016)

11	-0.0004	0.077* (0.017)
12	0.0002	0.077* (0.017)
13	0.003	0.080* (0.0177)
14	0.003	0.084* (0.018)
15	0.007	0.091* (0.019)

Table 4
Statistics for the control group – statistically significant cumulative average abnormal returns (caar) for any given day are denoted by an asterisk (*) and are found by dividing the caar by its standard error (se)

Time	aar	caar
-4	0.0104	0.010* (0.003)
-3	0.006	0.017* (0.004)
-2	0.002	0.018* (0.005)
-1	0.008	0.027* (0.006)
0	-0.004	0.022* (0.006)
1	0.004	0.027* (0.007)
2	0.0003	0.027* (0.007)
3	0.003	0.030* (0.008)
4	0.005	0.035* (0.008)

5	0.005	0.040* (0.009)
6	0.011	0.051* (0.009)
7	0.002	0.052* (0.010)
8	0.003	0.055* (0.010)
9	-0.003	0.052* (0.010)
10	0.003	0.056* (0.011)
11	0.006	0.062* (0.011)
12	-0.005	0.057* (0.011)
13	0.0001	0.057* (0.012)
14	0.001	0.058* (0.012)
15	0.005	0.063* (0.012)

From this, we can arrive at the following implications. First, value stocks did produce short-term abnormal returns in response to earnings surprises in the second quarter of 2022. Second, and more importantly, value stocks produce higher short-term abnormal returns in response to earnings surprises when compared to average stocks in this period. During the event window, the S&P 500 was mostly bullish and followed a short-term upward trend. Thus, coincident positive returns can be expected from most stocks. The results seem to be consistent with the existing literature; value stocks will outperform non-value stocks following earnings surprises. Had the market been bearish during the same event period, we would still expect the abnormal returns of value stocks to be higher than that of non-value stocks.

One limitation of this study is that, due to time constraints, it only investigated one earnings season: the second quarter of 2022. Further research could explore abnormal returns of value stocks using larger sample sizes and more earnings dates – looking at overall abnormal returns during all four earnings seasons of a given year, for example. Deeper research could look at many, if not all, historical earnings seasons and provide aggregated data.

Nonetheless, a trading strategy may potentially be derived from this data. An investor seeking short-term abnormal returns could long a portfolio of value stocks during an earnings season. The investor may open positions at around four trading days before and close the positions at around 15 trading days after each individual stock's earnings report. This strategy may lead to higher abnormal returns compared to portfolios of different compositions. During market up-trends, this portfolio may yield higher positive returns. During down-trends, this portfolio will likely yield lower negative returns or even positive returns. Thus, investors may use this strategy for short-term capital gains in bullish markets or hedging in bearish markets.

There are limitations to the described strategy. First, short-term trading strategies with frequent transactions will incur high commission costs. These costs significantly diminish total returns over time. Second, there is no evidence that following such a strategy will yield higher annualized returns versus a more passive value investing strategy, such as buying a value portfolio and rebalancing periodically. Further research can investigate these limitations.

One final area for further consideration is a comparison of short-term abnormal returns in value versus growth stocks. This paper compared value and average stocks since the existing literature demonstrates that average stocks serve as a control group. If short-term performance continues to be consistent with long-term performance results, we may expect the difference in short-term abnormal returns between value and growth stocks to be even greater than the difference between value and average stocks.

CONCLUSION

The findings of this paper support the hypothesis that value stocks outperform in the short run following earnings events. This paper demonstrates that value stocks yield higher short-term abnormal returns compared to average stocks following earnings surprises in the second quarter of 2022. Over a 20-trading day event period, a sample of 32 value stocks produced 44.75% higher cumulative average abnormal returns compared to a sample of 32 average stocks. The difference, in favour of value stocks, is expected to be larger if compared to growth stocks. All cumulative average abnormal returns for every event day are statistically significant at 5% for both groups.

These results provide evidence to support the proposition of superior capital returns to value stocks even in the short-term. It supplements the rigorous literature that has demonstrated, for many decades, that value stocks outperform non-value stocks in the long run. As such, taking advantage of the previously established high probability of analyst error (and therefore earnings surprise) for any given earnings event, a potential short-term value trading strategy may be explored and developed following further research.

Health And Labour Supply: The Effect Of Self-rated Health On Hours Worked

Derek Situ

ECON 490: Seminar in Applied Economics

ABSTRACT

This paper estimates the effect of self-rated health on weekly hours worked. Using National Health Interview Survey data from 2015 to 2018, I find that a unit increase in Likert-scale self-rated health is predicted to increase weekly hours worked by about 5-7 hours for individuals aged 21-67 in the U.S. This estimated effect is greater for individuals above the median age (45 years) at 6-8 additional hours per unit increase in Likert-scale self-rated health, and smaller for individuals not above the median age, at 3-5 additional hours. Endogeneity is addressed with an instrumental variables approach, and results suggest that justification bias is outweighed by other forms of bias in OLS estimates.

INTRODUCTION

Health is an important consideration in an individual's labour supply decisions. Many economists have studied the effect of health on labour force participation (Bound et al. 1999; Blundell et al. 2017; Cai 2010; Stern 1989), but few have studied its effect on hours worked; additionally, when hours worked is the focus, endogeneity of health is not deemed an issue and is not addressed (Pelkowski and Berger 2004) or U.S. data is not used (Cai, Mavromaras, and Oguzoglu 2014). This paper aims to estimate the effect of self-rated health on weekly hours worked for working-age individuals in the U.S. while addressing endogeneity concerns. In particular, I characterize the extent to which justification bias affects Ordinary Least Squares (OLS) estimates. Only individuals aged 21-67 years are considered, as the relation between health and hours worked for young children, teens, and the elderly is likely different in important ways. Specifically, these groups are likely to work relatively few hours regardless of their health status as they are less likely to be in the middle of their careers.

The principal challenge faced by scholars investigating the causal impact of health on labour supply is endogeneity. First, measurement error is likely to occur when survey respondents are asked to rate their own health. Not only are measurements unlikely to be consistent across individuals, as respondents may rate their health differently despite perceiving their health to be at the same level, Crossley and Kennedy (2002) show that 28% of respondents report their own health differently when asked the same question twice—before and after a set of additional health-related questions. Second, respondents not working may have psychological or economic incentives to report poor health or limitations due to disability, such as guilt, shame, or disability benefits (Black et al. 2017). This so-called "justification bias" may cause the impact of health on labour supply to be over-estimated. A third source of endogeneity is simultaneous causation. That is, health may affect labour supply, but labour supply may also affect health. The former effect is of interest in this study, but the latter would bias our estimates if true. Indeed, there are intuitively appealing reasons why it may be true. Stern (1989) suggests that boredom or lack of activity can possibly cause health deterioration when out of the labour force, as can bad working conditions or stress when in the labour force. Long working hours may also contribute to health deterioration.

Endogeneity of self-rated health due to measurement error, justification bias, and simultaneous causality is addressed by instrumenting self-rated health with objective health measures. Objective health measures are thought by many researchers to be robust to measurement error and justification bias, however Baker, Stabile, and Deri (2004) show that objective health measures suffer from the same weaknesses using a Canadian sample. While objective health measures may themselves be subject to measurement error (including measurement error due to justification bias), measurement errors in binary instrumental variables (IVs), such as the objective health measures in this study, do not bias estimates in general (Jiang and Ding 2020). Blundell et al. (2017) note that objective health is likely to be unrelated to measurement error and justification bias in subjective health measurements, while being strongly related to subjective health. This provides a theoretical basis for the exclusion restriction and relevance condition, suggesting that objective health measures may be appropriate instruments for self-rated health. A possible mechanism through which the exclusion restriction may be violated occurs if justification bias is correlated with obtaining diagnoses for objective health measures. Since obtaining diagnoses requires effort on the part of the respondent, it is possible that they may be more likely to do so only if they have to justify low hours worked. Such an IV approach also operates under the assumption that the measurement errors affecting each of objective and subjective health are uncorrelated.

In section 2, the existing literature on the subject is summarized. Section 3 outlines the methodology used, and section 4 presents the results of the analysis. Section 5 concludes the paper with a summary, interpretation of results, study limitations, and avenues for future research.

LITERATURE

Although the impact of health on labour supply has been studied, a consensus on the size of the effect and whether endogeneity affects estimates has not been reached. Pelkowski and Berger (2004) examine the effect of health problems on employment, hourly wages, and annual hours worked using the Health and Retirement Study dataset. They use OLS to investigate the effects of interest, incorporating fixed effects to account for unobserved heterogeneity, and a Heckman correction to account for the fact that only employed respondents had wage data. The authors perform Hausman tests, concluding from the results that endogeneity of health due to measurement error and simultaneity is not an issue for the equations estimated. They find that poor health leads to a reduction in annual hours worked.

Cai (2010) examines the effect of self-rated health on labour supply using the Household, Income and Labour Dynamics in Australia Survey. He tests for endogeneity and finds that simultaneous causation and unobserved individual fixed effects for females contribute to endogeneity. Cai addresses this concern using objective health measures as instruments for self-rated health and finds that health has a positive and significant effect on labour force participation.

Blundell et al. (2017) investigate the impact of health on employment status for adults near retirement. They compare estimates obtained from using U.S. data with estimates obtained from using English data, as well as estimates obtained from variations in methodology. The authors find that declines in health are able to explain between 3% to 15% of the decline in employment for adults aged 50 to 70, and that these effects are larger in the U.S. than in England. To address measurement bias, justification bias, and simultaneity bias, the authors use objective health measures as instruments for subjective health, and produce similar estimates as those obtained from using only one of objective measures or subjective measures. They provide empirical and theoretical evidence satisfying the exclusion restriction, using the Sargan test and arguing that objective health measures are likely robust to measurement error and justification bias.

Stern (1989) estimates the effect of disability on labour force participation using data from the Survey of Disability and the Health Interview Survey. Using objective health measures such as symptoms or diseases as instruments for reported disability, he finds that disability is able to explain a significant amount of variation in labour force participation. Three tests of endogeneity are performed, which provide only weak evidence of endogeneity of disability. Between measurement error, justification bias, and simultaneity bias, Stern finds that the only form of endogeneity consistent with the data is a specific case of simultaneity in which working deteriorates health.

Bound et al. (1999) analyze the relationship between health and labour force transitions such as exits, job changes, and applications for disability insurance. Using data from the Health and Retirement Survey, the authors find that poor health causes many older workers to exit the labour force. They describe possible sources of endogeneity of health measures which can be categorized under measurement error, justification bias, and simultaneous causation.

This paper provides two primary contributions to the health and labour supply literature. First, I provide an estimate of the impact of health on hours worked using recent U.S. data. Second, I assess whether justification bias has a significant role in altering OLS estimates.

METHOD

A. Data

The dataset used for this study comes from the 2015-2018 samples of the National Health Interview Survey (NHIS) hosted by IPUMS. The NHIS is a nationally representative, cross-sectional survey of about 40,000 U.S. individuals. The survey is administered annually and provides information on the demographics, health status, healthcare utilization, and health-related behaviours of a sample of the civilian U.S. population. Table 1 presents a summary of select variables from the final version of the dataset used in the analysis. The explanatory variable of interest, self-rated health, is originally available as a 5-point Likert-coded response ranging from 1 (Excellent) to 5 (Poor). The numbering has been inverted to achieve a more intuitive interpretation with greater numbers indicating greater self-rated health. The variables corresponding to health conditions—arthritis, asthma, chronic obstructive pulmonary disease (COPD), smoking (ever smoked 100 cigarettes), heart attack, heart condition, hypertension, vision problem, and stroke—are available as binary variables indicating past diagnosis, or in the case of smoking, the respondent's best estimate. I include only individuals aged 21-67 and remove all individuals missing information, leaving me with 86,293 observations.

Table 1. Summary of select variables

Hours Worked	Health	Arthritis	Asthma	Smoke	Age	Sex
Min: 0.0	1: 2526	0: 67766	0: 74244	0: 52615	Min: 21.0	Female: 46645
1st Qu: 0.0	2: 8402	1: 18595	1: 12117	1: 33746	1st Qu: 33.0	Male: 39716
Mdn: 40.0	3: 22647				Mdn: 45.0	
Mean: 28.9	4: 29107				Mean: 44.9	
3rd Qu: 40.0	5: 23679				3rd Qu: 57.0	
Max: 95.0					Max: 67.0	

The dependent variable, weekly hours worked, is not available for respondents reporting unemployed status. The value 0 is imputed in such cases. Income is reported for 89% of respondents and family income is reported for 79% of respondents. The dataset includes five imputed values in place of missing income and missing family income responses. The Centers for Disease Control and Prevention (2021b) provides information on the imputation methodology used. Each missing value of income or family income is replaced by the mean of the five imputed values. Under the assumption that the imputation model is correct, this produces unbiased point estimates, but underestimates variability (Centers for Disease Control and Prevention 2021b). Table 2 presents counts of missing values remaining in the dataset by variable, after accounting for imputed values for hours worked[1], income, and family income.

As seen in Table 2, each of the variables related to diagnosis of a health condition have about 121,000 missing values. According to the IPUMS NHIS website (Blewett et al. 2022), responses for these variables are only available for adults randomly sampled to respond to the survey on behalf of their family. Due to this random availability of health condition variables, they are likely missing completely at random. This justifies a complete-case analysis, in which all observations

1 The remaining missing values for hours worked are from participants who reported employed status, but refused to answer or reported not knowing the answer.

with any missing health condition variables are deleted, as the method has been shown to be the most appropriate when the data is missing completely at random (Zhu 2014).

Table 2. Summary of missing data

Variable	Missing (N)	Missing (%)
Smoke	121251	57.66
Hypertension	121074	57.58
Arthritis	121067	57.57
Heart Condition	121045	57.56
COPD	121043	57.56
Heart Attack	121033	57.56
Stroke	121031	57.56
Asthma	121030	57.55
Vision	120993	57.54
Education	2566	1.22
Hours Worked	2448	1.16
NAICS-2	1626	0.77
Married	706	0.34
Health	249	0.12
Region	0	0.00
Age	0	0.00
Sex	0	0.00
Number of Children	0	0.00
Income	0	0.00
Family Income	0	0.00

Table 2 shows that the remaining variables have a small proportion of their values missing—less than two percent in all cases. Given this, and the large size of the sample, the bias arising from a complete-case analysis is minimal (Zhu 2014), so observations with incomplete data for the remaining variables are removed. The affected variables are education, hours worked per week, two-digit North American Industry Classification System (NAICS), employment status, marital status, and self-rated health.

To protect the confidentiality of respondents, top-coding is used to censor the data. Income is top-coded at the 95th percentile of any given year. All income values corresponding to the top-coded value are multiplied by 1.4, which may result in a rough approximation for the mean income at the upper tail of a Pareto distribution. This follows a popular convention in the literature used by Card and

DiNardo (2002). Top-coded family income is treated similarly by multiplying other family members' combined income by 1.4, where other family members' combined income is calculated as the difference between family income and income. Hours worked per week is top-coded at 95 hours. The proportion of top-coded responses for this variable is small (less than 0.2%), and responses are not likely to be much greater than the top-coded value, which corresponds to almost 14 hours worked daily or 19 hours worked per day assuming a five-day work week. For these reasons, no adjustment is done for top-coded hours worked as the bias introduced is likely to be minimal.

Other data transformations include an inflation adjustment for both income variables, as well as an aggregation of responses for the education variable. Income and family income are converted to 2009 U.S. dollars by multiplying their value with a CPI multiplier included in the dataset. Doing so renders the estimated effect of the year variable to be statistically insignificant in the regression models, thus it is excluded. 22 different valid responses are available for the education variable, and these are grouped into four possible responses: did not graduate high school/no GED, high school/GED, bachelor's/associate's degree, and graduate/professional degree.

B. Empirical Strategy

Self-rated health and weekly hours worked are modelled as simultaneous equations estimated by two-stage least squares. Equations 1 and 2 specify the first stage and second stage respectively.

Controls is a vector of controls containing a set of region dummies, age, age squared, a sex dummy, inflation-adjusted family income, inflation-adjusted income, a marital status dummy, number of children, education dummies, and a set of 2-digit NAICS dummies. *Instrument* is a vector of dummies indicating diagnosis of various health conditions. Three sets of

$$Health_i = \alpha_0 + Instrument_i\delta + Controls_i\psi + \epsilon_i$$
$$HrsWorked_i = \beta_0 + \beta_1 Health_i + Controls_i\gamma + u_i$$

instruments are considered, resulting in three separate models. The first set of instruments—which contains diagnosis of arthritis, asthma, heart attack, heart condition, and vision problems—is constructed to obtain the smallest Sargan test statistic in order to appeal to the exclusion restriction. The second set of instruments—which contains asthma, COPD, and smoking—is constructed to represent conditions related to the lungs. The third set of instruments—which contains heart attack, heart condition, hypertension, and stroke—is constructed to represent conditions related to the heart. Estimates obtained from using different sets of instruments are compared to evaluate sensitivity of results.

For all three models, empirical tests of the relevance condition and the exclusion restriction are evaluated. Instrument relevance is tested by F-tests with the instruments excluded, and the exclusion restriction is tested by Sargan tests. However, the validity of the Sargan test for testing the exclusion restriction depends on the homogeneity of effects identified from each of the individual instruments (Angrist and Imbens 1995; Blundell et al. 2017). Furthermore, Lemieux (2019) remarks that in the case of heterogenous effects, rejection of the null hypothesis indicates the presence of heterogeneity in treatment effects as opposed to the lack of validity of one of the instruments. Therefore, the results of the Sargan tests should be interpreted with caution.

The set of instruments related to heart conditions may have a stronger theoretical basis for satisfying the exclusion restriction than the other sets of instruments. This is because they either relate to one-off shocks rather than long-term conditions that individuals base their decision to supply labour on; or in the case of hypertension, to a condition that usually has no symptoms (Centers for Disease Control and Prevention 2021a), and thus no direct reason to affect labour supply decisions.

Joint significance of sets of dummies for the categorical control variables (region, education, and NAICS) is evaluated by F-tests at the 5% significance level, and all control variables

---

[1] *Roth and Slotwinski attribute this to the lack of high-quality administrative data on relative incomes, particularly for women.*

with insignificant effects are removed from the final models presented. Notably, the year variable is excluded as a control, since its estimated effect is insignificant after adjusting income and family income for inflation. Hausman tests are conducted to test endogeneity of self-rated health, and OLS estimates are compared with IV estimates.

To identify potential differences in the effect of self-rated health for different ages, OLS estimates and estimates for the second-stage regressions of the three IV models are re-calculated using subsets of the dataset. One subset consists of respondents who are at most the median age (45 years) and the other subset consists of respondents over the median age.

RESULTS

Results from the regression analysis of the whole dataset are presented in Table 3. Estimates from the reduced form regressions of all three models show that diagnosis of any of the health conditions is associated with a reduction in hours worked. The estimated coefficients from the first-stage regressions show that diagnosis of any of the health conditions reduces predicted self-rated health, and the estimated intercepts show that self-rated health is predicted to be about 4.8 out of 5 in the absence of health conditions and with controls set to their reference level. The second-stage regressions of all three models show that a unit increase in Likert-scale self-rated health is predicted to increase weekly hours worked by about 5-7 hours; the estimates do not vary greatly depending on the instruments used. The estimated coefficient from the model with lung-related conditions as instruments is the greatest of the three at 7.13. The relatively large estimated effect appears mostly driven by the effect of diagnosis of a COPD, which can be observed in the reduced form and first-stage regressions. This is contrasted with the estimated coefficient from the model with heart-related conditions as instruments, which is the least of the three, at 5.23.

Test statistics for the Hausman tests and instrument relevance F-tests are presented at the bottom of Table 3. For all models, the null hypothesis that the instruments are weak is rejected at 1% significance, providing evidence for instrument relevance. From the Hausman test statistics, the null hypothesis that self-rated health is exogenous is rejected at 1% significance for all models. This provides evidence that the IV estimates shown perform better than the OLS estimate. The OLS estimate is considerably less than the IV estimates, suggesting that the overall bias is negative.

Table 3. Regression results: Whole dataset

		Sargan IVs			Lung IVs			Heart IVs		
	OLS	Reduced	Stage 1	Stage 2	Reduced	Stage 1	Stage 2	Reduced	Stage 1	Stage 2
(Intercept)	-27.28***	-15.71***	4.77***	-45.46***	-15.57***	4.79***	-49.68***	-15.59***	4.75***	-40.56***
	(0.81)	(0.76)	(0.04)	(1.19)	(0.76)	(0.04)	(1.54)	(0.76)	(0.04)	(1.24)
Health	2.45***			6.24***			7.12***			5.22***
	(0.06)			(0.19)			(0.28)			(0.20)
Arthritis		-2.86***	-0.47***							
		(0.15)	(0.01)							
Asthma		-1.45***	-0.26***		-1.58***	-0.31***				
		(0.16)	(0.01)		(0.16)	(0.01)				
COPD					-7.64***	-0.80***				
					(0.34)	(0.02)				
Smoking					-0.82***	-0.18***				
					(0.12)	(0.01)				
Heart Attk	-4.81***	-0.61***						-4.40***	-0.54***	
	(0.39)	(0.02)						(0.39)	(0.02)	
Heart Cond	-2.15***	-0.35***						-2.38***	-0.38***	
	(0.23)	(0.01)						(0.23)	(0.01)	
Hypertens								-1.53***	-0.49***	
								(0.14)	(0.01)	
Vision Prb	-2.43***	-0.40***								
	(0.19)	(0.01)								
Stroke								-5.90***	-0.51***	
								(0.39)	(0.02)	
Adj. R^2	0.41	0.41	0.26	0.38	0.40	0.22	0.37	0.40	0.25	0.40
F-test				1941.05			1445.27			1948.27
Hausman				482.68			327.04			206.29
Sargan				7.77			50.08			139.40

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 4. Young vs. old population - OLS and Stage 2 regression results

	Young				Old			
	OLS	Sargan IVs	Lung IVs	Heart IVs	OLS	Sargan IVs	Lung IVs	Heart IVs
(Intercept)	-23.97***	-38.64***	-36.08***	-33.43***	-112.90***	-140.24***	-149.10***	-131.33***
	(1.96)	(2.47)	(3.04)	(2.59)	(7.22)	(7.57)	(7.91)	(7.51)
Health	1.45***	4.59***	4.04***	3.48***	3.25***	6.96***	8.17***	5.75***
	(0.08)	(0.32)	(0.50)	(0.37)	(0.08)	(0.23)	(0.34)	(0.25)
Adj. R^2	0.35	0.33	0.34	0.34	0.45	0.43	0.40	0.44
F-test		659.88	433.77	608.71		1216.70	925.42	1236.87
Hausman		105.18	28.22	32.26		313.56	250.52	115.38
Sargan		13.26	23.35	27.85		1.84	8.67	99.53

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Sargan test statistics are presented at the bottom of Table 3. The null hypothesis that the instruments satisfy the exclusion restriction(2) is not rejected at the 10% significance level only for the first IV model. For the models with lung-related and heart-related instruments, the null hypothesis of the Sargan test is rejected at the 1% level. Thus, the instruments constructed to perform best in the Sargan test may have the strongest empirical appeal for satisfying the exclusion restriction. Furthermore, it is worth noting again that in the case of heterogenous effects, rejection of the Sargan test null hypothesis is an indication of heterogenous effects rather than a violation of the exclusion restriction. Thus, the results of the Sargan test should be considered with caution.

Table 4 presents estimates of the second-stage regressions for the three models, as well as OLS estimates, when modelling young respondents (at most 45 years of age) independently of old respondents (over 45 years of age). Depending on the instruments used, the estimated effect for young respondents ranges from 3.48 to 4.59 additional weekly hours per unit increase in Likert-scale self-rated health. For old respondents, the estimated effect ranges from 5.75 to 8.17—considerably greater than for young respondents. The OLS estimates for each subset are less than their IV counterparts, suggesting that the overall OLS bias is negative for both groups.

Under the assumptions of instrument relevance and instrument exogeneity, which are shown to have merit empirically and theoretically, interpretation of the estimated effect as causal is justified. Concerns of measurement error, justification bias, and simultaneity bias are addressed through the use of objective health measures as instruments for subjective health, making it possible to identify the effect of interest.

CONCLUSION

In this paper, the causal effect of self-rated health on weekly hours worked was estimated using National Health Interview Survey data from 2015 to 2018. Empirical tests suggested that the relationship was endogenous, and primary sources of endogeneity were likely to be measurement error, justification bias, and simultaneity bias. Endogeneity was addressed using a simultaneous equations model, instrumenting self-rated health with objective health measures. From a theoretical point of view, the objective health measures were likely robust to measurement error and justification bias, while being highly correlated with self-rated health, making them suitable instruments. Empirical tests confirmed instrument relevance, and provided evidence that the exclusion restriction was met for at least the first set of instruments presented. Failure of the empirical tests of the exclusion restriction for lung-related instruments and heart-related instruments did not necessarily indicate violation of the exclusion restriction, as heterogeneity of effects identified from individual instruments may have been the cause (Angrist and Imbens 1995; Blundell et al. 2017, Lemieux 2019). Instruments related to heart conditions may have had the strongest theoretical basis for satisfying the exclusion restriction.

A unit increase in Likert-scale self-rated health was predicted to increase weekly hours worked by about 5-7 hours, using data from respondents aged 21-67 years. This estimate was smaller for respondents of at most 45 years old (the median age in the sample) at about 3-4 hours. It was larger for respondents above the median age, at about 6-8 hours. This may have been observed because older individuals (especially those near retirement) may prioritize rest more than younger individuals in the middle of their careers, who may instead feel more incentivized to work despite poor health in order to advance in their careers, maintain relations with employers, or save money for retirement.

Although the effect of self-rated health driven by diagnosis of a COPD in particular was relatively large, the choice of instruments used did not affect the estimates greatly. IV estimates as a whole were greater than OLS estimates, suggesting that the OLS bias was negative. As justification bias is likely to bias estimates upwards, it may have been dominated by other forms of bias.

The findings of this study are limited by several factors. First, endogeneity of self-rated health is commonly addressed in health and labour supply literature using simultaneous equations models, but convincing intuitive arguments that the exclusion restriction is satisfied are rare. Most authors, myself included, use Sargan tests and explanations that objective health is unrelated to measurement error and justification bias (Blundell et al. 2017; Cai 2010), but ignore the possible direct causality between health conditions and labour supply. Second, it is conceivable that controlling for lagged health, lagged hours worked, or unobserved heterogeneity may alter estimates, so cross-sectional estimates such as those presented in this paper may be biased (Blundell et al. 2017). Reverse causality may contribute to this bias. Availability of panel data can allow future studies to control for lagged conditions and unobserved time-invariant characteristics of individuals, and offer protection against bias introduced by reverse causality, as long as the panel model is appropriately specified (Blundell et al. 2017; Leszczensky and Wolbring 2022). Third, the linear regression methods used in this study implicitly assume equal intervals between response options for Likert-scale self-rated health. That is, they assume for example that the increment from 1 (Poor) to 2 (Fair) is the same as the increment from 4 (Very good) to 5 (Excellent). Liddell and Kruschke (2018) note that while such metric models are routinely used to analyse Likert-scale data, this practice can lead to misinterpretations of the data, as the intervals between response options may not be equal in practice. Future design-based research on health should consider using binary measures of health or ordinal models, which do not rely on the assumption of equal intervals.

The Impact of Consumer Reviews on Product Sales: Evidence from Video Game Products on Amazon

Ting (Dongting) Guo

ECON 490: Seminar in Applied Economics

ABSTRACT

Consumer reviews are proven by most past literature to have an impact on product sales. This paper explores the impact of consumer reviews on product sales for video game products in particular, by using data extracted from Amazon and analyzing the relationship between review features and product Best Sellers Rank (BSR), an estimator of product sales. Several key findings emerge from analysis based on approximately 1.5 million reviews and 40,000 products. First, consumer reviews, especially ratings, have an impact on product sales: the higher the rating, the lower the product BSR and the higher the number of sales. More recent ratings have a larger impact than older ratings. In addition, product features are also found to have an impact on product sales, although the effects are not as large as ratings.

INTRODUCTION

Online retail is becoming increasingly prevalent in people's everyday life and is contributing a substantial portion to business revenue. In the United States, sales from e-commerce were $871 billion in 2021 and have been growing at 16% on average since 2011 (Wang, 2022). Online retail sales share out of total retail sales increased from 16% to 19% in 2020 due to Covid-19, further signifying the increasing importance of online retail (UNCTAD, 2021). Therefore, it is vital for businesses to understand the mechanisms behind consumers' online purchasing decisions in order to boost their online sales. Implemented as a common feature on most, if not all, of the e-commerce platforms globally, the consumer reviews feature has played a key role in helping consumers make buying decisions and in helping sellers increase online sales (Chen et al., 2022).

Many pieces of literature have proven the impact of consumer reviews on product sales. Research conducted by Consumer Focus finds that consumers commonly rely on others' reviews to guide their purchasing decisions (OECD, 2019). An analysis by Ahsan (2017) shows the correlation between consumer reviews and product sales. Sun et al. (2021) find that online reviews have a stronger influence on high-quality sellers than on low-quality sellers. Chevalier and Mayzlin (2006) analyze the data of books sold on Amazon and discover that an increase in a book's reviews leads to an increase in the relative number of sales for the site, and that one-star reviews have a greater impact than five-star reviews in most cases. For e-commerce in the fashion industry, Lohse, Kemper and Brettel (2017) find that positive online reviews have a greater impact on products with higher involvement. In the video games industry, Zhu and Zhang (2010) find that online reviews have a larger impact on less popular games.

However, several pieces of literature show different findings. In contrast to Sun et al. (2021)'s findings, Lohse, Kemper and Brettel (2017)'s study indicates that online reviews, both positive and negative, have a greater impact on weaker brands. Duan, Gu and Whinston (2008) show that ratings do not have a significant impact on movies' box office revenues. Chen, Wu and Yoon (2004) do not find product sales to be related to consumer ratings; however, they do find a positive association between sales and the number of reviews. Kim and Kim (2022) find that star ratings have an inverted U-shaped relationship with product sales; their results also show a positive relationship between rating volume and product sales.

Within this literature, there has been relatively little focus on the impact of customer reviews on product sales for video game products specifically. This paper will add value to the current literature by using video games data from Amazon. The hypothesis is that consumer reviews, especially ratings, will have an impact on product Best Sellers Rank (BSR), a rank that indicates how well a product is selling in the most recent period of time: a higher overall rating of a product will lead to a smaller product BSR, or higher product sales. A lower BSR implies a greater number of sales.

The data used in the analysis is obtained from the Amazon Review Data (2018) database which records product information and reviews information up until October 2018 (Ni, Li and McAuley, 2019). To explore the impact of consumer reviews on product sales, each product's number of sales is estimated by its Best Sellers Rank (BSR).

This paper applies natural language processing (NLP), a machine-learning technique which uses computers to study human language. A sentiment score is assigned to each review text using NLP, which measures how positive or how negative each review sounds to humans in a quantitative way. The empirical model evaluates the impact of each product's overall review sentiment on product sales. Not much literature in the research field uses this technique, and this paper adds value to the research area by using evidence from NLP.

In October 2016, Amazon banned incentivized reviews, which is an important milestone in its review system history. Before the ban, consumers were allowed to write reviews for products they purchased in exchange for a discounted or free product from product vendors. The original intention of this was to increase the frequency of ”honest” feedback such that consumers could get more product information from the reviews. However, research shows that incentivized reviews give generally higher ratings, which distorts the original purpose of the program (ReviewMeta, 2016). Amazon banned this program to increase review fairness, and this event is taken into account in the analysis. The results indicate that review features indeed have an impact on product BSR: the mean rating of reviews after the ban of incentivized reviews, total number of reviews, percentage of verified reviews, percentage of reviews after the ban, and the total number of review images together have the power to explain 30.1% of the variance in product BSR. All variables, except the percentage of reviews after the ban, have negative effects on product BSR: higher values in these variables leads to a lower BSR and hence, a higher product sales. The mean rating of reviews after the ban has the largest impact among all of the review features, providing support for my hypothesis

The empirical model is made more robust by conducting heterogeneity tests, where the impact of product features, such as brand popularity and number of product images, on product BSR are assessed in addition to review features. Product features are shown to have an impact on product sales, but the impact is smaller than the impact of consumer ratings. Higher product prices leads to lower sales, which partially explains vendors' motivation behind applying discounts on product prices.

The findings contribute to the existing literature by adding one more piece of evidence that consumer reviews indeed have an impact on product sales particularly for video game products. This suggests that businesses should pay attention to consumer reviews and react to consumer feedback promptly, since accumulating too many negative reviews in a recent period of time will likely affect the sales of a product negatively, which reduces business sales revenue for that period.

By understanding the factors that impact video game product sales, businesses in the video games industry can improve their business strategies. According to an online survey by Statista in 2022, 60% of video gamers in the United States are under the age of 34 (Clement, 2022b). As video game users are mostly young people, the finding of this paper can also partially imply how consumer reviews play a role in young people's purchasing decisions. Hence, businesses in other industries whose main retail platform is online and whose consumer age group skews young can also modify their business strategies based on the results of this paper.

The rest of the paper is structured as follows. Section I reviews past literature and provides background on Amazon's ranking system and its reviews system. Section II describes the data. Section III describes the empirical methodologies. Section IV presents the results. Section V discusses the findings. Finally, section VI concludes the entire paper and provides related implications.

BACKGROUND

A. Amazon Ranking System

Every product sold on Amazon is assigned a Best Sellers Rank (BSR) after it achieves one sale, and a lower BSR indicates higher sales. Each product can have multiple BSRs under different category levels. For instance, a product can hold a rank of 100 in the overarching video games category and a rank of 50 in the more specific PC games sub-category within the video games domain simultaneously. BSR is updated every hour and is largely based on a product's current sales tendency and its historical sales volume (Connolly, 2022).

Each product is also assigned an organic ranking, which is the rank it gets on Amazon's search engine results pages (SERP). A product's organic ranking does not necessarily reflect its BSR and its BSR does not necessarily reflect its organic ranking (Connolly, 2022).

B. Amazon Reviews System

Amazon has maintained a consumer reviews system since its inception. A few major changes happened to the reviews system before the latest date of the datasets used in this paper (October 2018), which includes the prohibition of incentivized reviews, the launch of the Early Reviewer Program, and the deactivation of a large number of accounts that are suspected members of underground review communities.

Although Amazon has long been prohibiting businesses to pay for fake reviews, it allowed incentivized reviews, where businesses offer free or discounted products to consumers to get ”honest” feedback, until this was banned in October 2016 (Perez, 2016). For incentivized reviews, reviewers write reviews of a product and in turn receive a large discount or free product from the vendor; all they need to do is disclose their affiliation with the vendors when writing the reviews. Although the original purpose of incentivized reviews was to encourage consumers to write truthful reviews that help other consumers make their purchasing decisions, in reality, incentivized reviews have been significantly biased toward higher ratings. An analysis conducted by ReviewMeta (2016) based on 7 million reviews observes that incentivized review texts are less critical than non-incentivized review texts, and incentivized ratings on average are 0.38 stars (out of 5 stars) higher than non-incentivized ratings, which strongly favors the vendors. By removing incentivized reviews, the reviews will likely reflect more honest opinions.

In February 2017, Amazon introduced the Early Reviewer Program, where shoppers are encouraged to write product reviews by receiving low-value gift cards in exchange (Masters, 2021). Unlike the incentivized review feature that was banned in October 2016, this program is only applicable to products with less than 5 reviews. This helps products with low sales to attract more customers by providing minimal consumer reviews, which mitigates the crackdown that the October 2016 policy change brought to the Amazon reviews system.

In April 2018, Amazon deactivated a large number of customer accounts that used the marketplace for commercial purposes. This was in effort to reduce underground reviews after the ban on incentivized reviews and make consumer reviews more credible.

It is important to account for the above changes in the analysis, as the removal of incentivized reviews could make reviews written after October 2016 more trustworthy to some consumers. Yet it is also important to keep in mind that incentivized reviews still exist after the ban under the table.

In addition, it is necessary to understand how product star rating is calculated. Amazon uses machine learning algorithms to detect fake reviews, and only authentic and verified[1] review ratings are used in calculating the star rating. Furthermore, more recent reviews are given more weight (Amazon, n.d.). Hence, the product star rating is different from the simple mean of all consumer ratings.

DATA

A. Data Sources

As an e-commerce giant, Amazon earned almost $470 billion USD in its net sales revenue in 2021, which makes it a good e-commerce platform to study (Coppola, 2022). This paper uses the Amazon Review Data (2018) provided by Ni, Li and McAuley (2019), which records reviews and product data on Amazon from May 1996 to October 2018. Amongst the 29 categories in the entire dataset, the video games category comprises 1.12% of all reviews and 0.56% of all products. Although it is not one of the largest categories in terms of review and product volumes on Amazon, Amazon is the most popular platform on which to purchase video games. According to the 2022 Statista Global Consumer Survey, a multi-selection survey, 44% of the 4,561 U.S. video gamers who have purchased digital video games made their purchases on Amazon within the past 12 months (Kunst, 2022). Furthermore, there are approximately 1.5 million reviews and 40,000 products available, which ensures that the sample size is large enough for analysis. Hence, it is convincing to analyze video game products and review data using data from Amazon even though the category does not make up a large portion of Amazon's overall product and review database. Lastly, Amazon has been an online seller from the very beginning and has always been online up until today, which eliminates the impact of store factors on product sales as well as the substitution effect between online sales and in-store sales.

Using data from the Amazon platform only removes e-commerce platform heterogeneity, since the user interface varies between different platforms, which may highlight different aspects of consumer reviews and result in different consumer purchasing decisions. In addition, different platforms also calculate product star ratings differently, which makes it hard to analyze the impact of star ratings on product

1 If a purchase is unverified, one needs to include more details in review texts for the review rating to get included in the star rating calculation

sales. Thus, it is the best choice to analyze data from just one platform.

Conducting analysis on video game data exclusively removes product category heterogeneity, since people's purchasing decisions can vary amongst different categories of products. For instance, consumer reviews may be less important to some consumers for necessities and/or very cheap products. Video games are not necessities, and their price range is also relatively small, which reduces heterogeneity.

This paper uses two datasets from the Amazon Review Data (2018). One dataset contains consumer reviews and the other contains product information; both datasets focus specifically on the category of video games and record information up until October 2018 (Ni, Li and McAuley, 2019). The two datasets are linked together by ASIN, Amazon Standard Identification Number which is assigned uniquely to each Amazon product.

The product data specifies the characteristics of each Amazon product under the video games category and includes information on product ASIN, name, price, brand, categories, BSR at different category levels, details, descriptions, technical specifications, images, high-resolution images, and other relevant products.

The reviews data specifies the details of each review of the products under the video games category and includes information on product ASIN, rating out of 5 stars, reviewer ID, review time, review summary, review text, whether the review is verified, number of votes, style of the product purchased, and images attached to the review.

B. Data Construction

15.22% of the product dataset are duplicate data, meaning that some products are recorded multiple times; these duplicates are removed. Products without a single review are also excluded. Products under accessories or console sub categories are excluded as the sales of accessories and consoles can be different from video games. Only the products that have a BSR at video games level are retained to ensure that the sales of all products in the analysis are compared at the same level. The above actions (including the removal of duplicates) remove 44,829 of 84,819 (52.85%) data points from the original product dataset, leaving approximately 40,000 data points remaining for analysis.

For the reviews data, 1,079,485 of 2,565,349 (42.08%) data points are excluded after removing reviews of products that are under accessories or console subcategories or do not have a rank in the video games level, leaving approximately 1.5 million data points remaining for analysis.

As summarized in Table 1, although the data cleaning process has removed a significant amount of data points from the original datasets, the process ensures that the final data to be used in the empirical model is clean, high-quality and more homogeneous. Moreover, the final data is still large in size despite the large exclusion of data points in percentage, hence the analysis based on the final data is still credible.

Table 1—: Data Cleaning Summary

Duplicate rows	12,908	15.22%
Products without BSR at video games level	10,296	12.14%
Products under accessories or console subcategories	21,623	25.49%
Products with no reviews	2	0.00%
Products remaining	**39,990**	**47.16%**
Total Products	84,819	100%
Reviews whose products do not have BSR at video games level, or are under accessories or console subcategories	1,079,485	42.08%
Reviews remaining	**1,485,864**	**57.92%**
Total Reviews	2,565,349	100%

Note: The table summarizes the data cleaning process for the product data and review data respectively. After data cleaning, around 40,000 products and 1.5 million reviews remain for analysis in Section III.

Natural language processing, a field of machine learning which focuses on linguistics, is used to analyze the sentiment of each review text in a quantitative way. This paper uses VADER (Valence Aware Dictionary for Sentiment Reasoning), a text sentiment analysis model that breaks down sentences into individual words and analyzes the lexical features of each word based on their semantics and suppositions in the sentence (Bajaj, 2021). Lexical features include sentiment polarity (positive, negative, or neutral) and sentiment intensity. The lexical features of individual words are combined together to produce an overall sentiment indicator of the texts. Each text is assigned a sentiment intensity score that ranges from -1 to 1, where a positive score indicates a positive valence, a negative score indicates a negative valence, and a score of 0 indicates neutral sentiment. The larger the magnitude of the sentiment score, the stronger the text is toward that valence; a sentiment score of 1 indicates absolutely positive sentiment and a sentiment score of -1 indicates absolutely negative sentiment. This paper uses the Natural Language Processing Tool Kit (NLTK) library in Python to perform sentiment analysis to each review text. The sentiment score not only quantifies the emotion of each review text, but also serves as an alternative measurement to review rating when analyzing the impact of reviews on product sales.

The Unscaled Section of Table 2 shows the summary statistics of the product features, aggregate review features, and product Best Sellers Rank at the video games level. Reviews features are aggregated by product ASIN; features include mean rating (MeanRating), mean sentiment intensity score (MeanSentiment), number of reviews (NumReviews), percentage of verified reviews (VerifiedPercentage), percentage of reviews after the ban of incentivized reviews (AfterBanPercentage), and total number of review images (TotalNumReviewImages). In addition, the meaning rating/ sentiment score of verified reviews, of reviews after the ban of incentivized reviews, and of verified reviews after the ban, are also computed to serve as alternative mean rating/sentiment values; this is because they may have a stronger impact on BSR than the rating/sentiment of all reviews. If, for instance, a product does not have any reviews that are written after the ban of incentivized reviews (October 1, 2016), then that product has a mean after-the-ban rating (MeanAfterBanRating) of 0. Product features include brand popularity (BrandPopularity), number of images (NumImages), number of high-resolution images (NumHighResImages), number of details (NumDetails), length of product description (DescriptionLength), and length of product feature description (FeatureLength). The popularity of product brand is calculated as the number of occurrences that each brand appears in the entire product dataset. For instance, if 20 out of the 39,990 products belong to the brand ABC, then each product under that brand is assigned a brand popularity of 20.

Product BSR at video games level serves as the dependent variable for the empirical model in Section III, and the aggregate review features are the independent

Table 2—: Summary Statistics

		Unscaled				Scaled			
		mean	std	min	max	mean	std	min	max
Dependent Variable	RankVideoGames[1]	66102.66	49716.75	2.00	241788.00	66102.66	49716.75	2.00	241788.00
Review Feature Variables	MeanRating	3.85	1.01	1.00	5.00	100.00	26.24	25.98	129.88
	MeanVerifiedRating	3.61	1.48	0.00	5.00	100.00	41.01	0.00	138.68
	MeanAfterBanRating	2.06	2.16	0.00	5.00	100.00	104.80	0.00	242.30
	MeanVerifiedAfterBan-Rating	2.03	2.17	0.00	5.00	100.00	107.09	0.00	246.83
	MeanSentiment	0.48	0.33	-1.00	1.00	100.00	67.26	-206.20	206.25
	MeanVerifiedSentiment	0.43	0.34	-1.00	1.00	100.00	77.37	-229.88	230.25
	MeanAfterBanSentiment	0.21	0.33	-1.00	1.00	100.00	152.24	-466.04	465.81
	MeanVerifiedAfterBan-Sentiment	0.21	0.32	-1.00	1.00	100.00	153.91	-475.20	475.72
	NumReviews	37.16	132.28	1.00	7630.00	100.00	356.02	2.69	20535.10
	VerifiedPercentage	0.67	0.33	0.00	1.00	100.00	48.61	0.00	148.87
	AfterBanPercentage	0.15	0.27	0.00	1.00	100.00	174.66	0.00	650.97
	TotalNumReviewImages	0.38	2.29	0.00	100.00	100.00	598.13	0.00	26138.96
Product Feature Variables	BrandPopularity	632.34	791.13	0.00	2387.00	100.00	125.11	0.00	377.48
	NumImages	3.04	3.09	0.00	14.00	100.00	101.62	0.00	460.95
	NumHighResImages	3.04	3.09	0.00	14.00	100.00	101.62	0.00	460.95
	NumDetails	0.01	0.21	0.00	7.00	100.00	2628.47	0.00	86665.63
	FeatureLength	264.13	356.85	0.00	7993.00	100.00	135.10	0.00	3026.18
	DescriptionLength	878.02	1242.89	0.00	18788.00	100.00	141.56	0.00	2139.81

Note: The table describes the summary statistics for all variables used in the empirical analysis, both before and after variable standardization. All review feature variables ($Feature_{rp}$'s in Equation 1) and product feature variables ($Feature_{ap}$'s in Equation 2) are standardized to the same mean value of 100 so that one can compare their magnitudes of coefficients to identify the variable that has the largest impact on product BSR.
1: $RankVideoGames$ represents the Best Sellers Rank (BSR) at video games level. It is the dependent variable in the empirical models. Unlike the independent variables, the BSR values are not scaled.

variables. Product features such as brand popularity are used in heterogeneity tests to increase model robustness.

C. Sample Description

Table 3 describes the rating distribution of the entire sample, verified vs. unverified purchases, as well as before vs. after the ban of incentivized reviews (October 1, 2016). The first four groups, as shown in columns (1) to (4) in the table, all follow the same rating distribution where 5-star rating reviews comprise the largest portion of the sample, followed by 4-star, 1-star, 3-star, and last 2-star ratings. The only exception is the sample of reviews after the ban of incentivized reviews, as indicated by column (5) in the table, whose 1-star ratings make up a larger portion than 4-star ratings. Rating distribution has become more extreme after the ban on incentivized reviews, where the proportion of 5-star and 1-star ratings both increased compared to before the ban. Verified reviews comprise a larger percentage of 5-star ratings and a smaller percentage of 1-star ratings than unverified reviews.

Figure 1 shows the relationship between the mean rating and number of reviews of each product after dividing products based on their verification status.

For products with less than ~ 300 reviews, review verification status does not play an important role in their mean rating. For products with a large number of reviews, verified reviews tend to lead to a higher mean rating. This suggests a

Table 3—: Rating Distribution

	(1)	(2)	(3)	(4)	(5)
		Verification Status		Before or After Ban	
	Full Sample	Verified	Unverified	Before	After
1 star	10.99%	9.37%	14.48%	10.63%	13.08%
2 stars	5.12%	4.04%	7.45%	5.31%	4.00%
3 stars	8.41%	7.54%	10.28%	8.77%	6.28%
4 stars	16.73%	15.12%	20.20%	17.68%	11.10%
5 stars	58.76%	63.93%	47.60%	57.61%	65.53%
Sample Size	1,485,864	1,015,507	470,357	1,269,926	215,938

Note: This table reports the rating distribution of all reviews in which their corresponding product is not under accessories or console subcategories and has a rank at video games level. Columns (2) and (3) restrict the sample based on review verification status, and columns (4) and (5) restrict the sample based on whether the review is written before or after the ban of incentivized reviews (October 1, 2016).

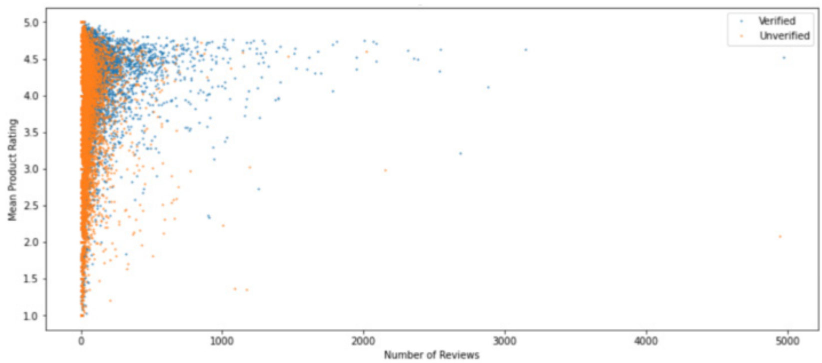


Figure 1. : Mean Product Rating vs. Number of Reviews

Note: This scatterplot shows the relationship between the mean rating and number of reviews of each product after dividing products based on their verification status.

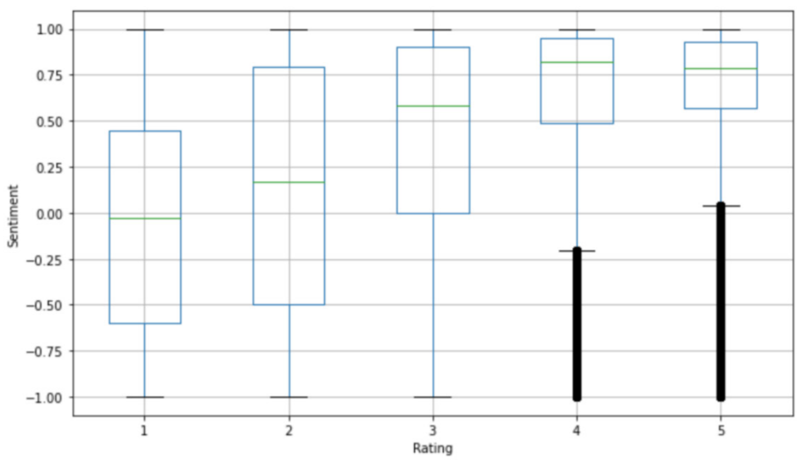


Figure 2. : Boxplot of Sentiment Score by Review Rating

Note: The boxplots show the distribution of the sentiment intensity score of each review, grouped by review ratings.

possibility to incorporate the combined effects of verification status and the number of reviews into the empirical analysis.

Figure 2 displays the distribution of sentiment intensity score of each review, grouped by review ratings. In general, sentiment score and rating show a positive correlation with each other. This correlation is stronger for lower ratings. On the other hand, sentiment scores show a similar distribution for 4-star and 5- star ratings, and it does not necessarily reflect the review rating. For instance, some reviewers show very negative sentiments in the review texts, yet give a 4- or 5-star rating at the same time. This can be explained by the inaccuracy of the sentiment analyzer (e.g.: the analyzer fails to detect sarcasm in the reviews) and/or the personality of reviewers (e.g.: some people tend to not give a low rating despite dissatisfaction with products). The above observations suggest that sentiment scores can be considered as an alternative variable for product rating. Although it has a positive correlation with review rating, the correlation is far from perfect; it might have a larger impact on BSR than product rating, which makes it worth comparing the impact of these two metrics, holding all else constant.

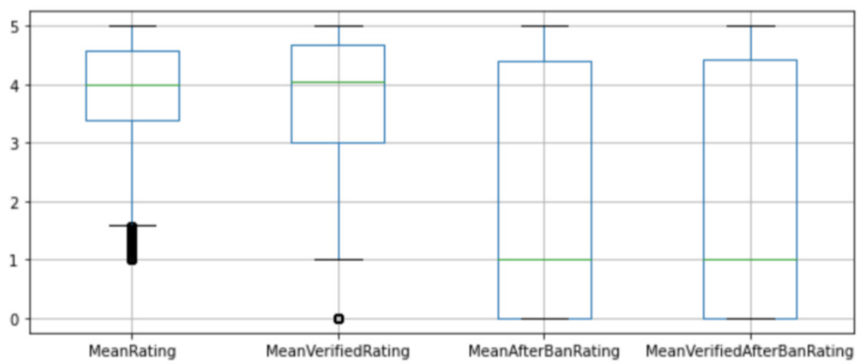


Figure 3. : Boxplots of Different Mean Ratings

Note: The boxplots show the distribution of mean rating of all reviews, mean rating of all verified reviews, mean rating of all reviews after the ban of incentivized reviews, and mean rating of all verified reviews after the ban of incentivized reviews, respectively. All means are aggregated by product.

Figure 3 shows the distribution of mean rating of all reviews, mean rating of all verified reviews, mean rating of all reviews after the ban of incentivized reviews, and mean rating of all verified reviews after the ban of incentivized reviews, all aggregated by product. Compared to the mean rating of all reviews, the mean rating of verified reviews shows a similar distribution, but the mean rating of reviews after the ban on incentivized reviews and that of verified reviews after the ban show a very different distribution, where the 75th percentile values are similar, but the median and the 25th percentile values are very low. This is because a large number of products do not have any reviews written after the ban of incentivized reviews, which results in their mean rating being 0 when conditioned on the reviews after the ban. As alternative means show different distributions from the simple mean of all reviews, it is worth testing which mean value impacts BSR the most.

EMPIRICAL METHODOLOGY

This paper uses OLS regression models to estimate the impact of a product's aggregate consumer review features, such as mean rating and total number of reviews, on its Best Sellers Rank (BSR) at the video games level.

The exact number of sales of a product is only available to the product vendor and is unavailable to the public. Hence, a proxy is needed for the product number of sales and it needs to be publicly available. Product Best Sellers Rank (BSR) serves as a good proxy, since BSR is publicly available under each product's Product Details section, is recorded in the datasets, and is determined largely by a product's current sales trend (Connolly, 2022). Several websites provide calculators that estimate the number of sales of an Amazon product based on its BSR at different category levels (Jungle Scout, n.d.). Unfortunately, these services are costly, hence product sales are estimated using product BSR, and in particular, using product BSR at the video games level. This ensures that the products are compared at the same level.

An advantage of using data from video game products is that it has a small number of returning consumers. Unlike grocery products such as tissue paper, video games are likely to be purchased once only for each consumer, which reduces the probability that a product's sales are constantly from the same group of repeat consumers without new consumers. Having repeat consumers can underestimate the impact of reviews on product sales, since repeat consumers will likely purchase products based on their user experience rather than other people's reviews. As video game products have a lower probability of getting repeat consumers, the impact of consumer reviews is likely not decreased by them.

A. Basic Model Setup

The general linear regression model is as follows:

$$BSR_p = \sum_{rp} \beta_{rp} Feature_{rp} + \epsilon_p \tag{1}$$

where BSR_p is the dependent variable which indicates the BSR of product *p* at video games level, $Feature_{rp}$ is the aggregate value for product p for review feature *r* (e.g.: mean rating, mean sentiment of verified reviews, number of reviews, percentage of verified reviews, etc.), β_{rp} is the coefficient of $Feature_{rp}$, and ϵ_p is the error term.

A lower BSR represents a higher number of sales, hence a very negative coefficient would indicate a large positive impact on the number of sales of a product. The hypothesis is that rating or sentiment will have a positive impact on product sales, meaning that their coefficients are hypothesized to be negative.

Values for $Feature_{rp}$ are standardized so that all of them share the same mean. This allows one to identify the aggregate review feature that has the largest impact on BSR_p, which is the $Feature_{rp}$ with the largest magnitude of the coefficients. In this model, all independent variables are standardized so that they share the same mean value of 100. Table 2 shows description of $Feature_{rp}$ variables and their summary statistics before and after the standardization; summary statistics before the standardization are under the Unscaled columns and summary statistics after the standardization are under the Scaled columns.

B. Heterogeneity Model Setup

This section explores heterogeneous factors that might have an impact on BSR by testing the effects of product attributes. The causal relationship between reviews and product sales is complex and can be influenced by different factors. By including various proxies for a product's quality (e.g.: product details, images, etc.), the model might get an estimate that is closer to the causal effect of reviews. The regression equation after adding heterogeneous factors is as follows:

$$BSR_p = \sum_{rp} \beta_{rp} Feature_{rp} + \sum_{ap} \beta_{ap} Feature_{ap} + \epsilon_p \tag{2}$$

where *Feature*$_{rp}$ is the value of product *p*'s product feature *a* (e.g.: number of images, brand popularity, etc.), and β_{ap} is the coefficient of *Feature*$_{rp}$. Again, all of *Feature*$_{rp}$ and *Feature*$_{ap}$ are standardized so that they all share the same mean (100 for this model), to assess the variable with the largest impact on BSR_p. The summary statistics of the variables are again shown in Table 2.

C. Multicollinearity and Variable Selection

OLS Technique requires variables to have no perfect multicollinearity since it will result in a large variance in the results. Hence, correlation and multicollinearity tests are conducted before determining the independent variables used in both the basic model and heterogeneity model.

First, the correlations between different variables are computed to detect pair wise collinearity. The correlation matrix in absolute values is shown in Figure 4. A larger absolute correlation coefficient (cells with darker backgrounds) indicates a stronger pairwise correlation, and a smaller absolute correlation coefficient (cells with lighter backgrounds) indicates a weaker pairwise correlation.

Most of the pairs are weakly correlated to each other, and most variables do not show a strong correlation to BSR (*RankVideoGames* in the heatmap). The mean rating of reviews after the ban of incentivized reviews (*MeanAfterBanRating* in the heatmap) is moderately correlated to BSR with an absolute correlation of 0.5, holds a stronger correlation strength compared to the mean of all ratings (*MeanRating* in the heatmap), and is in fact the variable with the strongest correlation to BSR out of all the variables. This suggests that the mean rating for reviews after the ban will likely fit the model better than the mean rating for all reviews.

Some pairs show a very strong correlation. For instance, a product's number of images (*NumImages*) and number of high-resolution images (*NumHighRes- Images* have an absolute correlation of 1. This suggests that the two variables should not be included in the same model simultaneously.

Multicollinearity checks are also performed to ensure the independent variables selected for the model show limited multicollinearity. Variance inflation factors (VIF) are used to measure multicollinearity. A VIF greater than 10 indicates strong multicollinearity, and at least one variable needs to be removed from the list of independent variables to ensure that the OLS assumption is satisfied.

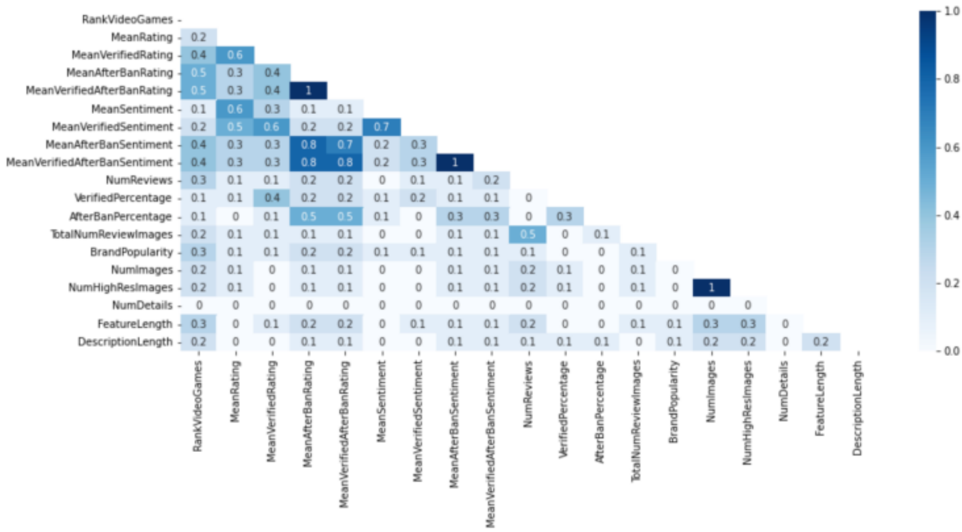


Figure 4. : Correlation Heatmap of Aggregate Review Features

Note: The heatmap shows pairwise correlations between different pairs of product features. All correlation coefficients are in absolute values. A larger absolute correlation coefficient (cells with darker backgrounds) indicates a stronger pairwise correlation, and a smaller absolute correlation coefficient (cells with lighter backgrounds) indicates a weaker pairwise correlation.

The results of multicollinearity tests of variables are shown in Table 4. In Table (a), *MeanRating* has a VIF larger than 10, which indicates strong multicollinearity between the set of variables. After removing *MeanSentiment*, the VIF of the rest of the variables are shown in Table (b); no VIF exceeds 10, which suggests limited multicollinearity between the variables, hence these variables can be used in the same regression model simultaneously.

Both the pairwise correlation and multicollinearity tests are performed based on non-standardized data, and only the regression model uses standardized data (i.e.: all independent variables are standardized to a mean value of 100 and the dependent variable stays unchanged).

RESULTS

Multicollinearity tests are performed over different combinations of variables and models that regress on those combinations are compared against each other in terms of R^2.

A. Basic Model

The review features of a product include its total number of reviews, percentage of verified reviews, percentage of reviews after the ban on incentivized reviews, total number of review images, as well as an overall rating/sentiment score. The choice of the overall rating/sentiment score is one of the mean value (rating or sentiment) of all reviews, mean value of all verified reviews, mean value of all reviews after the ban, and mean value of all verified reviews after the ban.

Table 4—: VIF of Sample Sets of Variables

(a) Significant Multicollinearity

Variables	VIF
RankVideoGames	2.42
MeanRating	10.88
MeanSentiment	4.76
NumReviews	1.45
VerifiedPercentage	5.04
AfterBanPercentage	1.50
TotalNumReviewImages	1.34

(b) Limited Multicollinearity

Variables	VIF
RankVideoGames	2.41
MeanRating	5.29
NumReviews	1.45
VerifiedPercentage	4.93
AfterBanPercentage	1.49
TotalNumReviewImages	1.34

Note: The tables show the variance inflation factor (VIF) values of two sample sets of variables to check multicollinearity of each set of variables. The variables in Table (a) are strongly multicollinear since *MeanRating* has a VIF larger than 10. The variables in Table (b) are limited in multicollinearity since all variables have a VIF less than 10.

The regression model that achieves the highest R^2 consists of the following independent variables: total number of reviews, percentage of verified reviews, percentage of reviews after the ban of incentivized reviews, total number of review images, and mean rating of all reviews after the ban. Compared to the most ordinary model that uses the mean rating of all reviews and has an R^2 of 0.122, the best-performing model gives an R^2 of 0.301, which significantly increases the power of the independent variables in explaining the variance in BSR at the Video Games level.

The coefficients of the best-performing basic model are summarized in Table 5. All coefficients, except the percentage of reviews after the ban (*AfterBan - Percentage*), have a negative coefficient, which indicates a negative relation ship with product ranking and hence, a positive relationship with product sales. The mean of all reviews after the ban, or *MeanAfterBanRating*, displays the strongest magnitude. As all of the independent variables are standardized to the same mean value, this suggests that *MeanAfterBanRating* has the largest impact on product sales. On the other hand, the total number of review images, or *TotalNumReviewImages*, exhibits a very small, although non-zero, impact on BSR, which implies that it does not impact product sales much.

An R^2 of 0.301 implies that the review features altogether play a non-trivial role in explaining product sales. This result supports the hypothesis, which predicts a positive impact of consumer reviews, especially consumer ratings, on product sales, or a negative impact on product ranking.

Table 5—: Basic Model Coefficient Summary

	Coefficient	Standard Error	2.5% CI	97.5% CI
MeanAfterBanRating	-258.06	2.35	-262.66	-253.46
NumReviews	-18.26	0.68	-19.59	-16.92
VerifiedPercentage	-54.69	4.50	-63.50	-45.87
AfterBanPercentage	45.84	1.43	43.04	48.64
TotalNumReviewImages	-2.39	0.40	-3.17	-1.61

Note: This table summarizes the coefficients of the independent variables of the best-performing basic model, as well as coefficient standard errors and 2.5% and 97.5% confidence interval values. All of the independent variables are standardized to the same mean value of 100 before conducting linear regression using the OLS method to ensure that the magnitudes of coefficients are unbiased for comparison.

B. Heterogeneity Model

The heterogeneity model adds heterogeneous factors, the product features of each product, to the basic model. Heterogeneous factors include brand popular ity (measured by the number of times the brand appears out of all products), number of high-resolution images, number of details, length of product features, and length of product description. VIF checks are done and no significant multi collinearity is found.

The coefficient summary of the heterogeneity model is shown in Table 6. The explanatory power of review features is diluted as additional variables with non zero coefficients are added in. Despite that, the mean rating of all reviews after the ban of incentivized reviews still appears to be the variable with the largest impact on product ranking. Brand popularity, or BrandCount, has the strongest impact out of the five product feature factors, and all product features except number of details (*NumDetails*) show a negative impact on product BSR, or a positive impact on product sales; number of details of a product exhibits a negligible impact.

Table 6—: Heterogeneity Model Coefficient Summary

	Coefficient	Standard Error	2.5% CI	97.5% CI
MeanAfterBanRating	-219.34	2.34	-223.93	-214.75
NumReviews	-11.51	0.67	-12.81	-10.20
VerifiedPercentage	-65.30	4.36	-73.84	-56.77
AfterBanPercentage	32.71	1.39	29.98	35.43
TotalNumReviewImages	-2.62	0.38	-3.37	-1.88
BrandCount	-61.84	1.65	-65.06	-58.61
NumHighResImages	-23.11	2.07	-27.18	-19.05
NumDetails	-0.05	0.08	-0.20	0.10
LenFeature	-49.38	1.59	-52.50	-46.26
LenDescription	-24.76	1.50	-27.69	-21.83

Note: This table summarizes the coefficients of the independent variables of the heterogeneity model, as well as coefficient standard errors and 2.5% and 97.5% confidence interval values. All of the independent variables are standardized to the same mean value of 100 before conducting linear regression using the OLS method to ensure that the magnitudes of coefficients are unbiased for comparison.

The R^2 of the heterogeneity model is 0.358, a small increase from the basic model. A regression model that explains BSR using only the product features is developed as a supplement, where the model formula is displayed in Equation 3. The R^2 of the product features-only model is 0.165, a much smaller R^2 compared to the heterogeneity model.

$$BSR_p = \sum_{ap} \beta_{ap} Feature_{ap} + \epsilon_p \tag{3}$$

The comparisons between R^2 of different models, along with the analysis of coefficients, indicate that although product features do have some impact on product sales, the impact overall is not large compared to review features. The impact of the product price on product sales is also analyzed.

Only 6,632 (16.58%) products have a valid price in the products dataset, so it is inappropriate to include product price in the heterogeneity model. To test the effect of price, the subset of products that have a valid price is selected and a comparison between the heterogeneity model without the presence of price and with the presence of price is made. The heterogeneity model without the presence of price has an R^2

of 0.281 and the heterogeneity model with the presence of price has an R^2 of 0.286. The coefficient of the price variable is 19.99, which indicates a positive impact on product BSR and thus a negative impact on product sales. This is expected since consumers are expected to be more willing to purchase cheaper products.

DISCUSSIONS

A. Explaining Mechanisms Behind the Results

The dataset does not record product star ratings, and due to the black box algorithm of Amazon star ratings, the star ratings that actually appear on each product page cannot be fully replicated. Different measures of mean rating and mean sentiment are estimates only, which may overstate or understate the actual product star rating. The lack of information on product star ratings makes it harder to analyze, which may result in either overestimation or underestimation of the impact of consumer ratings on product sales.

Keeping all other independent variables constant, the basic model that in cludes *MeanAfterBanRating* shows a greater explanatory power on product ranking than the basic model that includes *MeanRating*. This may be related to review recency: all reviews prior to the ban of incentivized reviews (October 2016) are relatively old compared to when the data was extracted (October 2018). If a product has a large number of reviews after the ban, consumers may not scroll far enough to see the reviews before the ban. If a product does not have a large number of reviews after the ban, consumers may consider the prod uct as relatively unpopular or of low quality, which may hinder their decision to purchase. Both of these possibilities suggest that the reviews written before the ban play a relatively trivial role on product ranking.

This result has implications for vendors: vendors should pay attention to con sumer reviews, especially recent reviews, since they can influence product sales in the short run. For example, if many recent reviews of a product complain about a product's packaging, the vendor should check out the problem with the packag ing of the product and develop solutions to fix the problem before the product's overall rating starts declining, which will likely lead to a decrease in product sales. The importance of customer reviews also suggests that businesses could increase the number of employees in the customer aftercare team to make sure customer feedback is received promptly and the business can adjust strategies to better suit consumers' needs.

The basic model which includes *MeanVerifiedAferBanRating*, the mean rating of all verified reviews after the ban, shows a slightly smaller R^2 of 0.300 than the basic model that includes *MeanAfterBanRating*, which has an R^2 of 0.301. This could be due to the following reasons: (1) Consumers do not care much about whether a review is verified or not; (2) There is only a very small number of unverified reviews after the ban; (3) The unverified reviews show a similar rating distribution as verified reviews.

For the basic model, different means of sentiment scores are not as good an indicator of the product BSR as means of ratings, suggesting that consumers care more about the overall rating than reading the actual review texts. Indeed, a quantified number is often more representative than texts and easier to un derstand for people who do not want to read long lines of reviews. It is also possible that reviews are written in a language that one does not understand, which makes it impossible for consumers to make purchase decisions based on review texts. Mean sentiment score is not included in the model since adding it to the model would increase VIF significantly due to multicollinearity. However, this also suggests that both rating numbers and review texts together have a joint effect on product sales.

Both the basic model and the heterogeneity model show that *MeanAfterBan- Rating* has the largest impact on product BSR, hence it has a greater effect than the total number of reviews. This indicates that the overall rating is more impor tant than the number of reviews. Both models also show that the total number of review images, or *TotalNumReviewImages,* has a very small impact on product sales. This may be explained by the video games category being the focus of this paper. Consumers can access game demonstrations and gameplay videos from the internet, which reduces the strength of consumer review images in impacting product sales. However, the impact of the number of review images can be very different for products under other categories. Further research can compare the impact of the number of review images for various categories of products.

In the heterogeneity model, a subset of the data that contains a valid price is used to test the impact of price on product sales. Although this subset does not make up a large percentage of the entire dataset, its size (6632) is still relatively large, which makes the result credible. The result shows that price does have a negative impact, although not too large, on product sales. This can be explained by the observation that the price range of video game products, which is shown in Figure 5, is not large. Most (over 75%) of the products have a price below $200, and half of the products have a price below $100. Nevertheless, the negative impact of price on product sales suggests that it is worthwhile for merchants to consider applying price discounts to their products if they want to increase the number of sales, depending on the current sales and price of the product. An important consideration is the trade-off between the increase in the number of sales and the decrease in the profit earned per product sold.

These findings can be used to explain why businesses are paying expensive advertisement fees to platforms and/or influencers to promote their products: advertising increases product exposure and encourages product reviews amongst consumers. By asking influencers to give positive feedback on their products, busi nesses can spur positive reviews, which leads to an increase in product sales. The cost of advertisement is lower than the revenue earned from the advertisements, which explains the motivation behind this action.

The findings also explain the reasons behind Amazon's decision to ban incentivized reviews and the continuing existence of underground review communities after the ban. Before the ban, sellers well understood the importance of consumer reviews, hence they were willing to offer discounted or free products in exchange for likely higher reviews, which could lead to an increase in sales. Understanding the unfairness created by this program, Amazon banned it to promote more hon est feedback. After the ban, some sellers were willing to illegally pay for positive reviews to boost their sales, which is why underground review communities still exist. In response to this, Amazon kept on detecting possibly fake reviews and hence in April 2018, it deactivated a large number of customer accounts deemed suspicious of using the marketplace for commercial purposes.

B. The Impact of Other Variables on Product Sales

The heterogeneity model result indicates that only 35.8% of the variance in product BSR is explained by review

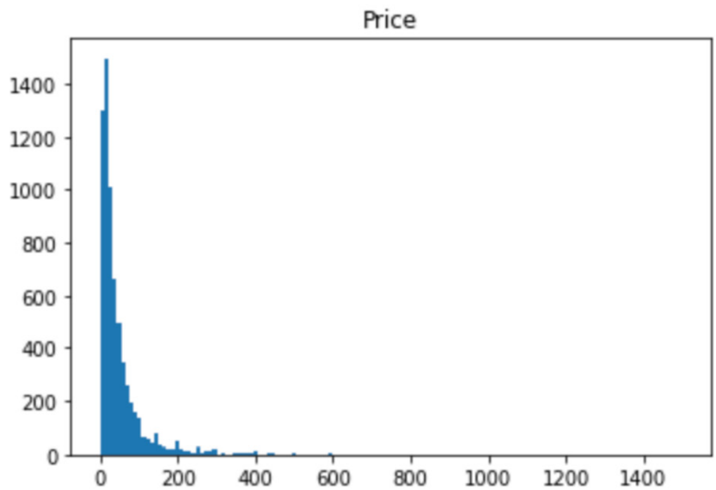


Figure 5. : Boxplot of Product Prices

Note: The histogram shows the distribution of product prices.

features and product features. This is not surprising since other, more external factors are expected to influence consumer purchasing decisions. The information on reviews and products is also not complete.

The datasets do not record whether the product prices are at discount or not, and the most recent price and the date the product is sold at that price. The price of the product that each review is based on is also unknown, hence the prices recorded in the datasets are only an estimate of the true prices. The shipping fee for each product is also unknown. The hypothesis is that the presence of shipping will lower product sales. Further research can investigate the impact of (the presence of) shipping fees on product sales.

As mentioned in Section I, each product has a Best Sellers Rank (BSR) and organic ranking. The hypothesis is that organic ranking will have a positive impact on product sales. In other words, the more a product appears at the front line on Amazon's search engine results pages, the more likely it will get higher sales, when keeping all other variables constant. The dataset does not record product organic rankings, hence the impact of organic ranking is unknown.

External factors such as consumer socioeconomic status can also play a huge factor. The purchasing power of consumers at a higher socioeconomic status is higher, which allows them to buy more video games at higher prices with less hesitation.

The gender of the consumer can also impact their purchasing decisions. For instance, Chen et al. (2022) find that female consumers, compared to male con sumers, are more affected by negative reviews than positive reviews. The reviews dataset does not record information on reviewer gender; using natural language processing to identify reviewer gender based on their name is not a good alterna tive, since people of different genders can share the same name and some reviewers do not use their real names, making it impossible for computers to identity re viewer gender. A controlled experiment is suggested if one aims to explore the impact of consumer reviews on sales by different genders. In addition, social media reviews and reviews by friends and family can also play a role in product sales, since they are also reviews just like the reviews written on Amazon.

Aside from the above factors, the region of each product may also be a useful factor. People from different countries hold different review evalu ation standards; Amazon also owns a different percentage of market share and faces different competitors. It would be valuable to analyze the degree of impact that reviews have on video game buying decisions in different countries, or even different regions of countries.

The analysis is based on the data from Amazon only, and results based on data from other platforms are unknown. Future research can focus on the analysis on another platform and compare the results from this paper to investigate the impact of reviews on various platforms.

As mentioned in Section I, product BSR is updated every hour to reflect its most recent sales. Hence, this paper analyzes the impact on product sales at a specific point in time only. Future research can consider keeping track of product BSR over time and performing a time series analysis to reach a more holistic conclusion.

In 2020, people's lives were significantly impacted by the outbreak of the Covid-19 coronavirus. People were forced to stay home during the lockdown, which boosted video game sales as people tried to entertain themselves without leaving home. In a survey, European gamers report that video games make them feel less detached and happier overall (Clement, 2022a). According to Statista, digital gaming sales on in-game content increased by 12% and the sales on paid downloads increased by 21% globally in 2020 (Clement, 2022a). The statistics indicate that it is worthwhile to study the impact of Covid-19 on product sales. However, the dataset only records data up to October 2018, which is long before the onset of Covid-19, hence the impact of Covid-19 is untestable. Nonetheless, it is possible to analyze the role of Covid-19 in consumer reviews and sales when sufficient data becomes available.

Finally, it is important to recognize that people extremely satisfied or ex tremely dissatisfied are more likely to begin word-of-mouth reviews, which could influence the purchasing decisions of people close to them (Anderson, 1998). This also makes online reviews biased.

CONCLUSION

Consumer reviews in general have an impact on product sales and explain ap proximately 30.1% of a product's Best Sellers Rank. Consumer ratings have the largest impact amongst all consumer review features: the higher the rating, the lower the product BSR, which indicates higher product sales. Ratings after the ban of incentivized reviews have a larger impact than the ratings of all reviews. While product features also have an impact, the impact is not as large as the impact of ratings; review features and product features altogether explains ap proximately 35.8% of product BSR. Altogether, product brand has a positive impact on sales. On the other hand, product price can influence sales negatively: the higher the price, the lower the product sales.

III

REFERENCES AND APPENDICES

Contents

REFERENCES AND APPENDICES

Body Shape and Career Outcomes in Entry-level Jobs: The Study of South Korean Young Adults

Mea Srisan

ECON 490: Seminar in Applied Economics

A. REFERENCES

Biddle, B. J., & Hamermesh, D. S. (1998). Beauty, Productivity and Discrimination: Lawyers' Looks and Lucre. Journal of Labor Economics, 16(1), 172–201. https://doi.org/10.3386/w5366

Brown, C. (1980). Equalizing Differences in the Labor Market. The Quarterly Journal of Economics, 94(1), 113. https://doi.org/10.2307/1884607

Brownell, K. D. (2001). Bias, Discrimination, and Obesity. Obesity Research, 9(12), 788–805. https://doi.org/10.1038/oby.2001.108

Carr, D., & Friedman, M. A. (2005). Is Obesity Stigmatizing? Body Weight, Perceived Discrimination, and Psychological Well-Being in the United States. Journal of Health and Social Behavior, 46(3), 244–259. https://doi.org/10.1177/002214650504600303

Clifford, M. M., & Walster, E. (1973). The Effect of Physical Attractiveness on Teacher Expectations. Sociology of Education, 46(2), 248. https://doi.org/10.2307/2112099

Efrain, M. G., & Patterson, E. W. J. (1974). Voters vote beautiful: The effect of physical appearance on a national election. Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement, 6(4), 352–356. https://doi.org/10.1037/h0081881

Fisher, M. L., & Voracek, M. (2006). The shape of beauty: determinants of female physical attractiveness. Journal of Cosmetic Dermatology, 5(2), 190–194. https://doi.org/10.1111/j.1473-2165.2006.00249.x

Flint, S. W., Čadek, M., Codreanu, S. C., Ivić, V., Zomer, C., & Gomoiu, A. (2016). Obesity Discrimination in the Recruitment Process: "You're Not Hired!" Frontiers in Psychology, 7. https://doi.org/10.3389/fpsyg.2016.00647

Frieze, I. H., Olson, J. E., & Russell, J. (1991). Attractiveness and Income for Men and Women in Management. Journal of Applied Social Psychology, 21(13), 1039–1057. https://doi.org/10.1111/j.1559-1816.1991.tb00458.x

Fuller, J. B., & Raman, M. (2017). Dismissed by Degrees: How Degree Inflation Is Undermining U.S. Competitiveness and Hurting America's Middle Class. Harvard Business School.

Hamermesh, D. S., & Biddle, J. E. (1994). Beauty and the Labor Market. The American Economic Review, 84(5), 1174–1194.

Jencks, C., Perman, L., & Rainwater, L. (1988). What Is a Good Job? A New Measure of Labor-Market Success. American Journal of Sociology, 93(6), 1322–1357. https://doi.org/10.1086/228903

Kim, T. H., & Han, E. (2015). Impact of body mass on job quality. Economics &Amp; Human Biology, 17, 75–85. https://doi.org/10.1016/j.ehb.2015.02.003

Lee, D. H., Shin, A., Kim, J., Yoo, K. Y., & Sung, J. (2011). Validity of Self-Reported Height and Weight in a Korean Population. Journal of Epidemiology, 21(1), 30–36. https://doi.org/10.2188/jea.je20100058

Lee, H., Ahn, R., Kim, T. H., & Han, E. (2019). Impact of Obesity on Employment and Wages among Young Adults: Observational Study with Panel Data. International Journal of Environmental Research and Public Health, 16(1), 139. https://doi.org/10.3390/ijerph16010139

Levine, E. E., & Schweitzer, M. E. (2015). The affective and interpersonal consequences of obesity. Organizational Behavior and Human Decision Processes, 127, 66–84. https://doi.org/10.1016/j.obhdp.2015.01.002

Li, P., Chen, X., Stafford, F., & Ou, J. (2022). Body shape and stable employment opportunity analysis of China's nonagricultural labor market. SSM - Population Health, 17, 101014. https://doi.org/10.1016/j.ssmph.2021.101014

Li, P., Chen, X., & Yao, Q. (2021). Body Mass and Income: Gender and Occupational Differences. International Journal of Environmental Research and Public Health, 18(18), 9599. https://doi.org/10.3390/ijerph18189599

Mason, K. (2012). The Unequal Weight of Discrimination: Gender, Body Size, and Income Inequality. Social Problems, 59(3), 411–435. https://doi.org/10.1525/sp.2012.59.3.411

Mobius, M. M., & Rosenblat, T. S. (2006). Why Beauty Matters. American Economic Review, 96(1), 222–235. https://doi.org/10.1257/000282806776157515

Park, R. H., Myers, P. L., & Langstein, H. N. (2019). Beliefs and trends of aesthetic surgery in South Korean young adults. Archives of Plastic Surgery, 46(06), 612–616. https://doi.org/10.5999/aps.2018.01172

Pfeifer, C. (2012). Physical attractiveness, employment and earnings. Applied Economics Letters, 19(6), 505–510. https://doi.org/10.1080/13504851.2011.587758

Salter, S. P., Mixon, F. G., & King, E. W. (2012). Broker beauty and boon: a study of physical attractiveness and its effect on real estate brokers' income and productivity. Applied Financial Economics, 22(10), 811–825. https://doi.org/10.1080/09603107.2011.627211

Schulte, P. A., Wagner, G. R., Ostry, A., Blanciforti, L. A., Cutlip, R. G., Krajnak, K. M., Luster, M., Munson, A. E., O'Callaghan, J. P., Parks, C. G., Simeonova, P. P., & Miller, D. B. (2007). Work, Obesity, and Occupational Safety and Health. American Journal of Public Health, 97(3), 428–436. https://doi.org/10.2105/ajph.2006.086900

Smith, S. M., McIntosh, W. D., & Bazzini, D. G. (1999). Are the Beautiful Good in Hollywood? An Investigation of the Beauty-and-Goodness Stereotype on Film. Basic and Applied Social Psychology, 21(1), 69–80. https://doi.org/10.1207/s15324834basp2101_7

Song, S., & Baek, S. (2021). Body shape matters: Evidence from machine learning on body shape-income relationship. PLOS ONE, 16(7), e0254785. https://doi.org/10.1371/journal.pone.0254785

Stiles, M. (2017, June 13). In South Korea's hypercompetitive job market, it helps to be attractive. Los

Angeles Times. Retrieved October 2, 2022, from https://www.latimes.com/world/asia/la-fg-south-korea-image-2017-story.html

Swami, V. (2006). The influence of body weight and shape in determining female and male physical attractiveness. In L. A. Ferrera (Ed.), Focus on body mass index and health research (pp. 1–28). Nova Science Publishers.

B. APPENDICES

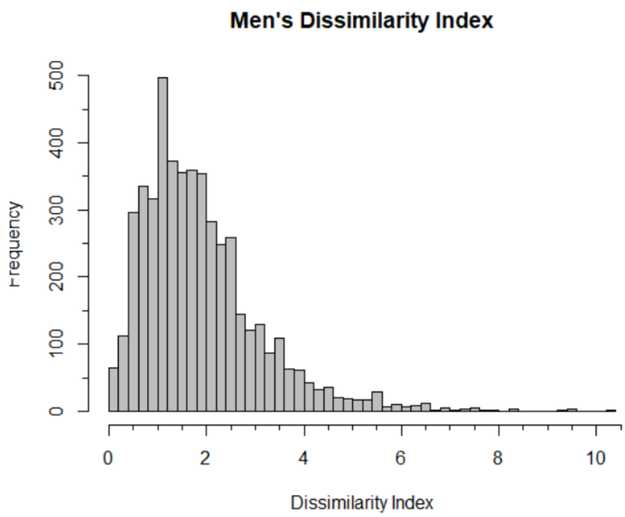


Figure 1: Histogram of the dissimilarity index for men

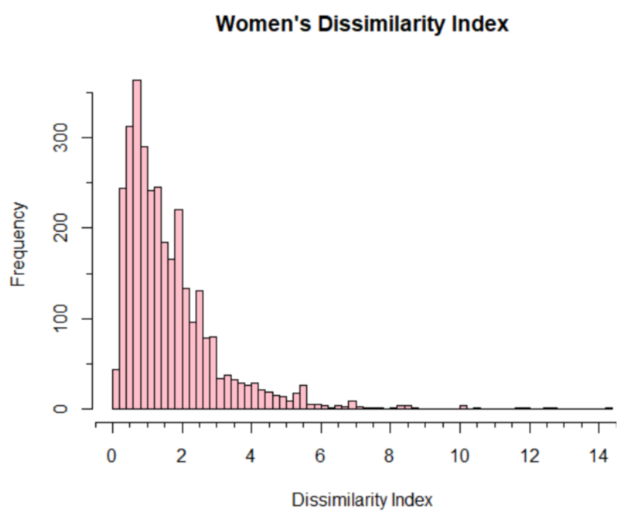


Figure 2: Histogram of the dissimilarity index for women

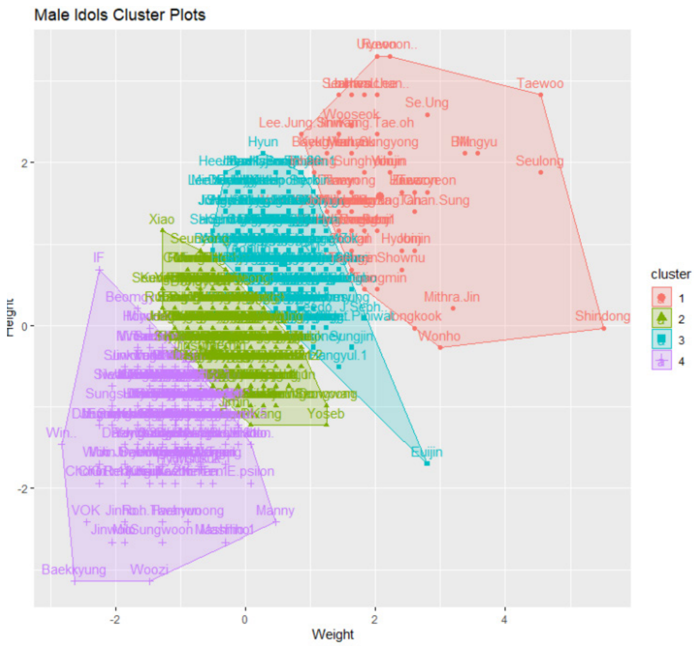


Figure 3: Scatter slot of weights and heights of male K-pop idols

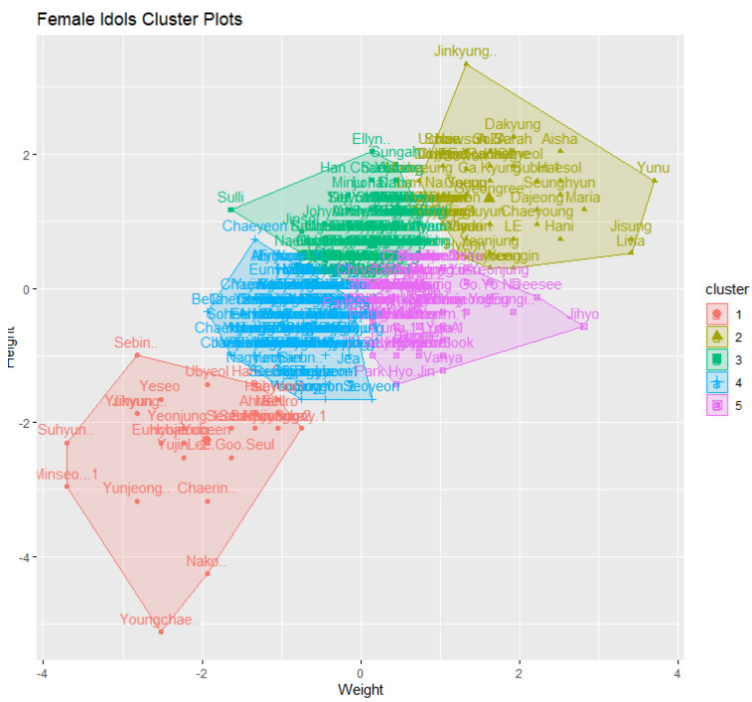


Figure 4: Scatter plot of weights and heights of female K-pop idols

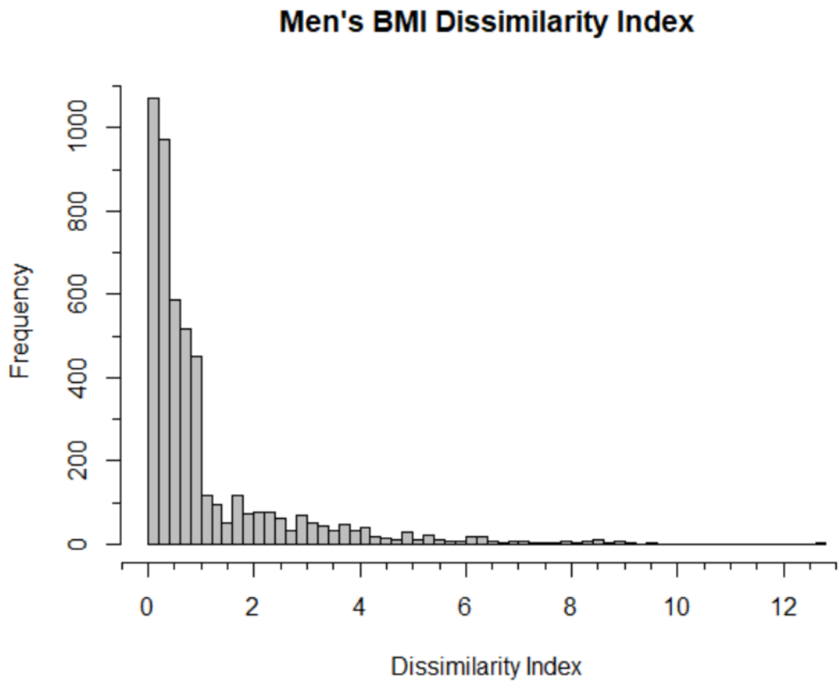


Figure 5: Histogram of the BMI dissimilarity index for men

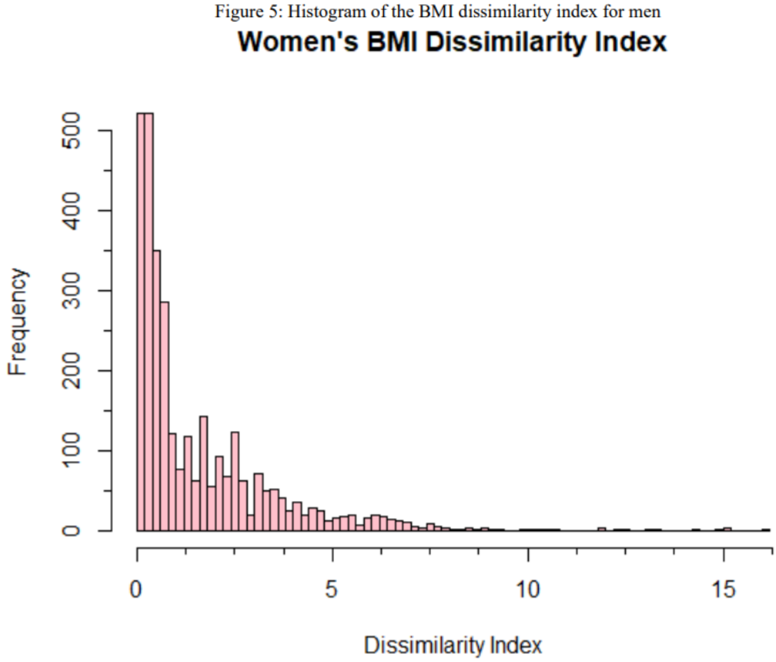


Figure 6: Histogram of the BMI dissimilarity index for women

Short-Term Value Stock Outperformance Following Earnings Surprises

Bi Yu Li

ECON 490: Seminar in Applied Economics

Basu, S. (1983). The relationship between earnings' yield, market value and return for NYSE common stocks. Journal of Financial Economics, 12(1), 129–156. https://doi.org/10.1016/0304-405x(83)90031-4

Dreman, D. N., & Berry, M. A. (1995). Overreaction, underreaction, and the low-P/E effect. Financial Analysts Journal, 51(4), 21–30. https://doi.org/10.2469/faj.v51.n4.1917

Dreman, D. N., & Lufkin, E. A. (1997). Do contrarian strategies work within industries? The Journal of Investing, 6(3), 7–29. https://doi.org/10.3905/joi.1997.7

Dreman, D. N. (1998). Contrarian Investment Strategies: The next generation: Beat the market by going against the crowd. Simon & Schuster.

Fama, E., & French, K. (1992). The cross-section of expected stock returns. The Journal of Finance, 47(2), 427–465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

Fama, E., & French, K. (1993). Common risk factors in the returns on stocks and Bonds. Journal of Financial Economics, 33(1), 3–56. https://doi.org/10.1016/0304-405x(93)90023-5

Fuller, R. J., Huberts, L. C., & Levinson, M. J. (1993). Returns to E/P Strategies, Higgledy-Piggledy growth, analysts' forecast errors, and omitted risk factors. The Journal of Portfolio Management, 19(2), 13–24. https://doi.org/10.3905/jpm.1993.13

McWilliams, J. D. (1966). Prices, earnings and P-E ratios. Financial Analysts Journal, 22(3), 137–142. https://doi.

org/10.2469/faj.v22.n3.137

Nicholson, S. F. (1960). Price-earnings ratios. Financial Analysts Journal, 16(4), 43–45. https://doi.org/10.2469/faj.v16.n4.43

Nicholson, S. F. (1968). Price ratios in relation to investment results. Financial Analysts Journal, 24(1), 105–109. https://doi.org/10.2469/faj.v24.n1.105

Nasdaq Data Link. data.nasdaq.com. (n.d.). Retrieved November 29, 2022, from https://data.nasdaq.com/data/MULTPL/SP500_PE_RATIO_MONTH-sp-500-pe-ratio-by-month

Nasdaq Data Link. data.nasdaq.com. (n.d.). Retrieved November 29, 2022, from https://data.nasdaq.com/data/MULTPL/SP500_PBV_RATIO_QUARTER-sp-500-price-to-book-value-by-quarter

Company earnings calendar. Yahoo! Finance. Retrieved November 29, 2022, from https://finance.yahoo.com/calendar/earnings/

Health And Labour Supply: The Effect Of Self-rated Health On Hours Worked

Derek Situ

ECON 490: Seminar in Applied Economics

Angrist, Joshua D., and Guido W. Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." Journal of the American Statistical Association 90 (430): 431-442.

Baker, Michael, Mark Stabile, and Catherine Deri. 2004. "What Do Self-Reported, Objective, Measures of Health Measure?" The Journal of Human Resources 39 (4): 1067-1093.

Black, Nicole, David W. Johnston, and Agne Suziedelyte. 2017. "Justification bias in self-reported disability: New evidence from panel data." Journal of Health Economics 54: 124-134.

Blewett, Lynn A., Julia A. Rivera Drew, Miriam L. King, Kari C.W. Williams, Natalie Del Ponte, and Pat Convey. 2022. IPUMS Health Surveys: National Health Interview Survey, Version 8.1 [dataset]. Accessed September 2022. https://nhis.ipums.org/nhis/.

Blundell, Richard, Jack Britton, Monica Costa Dias, and Eric French. 2017. "The impact of health on labor supply near retirement." Institute for Fiscal Studies Working Paper.

Bound, John, Michael Schoenbaum, Todd R. Stinebrickner, and Timothy Waidmann. 1999. "The dynamic effects of health on the labor force transitions of older workers." Labour Economics 6 (2): 179-202.

Card, David, and John E. DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." Journal of Labor Economics 20 (4):

733-783.

Cai, Lixin. 2010. "The relationship between health and labour force participation: Evidence from a panel data simultaneous equation model." Labour Economics 17 (1): 77-90.

Cai, Lixin, Kostas Mavromaras, and Umut Oguzoglu. 2014. "The effects of health status and health shocks on hours worked." Health Economics 23 (5): 516-528.

Centers for Disease Control and Prevention. 2021. High Blood Pressure Symptoms and Causes. Accessed September 2022. https://www.cdc.gov/bloodpressure/about.htm.

—. 2021. Multiple Imputation of Family Income in 2020 National Health Interview Survey: Methods. Accessed September 2021. https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2020/NHIS2020-imputation-techdoc-508.pdf.

Crossley, Thomas, and Steven Kennedy. 2002. "The reliability of self-assessed health status." Journal of Health Economics 21 (4): 643-658.

Jiang, Zhichao, and Peng Ding. 2020. "Measurement errors in the binary instrumental variable model." Biometrika 107 (1): 238-245.

Lemieux, Thomas. 2019. ECON 594: Lecture #2. Accessed September 2022. https://econ2017.sites.olt.ubc.ca/files/2019/05/pdf_594lecture2_thomas-lemieux.pdf.

Leszczensky, Lars, and Tobias Wolbring. 2022. "How to Deal With Reverse Causality Using Panel Data? Recommendations for Researchers Based on a Simulation Study." Sociological Methods and Research 51 (2): 837-865.

Liddell, Torrin M., and John K. Kruschke. 2018. "Analyzing ordinal data with metric models: What could possibly go wrong?" Journal of Experimental Social Psychology 79: 328-348.

Pelkowski, Jodi Messer, and Mark C. Berger. 2003. "The impact of health on employment, wages, and hours worked over the life cycle." The Quarterly Review of Economics and Finance 44 (1): 102-121.

R Core Team. 2022. "R: A Language and Environment for Statistical Computing." R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Stern, Steven. 1989. "Measuring the effect of disability on labour force participation." Journal of Human Resources 24 (3): 361-395.

Zhu, Xiaoping. 2014. "Comparison of Four Methods for Handing Missing Data in Longitudinal Data Analysis through a Simulation Study." Open Journal of Statistics 04 (11): 933-944.

The Impact of Consumer Reviews on Product Sales: Evidence from Video Game Products on Amazon

Ting (Dongting) Guo

ECON 490: Seminar in Applied Economics

A. REFERENCES

Ahsan, Ayesha. 2017. "Consumer ratings-reviews and its impact on consumer purchasing behavior."

Amazon. n.d.. "How Are Product Star Ratings Calculated?"

Anderson, Eugene W. 1998. "Customer Satisfaction and Word of Mouth." Journal of Service Research, 1(1): 5–17.

Bajaj, Aryan. 2021. "Can Python understand human feelings through words? – A brief intro to NLP and VADER Sentiment Analysis."

Chen, Pei-Yu, Shin-yi Wu, and Jungsun Yoon. 2004. "The Impact of Online Recommendations and Consumer Feedback on Sales." 58.

Chen, Tao, Premaratne Samaranayake, XiongYing Cen, Meng Qi, and Yi-Chen Lan. 2022. "The Impact of Online Reviews on Consumers' Purchas ing Decisions: Evidence From an Eye-Tracking Study." Frontiers in Psychology, 13.

Chevalier, Judith A., and Dina Mayzlin. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews." Journal of Marketing Research, 43(3): 345–354.

24 DEC 2022

Clement, J. 2022a. "COVID-19 impact on the gaming industry worldwide - Statistics & Facts."

Clement, J. 2022b. "Distribution of video gamers in the United States in 2022, by age group."

Connolly, Brian. 2022. "Amazon Best Sellers Rank: Everything You Need to Know in 2022."

Coppola, Daniela. 2022. "Annual net sales revenue of Amazon from 2006 to 2021, by segment."

Duan, Wenjing, Bin Gu, and Andrew B. Whinston. 2008. "Do online reviews matter? — An empirical investigation of panel data." Decision Support Systems, 45(4): 1007–1016.

Jungle Scout. n.d.. "Amazon Sales Estimator."

Kim, Da Yeon, and Sang Yong Kim. 2022. "The impact of customer generated evaluation information on sales in online platform-based markets." Journal of Retailing and Consumer Services, 68: 103016.

Kunst, Alexander. 2022. "Digital video game purchases by store brand in the U.S. in 2022."

Lohse, Tobias, Jan Kemper, and Malte Brettel. 2017. "HOW ONLINE CUSTOMER REVIEWS AFFECT SALES AND RETURN BEHAVIOR – AN EMPIRICAL ANALYSIS IN FASHION E-COMMERCE." In Proceedings of the 25th European Conference on Information Systems (ECIS), Guimar˜aes, Portugal, 2635–2644. ISBN: 978-0-9915567-0-0.

Masters, Kiri. 2021. "A Short History Of Amazon's Product Review Ecosystem, And Where We Are Today."

Ni, Jianmo, Jiacheng Li, and Julian McAuley. 2019. "Justifying Recom mendations using Distantly-Labeled Reviews and Fine-Grained Aspects." 188– 197. Hong Kong, China:Association for Computational Linguistics.

OECD. 2019. "Understanding online consumer ratings and reviews." OECD Dig ital Economy Papers, 289. Place: Paris.

Perez, Sarah. 2016. "Amazon bans incentivized reviews tied to free or dis counted products."

ReviewMeta. 2016. "Analysis of 7 million Amazon reviews: customers who receive free or discounted item much more likely to write positive review."

Sun, Miao, Jing Chen, Ye Tian, and Yufei Yan. 2021. "The impact of online reviews in the presence of customer returns." International Journal of Production Economics, 232: 107929.

UNCTAD. 2021. "Global e-commerce jumps to $26.7 trillion, COVID-19 boosts online sales."

Wang, Alex. 2022. “What Is the Share of E-Commerce in Overall Retail Sales?”

Zhu, Feng, and Xiaoquan (Michael) Zhang. 2010. “Impact of Online Con sumer Reviews on Sales: The Moderating Role of Product and Consumer Char acteristics.” Journal of Marketing, 74(2): 133–148.

B. APPENDIXES

The datasets and code for analysis can be found on: https://github.com/ ting486/Consumer-Reviews-And-Purchasing-Decisions.

Authors

Bi Yu Li

ECONOMICS

Bi Yu recently graduated from the University of British Columbia with a Bachelor of Arts in Economics with Distinction. His coursework focused on monetary theory and financial economics. He is a candidate for Level 2 of the Chartered Financial Analyst (CFA) designation. He currently works in corporate finance and is passionate about a career in finance, investment, and corporate management.

Derek Situ

ECONOMICS

Derek is a graduate of the University of British Columbia's Bachelor of Arts program, where he majored in Economics and Statistics. His interest in these subjects comes from his enthusiasm for applying empirical methods to the analysis of topics relevant to society. During his coursework and work placements, Derek developed a keen interest in international trade and finance. More recently, his research interests have included applied microeconomics, health, labour, and sociology. He is excited to start a Master's degree in Economics at Simon Fraser University in September 2023.

Ting (Dongting) Guo

ECONOMICS

Ting graduated from the University of British Columbia's Bachelor of Science program in 2023 with a combined major in mathematics and economics. Her research interests span various fields, including economics in marketing, finance, health and gender. Ting is currently working as a reporting and analytics analyst in Vancouver and will soon begin her Master of Science in Analytics degree at the Georgia Institute of Technology.

Mea Srisan

ECONOMICS

Mea graduated from the University of British Columbia in 2023, earning her Bachelor of Arts degree with a specialization in Economics. Originally from Bangkok, Thailand, she is currently working as a researcher at a policy research center in her hometown. Mea aims to apply the knowledge she gained during her time at UBC to help devise better policy solutions and make a positive impact on her beloved home country. Her interest lies in the intersection of economics, data, and technology. With this passion, she plans to pursue a Master's degree in data science.

IONA Team

EDITOR-IN-CHIEF

COLBY CHAMBERS

Bachelor of Arts - Honours Economics

Colby is a fourth year student in the honours economics program. He is specifically interested in theories of discrimination and its applications to behavioural and labour economics. To further his experience in this area, he is working as a predoctoral research assistant at the National Bureau of Economic Research and hopes to pursue a PhD. He sees IONA as an opportunity for undergraduate students to gain earlier and more detailed exposure to academic papers, while also building connections and friendships with classmates. He is committed to continuing the great work of the IONA Journal and its projects over the past 7 years.

DIRECTOR OF BUSINESS OPERATIONS

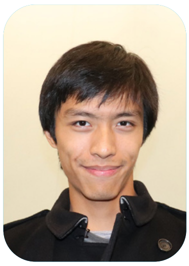
VALENTINA RAMIREZ

Bachelor of International Economics

Valentina, a third-year International Economics student, is captivated by the intricate quantitative methods driving economic development. With a history of spearheading her high school newspaper and a robust background in research, she's poised to elevate IONA's operations. Valentina's mission extends beyond numbers – she aims to bridge economics with other disciplines, fostering dialogues on historical context and modern trends in economic sustainability. Collaborative at her core, she eagerly anticipates the opportunity to lead a remarkable team toward impactful outcomes.

BI YU LI

Bachelor of Arts - Economics



Bi Yu Li is a distinguished fourth-year BA Economics Major on track for graduation in May 2023, with a primary focus on the intricate realms of monetary policy and asset prices. In addition to their academic pursuits, Bi Yu finds delight in the exhilaration of skiing throughout winter and the tranquility of hiking and backpacking during the summer months. Poised for new challenges, Bi Yu eagerly anticipates the prospect of collaborating with the esteemed IONA team, envisioning a mutual exchange of expertise that will undoubtedly shape their professional journey in a profound manner.

CHANYA CHAWLA

Bachelor of Arts - Honours Economics

Chanya, a distinguished fourth-year Economics Honors scholar at the VSE, embodies a fervent commitment to the field. Her profound interests lie within the realms of financial and monetary economics, driven by an unwavering aspiration to shape policy landscapes. Currently engaged in the pursuit of the CFA level 1 exam, her enthusiasm for finance remains unabated. Beyond academia, her passions are equally intriguing; an innate ability to embrace sleep at any time and any place complements her undeniable affection for culinary pleasures, particularly those of the sweet variety.

CHRIS HAUN

Bachelor of Arts - Economics and Political Science

Chris Haun, a distinguished individual, emerges as a resolute academic in the final throes of a rigorous academic journey, pursuing a dual major in Economics and Political Science. As he navigates his fifth and culminating year, Chris's scholarly pursuits center around the captivating realm of empirical economics, fueled by an undeniable fascination for the intricate interplay of political economy. With an unwavering gaze fixed upon the horizon of public policy, he aspires to contribute profoundly to the fabric of societal progress. The alluring prospect of engaging with the rich tapestry of undergraduate research at UBC beckons, and Chris eagerly awaits the intellectual odyssey that lies ahead, poised to enrich his academic expedition in unprecedented ways.

SENIOR EDITORS

JASMINE ARORA

Bachelor of Arts - Economics

Jasmine is a third year BA Economics student with an interest in environmental economics, policy and philosophy. She joined IONA to help relay unique and impactful ideas of her peers and to gain further first-hand experience in the process of writing and editing academic research papers. She has previously worked as a research assistant in the Vibrant Forests Landscapes lab at UBC where she explored the economic concerns of forestry dependent towns in British Columbia. She is excited to work with the IONA wonderful team this year.

MANAN AGGARWAL

Bachelor of Arts - Economics

Manan Aggarwal, a scholarly presence, illuminates the academic landscape as a fourth-year Economics major. His intellectual predilections orbit around the captivating realms of financial economics and game theory, where he finds both intrigue and depth. At the heart of his academic journey lies a profound fascination with the foundational premise set forth by Adam Smith in his seminal work, 'The Wealth of Nations'. Manan's scrutiny centers on Smith's assertion that humans are inherently self-serving beings, and yet, through the collective pursuit of individual self-interest, the intricate tapestry of societal material needs is woven and fulfilled. This captivating exploration marks just one facet of Manan's intricate intellectual pursuits.

MEGUMI BERNARDO

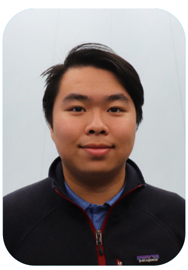
Bachelor of Arts - Honours Economics

Meg, a diligent third-year Honours Economics student, has been an integral part of the Iona Journal team since her very first year, fostering a connection that has only deepened over time. With the dawn of a new academic year, her excitement is palpable, fueled by the anticipation of not only witnessing the Journal's growth but also the maturation of the team that propels its success.

In parallel, Meg's heartwarming affinity for the Journal is underscored by her two-year engagement, a period during which its significance has become ingrained. This resonance is further amplified by her innate knack for refining academic prose. This editorial acumen positions her to contribute most effectively within the Journal, aligning her strengths harmoniously with its demands. Her conviction rests in the understanding that this portfolio is where her skills can most significantly contribute to the team's ongoing accomplishments.

PIUS LAU

Bachelor of International Economics



Pius Lau emerges as a dedicated third-year participant in the BIE program. As he embarks on a new chapter within the IONA fold, his aspirations echo a yearning for the expansion of his econometric expertise, coupled with an insatiable appetite for perusing impactful academic papers. Pius's intellectual curiosity converges at the juncture of trade, development, and history, where he seeks to unravel the intricate intersections that shape our global landscape.

Beyond the academic sphere, Pius finds a unique sense of engagement on the golf course, a newfound passion that has ignited his spirits. His fervor for the sport finds resonance in the serenity of the greens, where strategic prowess is equally matched by the camaraderie that accompanies the game. An open invitation stands, as Pius welcomes any inquiries, embracing the prospect of vibrant conversations that transcend his academic and personal pursuits.

JUNIOR EDITORS

ANDIE BARTOLOME

Bachelor of International Economics

Andie is a third year international economics student and is highly interested in policy, research and development economics. Having worked for the editorial board as a junior editor last year, she was motivated by the passion her fellow VSE peers had for the field of economics and their unique insights into the various ways economics could be applied. She's looking forward to seeing the talent of this year's contributors.

ANNA CHERNESKY

Bachelor of Arts - Honours Economics

Anna Chernesky, an accomplished third-year honours economics student, is characterized by an ardent affinity for the realms of mathematics and statistics, which stands harmoniously juxtaposed with her passion for the creative expressions of reading and crafting fiction during her leisure hours. Anna's intellectual canvas is further enriched by her steadfast dedication to the discipline of economics, a field that captivates her curiosity. Within the IONA journal, Anna envisions a learning journey that transcends the boundaries of her academic pursuits. Her aspirations orbit around the acquisition of insight into the intricate process of editing and publishing compelling economic research, an endeavor that promises to refine her skills and deepen her appreciation for the scholarly landscape.

CASANDRA LIM

Bachelor of International Economics

Casandra Lim, a distinguished individual, introduces herself as a proactive participant in the BIE program, marking her presence in the third year of her academic journey. In her personal pursuits, she finds solace in the captivating worlds unveiled through reading and the exhilaration of exploring new and uncharted landscapes. Fueled by an intrinsic thirst for knowledge, Casandra's enthusiasm resonates keenly with her aspirations within the IONA realm. With a palpable eagerness, she anticipates delving into the intricacies of publishing and immersing herself in the narratives woven within scholarly papers, nurturing a journey of both intellectual growth and discovery.

HIMANAYA BAJAJ

Bachelor of International Economics

Himanaya Bajaj steps forward as a dedicated second-year BIE student, her presence resonating within the academic tapestry. Amidst her pursuits, she embraces the exhilaration of tennis, the solace of literature, and the creative expression found in writing articles for the university newspaper. Rooted in her academic journey is an abiding fascination for behavioural economics, a field that ignites her intellectual curiosity. Himanaya's anticipation reverberates with genuine excitement as she envisions a symbiotic exchange within the IONA domain. Eager to immerse herself in the research papers crafted by her peers, she embarks on a quest of mutual growth, fervently seeking to glean insights and expand her horizons through the collaborative journey that awaits.

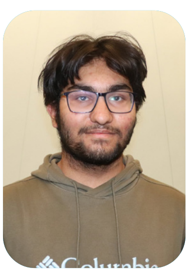
ISHAAN SAHAI

Bachelor of Arts - Economics

Ishaan is a third-year standing Economics major and is excited to grow his involvement with the undergraduate economics community at UBC. He sees the ultimate importance of economics as being rooted in its significance to people and is deeply interested in how the economic lens can be used as a tool to examine the issues of the world around him. He is particularly keen on economics' connections with policy, politics, climate change, and inequality, and enjoys film in his spare time.

JANESH TULSAN

Bachelor of Arts - Economics and Mathematics



Janesh Tulsan, a first-year entrant, embarks on his academic journey with a combined major in Economics and Mathematics. Outside the classroom, his passions extend to the harmonies of music and the excitement of basketball, often accompanied by the comfort of pizza. Enthralled by the prospect of learning, Janesh eagerly anticipates his engagement with IONA. With a desire to refine academic writing and explore advanced economic concepts, he envisions a path of growth and enrichment that aligns seamlessly with his pursuits.

JOSHUA EVANGELISTA

Bachelor of International Economics

Josh is a second year Bachelor of International Economics student with an interest in the capital market and quantitative analysis. He is particularly passionate about proposing reccomendations that will reduce Child Poverty. By working as a pro-bono consultant and assisting organizations in Lethbridge, Mexico and Colombia, Josh was able to effectively achieve his goal. Josh is excited to have a great year, working alongside an amazing team.

JULIA HEDDA LIMA PASTANA

Bachelor of Arts - Economics

Julia, a spirited second-year Arts student hailing from São Paulo, Brazil, introduces herself with enthusiasm. Stepping into the role of an incoming Junior Editor at IONA, Julia's aspirations gravitate towards a multifaceted exploration of economic research. With a resolute focus on its practical implications within the realms of public policy and development, she's equally captivated by the intricate intersections it forms with cutting-edge themes such as technology and artificial intelligence. Julia's journey is underscored by a thirst for knowledge and a fervent drive to uncover the intricate tapestry of economic dynamics that shape the contemporary world.

LEONA YU

Bachelor of International Economics

Leona Yu, a fresh presence in the academic landscape, introduces herself as a first-year student in the Bachelor of International Economics (BIE) program. Her origins trace back to Calgary, Alberta, although her formative years unfolded in Malaysia. Joining the IONA Journal, Leona aims to channel her passion for writing and editing. Looking ahead, she envisions graduate studies and research pursuits in her post-graduation path. Amidst her academic pursuits, Leona finds delight in animated films, novels, and a quest for raccoons. Her heart resonates deeply with cats, and she eagerly anticipates contributing her energies to the IONA Journal in the year ahead.

LUIGI VICENCIO

Bachelor of Arts - Economics

As a first-year student in the Bachelor of Arts program, Luigi Vicencio hopes to foster his economics knowledge and gain experience in academia. During high school in Brazil, Luigi earned silver and bronze medals in the Brazilian Economics Olympiad and co-founded a STEM-focused Blog in which he wrote introductory economics articles. His major economic interests are educational public policies and macroeconomics applied to finance. In his spare time, Luigi enjoys watching soccer and discovering new spots at UBC.

NOAH MONASH

Bachelor of International Economics

Noah, a vivacious presence, steps forward to share his narrative. As a first-year participant in the esteemed Bachelor of International Economics (BIE) program, his academic journey commences with zeal. Taking on the role of a junior editor within the IONA Journal, Noah aspires to not only refine his editorial skills but also entertain the possibility of contributing papers to its pages in the years to come. A dedicated reader, his heart resonates particularly with the allure of classic French literature. Outside the realm of academia, Noah embraces the vitality of sports, finding enjoyment in their diverse forms. Amidst these passions, he threads a unique path of growth within the IONA Journal, where his contributions promise to enrich the academic tapestry.

PRANAV VINOD

Bachelor of International Economics

Pranav, a dynamic presence, steps into the spotlight with an enthusiastic greeting. Embarking on his first year in the Bachelor of International Economics program, his academic journey begins with fervor. Fueled by an unwavering interest in monetary policy and macroeconomic stabilization, Pranav's intellectual pursuits are firmly grounded in economic dynamics. Building upon a foundation of diverse experiences, including research internships and journalism endeavors during high school, he stands poised to contribute to the pioneering realm of student research through the IONA Journal. Pranav's aspirations are twofold— to enrich his own perspectives and knowledge within economic research, and to eventually author a paper focused on macroeconomic stabilization and the intricacies of low-growth traps.

PRIYANKA SRINIVAS

Bachelor of International Economics

Priyanka steps forward as a third-year BIE student, her academic journey marked by depth and dedication. Beyond the classroom, her multifaceted interests extend to the realms of music, marketing, and the captivating embrace of literature. Within the tapestry of her pursuits, Priyanka's heart is drawn to developmental economics, a realm that sparks her intellectual curiosity. Embracing the opportunity within the IONA Journal, her aspirations echo a yearning to explore the interconnections that economics weaves with other fields of study. As she delves into this collaborative realm, Priyanka's passion is poised to contribute a unique perspective, enriching both her own academic journey and the collective efforts of the team.

SHREYA IYER

Bachelor of Arts - Economics

Shreya Iyer, a vibrant presence, introduces herself with exuberance. As a third-year Economics major, her academic path is marked by dedication and a thirst for knowledge. Enlisting herself within the IONA Journal, Shreya's aspirations gravitate towards the realm of practical applications within economic research. Eagerly anticipating this journey, she looks forward to forging connections and collaborative endeavors with the esteemed editorial team. With each step, Shreya embarks on a voyage that promises to expand her horizons, align her academic pursuits with real-world contexts, and contribute her unique perspective to the collective scholarly exploration of the Journal.

TEONA MARIUTA

Bachelor of International Economics

Teona, a fresh presence in the academic sphere, introduces herself with clarity. Stepping into her first year as a BIE student, her academic journey is marked by the promise of growth and exploration. Through her engagement with the IONA Journal, Teona aspires to extend her grasp of economics, nurturing an environment of continuous learning. A spirit of collaboration infuses her endeavors, as she eagerly anticipates the opportunity to work alongside peers and glean insights from upper-year students. In this voyage of knowledge and camaraderie, Teona embarks on a path that intertwines her individual pursuit of understanding with the collective pursuit of scholarly enrichment within the Journal.