UAB
Universitat Autònoma de Barcelona

Essays on Gender Earnings Gaps in the United States

Rui Yang

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  https://creativecommons.org/licenses/?lang=ca

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  https://creativecommons.org/licenses/?lang=es

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  https://creativecommons.org/licenses/?lang=en



Essays on Gender Earnings Gaps in the United States

By

Rui Yang

Supervisor

Maria Cervini-Plà

Submitted in partial fulfillment of the requirements for the doctoral degree of philosophy in Applied Economics

Department of Applied Economics
Universitat Autònoma de Barcelona
Bellaterra (Barcelona), Spain
June, 2026

ALL RIGHTS RESERVED

©2026

Rui Yang

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. María Cervini-Plá. I am sincerely thankful for the valuable time and effort she has devoted to guiding me throughout these four years. I know that I am not a very talkative person, and I often find it difficult to express my thoughts and feelings clearly. However, she has always been patient, understanding, and tolerant with me. What I appreciate most is the strong emotional support she has given me during this journey. No matter how my thesis, research progress, or presentations turned out, she always gave me recognition, encouragement, and confidence. Her belief in me often helped me believe in myself, especially in moments when I felt uncertain or discouraged. To me, she is not only my supervisor, but also someone who feels like family and a true friend. I will always be grateful for her kindness, patience, and trust.

I would also like to express my sincere gratitude to Prof. Rosella Nicolini. From the very beginning of my PhD application to the final submission of this thesis, she has helped me with countless questions, especially those related to university procedures, administrative requirements, and the submission of documents through the university system. I may have had more questions than many others, but she always replied to me with patience, clarity, and remarkable efficiency. Without her timely guidance and detailed explanations, many steps of this doctoral journey would have been much more difficult for me. Beyond this, I deeply admire her professionalism and dedication to her work. She carefully organized academic activities and training courses, regularly reminded us of important deadlines and schedules, and often shared useful information and academic opportunities with us, encouraging us to participate and grow as researchers. I was also impressed by her strong sense of responsibility and diligence; on several occasions, when I arrived at the office very early, she was already there working. I am truly grateful for all the time, effort, patience, and care she has given to me and to the PhD community.

My sincere thanks also go to Prof. Hector Sala Lorda, Prof. Sara Ayllón Gatnau, and Prof. Jenifer Ruiz-Valenzuela for their support during my PhD application. Their recommendation letters, trust, and encouragement were extremely important at the beginning of this journey. I am especially grateful to Prof. Jenifer Ruiz-Valenzuela, who supervised my Master's thesis. Her guidance during my Master's studies helped me build the foundation for my later doctoral research. Their support gave me confidence and encouraged me to continue my academic path.

I would also like to thank Prof. Antonio Di Paolo and all the professors who taught and

supported me during the Master's Programme in Labour Economics at the University of Barcelona. The knowledge, training, and encouragement I received during that period played an important role in my academic development. My experience in the Master's programme helped me understand my research interests more clearly and prepared me for the challenges of doctoral studies.

I am also deeply grateful to my PhD colleagues and friends. I am not a very talkative person, and I am often quiet. However, during these years, I was lucky to meet many kind, friendly, and supportive people. Their warmth and understanding made me feel less alone in a foreign country. They made my PhD life not only a period of hard work, but also a time filled with kindness and precious memories. I would like to especially thank Dalia Argudo Barrera, Camila Zamora Osorio, Adrián Moreno Maria, and Alejandro Gobierno García, as well as all the other colleagues and classmates who shared this journey with me. Their friendship, support, and kindness have meant a lot to me.

Above all, I would like to express my deepest gratitude to my beloved husband, Sen Miao. Without him, I would never have been able to come this far. His unconditional love, unbroken trust, and constant support have been my greatest source of strength, motivation, and courage throughout this doctoral journey. During these years, he took care of countless daily responsibilities and household matters so that I could focus on my research. He has been my caregiver, protector, cook, secretary, and, most importantly, my strongest emotional support. Whenever I felt tired, lost, discouraged, or close to giving up, he was always there, quietly staying by my side, listening to me, comforting me, and helping me regain the courage to continue. I am also deeply inspired by his own passion and dedication to academia. As an excellent researcher, he has taught me not only through his guidance in academic writing, data collection, coding, and research methods, but also through his seriousness, perseverance, and love for research. He believed in me when I doubted myself, and he supported me in every possible way. I will never be able to thank him enough. This doctoral dissertation belongs to him as much as it belongs to me.

Finally, I would like to thank my family, especially my parents. Their love, trust, and unconditional support have always been the foundation of my life. They gave me the courage to leave home, to study abroad, and to pursue a path that was never easy. Even when they were far away, their love was always with me. They believed in me when I doubted myself, and they supported me without asking for anything in return. Every step I have taken has been possible because of their sacrifices, patience, and encouragement. I owe this achievement to them as well.

Looking back on this doctoral journey, I feel grateful for every person who helped me,

supported me, and believed in me. These years were not always easy, but they have been meaningful and unforgettable. This thesis marks the end of one important chapter of my life, but it also carries the love, kindness, and strength that I received from all the people around me.

Table of contents

Acknowledgements **3**

General introduction **1**

1 Unpacking the U.S. Gender Wage Gap Across STEM-Classified Occupational Fields in the Context of COVID-19 Pandemic **6**

1.1 Introduction 6
1.2 Literature review 9
1.3 Impact of COVID-19 on labor market 12
1.4 Data and variables 14
1.4.1 Data 14
1.4.2 Variables 14
1.5 Descriptive evidence 16
1.5.1 Variable descriptives and sample characteristics 16
1.5.2 Evidence supporting the expanded STEM workforce 18
1.5.3 Evidence supporting the refined STEM occupational classification 19
1.5.4 Gender wage gap and the COVID-19 pandemic 20
1.6 Methodology 22
1.6.1 Unconditional quantile regression 22
1.6.2 RIF decomposition 23
1.7 Results and Discussion 25
1.7.1 Unconditional quantile regression results 25
1.7.2 Recentered influence function decomposition results 27
1.8 Conclusion 33
Appendix A: Estimation of the potential work experience 36
Appendix B: Unconditional quantile regression results 41

2 Gender Income Gap in the Creator Economy: Evidence from U.S. Top YouTubers **48**

2.1 Introduction 48
2.2 Literature review 52
2.3 YouTube monetization mechanics 55
2.3.1 YouTube partner program 55
2.3.2 Ads-based monetization structure 56
2.3.3 CPM and income estimation 57
2.4 Gender segregation and discrimination across content niches 60

2.5 Data . . . 61
2.5.1 Data source . . . 62
2.5.2 Sample construction . . . 64
2.5.3 Descriptive statistics . . . 65
2.6 Empirical strategy . . . 69
2.6.1 Model specification . . . 69
2.6.2 Frameworks for interpreting the gender gap . . . 70
2.7 Empirical Results . . . 72
2.7.1 Persistent gender income gap . . . 72
2.7.2 Explaining the gap . . . 76
2.7.3 Robustness and validation . . . 84
2.8 Conclusion . . . 89
Appendix A: Supplementary figures . . . 92
Appendix B: Supplementary tables . . . 93

3 Do Words Matter? Explaining the Gender Income Gap on YouTube through Top Content 95
3.1 Introduction . . . 95
3.2 Literature review . . . 98
3.3 Data . . . 101
3.3.1 Data sources . . . 101
3.3.2 Sample construction . . . 102
3.3.3 Text analysis . . . 104
3.3.4 Descriptive statistics . . . 108
3.3.5 Emotion disaggregation . . . 112
3.4 Empirical strategy . . . 118
3.4.1 Model specification . . . 119
3.4.2 Gelbach decomposition . . . 120
3.5 Results . . . 121
3.5.1 A persistent baseline gender income gap . . . 122
3.5.2 Sequential model specification . . . 123
3.5.3 Gelbach decomposition results . . . 127
3.5.4 Separate block inclusion results . . . 129
3.5.5 Variable-level contributions . . . 131
3.5.6 Robustness check . . . 139
3.6 Conclusion . . . 142
Appendix A: Supplementary figures . . . 146
Appendix B: Supplementary tables . . . 147

General conclusion **167**

References **172**

General introduction

Gender gaps in earnings remain a central concern in labour economics. Although the gender wage gap has narrowed considerably over the long run, the mechanisms underlying such disparities have also been evolving. A large body of research shows that traditional human-capital factors, such as education and work experience, now explain only a limited share of the remaining gap. Instead, occupational segregation, workplace institutions, promotion and pay-setting processes, and differences in temporal flexibility have become increasingly important (Blau and Kahn, 2017; Goldin, 2014). Meanwhile, a growing literature on psychological and behavioral mechanisms suggests that gender differences in preferences and behavior—particularly in relation to competition, risk-taking, and other experimentally measured traits—may also contribute to differences in educational choices, labour-market sorting, and earnings outcomes (Croson and Gneezy, 2009; Niederle and Vesterlund, 2007). Taken together, existing research suggests that gender inequality in earnings is a complex phenomenon shaped by a combination of structural, institutional, and behavioral mechanisms, rather than by any single factor alone.

At the same time, technological change and digitalization are reshaping the organization of work itself and accelerating the diffusion of flexible and digitally mediated forms of labour. This transformation matters not only because it has expanded flexibility within conventional employment, but also because it points to a broader reorganization of labour markets in which work is increasingly organized, evaluated, and monetized through digital technologies and platform (Mas and Pallais, 2020; Vallas and Schor, 2020). This thesis takes this transformation as both a substantive motivation and an analytical bridge. It starts from the idea that changes in the organization of work can reshape gender inequality even within conventional employment, suggesting that such effects may be still more consequential in labour markets built around digital coordination from the outset. As digital technologies make work more flexible, it becomes natural to ask whether similar, and potentially stronger, mechanisms operate in emerging labour markets where flexibility

and digital intermediation are not peripheral features but fundamental ones. This leads to a central question about labour-market change: when work is less tied to standard work settings, do gender gaps in earnings narrow, or do they instead take new forms through different mechanisms?

To answer this question, the thesis follows a comparative progression across labour-market settings in the United States, from conventional employment to the creator economy. The three chapters are linked by a common concern with how labour-market transformation changes the mechanisms that sustain gender inequality. Chapter 1 begins in the conventional labour market and focuses on STEM, a representative domain in which gender disparities in earnings have long persisted. By distinguishing between core STEM, STEM-related, and non-STEM occupations, and by comparing the pre-pandemic and pandemic periods, it examines both heterogeneity across occupational fields and the role of remote work in reshaping the gender wage gap. Its findings suggest that changes in work organization and flexibility can matter for gender inequality even within standard employment settings. If that is the case, then such mechanisms may be even more consequential in labour markets built around flexibility, digital coordination, and platform intermediation from the outset. Chapter 2 therefore turns to the creator economy and asks whether flexible, platform-mediated work reduces or reproduces gender income inequality among top U.S. YouTube creators. Chapter 3 builds on that analysis by introducing a much richer set of video-level and text-based measures, including creator-side language from titles, descriptions, and captions as well as audience-side responses expressed in comments, to investigate which additional content- and audience-side factors help explain the residual gap that remains after rich observable controls are introduced.

In this progression, Chapter 1 first studies gender wage inequality in STEM-classified occupations and related traditional labour markets before and after the COVID-19 pandemic. Its main objective is twofold: first, to compare the magnitude and determinants of the gender wage gap across core STEM, STEM-related, and non-STEM occupations; and second, to examine whether the gap widened or narrowed after the outbreak of COVID-19

and to identify the factors that may account for that change. Using U.S. labour-market microdata from IPUMS, the chapter distinguishes between core STEM, STEM-related, and non-STEM occupations and compares not only the size of the gender wage gap across occupational groups and across the pre-pandemic and pandemic periods, but also the mechanisms underlying these differences. Methodologically, it combines year-by-year analysis, distributional evidence across the wage distribution, and decomposition methods to assess both changes in the gap and the role of specific explanatory factors, including work-from-home arrangements. The chapter shows that the gender wage gap is smallest in core STEM occupations and largest in STEM-related occupations, and that its determinants also differ markedly across occupational categories. It further finds that the gap narrowed after the outbreak of COVID-19 across parts of the wage distribution, with work-from-home emerging as one plausible mechanism behind that change. The chapter therefore suggests that changes in work organization and flexibility can reshape gender inequality even within conventional employment, although their effects depend strongly on occupational context.

Building on this bridge, Chapter 2 turns to the creator economy as an emerging digitally mediated labour market and examines the gender income gap among top U.S. YouTube creators. In this setting, earnings are not determined by employer-set pay schedules, but by platform monetization, audience demand, content specialization, and user engagement. The main objective of the chapter is to measure the size of the gender income gap in this new labour-market setting and to identify the observable factors that account for it. To do so, the chapter constructs a novel creator-level dataset by combining estimated income and channel-performance information with a rich set of demographic, channel, content, production, and engagement variables. In particular, demographic characteristics are hand-collected and cross-validated across multiple public platforms and online sources, then integrated with platform-level channel data to build a uniquely detailed dataset on top U.S. YouTube creators. Methodologically, the chapter combines baseline regression analysis with sequential inclusion, Gelbach decomposition, Blinder–Oaxaca

decomposition, and a range of robustness checks. The results reveal a substantial and persistent income gap: female creators earn markedly less than comparable male creators even after conditioning on a broad set of observables. Traditional demographic and human-capital variables explain little of this disparity and in some specifications widen it, whereas content niche, annual content production, and especially audience engagement account for much of the explained component. At the same time, a meaningful residual remains. The chapter therefore shows that flexibility and digital intermediation do not automatically eliminate gender inequality; instead, they can reproduce it in new forms through platform monetization, category-specific returns, and audience-related mechanisms.

Chapter 3 investigates the residual female income penalty identified in Chapter 2. Its main objective is to determine whether richer content- and audience-side measures, especially video-level textual and language-based variables, can account for the part of the gender gap that remains unexplained after conditioning on creator-, channel-, content-, production-, and engagement-related characteristics. To do so, the chapter builds on the baseline creator-level dataset from Chapter 2 and enriches it with newly collected video-level information from YouTube-related APIs. For the final analytical sample of 485 U.S. creators, it constructs measures from each creator's most visible output, using up to 50 most popular long-form videos and up to 50 most popular shorts per creator. The resulting dataset combines information from 41,180 videos and 17.4 million comments, including general top-content characteristics, creator-generated text from titles, descriptions, and captions, audience-generated text from comments, and commercial and interaction signals. Methodologically, the chapter relies primarily on sequential inclusion regressions and Gelbach decomposition, complemented by separate block inclusion exercises and supplementary robustness checks. The findings show that the residual gender gap can be fully explained once these richer top-content, textual, and audience-response measures are taken into account. General top-content characteristics explain an important part of the gap, but the strongest explanatory role is concentrated on the audience side. Creator-controlled language explains relatively little, whereas audience-side indicators—especially average

likes and the evaluative tone of comments, more broadly reflected in sentiment-related measures—are the most informative factors. The chapter therefore shows that gender inequality in the creator economy is shaped not only by platform monetization structures or differences in content production, but also by how audiences publicly evaluate and reward the content that becomes visible.

Overall, this thesis seeks to understand how gendered earnings gaps persist and change as labour markets become more flexible, more digitally mediated, and less tied to standard work settings. It combines evidence from both the conventional labour market and the creator economy, using complementary empirical strategies to trace not only whether gender gaps remain, but also which mechanisms sustain or reshape them across different institutional settings. Through the findings, the thesis highlights that labour-market transformation does not simply eliminate gender inequality. Rather, it changes the channels through which inequality is produced, shifting attention from occupational structures, work arrangements, and the devaluation of women's work in conventional employment toward content specialization, platform monetization, user engagement, and audience interaction in digital, creator economy. In doing so, the thesis contributes to a more integrated understanding of gender inequality in contemporary labour markets by tracing how gendered earnings gaps are sustained, transformed, and reproduced from conventional employment to digitally mediated platform work.

Chapter 1

Unpacking the U.S. Gender Wage Gap Across STEM-Classified Occupational Fields in the Context of COVID-19 Pandemic

1.1 Introduction

In the knowledge-based economy of the 21st century, STEM (Science, Technology, Engineering, and Mathematics) fields play a vital role in driving global economic growth and technological innovation (Poggesi et al., 2019; Kuschel et al., 2020). As one of the world's leading innovation-driven economies, the United States relies heavily on the STEM sector to sustain growth and maintain competitiveness (Peri et al., 2015), making it a critical context in which to explore and examine gender inequality in this high-skilled labor market. Despite women's rising participation in STEM occupations and the progress they have achieved over the past few decades, substantial gender disparities in employment and wages still persist (Jiang, 2021; Fry et al., 2021): According to Jiang (2021), the gender wage gap (GWG) for women with STEM degrees and employed in STEM occupations is equivalent to approximately one year's wage over the first ten years of their careers. Although existing research has contributed valuable insights for promoting women's employment and career advancement in the STEM sector, recent developments underscore the need for a continued and in-depth study.

On one hand, the definition of the STEM workforce has evolved over time. Traditionally, STEM workers were understood as professionals holding at least a bachelor's degree. More recent definitions, however, have increasingly recognized the importance of workers who possess science and engineering skills and technical expertise, regardless of formal educational attainment. According to the National Science Foundation's report (National

Science Foundation, 2021) *"The STEM labor force of today: Scientists, engineers and skilled technical workers"*, more than half of STEM workers in the United States do not hold a bachelor's degree. The report therefore expands the definition to include not only degree holders but also workers without higher education qualifications, whose skills and technical competencies are also essential to innovation and technological development. From this perspective, re-considering and including the lower-educated segment is crucial for understanding a full picture of the STEM workforce.

On the other hand, the occupational classification of the STEM sector has also been refined, particularly through the distinction between core STEM and STEM-related occupations, given the markedly uneven representation of women across these fields. According to the *"STEM, STEM-related, and non-STEM related code list"* published by the U.S. Census Bureau (U.S. Census Bureau, 2018), STEM-related occupations now include a broad range of healthcare practitioners and technical occupations, such as nurse practitioners, pharmacy technicians, and surgical technologists, in which women are very well represented. This contrasts sharply with traditional STEM fields such as physical sciences, computer science, and engineering, where women remain substantially underrepresented (González-Pérez et al., 2020; Cho, 2021; Fry et al., 2021). This more refined classification is crucial not only for accurately capturing the composition of the STEM workforce but also for revealing distinct patterns in the GWG across STEM and STEM-related fields.

Beyond changes in the definition and classification of the STEM workforce, it is also important to consider the effects of the recent global crisis triggered by the COVID-19 pandemic. The pandemic profoundly altered the work organization across sectors. Economic disruption, quarantine, and social distancing measures not only caused sharp declines in labor force participation and surges in unemployment in many industries, but also reshaped occupational structures and the gender wage distribution (Coibion et al., 2020; Gallant et al., 2020; Albanesi and Kim, 2021). Importantly, these effects were highly uneven across sectors. Industries such as tourism and hospitality were severely affected, whereas information technology and other telework-related sectors remained relatively

resilient (Dube et al., 2020; Assaf and Scuderi, 2020). At the same time, the rise of work from home (WFH) as a widespread labor market practice widened inequalities across occupations, as reflected in growing disparities in wages and employment opportunities (Bolisani et al., 2020; Dockery and Bawa, 2020; Bick et al., 2023). These developments provide a distinct context for re-examining the gender wage gap (GWG) across core STEM, STEM-related, and non-STEM occupations.

Grounded in this context, this paper uses the IPUMS data to examine recent trends in the U.S. gender wage gap (GWG) before and during the COVID-19 pandemic, with particular attention to differences across core STEM, STEM-related, and non-STEM occupations. Using Unconditional Quantile Regression (UQR) and Recentered Influence Function (RIF) decomposition, we identify and analyze the main factors underlying the observed GWG. This paper makes several contributions to the literature. First, it provides up-to-date empirical evidence on recent GWG trends, offering new insights into evolving labor market dynamics in the context of the COVID-19 pandemic. Second, by distinguishing STEM-related occupations from core STEM fields, it reveals distinct gender disparities across these groups that have often been overlooked in previous research, while providing new empirical evidence on the structural drivers of the GWG in STEM and STEM-related occupations. These findings carry clear policy implications for promoting gender equality in the U.S. labor market and highlight the importance of field-specific interventions, educational inclusion, and flexible work arrangements in addressing persistent gender wage disparities.

The subsequent sections are structured as follows: Section 1.2 reviews the relevant literature. Section 1.3 depicts the effects of the COVID-19 pandemic on the U.S. labor market. Section 1.4 describes the data sources and the variables for analysis. Section 1.5 presents the descriptive evidence. Section 1.6 explains the methodologies employed in the empirical analysis. Section 1.7 analyzes and discusses the results, while Section 1.8 concludes with key findings and policy recommendations.

1.2 Literature review

This paper relates to several strands of the literature. We begin with the historical evolution of the gender wage gap (GWG) and the current state of research on STEM fields, highlighting the main gaps in the existing literature. We then reviews human capital theory as the classical analytical framework for explaining the GWG, while also discussing its limitations. Subsequently, we discuss the effects of the COVID-19 pandemic on the labor market and the GWG. Taken together, these gaps motivate the contributions of this paper.

Existing research on the GWG in the United States shows that, although the gap has narrowed over the long run, progress has slowed. Based on an extensive review of the literature, Blau and Kahn (2017) show that the GWG declined from about 36-38% in 1970 to 18-21% in 2010, while the increase in the female-to-male wage ratio slowed after peaking in the 1980s. They argue that although women have overtaken men in educational attainment and the gender gap in work experience has narrowed substantially, the GWG remains persistent. In the STEM domain, the existing literature generally finds that the GWG, while still present, is smaller than the roughly 20% gap observed in the overall labor market, typically around 14-15% (Xu, 2015; Kahn and Ginther, 2017; Xie and Connor, 2022). Speer (2020), however, notes an important limitation of existing research: estimates of the GWG vary substantially depending on how STEM is defined. The definition of the STEM workforce has evolved over time and is no longer confined to workers with higher education degrees, but now also includes those with scientific and engineering skills who do not hold formal higher education credentials (National Science Foundation, 2021). Likewise, the classification of STEM occupations has become more refined, distinguishing STEM-related occupations from traditional STEM fields (U.S. Census Bureau, 2018). Most existing studies, however, have subsumed STEM-related occupations within the broader STEM category and rely on a simple STEM/non-STEM distinction. For example, Xu (2015) treated medical professionals as part of the STEM category when comparing wage gaps between STEM and non-STEM occupations, while

Orozco (2021) classified science- and engineering-related occupations as a subfield of STEM. Xie and Connor (2022), meanwhile, excluded health-related occupations from the analysis on the grounds that they differ from core science and engineering occupations in qualification requirements and job responsibilities, even though these occupations clearly fall within the STEM-related category. These limitations suggest that adopting a more recent, standardized, and refined definition of STEM and its occupational categories may improve the analysis of the GWG across STEM and STEM-related occupations and shed further light on its underlying mechanisms.

For understanding and decomposing the GWG, human capital theory provides an important analytical framework. Originally proposed by Becker (1962) and Rosen (1976), the theory attributes wage disparities to differences in education, training, skills, and work experience. However, related literature has identified important limitations of this framework. While acknowledging its role in explaining long-run trends in the GWG, Polachek (2004) noted that it is less effective in accounting for short-run disparities. Similarly, Lips (2012) argued that the GWG as observed in practice is more complex than traditional human capital models can capture. Using Oaxaca–Blinder decomposition, supplemented by the methods of Chernozhukov et al. (2013) and Firpo et al. (2009), Blau and Kahn (2017) found that education and work experience explain only a limited share of the overall GWG. Later, they also found that gender differences in educational attainment account for only a small portion of the overall pay gap (Blau and Kahn, 2020). These findings suggest that factors beyond human capital, including discrimination and occupational segregation, play an important role in shaping gender wage inequality. They also emphasize the importance of handling work experience carefully, particularly when actual work experience is unavailable and researchers must rely on estimated potential experience instead.

Polachek (1975) was among the first to emphasize the importance of potential work experience in explaining the gender wage gap (GWG). Later research, for example, Regan and Oaxaca (2008), further examined the bias inherent in using potential work experience,

derived from the Mincer model (Mincer and Polachek, 1974), rather than actual work experience in human capital earnings models, and explored its implications for wage decomposition. They found that potential experience affects the effects of schooling and the returns to work experience, and using this measure in earnings models underestimates the explained portion of the GWG. Zveglich et al. (2019) also pointed out a key limitation of the Mincer and Polachek (1974) measure: it assigns the same level of potential experience to individuals with the same age and education, regardless of gender. This is problematic because many women experience career interruptions due to childbirth and child-rearing, so potential experience may overstate their actual experience and understate the gender gap in accumulated work experience. Blau and Kahn (2017) likewise argued that work interruptions and shorter working hours are more important in explaining the GWG in high-skilled occupations (such as those in the STEM occupational fields), relative to the labor market as a whole. As an alternative, Zveglich et al. (2019) used age-sex-specific labor force participation rates to approximate expected work experience and showed that this measure more closely tracks the actual median experience of men and women. Similarly, Meara et al. (2020) estimated potential work experience using a probit model of employment probabilities at each age between 15 and 65 and then aggregated these probabilities into years of expected work experience.

The importance of work experience measurement in explaining the gender wage gap (GWG) has become even more salient in the context of the COVID-19 pandemic, considering the changes it induced in work patterns across the global labor market. As mentioned earlier in the Introduction, the pandemic affected occupational groups unevenly. Dockery and Bawa (2020) argued that understanding which jobs are most compatible with work from home (WFH) is crucial, as it determines which workers can remain productive during lockdowns and thus avoid complete unemployment. Likewise, Adams et al. (2025) suggested that WFH is likely to be a key factor shaping GWG during the pandemic. Bick et al. (2023) further examined the effects of WFH and argued that its expansion may exacerbate existing inequalities by altering the relative position of occupations in the labor market.

Although WFH promotes employment stability during the pandemic, its implications for gender equality are complex (Carlson et al., 2021). On the one hand, WFH may enable women to work longer hours and access remote jobs that would otherwise be less feasible, thereby improving their career prospects and helping to narrow the gender wage gap (GWG). On the other hand, greater reliance on WFH may reinforce an unequal division of household responsibilities, with one partner focusing more on career and the other on family obligations, while women often assume a larger share of childcare and domestic work. This, in turn, may widen the GWG over the long run. These findings suggest that the effects of the COVID-19 pandemic on the GWG, particularly through differences in the feasibility of WFH, should be carefully examined across occupational categories.

1.3 Impact of COVID-19 on labor market

According to the U.S. official document[1], the federal COVID-19 public health emergency (PHE) officially ended on May 11, 2023. Since the formal termination of the PHE did not necessarily mark the end of the pandemic's adverse labor market effects, we examined employment and unemployment rates from January 2017 to July 2023 using Current Population Survey (CPS) data published by the U.S. Census Bureau (U.S. Census Bureau, 2023). The results are presented in Figure 1.1.

As shown in Figure 1.1, a sharp break occurred in April 2020, when employment rates for both men and women fell abruptly and unemployment rates surged, marking the most severe labor market disruption of the pandemic. Female unemployment increased more steeply and remained elevated until the end of 2020, indicating that women were disproportionately affected. Notably, although the U.S. government officially declared the end of the pandemic in May 2023, the data suggest that the labor market had largely returned to its pre-pandemic condition by mid-2022.

Accordingly, we primarily focus on the period from 2018 to 2022 and divides it into

[1]The Executive Office of the President,Office of Management and Budget's publication of "A bill to terminate the public health emergency declared for COVID-19" (United States Congress, 2023)

three phases for subsequent analysis:

(1) Before pandemic (2018–2019): the baseline period before the outbreak of COVID-19;

(2) Pandemic outbreak (2020): the initial shock period marked by the most severe labor market disruption;

(3) During pandemic (2021–2022): the adjustment period reflecting labor market adaptation and recovery under pandemic conditions.

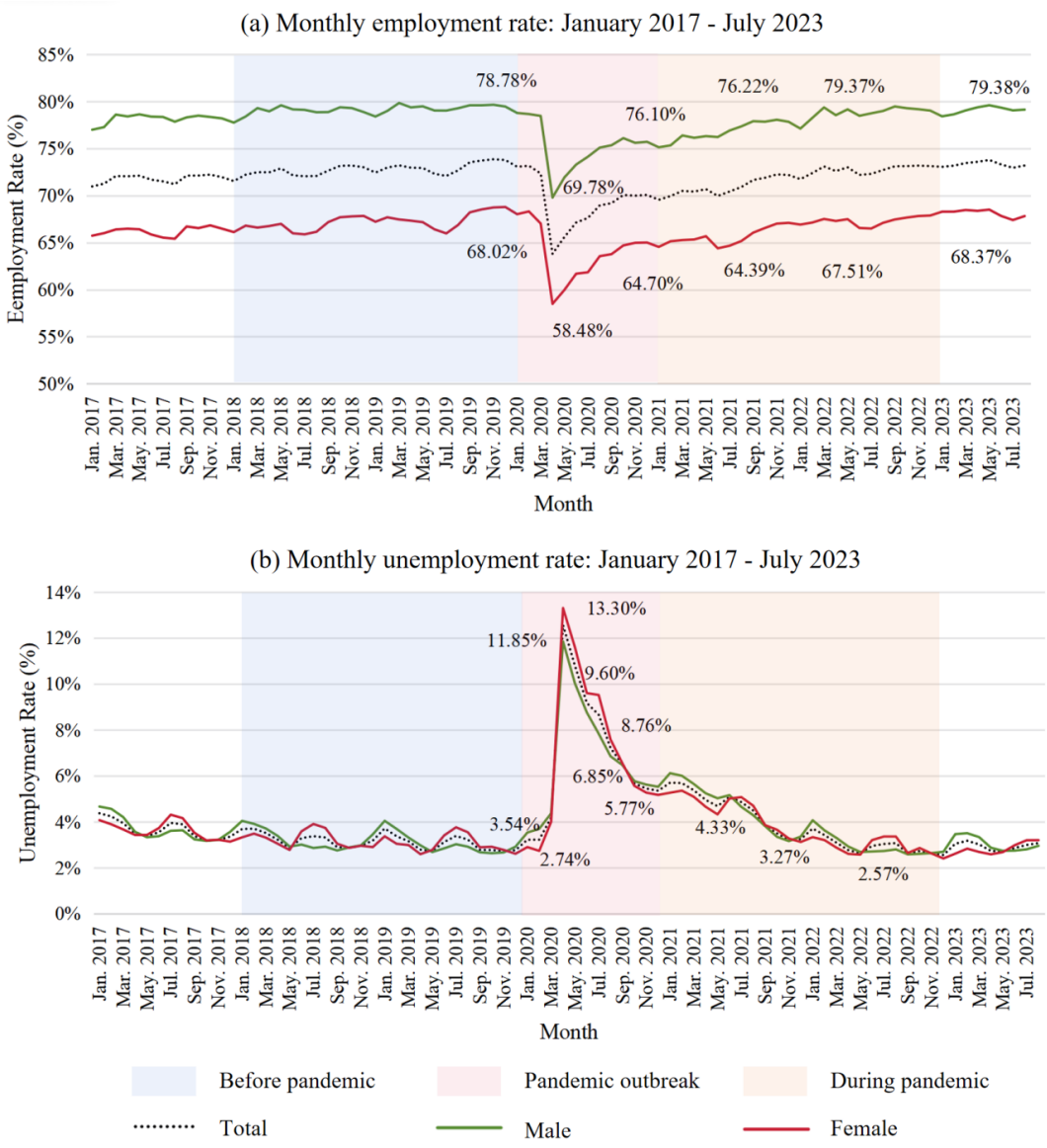


Figure 1.1. Monthly employment and unemployment rates January 2017 to July 2023 in the United States (CPS 2017-2023)

1.4 Data and variables

1.4.1 Data

The primary data source is the Integrated Public Use Microdata Series, USA (IPUMS, 2023), a large and widely used microdata database maintained by the Minnesota Population Center. We focus on IPUMS data from 2000 to 2022 for individuals aged 25 to 65, with particular focus on pooled cross-sectional data for 2018 to 2022. This period is especially relevant because it captures the key trends in the U.S. gender wage gap (GWG) before and during the COVID-19 pandemic that are central to the analysis.

The initial dataset contains a total of 8,130,981 observations, including 4,113,648 female and 4,017,333 male samples. After excluding the self-employed and part-time workers, military personnel, and non-meaningful samples (i.e., individuals with zero wages and negative potential work experience), the final sample comprises 3,883,551 observations in total, of which 1,689,712 are women and 2,193,839 are men.

1.4.2 Variables

The analysis primarily focuses on individual's wages by gender. Accordingly, the dependent variable is the logarithm of annual wages — *lnwage*, and the key explanatory variable is gender (0 for men and 1 for women). Additional covariates include a range of indicators capturing demographic characteristics, family background, educational background, job-related characteristics, as well as pandemic-related factors, as detailed below:

- Demographic Characteristics: This includes age (*age*), race (*white, black, indian, asian, other, more_than_one*), geographical location (*midwest, northeast, south, west*) indicating the individual's region within the United States, and migration (*migration*) identifying individuals living in a region different from their birthplace.

- Family Background: This involves marital status (*unmarried, married, single*), the number of children (*child0, child1_2, child3_4, child5_*), and the number of parents

in the household (*nparents*). In addition, there are dummy variables for children under five years old (*childage0_5*=1 if there are children under age five, otherwise *childage0_5*=0) and for school-age children (*childage6_17*=1 for children aged 6-17, otherwise *childage6_17*=0), along with their respective interaction terms with the single variable.

- Educational Background: This consists of the categorical variables indicating the highest level of education (*loweduc*[2]*, bachelor, master, professional*[3]*, doctoral*), and a second-degree dummy variable (*degree2*) to identify if the individual has a second undergraduate degree. Additionally, there is a variable (*school*) to identify those who may have attended school or college in the last 3 months[4].

- Job-related Characteristics: This includes the job sector (*private, government, non-profit*), occupational fields (*stem, stem-related, non-stem*), along with their respective interaction terms with loweduc, and the variable *fratio*[5] to measure the "female devaluation" in various occupations. The typical weekly working hours (*wkhours*) and potential work experience variables (*wkexp*[6]) are also included.

- Pandemic-related Variables: This consists of a variable indicating specific years of the pandemic (*covid2020, covid21&22*), a dummy variable indicating work from home (WFH) (*workhome*=1 if the individual works from home, otherwise *workhome*=0), along with their respective interaction terms with the two pandemic-specific years.

[2]This variable identifies individuals in the labor force who have not obtained a bachelor's degree.

[3]A professional degree is a credential designed to prepare individuals for specialized professions, practices,or industries,often fulfilling the academic requirements for licensure or accreditation.

[4]This variable identifies individuals who are recent graduates with little experience,or who may have just experienced a disruption in the workforce

[5]This variable measures the proportion of females in each occupation.

[6]In this study,potential work experience is estimated using the cumulative probability method (CPM), which better captures labor market interruptions and therefore provides a more appropriate measure for explaining the GWG.This variable is labeled *"wkexp"* and is used throughout the main analysis. As a robustness check,the alternative and widely used Mincer measure is also examined in Appendix 1.8, which shows why the CPM is the preferred specification.

1.5 Descriptive evidence

1.5.1 Variable descriptives and sample characteristics

Table 1.1 reports the mean values of the main variables by gender, together with significance tests for the observed differences. A significant gender wage gap (GWG) is evident: women's annual wages amount to 76.60%[7] of men's.

The sample shows similar average ages for men and women, although their demographic and family characteristics differ. The male sample has a higher proportion of White individuals, whereas the female sample has a higher proportion of Black individuals. Unmarried and single women are more prevalent in the sample, reflecting potential underrepresentation of married women due to greater family responsibilities. In addition, women exhibit higher average educational attainment than men, suggesting that they often need to obtain higher qualifications to access job opportunities and positions comparable to those of men.

Job characteristics also differ markedly by gender. Men are more likely to work in higher-paying private-sector jobs, whereas women are more concentrated in lower-paying sectors such as government and non-profit organizations. Men also tend to work longer hours and accumulate more work experience. Gender differences are also evident in working hours and work experience, with men typically working longer hours and accumulating more work experience.

As a key concern of this study, the occupational distribution reveals limited gender differences in non-STEM occupations but substantial disparities in core STEM[8] and STEM-related fields: men are disproportionately represented in core STEM fields, whereas women are more concentrated in STEM-related fields. Another key concern, work from home (WFH) variable, suggests that women were more likely to work remotely during

[7]The value is calculated as the ratio of women's average annual wage (65,473.14) to men's average annual wage (85,475.89) reported in Table 1.1.

[8]The term "core STEM" refers specifically to traditional occupational fields in Science, Technology, Engineering, and Mathematics, excluding STEM-related occupations.

the pandemic, reflecting their greater household and care-giving responsibilities when families faced such crises.

Table 1.1. Mean Values of the Main Variables by Gender (IPUMS 2018–2022)

Variables	**Female (N = 1,689,712)**	**Male (N = 2,193,839)**	**Mean Difference**
Annual wage	65,473.144	85,475.891	19,989.651***
Demographic Characteristics			
age	44.552	44.410	−0.142***
white	0.706	0.736	0.030***
black	0.101	0.070	−0.031***
indian	0.010	0.009	−0.001***
asian	0.022	0.018	−0.003***
migration	0.114	0.116	0.002***
Family Background			
unmarried	0.231	0.226	−0.005***
married	0.565	0.644	0.079***
single	0.204	0.130	−0.074***
nchild	0.849	0.926	0.077***
nparents	0.128	0.124	−0.003***
Educational Background			
degree2	0.059	0.046	−0.014***
low_educ	0.507	0.590	0.083***
bachelor	0.284	0.253	−0.031***
master	0.159	0.109	−0.050***
professional	0.029	0.026	−0.003***
doctoral	0.022	0.022	−0.000
Job-related Characteristics			
private	0.642	0.775	0.132***
government	0.224	0.162	−0.062***
nonprofit	0.133	0.063	−0.070***
stem	0.061	0.140	0.079***
stem-related	0.115	0.035	−0.080***
non-stem	0.825	0.825	0.000
fratio	0.610	0.301	−0.309***
wkhours	43.163	45.040	1.877***
wkexp	9.191	11.803	2.612***
Pandemic-related Variable			
workhome	0.138	0.110	−0.029***

Note: Table presents only the key variables,while information on the full set of variables is reported in the unconditional quantile regression results in Appendix 1.8. *, **, and *** denote significance at the 5%, 1%, and 0.1% levels, respectively, based on two-tailed tests.

1.5.2 Evidence supporting the expanded definition of the STEM workforce

Figure 1.2 presents a Sankey diagram illustrating the relationship among gender, educational attainment, and occupational choice. Consistent with the descriptive statistics in Table 1.1, women tend to have higher educational attainment than men, whereas men are more concentrated in the low-education group. This gender difference narrows as educational attainment rises and nearly disappears at the doctoral level.

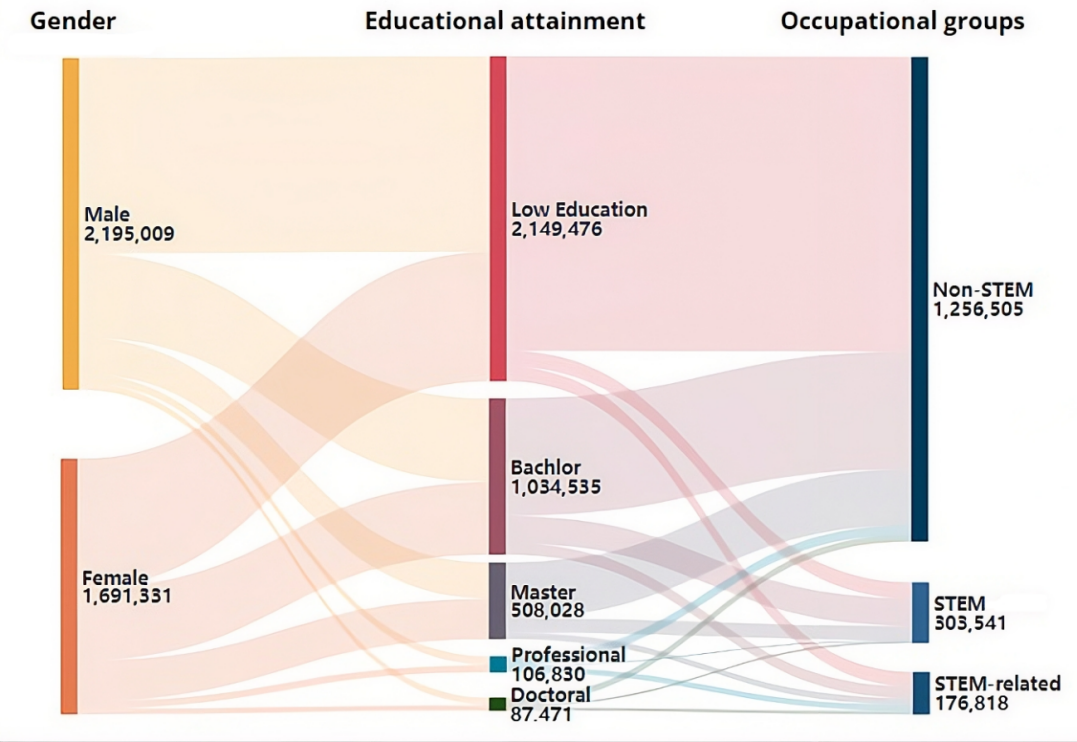


Figure 1.2. Sankey diagram of gender differences in education and occupation (IPUMS 2018-2022)

Notably, a substantial share of individuals (often more than half) with higher education degrees work in non-STEM occupations, while the low-education group accounts for a significant portion of the workforce in STEM and STEM-related occupations. This evidence further supports the National Science Foundation's (NSF) decision to broaden the definition of the STEM workforce to include workers with lower levels of formal education.

1.5.3 Evidence supporting the refined STEM occupational classification

Using IPUMS data from 2000 to 2022, we calculated the Duncan Index[9] under two occupational classification schemes: one that distinguishes between core STEM and STEM-related occupations, and one that does not. The results are presented in Figure 1.3.

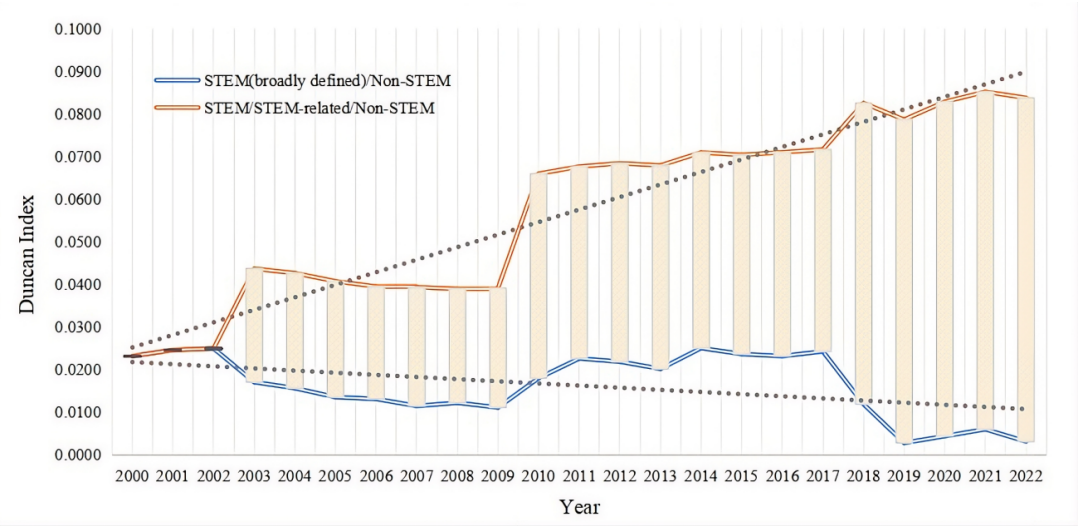


Figure 1.3. Duncan Index under two occupational classification schemes (IPUMS 2000-2022)

The findings show that, since 2002, the Duncan Index derived from the two classification schemes have continued to diverge. When STEM and STEM-related occupations are differentiated, the Duncan Index is substantially higher and displays an upward trend. By contrast, when STEM and STEM-related occupations are treated as a single category, the index follows the opposite pattern. This suggests that men and women have become increasingly concentrated in different types of STEM occupations over time, providing evidence in support of a more refined occupational classification.

[9]The Duncan Index is a widely used measure in labor market research for assessing the extent to which men and women are distributed across occupations. It is calculated as: $D = \frac{1}{2}\sum_{s=1}^{N}\left|\frac{f_s}{F} - \frac{m_s}{M}\right|$, where s denotes the occupation category, f_s and m_s represent the numbers of women and men employed in occupation s, respectively, and F and M denote the total numbers of women and men in the sample.



Figure 1.4. Trends and annual changes in the gender wage gap across quantiles (IPUMS 2000-2022)

1.5.4 Gender wage gap and the COVID-19 pandemic

Figure 1.4 illustrates the gender wage gap (GWG) and its rate of change across quantiles from 2000 to 2022. Consistent with the existing literature, Figure 1.4(a) shows that the

GWG narrowed across the wage distribution since 2000, but remained larger at higher quantiles, indicating a persistent glass ceiling among top earners. Figure 1.4(b) illustrates annual changes in the GWG and shows that the pandemic years (2020–2022) were associated with the most pronounced shifts for the lower-middle wage group (the 25th and 50th quantiles) in nearly two decades. Figure 1.4(c) compares GWG change rates before and during the pandemic. Relative to 2018-2019 (Before pandemic), the GWG narrowed across all quantiles in 2020 (Pandemic outbreak). During 2021-2022, however, the pattern became more heterogeneous across the wage distribution, with the largest changes concentrated at lower quantiles. This suggests that low-wage workers, as a vulnerable group, were more severely affected by pandemic-induced labor market disruption and economic uncertainty, including reductions in working hours and job loss. These findings highlight the need for targeted policies to protect vulnerable workers during and after such economic shocks.

The GWG before and during the pandemic (2018–2022) is further examined by occupational category in Figure 1.5. The results show that the GWG remains consistently lowest in core STEM occupations and highest in STEM-related occupations, especially at higher wage quantiles due to a stronger glass ceiling effect. Before the pandemic (2018-2019), the GWG in core STEM was relatively small and fairly uniform across quantiles. Since the pandemic outbreak in 2020, however, the gap widened at lower quantiles and narrowed at higher quantiles. This pattern suggests the emergence of a "sticky floor" effect, reflecting the greater vulnerability of low-wage women and the relative advantages of high-earning women (such as scarce skills and greater adaptability to teleworking). During the pandemic (2021-2022), both sticky floor and glass ceiling effects are evident, pointing to the lasting impact of the pandemic and related restrictions (e.g., lockdowns) on labor market inequality. By contrast, the glass ceiling effect remained persistent in STEM-related occupations throughout the period, highlighting deeper structural inequalities associated with occupational gender segregation.

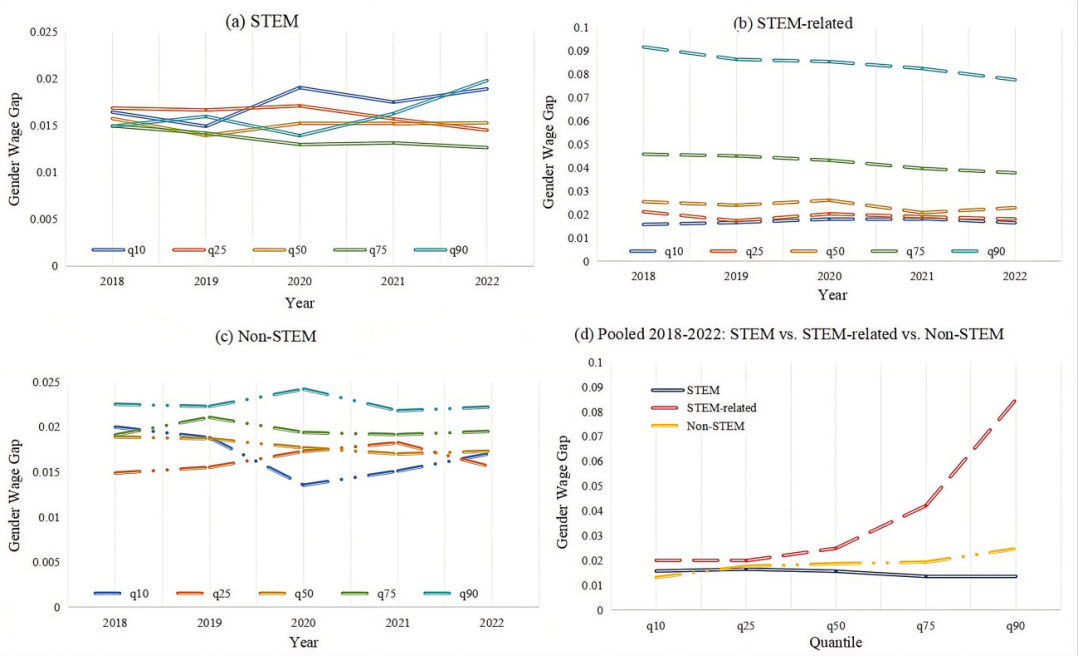


Figure 1.5. Gender wage gap in STEM, STEM-related, and non-STEM occupations (IPUMS 2018-2022)

1.6 Methodology

1.6.1 Unconditional quantile regression

Quantile regression, first introduced by Koenker and Bassett (1978), examines how explanatory variables affect different quantiles of the outcome distribution rather than only its mean. Although informative, this approach can become difficult to interpret when the model includes a large number of highly specific individual characteristics. To address this limitation, Unconditional Quantile Regression (UQR) was developed to estimate the overall, or unconditional, effect of covariates on the outcome distribution without conditioning on specific sample characteristics. Firpo et al. (2009) further extended this approach through the Recentered Influence Function (RIF) regression, which allows the direct estimation of covariate effects at different quantiles of the wage distribution. This makes the method particularly useful for labor economics and policy analysis. As a robust method, RIF regression captures how slight changes in variables affect the overall distribution, thereby providing a clearer picture of shifts in the wage distribution. For the RIF

regressions used below, whether linear or logit, unconditional quantiles can be obtained by integration:

$$\text{RIF}(Y; q_\tau, F_Y) = q_\tau + \text{IF}(Y; q_\tau, F_Y) = q_\tau + \frac{\tau - \mathbf{1}\{Y \leq q_\tau\}}{f_Y(q_\tau)} \tag{1.1}$$

where $RIF(y; q_\tau, F_Y)$ represents the Recentered Influence Function corresponding to the τ-quantile of F_Y, and q_τ is the unconditional quantile of Y, which satisfies $F_Y(q_\tau) = \tau$, and $f_Y(\cdot)$ is the density function of Y. Using the iterative rule of conditional expectation[10], equation (1) can be written as:

$$q_\tau(Y) = \int E\left[\text{RIF}(Y; q_\tau, F_Y) \mid X = x\right] dF_X(x) \tag{1.2}$$

Firpo et al. (2009) makes each component of X undergo a virtual infinitesimal translation transformation (location shift) to obtain the influence of the marginal change of F_X on $q_\tau(Y)$, so the right side of (2) becomes:

$$q_\tau(Y') = \int E\left[\text{RIF}(Y; q_\tau, F_Y) \mid X = x\right] dF_X(x - \Delta x) \tag{1.3}$$

(3)-(2), divide by the increment Δx and let Δx tend to 0, we can get the marginal impact of the unit translation of X on the τ-unconditional quantile of Y, that is, the unconditional quantile partial effect (UQPE):

$$\text{UQPE}(\tau) = \int \frac{\partial E\left[\text{RIF}(Y; q_\tau, F_Y) \mid X\right]}{\partial X} dF_X \tag{1.4}$$

1.6.2 RIF decomposition

According to Firpo et al. (2007a,b, 2009), when the distribution statistics are quantiles, the RIF regression model can directly estimate the effect of covariates on the dependent variables, making it particularly useful for decomposing changes in the wage distribution.

[10]The unconditional expectation of the RIF is the statistic itself. This property forms the basis for RIF regression.

RIF decomposition is exactly the method of wage difference distribution decomposition with the RIF regression model at its core. Based on the above introduction to RIF regression, RIF can be linearly expressed as a function of other explanatory variables:

$$\mathrm{RIF}(\ln Y_i, Q_\tau) = X_i\beta + \varepsilon \tag{1.5}$$

where $\ln Y_i$ is the logarithm of the dependent variable (annual wage), Q_τ is the quantile, X_i is the demographic characteristics, family background, educational background, and job-related explanatory variables. The parameter β can be estimated through the ordinary least squares (OLS) equation.

Decomposition methods based on RIF quantile regression require the construction of counterfactual distribution functions. This process involves using the reweighting function proposed by DiNardo et al. (1996), which allows differences in the wage distribution to be decomposed into two main components: composition effects and wage structure effects. Firstly, after estimating the wage equation with the UQR model, the gender wage difference can be decomposed into two parts:

$$Q_\tau(\ln Y_m) - Q_\tau(\ln Y_f) = [Q_\tau(\ln Y_m) - Q_\tau(\ln Y_c)] + \left[Q_\tau(\ln Y_c) - Q_\tau(\ln Y_f)\right] \tag{1.6}$$

where $\ln Y_c$ represents the statistic of the constructed counterfactual distribution. We construct the counterfactual wage distribution using female characteristic variables and male wage structures, that is, the distribution function of Y_f when the return rates of female labor characteristics are the same as those of males. By substituting Equation (5) into Equation (6), we can obtain:

$$Q_\tau(\ln Y_m) - Q_\tau(\ln Y_f) = \left[(X_m - X_f)\beta_m + \varepsilon_{mc}\right] + \left[(\beta_m - \beta_f)X_f + \varepsilon_{cf}\right] \tag{1.7}$$

Therefore, the first term on the right side of Equation (7) can be regarded as the difference arising from the distinct characteristic variables between men and women, which is the characteristic effect or composition effect. The second term can be understood as the difference caused by the variations in return rates, which is the coefficient effect or wage structure effect.

In summary, RIF decomposition is a method that Firpo et al. (2007a,b, 2009) construct the counterfactual wage distribution with the reweighting function proposed by DiNardo et al. (1996) to decompose the wage distribution into composition effects and wage structure effects, and then regresses the recentered influence function (RIF) of the distribution statistics, and finally subdivides the above two effects further on each covariate.

1.7 Results and Discussion

1.7.1 Unconditional quantile regression results

The results of the unconditional quantile regression (UQR) for male and female workers are reported in Appendix 1.8(Tables B1 and B2, respectively). The general findings regarding demographic characteristics, family background, educational, and job-related characteristics are consistent with the documented facts in the established literature. For example, wages increase with age regardless of gender, though the effect is less pronounced at higher quantiles. Ethnic minorities, particularly Black workers, generally earn lower wages than Whites. Women face a stronger parenthood penalty than men across the wage distribution, regardless of the number of children. Marital status has no negative effect on men's wages at any quantile, and high-earning single women appear to face a penalty that may reflect prejudice. Obtaining an additional degree is associated with higher wages, while further education offers higher returns for women, etc. Therefore, this section focuses primarily on several aspects that have often been overlooked in the existing literature: the role of the pandemic and work from home (WFH), core STEM and STEM-related occupations, as well as low-education workers. The results are presented

in Figure 1.6.

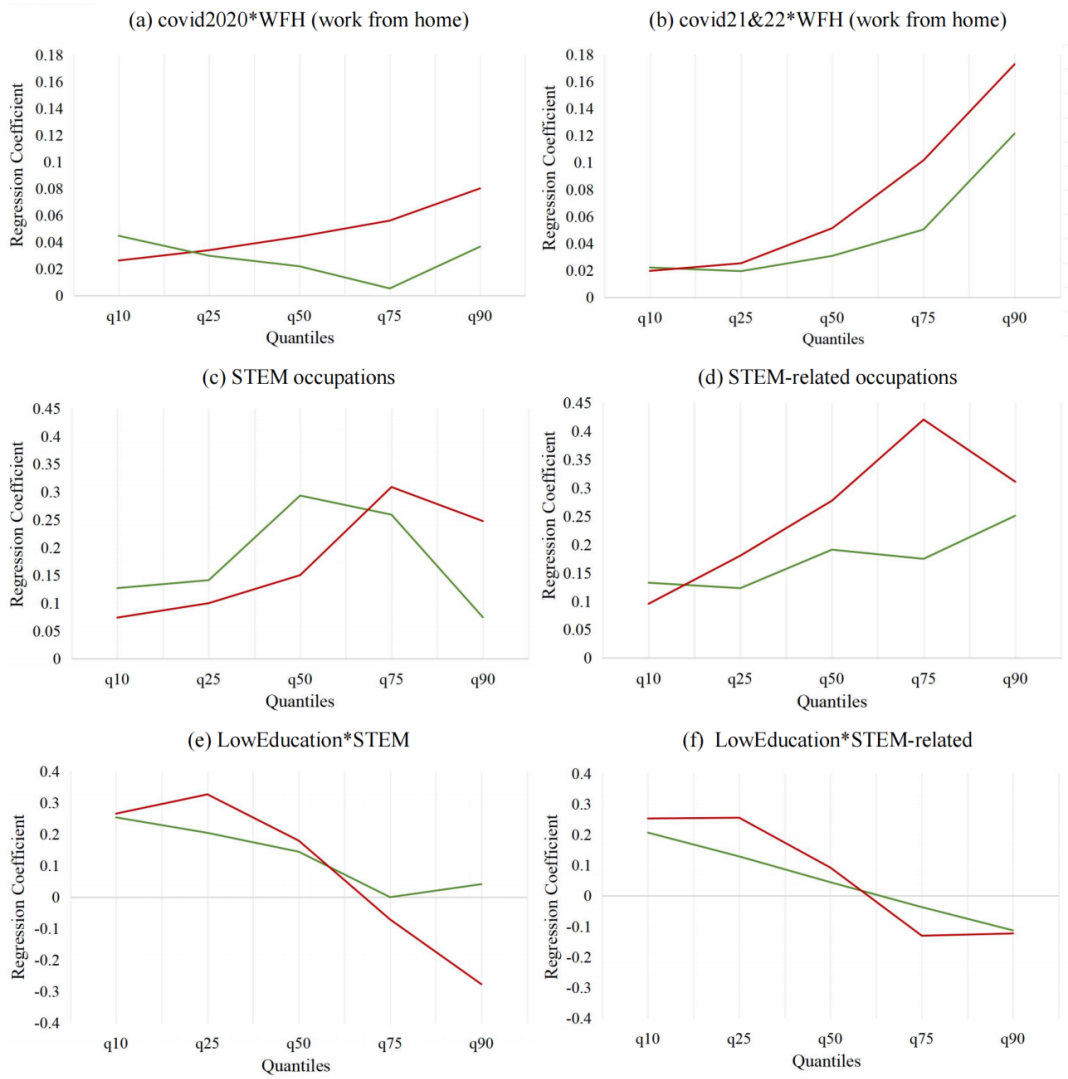


Figure 1.6. Unconditional quantile regression results of the key variables by gender

Figures 1.6(a) and 1.6(b) show that WFH had different effects on male and female wages during the COVID-19 pandemic. The regression results indicate that WFH had a positive effect on wages for both men and women, with particularly pronounced gains for women above the 10th wage percentile. Notably, this positive effect generally became stronger at higher wage quantiles, except for men between the 10th and 75th percentiles in 2020 (Pandemic outbreak), where the pattern was reversed and the WFH wage premium declined as wage levels increased. This suggests that WFH provided a relative advantage to women during the pandemic, potentially contributing to a narrowing of the gender wage gap (GWG). Moreover, during the pandemic period from 2021 to 2022, the largest

positive effect of WFH was observed among high-wage workers of both genders. This suggests that, as the pandemic progressed, higher earners were better positioned to benefit from teleworking, likely because they had greater occupational flexibility and were better able to maintain productivity outside traditional office settings. Overall, these findings highlight the significance of teleworking in shaping wage dynamics during crises such as the COVID-19 pandemic, particularly among higher-income workers.

Figures 1.6(c) and 1.6(d) present the UQR results for STEM and STEM-related occupations. Across the wage distribution, both men and women earn higher wages in STEM and STEM-related occupations than in non-STEM fields. However, the patterns vary across wage quantiles. In STEM occupations, women exhibit smaller wage premiums at the lower and middle quantiles but exceed men at the upper quantiles. Similarly, in STEM-related fields, women generally receive higher wage returns than men, except at the lowest quantile, suggesting that these occupations may offer women more favorable opportunities. Nevertheless, men exhibit stronger wage growth at the top quantile in STEM-related fields, while women at the upper end of the wage distribution experience a marked decline in wage returns in both STEM and STEM-related occupations. This explains the "glass ceiling" effect observed in Figures 1.5(c) and 1.5(d) and suggests that women continue to face persistent barriers to advancement and promotion in these fields.

Furthermore, the interaction effects shown in Figures 1.6(e) and 1.6(f) suggest that, at the lower and middle parts of the wage distribution, lower-educated women earn higher returns than men in both STEM and STEM-related occupations. By contrast, at higher wage levels, low educational attainment generally has a negative effect on wage growth, highlighting the critical role of education in women's career advancement and promotion in these technical occupations.

1.7.2 Recentered influence function decomposition results

This subsection first examines the gender wage gap (GWG) in the U.S. labor market from 2018 to 2022. It then performs a year-by-year analysis to identify the key factors

contributing to the narrowing of the GWG during the pandemic. Finally, it examines how the GWG varies across occupational categories (STEM, STEM-related, and non-STEM).

Table 1.2. RIF Decomposition Results: Pooled Data 2018–2022

Wage Quantiles	**(1)** **q10**	**(2)** **q25**	**(3)** **q50**	**(4)** **q75**	**(5)** **q90**
Gender wage gap	0.156***	0.201***	0.184***	0.256***	0.294***
(Unadjusted gap)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Endowments Effect					
Total endowments effect	0.013***	0.049***	0.101***	0.214***	0.305***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Pandemic-related	−0.001	−0.001	0.000	0.000	0.000
Demographic	0.005	0.004	0.003	0.002	0.001
Family background	0.003	0.005	0.005	0.008	0.012
Education background	−0.029	−0.043	−0.042	−0.047	−0.047
Job-related	0.035	0.083	0.135	0.250	0.338
Coefficients Effect					
Total coefficients effect	0.142***	0.152***	0.083***	0.042***	−0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)

Note: The table reports endowment effects by five variable groups: pandemic-related, demographic, family background, educational, and job-related characteristics. Category-level endowment effects are calculated by summing the corresponding variable-level endowment coefficients; therefore, their standard errors and significance markers are not shown. Bootstrapped standard errors, based on 50 replications, are reported in parentheses. *, **, and *** denote significance at the 5%, 1%, and 0.1% levels, respectively.

Overall gender wage gap decomposition: 2018-2022 pooled data analysis

Table 1.2 reports the recentered influence function (RIF) decomposition results based on pooled cross-sectional data from 2018 to 2022. The results show that the GWG at the lower wage levels is driven primarily by the coefficient effect, whereas the endowment effect becomes dominant at higher wage levels. As the earlier UQR results indicate, women's educational advantage largely disappears at higher wage quantiles. Therefore, this pattern observed in the RIF results is not driven by the observed characteristics that explain the GWG, but rather by the diminishing negative endowment effect of educational characteristics as wage levels increase. These findings help explain the observed "glass ceiling" effect, whereby wage disparities are most pronounced among higher earners. Notably, the results are consistent with previous studies (e.g., Blau and Kahn, 2017; Rotman and Mandel, 2023), which show that traditional human capital variables alone

cannot explain the GWG in the U.S. labor market. Therefore, rather than revisiting these well-established findings, the following analysis focuses on identifying the key factors contributing to the narrowing of the GWG since the onset of the pandemic in 2020 through a year-by-year RIF decomposition.

Annual gender wage gap trends: year-by-year analysis

To identify the factors contributing to the narrowing of the gender wage gap (GWG) across the wage distribution, the analysis examines year-by-year changes in the endowment effects of all observable variables and finds that the endowment effect of work from home (WFH) exhibits a distinctive trend over time.

As shown in Figure 1.7(a), there is a striking step-like decline after the outbreak of the pandemic, a pattern not observed for any other variable. This negative endowment effect became more pronounced during the pandemic, thereby reducing the total endowment effect during this period. To further explore the mechanism underlying this pattern, the proportion of workers engaged in WFH was calculated by occupational category and gender. Figure 1.7(b) shows the highest WFH share in core STEM occupations and the lowest in STEM-related occupations, especially after the pandemic began. Figure 1.7(c) shows that WFH participation rose sharply for both genders after 2020, with a larger increase among women. Together with the UQR results, these findings suggest that WFH helped narrow the GWG during the pandemic, especially at higher wage levels.

Occupation-based gender wag gap: STEM, STEM-related and non-STEM

The gender wage gap (GWG) was then examined across occupational categories, and its underlying determinants were analyzed. The earlier descriptive statistics and UQR analyses reveal an important pattern: substantial gender differences exist between core STEM and STEM-related fields in terms of the distribution of male and female characteristics, the extent of the GWG, and gender-specific wage returns. For example, the GWG in STEM-related occupations is consistently larger than in the other two occupational categories

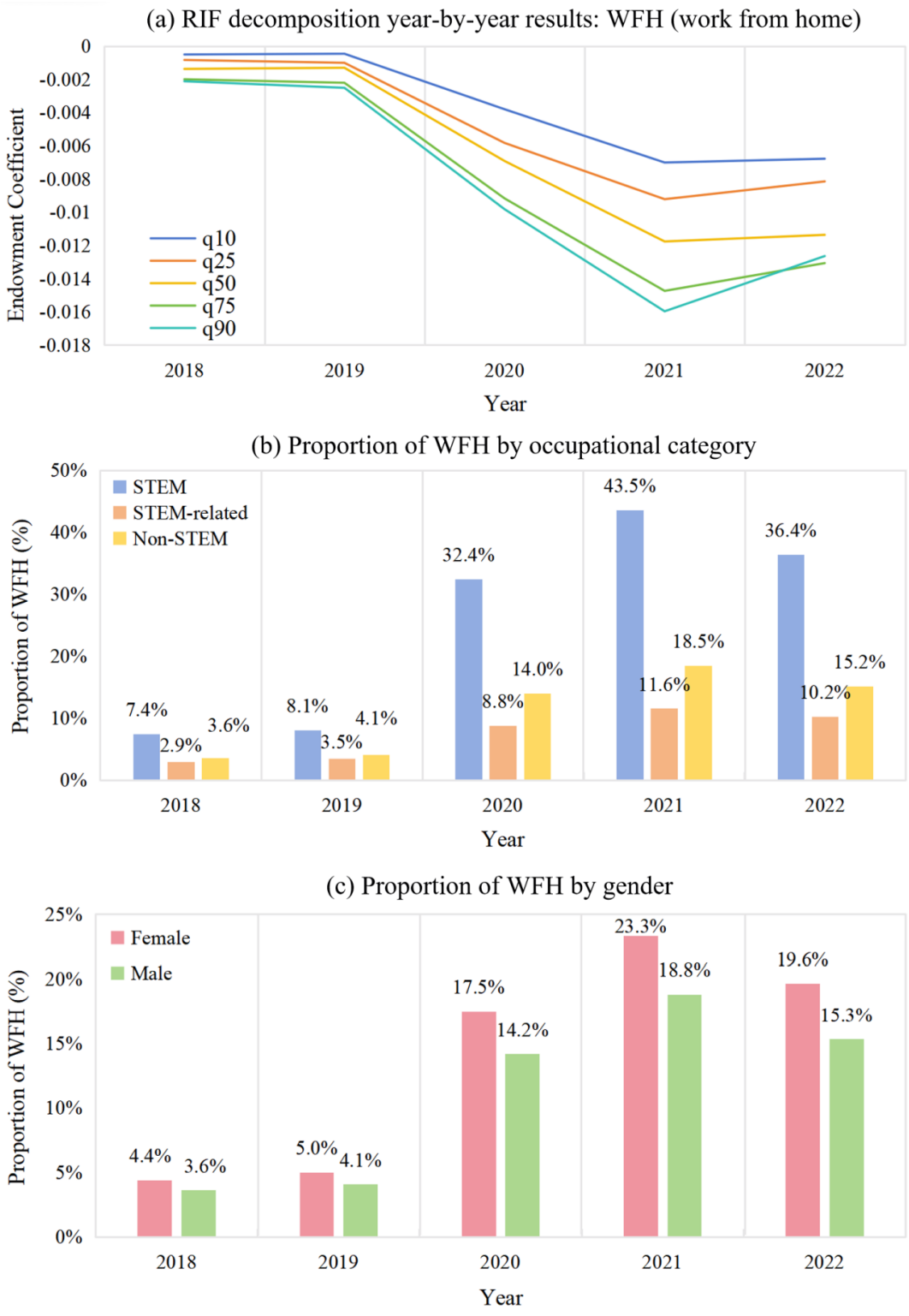


Figure 1.7. The effect of work from home (WFH) on the narrowing of gender wage gap during the pandemic

across all quantiles and exhibits a pronounced "glass ceiling" effect. In contrast, the GWG in core STEM occupations is the smallest at almost all quantiles, suggesting a relatively higher level of gender equality. These findings motivate a deeper exploration of whether the determinants of the GWG also differ across occupational categories.

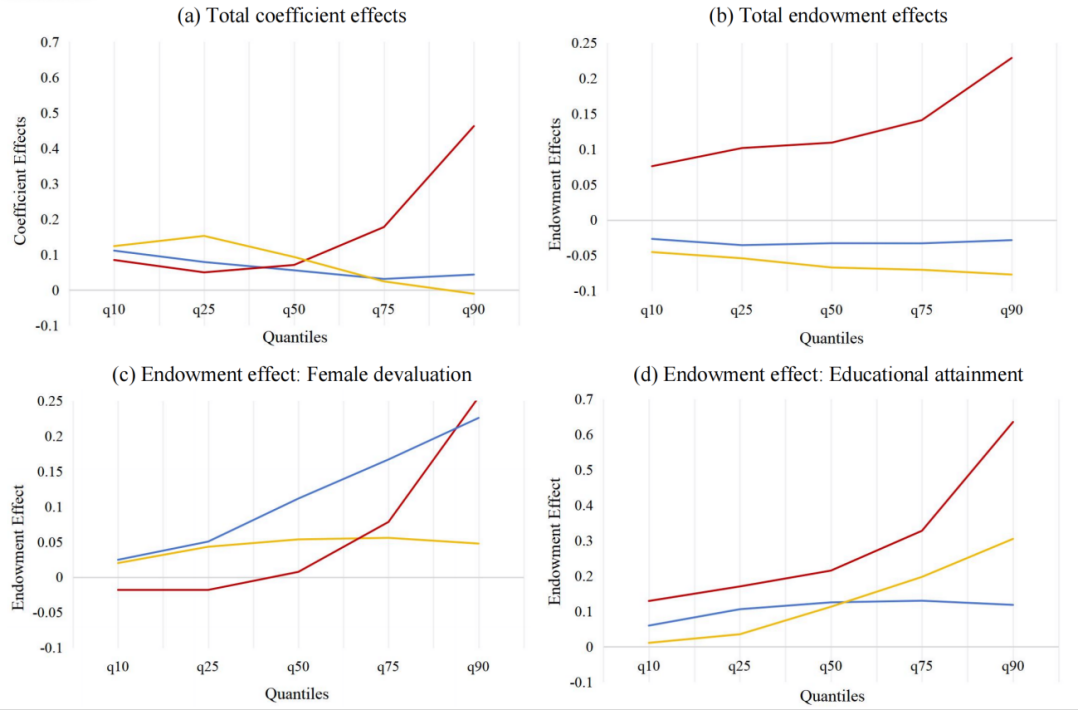


Figure 1.8. Occupation-specific RIF decomposition results

As shown in Figure 1.8(a), the coefficient effect at the upper quantiles is substantially larger in STEM-related fields than in the other two occupational categories. This suggests that high-wage women in STEM-related occupations face stronger gender-based discrimination and greater barriers to advancement into high-paying positions, thereby contributing to the pronounced "glass ceiling" effect observed in these fields. The total endowment effect (Figure 1.8(b)) rises with the wage quantiles, following a pattern similar to that observed in non-STEM occupations. In addition, female devaluation[11] appears to best explain the GWG across all quantiles in both core STEM and non-STEM occupations and also explains a substantial share of the GWG at higher quantiles in STEM-related occupations, as shown in Figure 1.8(c). Notably, although educational attainment is the main factor explaining the GWG in STEM-related occupations, it does not explain the

[11]This is measured as the proportion of women in each occupation.

GWG in either core STEM or non-STEM occupations at any quantile, as reflected in the persistently negative endowment-effect coefficients shown in Figure 1.8(d), highlighting structural differences in how educational credentials are rewarded across occupational categories.

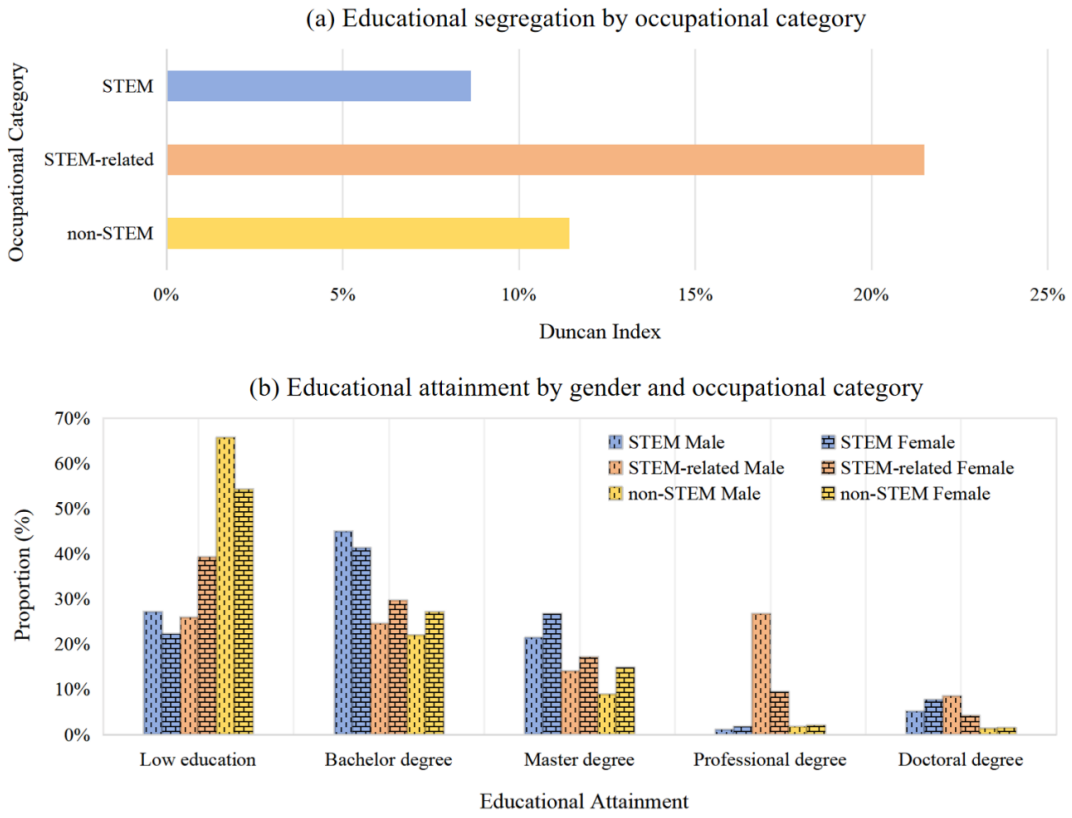


Figure 1.9. Educational segregation and educational attainment by occupational category

To further explore the underlying reasons for this significant difference, educational segregation[12] was examined across occupational categories, as shown in Figure 1.9. As expected, the distribution of educational attainment between men and women is more uneven in STEM-related occupations, where men are more likely to hold professional or doctoral degrees. By contrast, women tend to have higher educational attainment in the other two occupational categories.

In summary, the above analyses reveal that the determinants of the GWG differ across STEM, STEM-related, and non-STEM occupations. The female devaluation effect consistently explains the GWG across occupational categories, particularly at higher quantiles in

[12]This is measured as the Duncan Index (see footnote 10).

STEM-related occupations, highlighting a pervasive and persistent ”glass ceiling” effect. Moreover, although educational attainment is negatively associated with the GWG in core STEM and non-STEM occupations, it remains the primary explanatory factor in STEM-related occupations because of greater educational segregation. These findings underscore the need for targeted policies not only to promote gender equality in career advancement, but also to address the structural educational barriers that continue to perpetuate wage disparities across these fields.

1.8 Conclusion

This study provides an in-depth analysis of the gender wage gap (GWG) in the U.S. labor market over the period from 2018 to 2022, with particular attention to STEM and STEM-related occupations in the context of the COVID-19 pandemic. The main findings are as follows.

First, the year-by-year analysis reveals a marked narrowing of the GWG across the wage distribution since the outbreak of the COVID-19 pandemic, with work from home (WFH) playing a key role in this trend. Second, while traditional human capital variables hardly explain recent changes in the GWG, job-related characteristics, such as potential work experience, working hours, and occupational categories, emerged as the main explanatory factors, with female devaluation playing a particularly important role. Third, as the core of this study, the analysis across occupational categories (STEM, STEM-related, and non-STEM) reveals substantial differences in both the magnitude and determinants of the GWG. The GWG is the largest in STEM-related occupations, accompanied by a pervasive and persistent ”glass ceiling” effect. In contrast, the gap is considerably smaller in core STEM occupations. Across occupational categories, female devaluation emerges as the key explanatory factor. Although educational attainment is a major determinant of the GWG in STEM-related occupations, it does not explain the gap in core STEM and non-STEM occupations, where women generally have higher levels of educational attainment

than men. The large GWG in STEM-related occupations appears to be driven primarily by educational segregation, with men more likely to hold advanced professional degrees that are especially highly rewarded in these occupations.

In summary, these findings highlight the complex interplay between educational and occupational characteristics in shaping the GWG, while also underscoring the positive role of teleworking during the pandemic in narrowing the gap. They further point to the importance of occupation-specific policies for advancing gender equality in the labor market.

Nonetheless, the study also has several limitations. First, the IPUMS data may not fully capture all factors affecting the GWG in the U.S. labor market, particularly during the pandemic period. Second, selection bias remains a potential concern and may affect the robustness of the results. In addition, the RIF decomposition method has certain limitations. Some previous studies using this approach have found a large "unexplained" component of the GWG, which may be attributable to heterogeneity within the sample (Kassenboehmer and Sinning, 2014; Töpfer, 2017; Meara et al., 2020). Therefore, future research is recommended to use more comprehensive data, where available, and further refine the empirical approach to improve the accuracy and generalizability of the findings.

Based on the above findings, we propose several policy recommendations:

(1) Develop occupation-specific policies: In STEM-related occupations, policy efforts should focus on breaking the glass ceiling, whereas in STEM and non-STEM occupations, policies aimed at raising the minimum wage for women may be more effective.

(2) Address educational segregation: Although women have achieved notable educational success in STEM and non-STEM occupations, they remain underrepresented in advanced professional degree programs associated with STEM-related occupations, such as healthcare. Policies should therefore prioritize support for women's pursuit of professional qualifications that facilitate career advancement in these fields.

(3) Promote gender equality across all STEM-related occupations: Although core STEM occupations often receive greater attention, STEM-related occupation exhibit more

pronounced gender inequality and therefore deserve equal policy attention and resource allocation.

(4) Encourage flexible work arrangements: Policies should also promote flexible and hybrid work arrangements, given the observed role of teleworking in mitigating the GWG, particularly benefiting for women.

Appendix A: Estimation of the potential work experience

The study examined two main methods of estimating potential work experience in the preliminary analysis. The first method was proposed by Mincer and Polachek (1974), which simply estimates work experience by subtracting years of schooling and the school-entry age (generally 5-6 years old in the U.S.) from an individual's reported age:

$$wkexp1 = Age - Schooling - 5 \tag{A1}$$

The second method is the cumulative probability method (CPM), which was adopted in the main analysis because it has been shown to yield more accurate and robust estimates by accounting for labor market interruptions. More specifically, a probit model was used to estimate the probability of employment across age profiles separately for men and women in each year from 2018 to 2022.

The model used to estimate the probability of employment is defined as follows:

$$\begin{aligned} P(employ_{t,g} = 1 \mid x) = {} & \alpha_{0,t,g} + \alpha_{1,t,g} married + \alpha_{2,t,g} single + \alpha_{3,t,g} parent \\ & + \alpha_{4,t,g} migrant + \alpha_{5,t,g} race + \alpha_{6,t,g} region + \alpha_{7,t,g} educ \end{aligned} \tag{A2}$$

where $employ_{t,g}$ is a binary variable indicating whether an individual of gender g is employed in year t (t=2018, 2019, 2020, 2021, 2022). The explanatory variables including marital status, parental status, immigration status, race, region and education level. $\alpha_{0,t,g}, \alpha_{1,t,g}, \alpha_{2,t,g}, ...$ are the parameters to be estimated.

By summing the estimated employment probabilities across ages, the study then computes the expected work experience for an individual, thereby providing a comprehensive measure of work experience for each year:

$$wkexp2 = wkexp_g = \sum_{t=25}^{65} P(employ_{t,g} = 1 \mid x) \tag{A3}$$

where $wkexp_g$ represents the cumulative expected work experience for individuals of gender g throughout their working life, from age 25 to 65.

The study then performed a preliminary analysis to determine the most appropriate method for the main analysis. The work experience variable estimated using the Mincer method is denoted by *"wkexp1*," whereas the variable estimated using the CPM method is denoted by *"wkexp2"*. Table A1 reports the descriptive statistics for the work experience estimated by these two methods.

Table A1. Means for the Two Estimated Work Experiences, Females versus Males

Variables	**Female (N = 1,689,712)**	**Male (N = 2,193,839)**	**Mean Difference**
wkexp1	24.5987	24.9483	0.3496***
wkexp2	9.1912	11.8032	2.6120***

Note: *, **, and *** denote significance at the 5%, 1%, and 0.1% levels, respectively, based on two-tailed tests.

As shown, the gender difference is much smaller when work experience is estimated using the Mincer method. By contrast, the CPM accounts for labor force interruptions, gender-specific characteristics, and year-specific economic conditions, thereby revealing a more substantial gender difference in work experience.

The analysis then compares the unconditional quantile regression (UQR) results obtained using the two work experience estimation methods. Figure A1 shows the effects of work experience, estimated using the two different methods, on wages across quantiles for both men and women. Overall, the cumulative probability method (CPM) yields higher returns than the Mincer method, especially for women. The difference in estimated returns between the two methods is more pronounced at the lower and middle quantiles for both men and women. This suggests that the CPM, by accounting for employment interruptions, captures work experience and its wage returns more accurately for individuals with less stable employment histories. However, this difference narrows at the 75th and 90th percentiles, possibly reflecting the fact that higher earners experience fewer labor force interruptions, resulting in a more consistent valuation of experience across the two methods. Overall, this pattern underscores importance of accounting for work interruptions

when assessing the wage effects of experience, especially among workers in the lower and middle parts of the wage distribution.

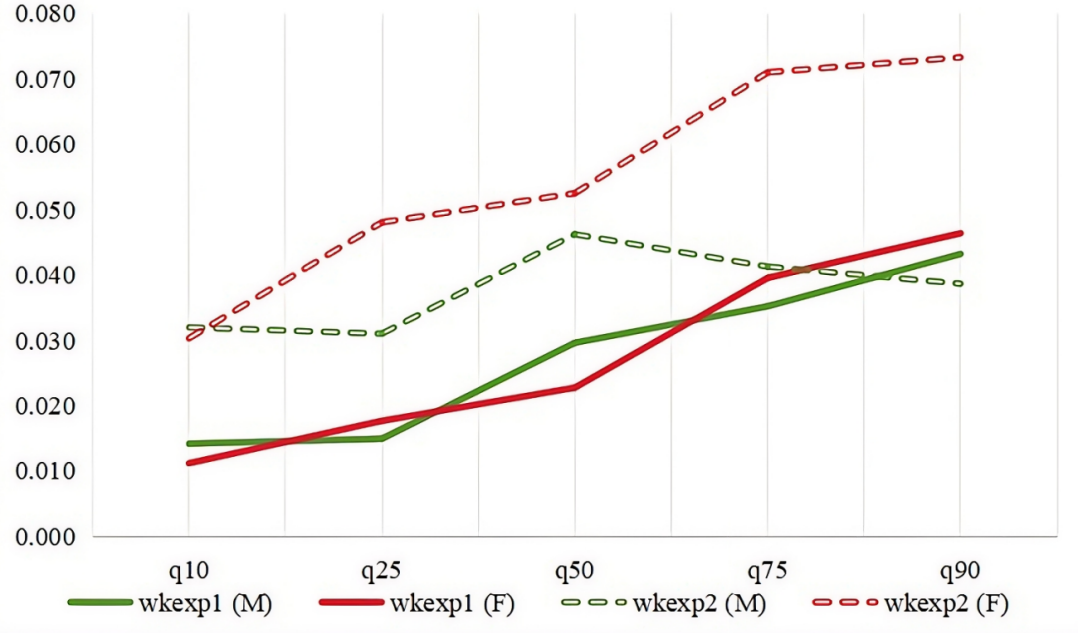


Figure A1. UQR results for the two estimated work experiences by gender

Moreover, the results show that returns to work experience are generally higher for women than for men across most wage quantiles. This suggests some progress toward gender equity in the recognition and reward of work experience, as well as a possible narrowing of the experience-related component of the GWG. Moreover, these findings suggest that job interruptions and career continuity affect returns to experience differently by gender, with important implications for the analysis of the gender wage gap and highlight the careful measurement of potential work experience.

Figure A2 presents the RIF decomposition results, allowing comparison of changes in the different components of the GWG across the three defined phases of the pandemic. The endowment effect declines slightly over time across wage levels, suggesting that gender differences in observable characteristics have narrowed, possibly because women have made greater progress in education, work experience, remote work, and related dimensions. As for the coefficient effect, a decline is observed in 2020 at lower wage levels, suggesting that gender gaps associated with unobservable characteristics and discrimi-

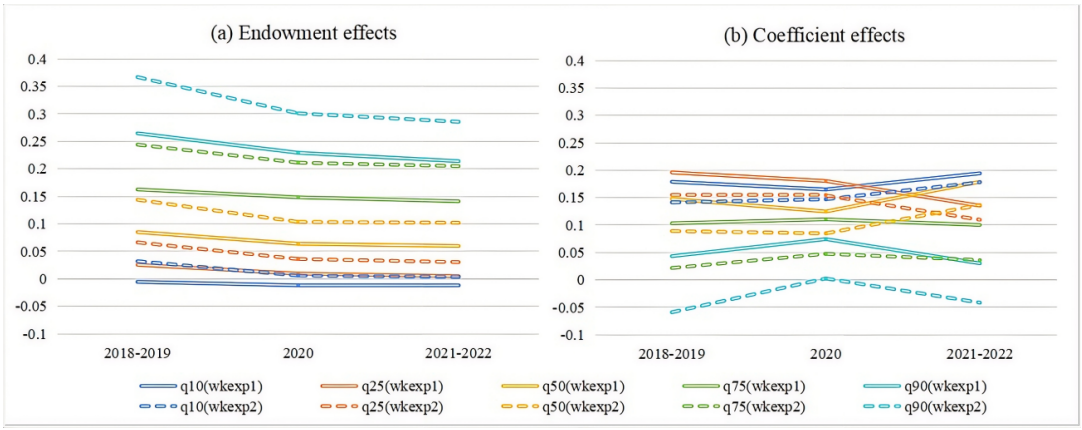


Figure A2. RIF decomposition results using two methods of work experience across the three defined phases of the pandemic: Endowment and coefficient effects

nation against women may have narrowed among low earners. However, the coefficient effect increases in 2020 at higher wage levels, specifically at the 75th and 90th quantiles. This suggests that, among higher earners, the GWG persists because of deeper structural factors and discriminatory practices, possibly linked to the ”glass ceiling.” Comparing the two measures of potential work experience, the CPM method (*wkexp2*) yields a larger endowment effect and a smaller coefficient effect at higher wage levels, particularly at the 90th quantile. By contrast, the Mincer method (*wkexp1*) shows a similar pattern at the lower and middle parts of the wage distribution.

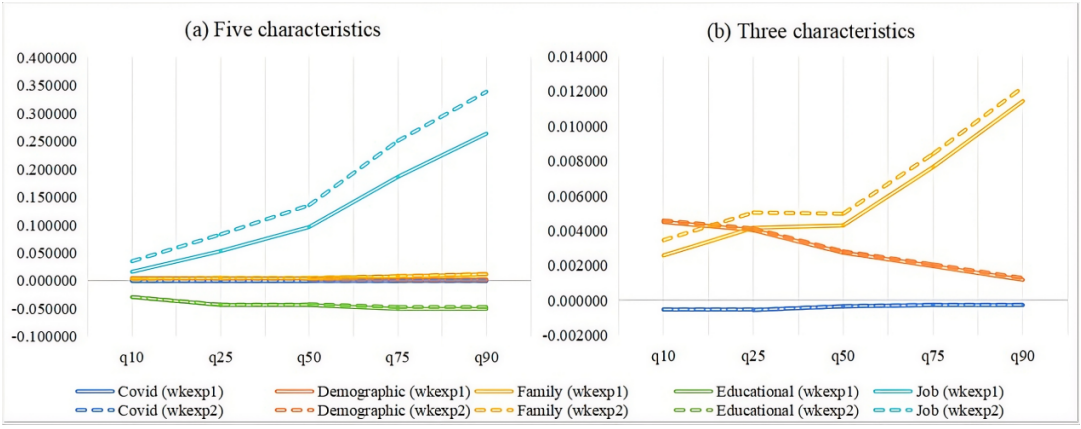


Figure A3. Endowment effects with the two work experience estimation methods

Figure A3 further shows the endowment effects of various characteristics obtained by alternatively including the two measures of potential work experience. The results show that the CPM yields a larger endowment effect of work experience at all quantiles. In other

words, the CPM-based estimation of potential work experience has greater explanatory power for the GWG.

Given the substantial endowment effect observed for job-related characteristics, the analysis further examines the endowment effects of individual job-related variables to identify the most important factors underlying the GWG. As shown in Figure A4, female devaluation best explains the GWG among all job-related characteristics. In addition, with respect to work experience, the CPM-based estimation explains the GWG far better than the Mincer-based estimation.

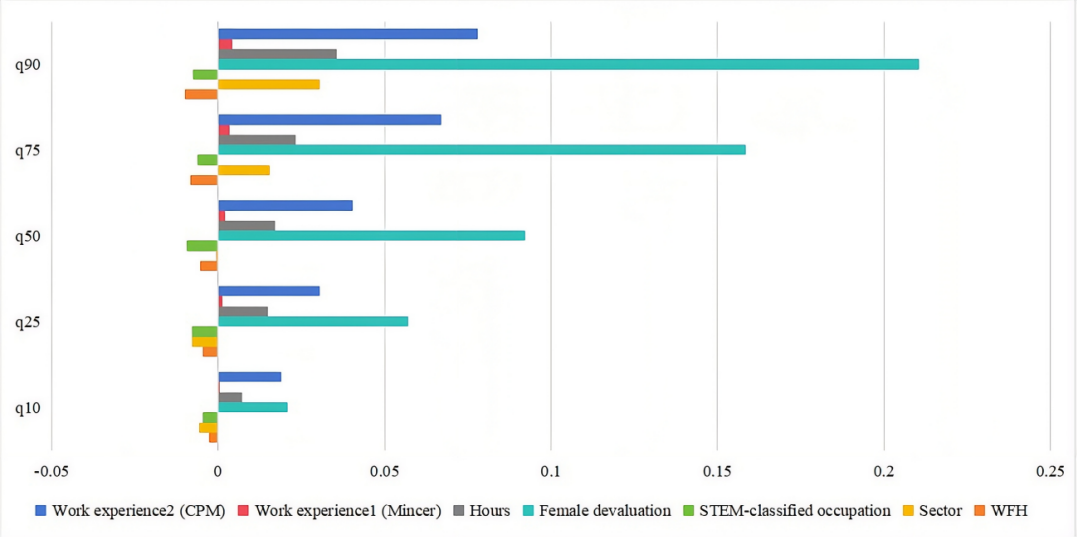


Figure A4. Endowment effects of job-related characteristics

In summary, the method used to estimate potential work experience can lead to different results in both the UQR and RIF decomposition analyses. Potential work experience estimated using the CPM is substantially lower for both men and women than that estimated using the Mincer method. The UQR results show that, using CPM-based estimates, women receive higher returns to experience than men at almost all quantiles. The RIF decomposition further shows that CPM-based work experience, which accounts for labor force interruptions, better explains the GWG across the wage distribution during 2018–2022. These findings highlight the importance of accounting for labor force interruptions when examining the GWG in the U.S. labor market, especially during the pandemic.

Appendix B: Unconditional quantile regression results

Table B1. Unconditional Quantile Regression Results for Male Workers in the United States, Using 2018–2022 Pooled Cross-sectional Data

Wage Quantile	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
	q10	**q25**	**q50**	**q75**	**q90**
Panel A. Pandemic-related Variables					
covid2020	0.054***	0.040***	0.043***	0.028***	0.033***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
covid21&22	0.116***	0.080***	0.073***	0.043***	0.053***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
workhome	0.027***	0.055***	0.161***	0.226***	0.243***
	(0.006)	(0.003)	(0.004)	(0.004)	(0.006)
Panel B. Demographic Characteristics					
black	−0.235***	−0.167***	−0.162***	−0.108***	−0.104***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
indian	−0.256***	−0.200***	−0.188***	−0.110***	−0.088***
	(0.007)	(0.004)	(0.005)	(0.005)	(0.007)
asian	−0.127***	−0.070***	−0.046***	0.046***	0.112***
	(0.005)	(0.003)	(0.003)	(0.003)	(0.005)
other	−0.202***	−0.159***	−0.149***	−0.079***	−0.067***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
more_than_one	−0.103***	−0.089***	−0.091***	−0.064***	−0.074***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
migration	−0.040***	−0.024***	−0.020***	−0.007**	0.005**
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
northeast	0.042***	0.053***	0.122***	0.137***	0.185***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
south	−0.103***	−0.057***	0.004**	0.064***	0.092***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)

Continued on next page

Table B1. Unconditional Quantile Regression Results for Male Workers in the United States, Using 2018–2022 Pooled Cross-sectional Data (continued)

Wage Quantile	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
	q10	**q25**	**q50**	**q75**	**q90**
west	0.007***	0.027***	0.116***	0.160***	0.218**
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Panel C. Family Background					
child1_2	−0.006**	−0.002	0.010***	0.015***	0.012***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
child3_4	−0.050***	−0.031***	−0.007***	0.022***	0.040***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
*child5_*	−0.114***	−0.082***	−0.061***	−0.027***	−0.035***
	(0.008)	(0.005)	(0.005)	(0.005)	(0.008)
married	0.224***	0.176***	0.179***	0.120***	0.113***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
single	0.060***	0.049***	0.035***	0.019***	0.008**
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
childage0_5	0.022***	0.023***	0.030***	0.044***	0.061***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
*sin*chilage0_5*	0.004	−0.016*	−0.013**	−0.006	0.020*
	(0.010)	(0.006)	(0.007)	(0.007)	(0.010)
childage6_17	0.005**	0.011***	0.043***	0.078***	0.141***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
*sin*chilage6_17*	0.081***	0.059***	0.033***	−0.021***	−0.068***
	(0.006)	(0.004)	(0.004)	(0.004)	(0.006)
nparents	−0.153***	−0.112***	−0.084***	−0.025***	0.002
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Panel D. Educational Background					
school	−0.052***	−0.050***	−0.075***	−0.075***	−0.098***
	(0.004)	(0.002)	(0.002)	(0.002)	(0.004)

Continued on next page

Table B1. Unconditional Quantile Regression Results for Male Workers in the United States, Using 2018–2022 Pooled Cross-sectional Data (continued)

Wage Quantile	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
	q10	**q25**	**q50**	**q75**	**q90**
degree2	0.008***	0.007***	0.018	0.022***	0.047***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
bachelor	0.325***	0.306***	0.451***	0.427***	0.490***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
master	0.340***	0.342***	0.577***	0.638***	0.823***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
professional	0.321***	0.335***	0.644***	0.867***	1.477***
	(0.005)	(0.003)	(0.003)	(0.003)	(0.005)
doctoral	0.324***	0.338***	0.614***	0.766***	1.065***
	(0.005)	(0.003)	(0.003)	(0.003)	(0.005)
Panel E. Job-related Characteristics					
*wkhm*cov2020*	0.044***	0.029***	0.021***	0.005	0.037***
	(0.007)	(0.004)	(0.005)	(0.005)	(0.007)
*wkhm*cov21&22*	0.021***	0.019***	0.030***	0.050***	0.121***
	(0.006)	(0.004)	(0.004)	(0.004)	(0.006)
nonprofit	−0.056***	−0.069***	−0.157***	−0.219***	−0.308***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
government	0.128***	0.088***	−0.004***	−0.157***	−0.345***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
stem	0.124***	0.139***	0.292***	0.266**	0.075***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
stem-related	0.133***	0.123***	0.191***	0.174***	0.250***
	(0.005)	(0.003)	(0.003)	(0.003)	(0.005)
*leduc*stem*	0.254***	0.205***	0.145***	0.000	0.041***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.005)
*leduc*stemr*	0.206***	0.128***	0.044***	−0.036***	−0.112***
	(0.009)	(0.005)	(0.006)	(0.006)	(0.009)

Continued on next page

Table B1. Unconditional Quantile Regression Results for Male Workers in the United States, Using 2018–2022 Pooled Cross-sectional Data (continued)

Wage Quantile	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
	q10	**q25**	**q50**	**q75**	**q90**
fratio	−0.158***	−0.111***	−0.086***	−0.060**	−0.072***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
wkhours	0.008***	0.009***	0.012***	0.011***	0.015***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
wkexp1	0.032***	0.031***	0.046***	0.041***	0.039***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
wkexp2	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
constant	9.456***	9.817***	9.790***	10.340***	10.502***
	(0.005)	(0.003)	(0.004)	(0.003)	(0.005)

Note: Standard errors are reported in parentheses. *, **, and *** denote significance at the 5%, 1%, and 0.1% levels, respectively.

Table B2. Unconditional Quantile Regression Results for Female Workers in the United States, Using 2018–2022 Pooled Cross-sectional Data

Wage Quantile	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
	q10	**q25**	**q50**	**q75**	**q90**
Panel A. Pandemic-related Variables					
covid2020	0.044***	0.038***	0.023***	0.016***	0.019***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)
covid21&22	0.082***	0.084***	0.052***	0.041***	0.040***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
workhome	0.072***	0.133***	0.137***	0.210***	0.207***
	(0.005)	(0.004)	(0.003)	(0.004)	(0.006)
Panel B. Demographic Characteristics					

Continued on next page

Table B2. Unconditional Quantile Regression Results for Female Workers in the United States, Using 2018–2022 Pooled Cross-sectional Data (continued)

Wage Quantile	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
	q10	**q25**	**q50**	**q75**	**q90**
black	−0.104***	−0.097***	−0.066***	−0.037***	−0.033***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)
indian	−0.145***	−0.159***	−0.120***	−0.093***	−0.050***
	(0.007)	(0.005)	(0.004)	(0.006)	(0.009)
asian	−0.088***	−0.050***	0.001	0.106***	0.261***
	(0.005)	(0.004)	(0.003)	(0.004)	(0.006)
other	−0.145***	−0.143***	−0.088***	−0.038***	−0.011**
	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)
more_than_one	−0.060***	−0.063***	−0.039***	−0.031***	−0.034***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.004)
migration	−0.032***	−0.020***	−0.010***	−0.005**	0.001
	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)
northeast	0.051***	0.098***	0.129***	0.196***	0.202***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)
south	−0.088***	−0.048***	0.003**	0.041***	0.081***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
west	0.052***	0.098***	0.126***	0.206***	0.240***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)
Panel C. Family Background					
child1_2	−0.049***	−0.054***	−0.037***	−0.028***	−0.023***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
child3_4	−0.128***	−0.125***	−0.080***	−0.056***	−0.033***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.004)
*child5_*	−0.234***	−0.210***	−0.143***	−0.113***	−0.098***
	(0.010)	(0.008)	(0.006)	(0.008)	(0.013)
married	0.073***	0.076***	0.045***	0.049***	0.062***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)

Continued on next page

Table B2. Unconditional Quantile Regression Results for Female Workers in the United States, Using 2018–2022 Pooled Cross-sectional Data (continued)

Wage Quantile	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
	q10	**q25**	**q50**	**q75**	**q90**
single	0.014***	0.014***	−0.002	−0.012***	−0.018***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
childage0_5	0.036***	0.064***	0.074***	0.134***	0.190***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.004)
*sin*chilage0_5*	−0.060***	−0.063***	−0.034***	−0.027***	−0.034***
	(0.008)	(0.006)	(0.005)	(0.007)	(0.010)
childage6_17	0.005**	0.016***	0.030***	0.087***	0.147***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
*sin*chilage6_17*	0.024***	0.017***	0.004	−0.018***	−0.037***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.006)
nparents	−0.052***	−0.069***	−0.052***	−0.036***	−0.018***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Panel D. Educational Background					
school	−0.019***	−0.034***	−0.044***	−0.062***	−0.071***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.004)
degree2	0.005	0.011***	0.022***	0.042***	0.057***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.004)
bachelor	0.330***	0.461***	0.403***	0.424***	0.393***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
master	0.361***	0.543***	0.551***	0.637***	0.626***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)
professional	0.347***	0.508***	0.561***	0.886***	1.458***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.006)
doctoral	0.340***	0.511***	0.578***	0.892***	1.153***
	(0.005)	(0.003)	(0.003)	(0.004)	(0.006)
Panel E. Job-related Characteristics					

Continued on next page

Table B2. Unconditional Quantile Regression Results for Female Workers in the United States, Using 2018–2022 Pooled Cross-sectional Data (continued)

Wage Quantile	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
	q10	**q25**	**q50**	**q75**	**q90**
*wkhm*cov2020*	0.026***	0.033***	0.044***	0.056***	0.080***
	(0.007)	(0.005)	(0.004)	(0.006)	(0.008)
*wkhm*cov21&22*	0.019***	0.025***	0.051***	0.101***	0.172***
	(0.006)	(0.005)	(0.003)	(0.005)	(0.007)
nonprofit	0.022***	0.016***	−0.030***	−0.113***	−0.178***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)
government	0.063***	0.103***	0.031***	−0.127***	−0.296***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
stem	0.073***	0.099***	0.149***	0.308***	0.247***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.004)
stem-related	0.096***	0.180***	0.277***	0.420***	0.310***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.004)
*leduc*stem*	0.266***	0.328***	0.180***	−0.070***	−0.277***
	(0.007)	(0.005)	(0.004)	(0.006)	(0.009)
*leduc*stemr*	0.253***	0.255***	0.092***	−0.130***	−0.122***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.006)
fratio	−0.065***	−0.180***	−0.289***	−0.508***	−0.678***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.004)
wkhours	0.004***	0.008***	0.009***	0.012***	0.019***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
wkexp1	0.030***	0.048***	0.052***	0.071***	0.073***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
wkexp2	−0.001***	−0.002***	−0.002***	−0.003***	−0.002***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
constant	9.609***	9.620***	10.026***	10.215***	10.390***
	(0.006)	(0.005)	(0.004)	(0.005)	(0.007)

Note: Standard errors are reported in parentheses. *, **, and *** denote significance at the 5%, 1%, and 0.1% levels, respectively.

Chapter 2

Gender Income Gap in the Creator Economy: Evidence from U.S. Top YouTubers

2.1 Introduction

A large empirical literature documents persistent gender differences in earnings across labor markets. Although the gender wage gap has narrowed over time, women in the United States still earn roughly 15-20% less than men on average (Blau and Kahn, 2017; Kochhar, 2023). Differences in education and observable human capital characteristics now explain only a limited portion of this disparity, while occupational segregation, career interruptions associated with family responsibilities, and the "child penalty" remain central mechanisms (Blau and Kahn, 2017; Kleven et al., 2019). High-paying occupations also tend to reward long and inflexible working hours disproportionately, further widening gender differences in earnings (Goldin, 2014). Beyond these institutional structures, behavioral responses and social norms may also contribute to the gap: women are less likely to enter competitive environments and to engage in self-promotion even when their ability is similar to that of men (Croson and Gneezy, 2009; Bertrand, 2011; Niederle and Vesterlund, 2007, 2011; Fortin, 2015; Exley and Kessler, 2022). Despite extensive research, identifying the sources of the gender wage gap remains challenging because most labor markets combine multiple mechanisms simultaneously. Workers' productivity is imperfectly observed, earnings are mediated by employers, and wages depend on promotion systems, negotiation, and internal evaluation (Dohmen, 2004; Lazear and Oyer, 2004). Consequently, it is difficult to separate whether gender differences arise from productivity, labor supply choices, institutional constraints, or discrimination.

The expansion of digital labor markets now offers new opportunities to study these

mechanisms. Gig economy platforms reorganize work through algorithmic coordination and flexible labor supply, thereby reducing some traditional workplace frictions. Evidence from these settings shows that gender pay gaps persist but operate through different channels. For example, male Uber drivers earn about 7% more per hour than female drivers, a gap largely explained by differences in experience accumulation, location choices, and work patterns rather than direct discrimination (Cook et al., 2020). Similarly, women on online labor platforms often sort into lower-paying tasks or contracts (Litman et al., 2020; Teutloff et al., 2025), and even consumer-to-consumer markets exhibit lower prices for identical products sold by women (Kricheli-Katz and Regev, 2016). These findings suggest that removing traditional employment relationships does not eliminate gender inequality, but instead changes its mechanisms. The emergence of the creator economy represents a further transformation of labor market organization. Digital platforms such as YouTube, Instagram, TikTok, and Facebook increasingly function as labor infrastructures that enable income generation, entrepreneurship, and self-employment (Berg et al., 2018; Cunningham and Craig, 2019; ILO, 2021; OECD, 2023). Among these platforms, YouTube stands out as one of the largest and most established markets for digital content creation, reaching over 2.5 billion monthly active users worldwide and approximately 238 million in the United States (Analyzify, 2025). Consistent with its scale, YouTube's creative ecosystem contributed over $55 billion to U.S. GDP in 2024 and supported nearly 490,000 full-time equivalent jobs (YouTube, 2025).

Unlike traditional employment, or even gig work, creators are neither hired nor paid wages by firms. Instead, their earnings depend primarily on audience demand, measured through publicly observable performance metrics such as views, engagement, and advertising monetization (YouTube, 2021). Compensation is therefore not determined by promotion, bargaining, or employer evaluation, and individual output is visible to both researchers and consumers. In principle, such a market could approximate a meritocratic system in which income closely reflects performance. However, digital media markets also exhibit pronounced "superstar" effects, a winner-take-all dynamic where a small mi-

nority captures a disproportionate share of attention and revenue (Rosen, 1981; Frank and Cook, 1996; Krueger, 2005; Koenig, 2023). On YouTube, approximately 3% of channels account for roughly 85% of all views, while a small elite of top creators receive the vast majority of revenues and most creators earn little (Bärtl, 2018; Verwiebe et al., 2026). At the same time, emerging evidence points to systematic inequalities in participation and visibility: female creators are underrepresented among top creators and often experience lower engagement despite greater effort and persistence on the platform (Leighton, 2020; Eilon, 2024; Gioia and Morabito, 2025; Eizmendi-Iraola et al., 2025). Algorithmic recommendation systems may further reinforce these disparities by amplifying already dominant creators (Ribeiro et al., 2023; Ionescu et al., 2023; Corsi, 2024; Lou et al., 2025), while audience and advertiser responses may also vary by gender (Wotanis and McMillan, 2014; Nakandala et al., 2017; Döring and Mohseni, 2020; Cambronero-Saiz et al., 2023). Taken together, these multiple strands of evidence suggest that although YouTube, as a leading creator platform, is often portrayed as a meritocratic and decentralized market in which anyone with talent and effort can succeed, such digital labor platforms may nonetheless replicate the gender inequalities long observed in traditional labor markets.

Grounded in this context, this paper studies gender differences in income among top U.S.-based YouTube creators, a labor market in which compensation depends directly on audience demand and performance metrics are publicly observable. Understanding gender inequality in this setting is important because platform-mediated work is becoming an increasingly relevant form of labor market organization. If disparities persist even when output is publicly observable and compensation depends directly on consumer response, then explanations based solely on firm-level institutions, such as promotion structures, bargaining, or employer discrimination, are incomplete. The creator economy therefore provides a useful setting in which to assess which mechanisms underlying the gender wage gap persist when conventional workplace structures are largely absent. We assemble a novel dataset combining channel-level earnings estimates from *Social Blade* with hand-collected information on creator demographics and platform activity, including content

niche, upload frequency, channel size, and audience engagement. Our sample consists of 518 high-performing non-celebrity creators identified from top 2,000 U.S.-based YouTubers, for whom YouTube represents a meaningful source of income. We document a substantial gender income gap: female creators earn 22.8% less than comparable male creators.

To investigate the sources of this disparity, we apply the Blinder-Oaxaca and Gelbach decomposition methods that incorporate both traditional demographic characteristics and platform-specific performance measures. Observable demographic variables explain little of the gap. Platform features, particularly content niche and audience engagement, account for part of the gap, but a substantial residual disparity remains even after controlling for channel size, activity, and monetization proxies. The persistence of this unexplained gap is informative. Because creators are neither hired, promoted, nor paid wages by employers, conventional institutional explanations for the gender wage gap are unlikely to fully account for the remaining difference. Instead, the evidence points to mechanisms operating on the demand side of the market. Gender differences in content specialization may reflect social norms or preferences that lead women to concentrate in categories with different monetization potential, while audiences may differentially select or engage with creators across gender. Taken together, these findings suggest that digital platforms do not eliminate gender inequality, but rather shift its sources toward market-mediated attention and consumer response.

This paper contributes to three strands of literature. First, it contributes to research on the gender wage gap by examining a labor market in which performance is publicly observable and income depends directly on consumer demand. Second, it contributes to the literature on digital and platform labor by studying a setting that differs from gig work, where labor supply decisions and experience accumulation remain central. Third, it contributes to the emerging literature on the creator economy by providing systematic quantitative evidence on income inequality among high-earning content creators.

The remainder of the paper proceeds as follows. Section 2.2 reviews the related

literature on gender inequality in both traditional labor markets and emerging digital labor markets. Section 2.3 explains YouTube's monetization mechanisms. Section 2.4 examines gender segregation across content niches and its implications for income disparities. Section 2.5 describes the dataset constructed for the analysis. Section 2.6 outlines the empirical methodology, including the regression model and the Blinder-Oaxaca decomposition. Section 2.7 presents the main results and robustness checks. Section 2.8 concludes by summarizing the key findings and discussing the study's limitations as well as directions for future research.

2.2 Literature review

In this section, we review the recent literature on gender income disparities in both traditional and emerging digital labor markets. We begin with conventional labor markets, where decades of empirical research have documented the persistence and underlying determinants of the gender wage gap. We then turn to behavioral and normative explanations, which help explain how individual choices and social norms interact with institutional structures to shape earnings. Subsequently, we examine evidence from digital platform-based labor markets, where algorithmic coordination and flexible labor supply reshape, but do not eliminate, existing inequalities. Finally, we review the emerging literature on the creator economy, where gender disparities re-emerge through distinctive market segmentation, visibility dynamics, and monetization structures. Taken together, this structured literature review provides the conceptual framework for our analysis of the gender income gap among YouTubers and situates it within the broader evolution of gender inequality in modern labor markets.

A vast body of research documents the persistence of the gender wage gap in traditional labor markets. In the United States, women's earnings have remained approximately 15-20% below those of men for nearly two decades, with women earning about 82 cents for every dollar earned by men in 2022 (Blau and Kahn, 2017; Kochhar, 2023). Despite

substantial gains in women's educational attainment, progress in narrowing the gap has largely stalled (Blau and Kahn, 2017). Human capital factors, such as differences in education and work experience, now explain only a small share of the disparity, as women have caught up with or surpassed men along these dimensions (Lips, 2012). Instead, occupational and industrial segregation continue to play a central role, explaining roughly one-third to one-half of the gap (Blau and Kahn, 2017, 2020). Women's more frequent career interruptions and shorter working hours, often associated with family responsibilities, further constrain earnings growth. These patterns are reinforced by the "child penalty" documented by Kleven et al. (2019), whereby childbirth leads to persistent income losses for women but not for men. Goldin (2014) further showed that high-paying occupations disproportionately reward long and inflexible hours. When firms restructure jobs to allow greater temporal flexibility, gender pay gaps narrow substantially.

Beyond these structural mechanisms, a complementary strand of literature emphasizes behavioral and normative mechanisms that interact with institutional constraints. Bertrand (2011) reviewed evidence showing that women, on average, tend to exhibit greater risk aversion, lower willingness to compete, and less aggressive negotiation behavior, with these differences shaped largely by socialization rather than innate traits. Such behavioral patterns can be particularly costly in tournament-like environments, where rewards depend strongly on relative performance. Sent and van Staveren (2018) argued that these gendered preferences are context-dependent and mediated by stereotypes as well as institutional design. Evidence from both laboratory and field experiments supports this view: Croson and Gneezy (2009) documented stronger social preferences and greater risk aversion among women, while Niederle and Vesterlund (2007, 2011) show that, even at similar levels of ability, men are more likely than women to enter competitive compensation schemes. At the cultural level, Fortin (2015) links traditional gender-role attitudes to lower female labor-force attachment and slower occupational advancement. Exley and Kessler (2022) further showed that women engage in less self-promotion even when incentives are identical, leading to lower performance evaluations and slower wage growth.

Since digital labor markets emerged, scholars have examined whether such flexible, algorithmically managed, and decentralized work settings with fewer traditional workplace frictions, mitigate or reproduce gender inequality. Evidence from the gig economy suggests that gender pay gaps persist, although the underlying mechanisms differ from those in traditional labor markets. Kricheli-Katz and Regev (2016) discovered that, even on consumer-to-consumer platforms such as eBay, products sold by women receive about 20% lower prices than identical products sold by men. Cook et al. (2020) found that male Uber drivers earn about 7% more per hour than female drivers, with the gap explained by differences in experience, driving speed, and location preferences rather than direct discrimination. On Amazon Mechanical Turk, women earn roughly 10% less per hour, largely because they sort into simpler, less competitive, and lower-paying tasks and contracts (Litman et al., 2020). Similarly, Teutloff et al. (2025) reported an 8-12% gender income gap in online freelancing markets even after controlling for skill and bidding behavior. Taken together, these estimates, ranging from 7% to 20%, indicate that gender differences in economic returns persist even in decentralized, digital-based markets that appear to reward performance fairly.

Emerging evidence from the creator economy extends these findings. Gioia and Morabito (2025) analyzed more than 18,000 YouTube channels and found that female creators are disproportionately concentrated in lifestyle and beauty content, categories typically associated with smaller audiences and lower advertising rates, whereas male creators are more prevalent in gaming and technology, where monetization opportunities are substantially greater. Industry evidence likewise points to sizable disparities in sponsorship income: male creators earn about 30% more per paid post, even though women account for the majority of brand collaborations (IZEA, 2021; Collabstr, 2025). Platform algorithms further exacerbate these inequalities by promoting content with higher engagement, which tends to favor male-dominated genres (Ribeiro et al., 2023; Ionescu et al., 2023; Corsi, 2024; Lou et al., 2025). Audience and advertiser biases compound these effects: Viewers are more likely to evaluate female creators on the basis of appearance and to express

hostility in comment interactions, while advertisers may undervalue women's content or confine them to lower-paying niches (Wotanis and McMillan, 2014; Nakandala et al., 2017; Döring and Mohseni, 2020; Cambronero-Saiz et al., 2023). Taken together, this evidence points to a multifaceted structure of gender inequality shaped by algorithmic visibility, social norms or preferences, and demand-side responses.

Taken together, the evidence from traditional, digital, and creator labor markets points to a consistent pattern: technological advancement has not eliminated gender disparities, but has instead transformed the mechanisms through which they arise. The 22.8% gender income gap we estimate among the top U.S. YouTubers lies near the upper end of the range documented in both traditional and digital labor markets. This suggests that the creator economy is not simply a neutral, merit-based arena, but rather a new setting in which long-standing inequalities re-emerge through distinct market, behavioral, and algorithmic pathways.

2.3 YouTube monetization mechanics

Before turning to the empirical analysis, we first explain YouTube's monetization mechanisms to clarify the revenue structure of this creator platform.

2.3.1 YouTube partner program

YouTube's monetization framework operates primarily through the YouTube Partner Program (YPP), which governs how creators access and earn income from the platform. Established in 2007, the YPP provides a standardized mechanism through which creators can monetize their content once they meet specific eligibility criteria. To be qualified, creators must have at least 1,000 subscribers and achieve either 4,000 valid public watch hours in the past 12 months or 10 million valid shorts views in the past 90 days (see Figure 2.1). These thresholds ensure that monetization privileges are reserved for active and established creators who generate sustained audience engagement.

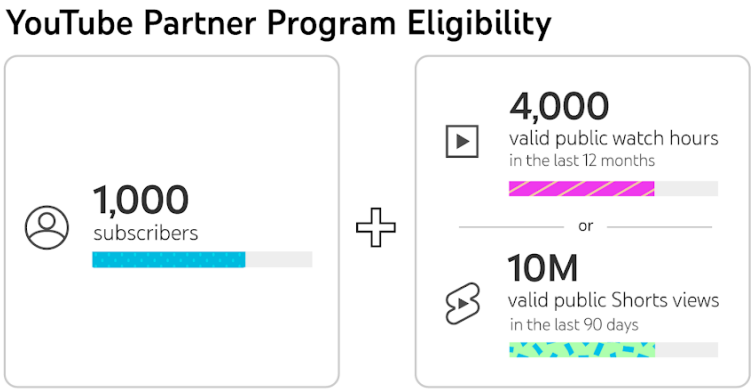


Figure 2.1. YouTube Partner Program (YPP) eligibility requirements (Source: YouTube Help (n.d.))

2.3.2 Ads-based monetization structure

Creators who qualify for the YouTube Partner Program (YPP) can monetize their content through a combination of platform-based and external revenue streams, as illustrated in Figure 2.2. On-platform monetization channels include advertising revenue, channel memberships, Super Chat and Super Stickers, YouTube Premium revenue sharing, and the merchandise shelf. Off-platform income sources include brand partnerships, sponsorships, affiliate marketing, and direct fan funding mechanisms such as Patreon. Although these alternative revenue streams can be highly lucrative for some creators, their magnitude varies substantially across content niches, audience engagement levels, and individual branding strategies, making them difficult to observe and quantify systematically.

Under YouTube's standard ads-based revenue-sharing model, creators receive 55% of the net advertising revenue generated by their videos, while the remaining 45% is retained by the platform to cover operational and distribution costs (YouTube Help, n.d.). This revenue-sharing arrangement creates a rule-based, platform-wide mechanism that transforms audience attention into measurable economic value, thereby effectively making advertising as the foundational income source for creators.

Advertising revenue therefore constitutes the dominant and most standardized stream of income, applicable to nearly all creators regardless of channel size or content niche.

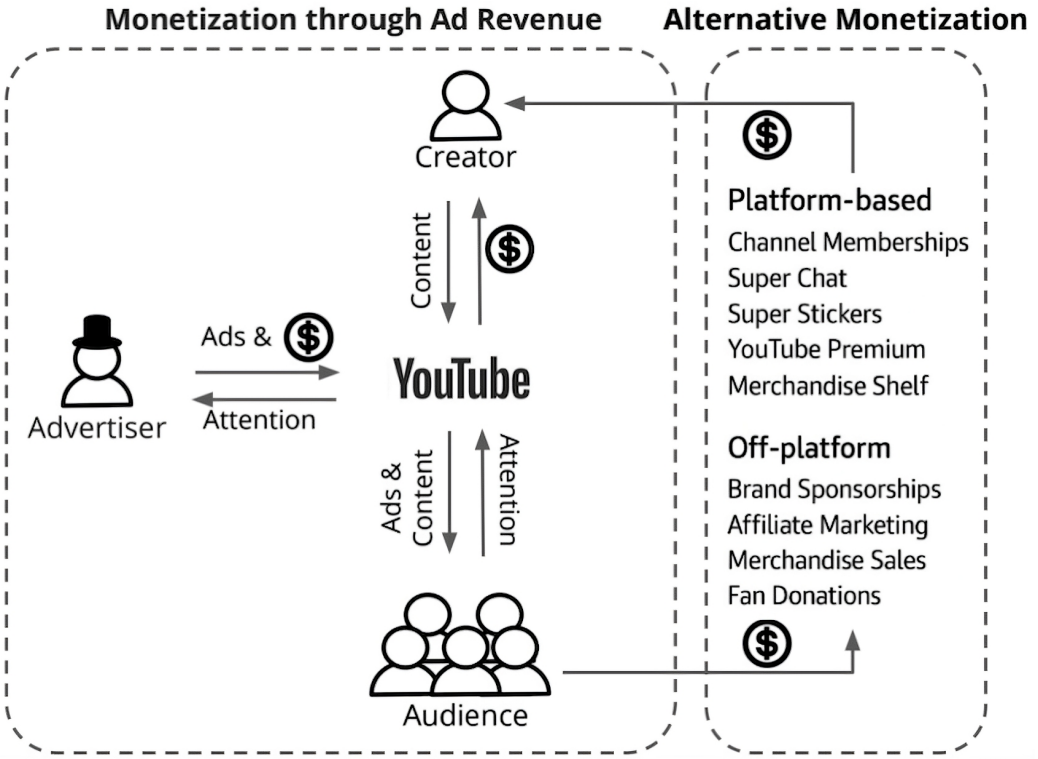


Figure 2.2. YouTube's monetization structure: ad-based and alternative revenue channels

According to (YouTube, 2021), advertising has long been at the core of creators' revenue streams and remains the primary way creators earn money on the platform. Within the YPP framework, advertisers pay to place ads before, during, or after videos, and creators receive a proportional share of the resulting revenue based on monetized views. This advertising-based model formalizes the conversion of user attention into monetary value, making ad revenue both the baseline measure in empirical studies of creator income and a consistent indicator of creators' financial performance (Lobato, 2016; Cunningham and Craig, 2019; Rieder et al., 2023).

2.3.3 CPM and income estimation

The key determinant of creators' advertising income is the Cost Per Mille (CPM), defined as the advertiser's payment per 1,000 ad impressions. Mathematically, ad-based revenue can be approximated by:

$$\text{Ad Income} = \frac{\text{Views}}{1000} \times \text{CPM} \times \text{Revenue Share (55\%)} \tag{2.1}$$

However, CPM rates are confidential between YouTube, Advertisers, and Creators, which vary substantially across content categories, audience demographics, geographic regions, and seasons. These variations complicate direct comparisons of creator income across channel categories. In this case, third-party analytics platforms such as *Social Blade* (the most widely used and representative one) provide estimated income ranges by applying a constant CPM range (commonly between USD 0.25 and 4.00)[1] to a channel's total annual views. Thus, the estimated income reported by *Social Blade* (SB) can be expressed as:

$$\text{SB Estimated Income} = \frac{\text{Views}}{1000} \times \text{Constant CPM Range (0.25–4.00)} \times \text{Revenue Share (55\%)} \tag{2.2}$$

Although *Social Blade* offers a convenient benchmark, its accuracy in reflecting actual YouTuber income remains uncertain. To address this concern and to evaluate the credibility of using *Social Blade* estimates for empirical research, we constructed a validation sample of approximately 50 YouTube creators who had publicly disclosed their actual income in their own videos. These self-reported data were then compared against the lower bound, upper bound, and median estimates given by *Social Blade* for the same YouTubers over the same time period. Using these four variables, we calculated Spearman rank-order correlation coefficients to assess the monotonic relationship between actual income and *Social Blade*'s estimated income.

As shown in Figure 2.3, the correlations between actual income and the three *Social*

[1] Social Blade applies a fixed CPM range primarily for comparability across channels. This range reflects long-term platform-wide averages derived from YouTube's public advertising data, encompassing both low-CPM and high-CPM markets. The lower bound (USD 0.25) approximates minimal monetization rates observed in low-demand regions or non-premium inventory, while the upper bound (USD 4.00) reflects typical CPMs for general U.S.-based content excluding premium niches such as finance or technology. This generalized approach allows consistent cross-channel estimation in the absence of direct access to actual advertiser bidding data.

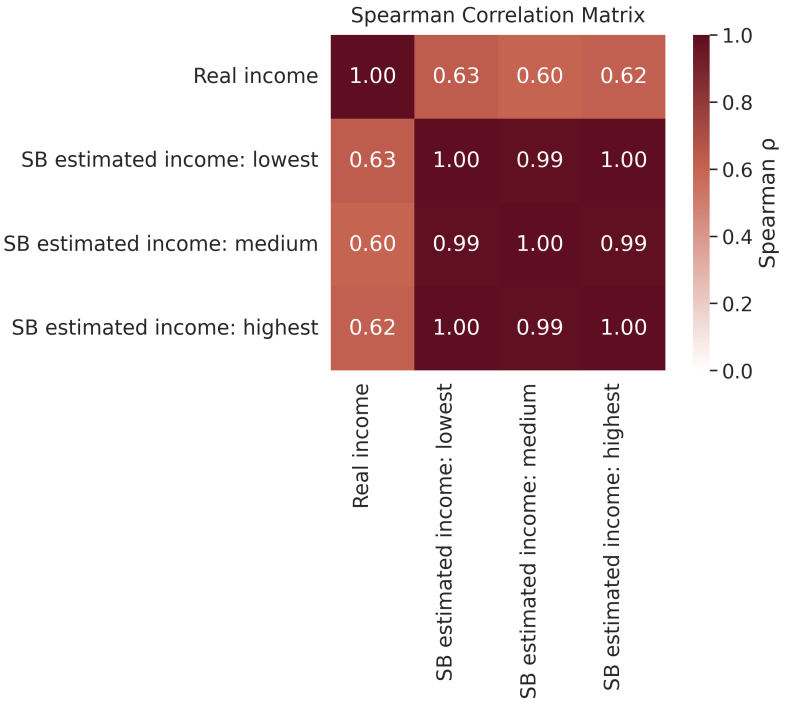


Figure 2.3. Spearman correlation matrix: Real vs Social Blade estimated income

Blade estimates range from 0.60 to 0.63, indicating a moderate to strong monotonic relationship. Although *Social Blade*'s estimates do not recover precise income magnitudes, they reliably preserve the ordinal ranking of creators' earnings. Moreover, the nearly perfect internal correlations (ρapproximately 0.99–1.00) among the lower, median, and upper estimates confirm that these measures are effectively scaled variants of the same underlying view-based metric. The median estimate therefore serves as a statistically defensible and representative proxy for YouTuber income in comparative analyses of income inequality. This validation supports the use of *Social Blade*'s median income estimate as a robust measure of creators' relative earnings in our study, enabling consistent comparisons across the top U.S.-based YouTubers by subscriber count.

2.4 Gender segregation and discrimination across content niches

Similar to patterns of occupational segregation in traditional labor markets, male and female YouTubers tend to cluster in different content categories that vary systematically in advertising value. To quantify the degree of this channel segregation, we compute a gender-based segregation index across a total of 18 YouTube channel categories, capturing how unevenly male and female creators are distributed across high- and low-CPM niches[2]:

$$D = \frac{1}{2} \sum_{s=1}^{N} \left| \frac{f_s}{F} - \frac{m_s}{M} \right| \tag{2.3}$$

where s indexes the YouTube channel category. f_s denotes the number of female creators in channel category s, and m_s is the number of male creators in channel category s. F is the total population of female creators in the sample, while M is the total number of male creators in the sample. Traditionally used to evaluate occupational segregation, the Duncan index is conceptually well-suited to the YouTube context, where each channel category represents a distinct content niche with a different economic value. Applying this measure enables us to summarize how unevenly male and female creators are distributed across higher- and lower-CPM content categories, providing a single, interpretable indicator of structural sorting.

The computed segregation index of 36.92% indicates a moderate to high degree of gender-based separation across YouTube content categories, implying that approximately 37% of creators would need to switch categories to achieve a gender-neutral distribution. This magnitude is comparable to the level of occupational segregation observed in traditional U.S. labor markets, which typically ranges from 40% to 50% (Blau and Kahn, 2017). Thus, despite the decentralized and self-directed nature of content creation, YouTube ex-

[2]The segregation index is calculated using the same dataset employed in the subsequent empirical analyses, namely the top U.S.-based individual YouTubers ranked by subscriber count. Detailed data construction procedures and sample characteristics are described in Section 5.

hibits patterns of gendered specialization similar to those found in conventional labor markets. This pattern reveals that male and female YouTubers are not randomly dispersed but tend to cluster in distinct niches, many of which are associated with varying CPM levels and monetization potential.

Notably, our income data drawn from *Social Blade*, apply a uniform CPM range across all content categories, which may overlook variation in advertising value across niches. As a result, the category-specific monetization differences implied by the segregation analysis are not directly reflected in these income estimates. In this case, we compile average CPM values[3] from other nine additional independent third-party analytics platforms to approximate the relative monetization potential of each content niche. Incorporating these category-level CPM differentials allows for a more accurate characterization and interpretation of gendered income variation on the platform.

Figure 2.4 illustrates the relationship between female representation and monetization potential across 18 content niches. A clear inverse relationship[4] emerges: categories with higher female representation, such as Entertainment, Music, and Kids & Family, tend to be associated with lower average CPM values, whereas high-CPM niches such as Finance, Technology, and Education exhibit markedly lower female shares. This pattern suggests that gender-based segregation across content niches systematically places female creators in categories with lower monetization potential, thereby reinforcing structural income disparities on the platform.

2.5 Data

We focus on the top 2,000 U.S.-based YouTube creators in this study for three main reasons. First, this group of creators are more likely to rely on YouTube as a primary source of

[3]These CPM values are calculated as the mean across nine independent third-party analytics platforms listed in Appendix 2.8 (Table B2). These averages reflect estimated monetization potential by niche.

[4]This inverse relationship reflects the well-documented phenomenon of female devaluation, whereby the economic value of work tends to decline as the proportion of women in an occupation rises (England, 1992; Levanon et al., 2009). In the literature, this pattern is commonly viewed as a proxy for gender-based discrimination, capturing the systematic undervaluation of female-dominated labor.

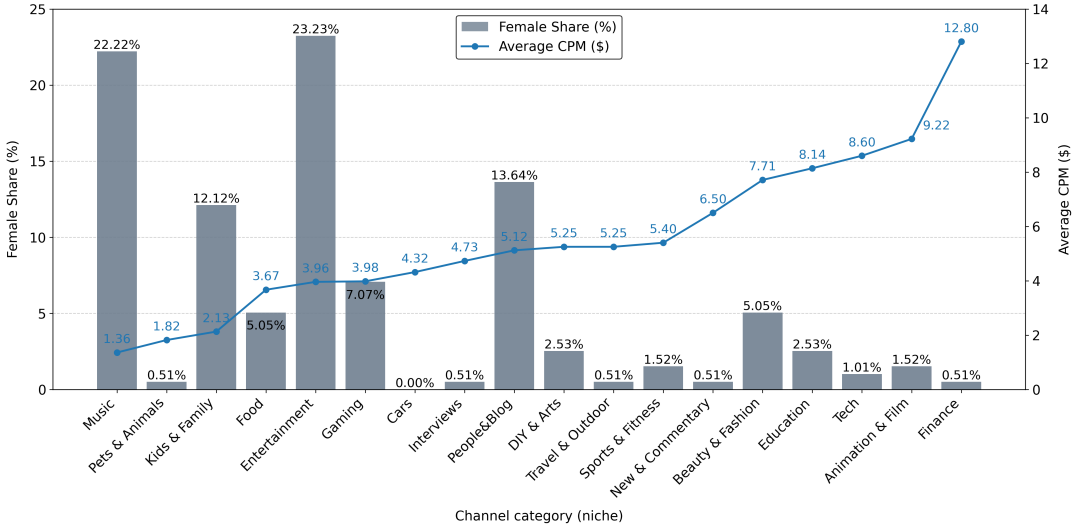


Figure 2.4. Female share (%) and average CPM ($) by YouTube channel category

income, and their engagement and monetization data are generally more reliable at scale. Second, similar to traditional labor markets, disparities in pay and opportunity in digital labor markets are likewise often most pronounced at the top of the distribution, where earnings increase disproportionately with rank. Third, for highly ranked YouTubers, demographic information such as gender, age, marital status, and education is more accessible and credible because of their public visibility and the availability of multiple information sources such as personal websites, interviews, and media coverage. This section describes how the dataset was constructed step by step.

2.5.1 Data source

The dataset of this study was developed using multiple complementary sources, integrating platform-level metrics, creator-level demographics, and video-level performance data. Table 2.1 provides an overview of the primary data sources, collection methods, and types of information obtained. The primary dataset originates from *Social Blade*, the most widely used analytics platform for online creators. Through a paid subscription, we obtained a complete list of the top 2,000 U.S.-based YouTube channels ranked by subscriber count as of November 2024, which together represent the majority of viewership and revenue concentration on the platform.

While *Social Blade* provides standardized quantitative indicators, it lacks detailed creator-level demographic information and video-level content statistics. Therefore, we developed automated Python scripts and a custom-built Chrome extension to extensively scrape relevant statistics directly from YouTube. This was complemented by a holistic manual verification process using multiple public and social media sources to collect personal information such as gender, age, education, and family status.

Table 2.1. Overview of Data Collection

Data Level	**Data Source & Collection**	**Information Obtained**
Channel list	*SocialBlade* — Paid subscription	List of the top 2,000 U.S.-based YouTube channels ranked by subscriber count.
Channel-level	*SocialBlade* — Automated Python-based web scraping with API	Channel statistics: total subscribers, total views, uploads, account creation date, channel type, and estimated annual income.
Video-level	*YouTube* — Google Chrome extension (JavaScript + HTML)	Video performance metrics: upload date, video length, view count, likes, comments, and content format (long-form videos vs. shorts).
Creator-level	*Public Databases* — Manual search and verification	Creator demographic characteristics: gender, age, education, occupation, nationality, and celebrity status.
	Social Media — Manual search and verification	Supplementary personal details: marital status, number of children, and life events, such as weddings, graduations, and birthdays.
	Interviews — Manual review	Verification of uncertain demographic attributes and professional background.

Note: Public databases include *Google*, *Wikipedia*, *Famous Birthdays*, *The Famous People*, *YouTube Fandom*, and *LinkedIn*. Social media sources include *Instagram*, *X (Twitter)*, and *Facebook*. Interviews and media coverage include publicly available videos, Q&A sessions, and interviews posted by YouTubers or online media outlets. Only information consistent across multiple sources was considered reliable and retained in the dataset.

The entire data collection process spanned approximately seven months, from November 2024 to May 2025, involving multiple automated and manual stages of verification. A detailed summary of the process, including the timeline and key stages of data acquisition, is provided in Appendix 2.8 (Figure A1). This multi-source approach ensures that our dataset captures both platform-level performance metrics and contextual creator characteristics, thereby providing a robust empirical foundation for analyzing gender income

disparities in the YouTube creator economy.

2.5.2 Sample construction

Based on the list of the top 2,000 U.S.-based YouTube channels obtained from *Social Blade*, we manually screened all channels to identify those eligible for our analysis. This step ensured that the dataset focused exclusively on individual creators, defined as channels whose content and brand identity are clearly tied to a single identifiable person who regularly appears in videos, including celebrity channels when the channel is centered on the creator's personal identity. All channels were classified into five operational types based on ownership, management, and on-screen identity: (1) Individual-operated, (2) Company-operated, (3) Group-operated, (4) Family-operated, and (5) Unknown. Company-operated channels refer to organization- or media-owned channels without a fixed individual host; group-operated channels involve two or more recurring individuals; family-operated channels are family-branded or family-centered; and unknown channels lack identifiable recurring hosts. Figure 2.5 summarizes the sample classification and refinement process. Nearly half of the top channels (48.85%) were operated by individuals, while 32.25% were company-affiliated, 8.90% were group-managed, 5.50% were family channels, and 4.50% could not be clearly identified.

From the 977 individual-operated channels, we conducted a four-stage demographic data collection and verification process (see Figure A2 in Appendix 2.8). Demographic information was collected from public databases, social media platforms, and publicly available interviews or Q&A videos, and was cross-checked across sources to improve consistency and fill missing values.

After excluding channels with unverifiable gender or other key demographic information, we obtained 730 individual YouTubers with verified profiles, representing 74.72% of the 977 individual-operated channels (Figure 2.5). Of these, 528 were male (72.20%) and 202 were female (27.80%), indicating a clear underrepresentation of female creators among top-ranked YouTubers by subscriber count. We then further excluded 185 celebrity

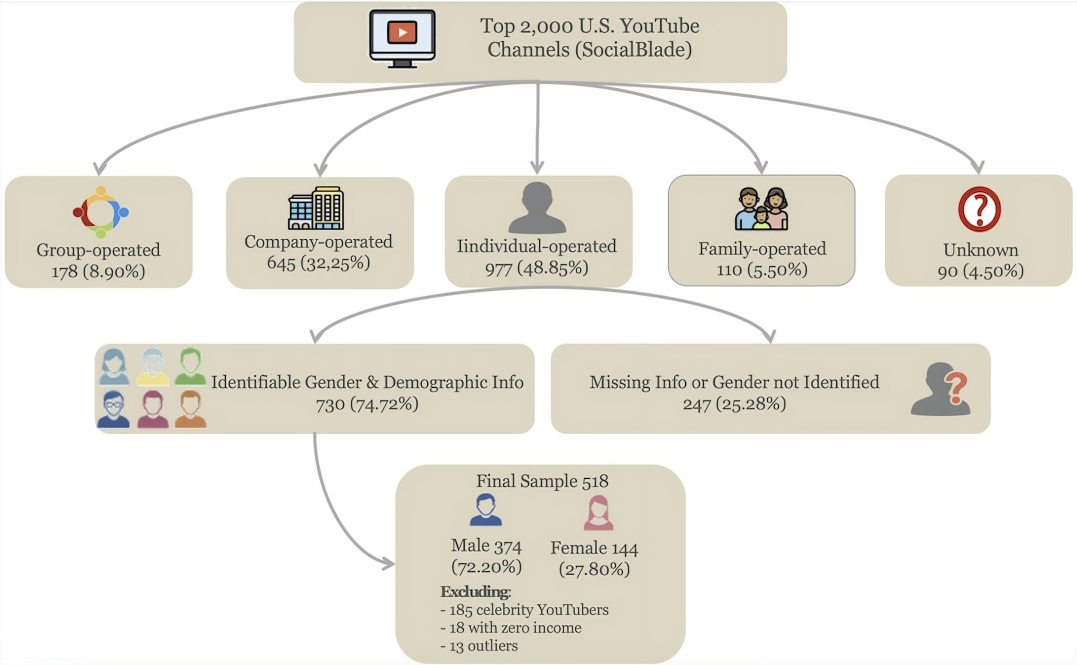


Figure 2.5. Sample selection for final dataset

YouTubers[5], 18 channels with zero estimated income, and 13 income outliers. The final sample consists of 518 individual creators: 374 men and 144 women.

2.5.3 Descriptive statistics

Following the above data collection and sample selection procedures, we constructed a comprehensive dataset and organized the variables into five categories, as detailed in Table 2.2. These categories include: (1) Demographics characteristics, (2) Channel basic characteristics, (3) Annual income & yearly content production, (4) Content frequency, and (5) Audience engagement. Each variable is clearly labeled with its corresponding unit to ensure clarity and consistency.

To further examine gender differences in the dataset, Table 2.3 presents descriptive statistics for the key variables by gender. The table reports the overall sample mean, the means for male and female creators separately, and the mean difference between the two groups (male minus female), with the corresponding statistical significance indicated. As seen, female YouTubers tend to be significantly younger than their male counterparts,

[5]Celebrity YouTubers are creators whose fame largely predated their YouTube activity, such as actors, musicians, athletes, or public figures, and whose channels are typically professionally managed.

suggesting that younger women are more likely to attain top-ranked visibility on the platform. Both male and female creators are predominantly White, although this pattern is more pronounced among men. Marital status does not differ significantly by gender, with the majority of both men and women being unmarried. Nevertheless, male YouTubers report a significantly higher average number of children than female YouTubers. This pattern suggests that, even in this emerging occupation, women may still face constraints associated with greater childcare responsibilities, which could limit the time and flexibility available for content production and audience engagement.

In terms of educational attainment, more than half of both male and female YouTubers do not hold a college degree, underscoring the limited relevance of formal education for success in this profession. Dropout rates are markedly higher among male YouTubers, consistent with anecdotal evidence that many high-earning creators leave school early to pursue full-time content production. Overall, while female YouTubers in this elite group tend to have slightly stronger educational backgrounds, YouTube as an occupation imposes no formal credential requirements: Success is instead driven more by creativity, audience engagement, and platform-specific skills rather than by academic qualifications.

Turning to basic channel characteristics, clear gender differences emerge across several dimensions. Female YouTubers tend to operate younger channels on average and to have fewer total uploads than male YouTubers, indicating shorter platform experience and possibly lower content production intensity. In terms of content specialization, women are more concentrated in lifestyle-oriented categories such as Beauty & Fashion, People & Blogs, and Kids & Family[6], which typically have lower CPM rates, whereas men are more concentrated in Gaming, Technology, and Education, categories generally associated with higher CPMs and audience engagement.

In addition, gender differences extend to production intensity, audience engagement, and monetization outcomes. Overall, male YouTubers earn higher annual income (al-

[6]This category, also referred to as minors' channels, includes those in which the primary on-screen figure is a child, typically managed by parents or guardians. Such channels often feature toy reviews, family activities, or educational entertainment content.

Table 2.2. Data description of constructed variables

Variable	Description
Demographic Characteristics	
gender	Gender of the YouTuber (Male, Female).
age	Age of the YouTuber, measured in years.
race	Racial identity (White, Black, Asian, Latino, Other).
marital_status	Marital status (Unmarried, Married, Divorced/Widowed).
nchild	Number of children, measured in single-digit units.
edu	Highest education level (low_educ, bachelor, master, Ph.D.).
dropout	Indicator for whether the YouTuber dropped out of school.
Channel Basic Characteristics	
chan_age	Years since the YouTube channel was created, measured in years.
total_subs	Total number of channel subscribers, measured in millions.
language	Indicator for whether the primary channel language is English.
face	Indicator for whether the YouTuber appears on screen.
chan_type	Channel category.
Annual Income & Yearly Content Production	
yincome	Estimated yearly income (median of *SocialBlade*'s range), measured in U.S. dollars (typically single digits).
numb_yuploads	Total number of uploads in the last 365 days, measured in single-digit units.
length_yuploads	Total duration of all uploads in the last 365 days, measured in seconds.
Content Frequency	
50_avg_days	Average number of days between uploads (based on 50 most recent videos), measured in days.
non_active	Indicator for whether the channel had zero uploads in the last 365 days.
Audience Engagement	
avg_likes	Average number of likes per video (based on 50 most recent videos), measured in single-digit units.
avg_comments	Average number of comments per video (based on 50 most recent videos), measured in single-digit units.

Note: low_educ refers to education below a bachelor's degree. The *chan_type* variable follows *SocialBlade*'s classification; misclassified channels were manually reviewed, and broad categories were further refined where necessary. Uploads include both long-form videos and shorts. Full variable definitions are reported in Appendix B, Table B1. Variable units are provided in the descriptions.

though not statistically significant) and produce both a larger number of videos and longer videos on average. Female YouTubers, by contrast, tend to upload slightly more frequently (although insignificantly), suggesting a strategy centered on more regular but shorter content production. At the same time, male YouTubers achieve significantly higher audience engagement, as reflected in their higher *avg_likes*, indicating that male-dominated content, particularly in Gaming and Technology, may generate stronger audience interaction per view.

Taken together, these results reveal notable gender differences across several key

Table 2.3. Descriptive statistics by gender

Variables	Mean (N=518)	Male (N=374)	Female (N=144)	Difference
Demographic Characteristics				
Age	30.685	31.808	27.771	4.037***
Race: white	0.666	0.695	0.590	0.105*
Race: black	0.056	0.070	0.021	0.049*
Race: asian	0.112	0.094	0.160	–0.066*
Race: latino	0.073	0.059	0.111	–0.052*
Race: other	0.004	0.003	0.007	–0.004
Race: more_than_one	0.089	0.080	0.111	–0.031
Marriage status: unmarried	0.641	0.642	0.639	0.003
Marriage status: married	0.332	0.334	0.326	0.008
Marriage status: divorced	0.027	0.024	0.035	–0.011
Number of children	0.498	0.556	0.347	0.209*
Educational level: low-edu	0.523	0.529	0.507	0.023
Educational level: bachelor	0.392	0.380	0.424	–0.044
Educational level: master	0.041	0.040	0.042	–0.002
Educational level: PhD	0.044	0.051	0.028	0.023
Drop-out-of school	0.220	0.249	0.146	0.103*
Channel Basic Characteristics				
Channel age	10.178	10.495	9.354	1.141**
Total subscribtion	10.944	11.044	10.686	0.358
Channel category: Entertainment	0.378	0.412	0.292	0.120**
Channel category: Education	0.060	0.072	0.028	0.044*
Channel category: Animation_Film	0.021	0.021	0.021	0.001
Channel category: Beauty_Fashion	0.027	0.011	0.069	–0.059***
Channel category: Cars	0.008	0.011	0.000	0.011
Channel category: DIY_Arts	0.019	0.013	0.035	–0.021
Channel category: Finance	0.006	0.005	0.007	–0.002
Channel category: Food	0.058	0.054	0.069	–0.016
Channel category: Gaming	0.160	0.187	0.090	0.097***
Channel category: Interviews	0.010	0.013	0.010	0.004
Channel category: Music	0.029	0.027	0.035	–0.008
Channel category: News_Commentary	0.010	0.013	0.000	0.013
Channel category: People_Blog	0.058	0.019	0.160	–0.141***
Channel category: Pets_Animals	0.019	0.024	0.007	0.017
Channel category: Sports_Fitness	0.039	0.048	0.014	0.034
Channel category: Tech	0.027	0.032	0.014	0.018
Channel category: Travel_Outdoor	0.012	0.013	0.007	0.006
Channel category: Kids_Family	0.060	0.024	0.153	–0.129***
Annual Income & Yearly Production				
Annual income	1,484,382	1,517,894	1,397,344	120,550
Number of videos per year	69.494	78.949	44.938	34.012
Length of videos per year	108,145	126,620	60,163	66,457*
Content Frequency Variables				
Average upload frequency (days)	8.667	9.001	7.805	1.195
non_active	0.056	0.056	0.056	0.001
Audience Engagement Variables				
avg_comments	163.447	160.947	166.663	18.161
avg_likes	81.737	94.253	49.229	45.024**

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed tests). "Difference" is Male minus Female.

dimensions of YouTubers' profiles. While demographic and educational characteristics are broadly comparable, male YouTubers tend to operate larger and more established channels, often in higher-CPM categories such as Gaming and Technology. Female YouTubers are more concentrated in lifestyle-oriented content such as Beauty & Fashion and Kids & Family, and generally produce shorter but more frequent uploads. Overall, gender disparities in the sample are more evident in channel scale, content specialization, and engagement metrics than in background characteristics.

2.6 Empirical strategy

2.6.1 Model specification

The empirical analysis aims to quantify the gender income gap among the top individual YouTubers in the United States, ranked by subscriber count. We estimate a series of regression models in which the dependent variable is the estimated annual income of YouTubers, obtained from *Social Blade*. The key explanatory variable of interest is a binary indicator, Female, equal to one if the primary operator or main on-screen figure of the channel is female, and zero otherwise. The baseline specification is formulated as follows:

$$\ln(Y_i) = \alpha + \beta\,\text{Female}_i + X_i'\gamma + \varepsilon_i \tag{2.4}$$

where $\log(Y_i)$ denotes the natural logarithm of the estimated annual income for YouTuber i. X_i represents a vector of control variables capturing five major categories of observable characteristics detailed in Section 5. The coefficient β measures the average percentage difference in income between female and male YouTubers, conditional on observable characteristics.

2.6.2 Frameworks for interpreting the gender gap

Sequential inclusion: To identify which types of characteristics contribute most to the observed gender income gap, we adopt a sequential inclusion framework. The general model can be expressed as:

$$\ln(Y_i) = \alpha + \beta_k \,\text{Female}_i + X_i^{(k)}\gamma_k + \varepsilon_i \tag{2.5}$$

where $X_i^{(k)}$ denotes the cumulative set of control variables included up to stage k, and ε_i is an idiosyncratic error term. The coefficient β_k captures the conditional gender income gap after controlling for all variables up to block k. The difference $\beta_{k-1} - \beta_k$ thus represents the proportion of the baseline gap explained by the newly added covariate group. Specifically, variables are introduced in the following order:

M0 (Baseline):	gender only
⇓	+ G1 Demographic Characteristics (*age, race, marital status, number of children, education level, dropout of school*)
M1:	gender + G1
⇓	+ G2 Channel Basic Characteristics (*channel age, total subscribers, language (English), on-screen presence*)
M2:	gender + G1 + G2
⇓	+ G3 Content Category (*18 content-type dummies*)
M3:	gender + G1 + G2 + G3
⇓	+ G4 Yearly Production (*annual uploads, annual video length*)
M4:	gender + G1 + G2 + G3 + G4
⇓	+ G5 Content Frequency (*average upload interval, inactivity indicator*)
M5:	gender + G1 + G2 + G3 + G4 + G5
⇓	+ G6 Audience Engagement (*average likes, average comments*)
M6:	gender + G1 + G2 + G3 + G4 + G5 + G6

The empirical framework begins with a baseline specification that includes only the gender indicator (*Female*) and then gradually adds six additional groups of explanatory variables one by one to how the estimated coefficient on *Female* (β_{Female}) changes as ad-

ditional information is incorporated. This approach allows us to account for progressively richer sources of heterogeneity and provides insight into which dimensions explain the largest share of the income differential between male and female YouTubers.

Gelbach decomposition: The stepwise design of the sequential inclusion framework allows us to trace how different dimensions of observable characteristics contribute to narrowing or widening the estimated gender income gap. However, the approach remains path-dependent, as the contribution attributed to each group may vary with the order in which variables are introduced. To address the path dependence of the sequential inclusion, we implement the Gelbach decomposition (Gelbach, 2016), which provides an exact and order-invariant decomposition of how each covariate group contributes to the change in the estimated gender coefficient between the baseline and full models:

$$\beta_1 - \beta_0 = \sum_{g=1}^{G} \delta_g \tag{2.6}$$

Each δ_g represents the marginal contribution of group g, computed as the product of the gender-covariate association and the group's partial effect on income, conditional on the full model. Positive values indicate factors that widen the gap when controlled for, whereas negative values indicate groups that help explain it.

Blinder–Oaxaca decomposition: While the Gelbach method attributes changes in the gender coefficient to covariate groups, it does not distinguish whether disparities stem from different endowments or different returns to those endowments. To capture this distinction, we employ the Blinder–Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973), which decomposes the total gender income gap into an *explained component* (differences in observable characteristics) and an *unexplained component* (differences in returns to those characteristics). The former reflects compositional effects, while the latter captures structural or unobserved mechanisms such as discriminatory audience preferences, unequal sponsorship opportunities, or algorithmic visibility biases. Formally, the decomposition can be expressed as:

$$\overline{Y}_m - \overline{Y}_f = (\overline{X}_m - \overline{X}_f)\hat{\beta}_f + \overline{X}_f(\hat{\beta}_m - \hat{\beta}_f) \tag{2.7}$$

where $\overline{Y}_m$ and $\overline{Y}_f$ denote the mean (log) incomes of male and female YouTubers, respectively; $\overline{X}_m$ and $\overline{X}_f$ are the mean vectors of characteristics; and $\hat{\beta}_m$ and $\hat{\beta}_f$ are the estimated return coefficients from gender-specific regressions. The first term represents the *explained component*, capturing the part of the gap attributable to differences in characteristics, while the second term represents the unexplained component, associated with differences in returns.

Together, these three frameworks provide complementary perspectives: the sequential inclusion analysis illustrates how the gender gap evolves as controls are progressively introduced; the Gelbach decomposition offers an order-invariant assessment of the marginal contribution of each covariate group; and the Blinder–Oaxaca decomposition distinguishes between differences in characteristics and differences in returns. Collectively, these complementary approaches provide a comprehensive and robust understanding of the mechanisms underlying the gender income gap among YouTubers.

2.7 Empirical Results

2.7.1 Persistent gender income gap

Table 2.4 reports the ordinary least squares (OLS) estimates of the YouTuber income equation, separately for male, female, and pooled samples under the full specification that controls for all observable characteristics. The OLS result of the pooled sample indicates that the coefficient on *Female* remains negative even after full specification ($\beta = -0.23$, $p < 0.10$), suggesting that female YouTubers earn roughly 23% less than their male counterparts. Among demographic variables, age is the only factor that is significantly associated with income ($\beta = -0.017$, $p < 0.10$), whereas variables capturing channel characteristics and video uploads are more strongly related to income.

Newer (younger) channels exhibit higher income levels, total subscribers displays a large elasticity (about 0.89), and producing in English is positively related to income (0.323, $p < 0.10$), consistent with greater market reach. Production intensity correlates positively with earnings: more uploads and larger annual content volumes are associated with higher income. By contrast, longer average upload interval (in days) are negatively associated with earnings (−0.056, $p < 0.01$), and inactivity is penalized. Engagement measures are also informative: average likes are strongly, positively associated with income (0.268, $p < 0.01$), while average comments are weakly negative, suggesting that not all interactions translate into monetizable attention.

Columns (1)–(2) reveal asymmetric returns to the same characteristics across gender. The age premium is more negative for women than men (−0.048 vs. −0.010), and "face showing" is associated with a significant penalty for only female YouTubers (−0.828, $p < 0.10$) but not for males. Larger subscriber bases are associated with higher income for both groups (0.92 for men, 0.83 for women), but the advantage of producing content in English appears to be concentrated among men, as the coefficient is positive and significant only in the male and pooled samples. Returns to content categories also differ by gender: relative to the reference category People & Blog, where women are most represented, female YouTubers earn higher returns in Education, Finance, Technology, and Travel & Outdoor, with the coefficients positive and often statistically significant. These niches typically feature higher average CPMs, yet women remain markedly underrepresented in these profitable domains (see Figure 2.4). In contrast, men perform relatively worse in Beauty & Fashion and Gaming compared with the People & Blog baseline. Posting frequency has similar effects across gender, with shorter intervals between uploads associated with higher income in both groups. Average likes are strongly predictive of earnings for both men and women. Taken together, these results indicate not only a persistent gender income gap, but also systematic differences in the returns to visibility, language, and niche positioning, highlighting that men and women do not benefit equally from comparable observable characteristics.

Table 2.4. OLS regression results by gender

Variables	Male (N=374)		Female (N=144)		Combined (N=518)	
Demographic Characteristics						
Female					–0.228*	(0.132)
Age	–0.010	(0.002)	–0.048**	(0.024)	–0.017*	(0.010)
Race:Black	–0.088	(0.241)	0.261	(0.843)	–0.067	(0.228)
Race:Asian	0.239	(0.193)	0.270	(0.377)	0.216	(0.172)
Race:Latino	0.142	(0.230)	–0.862**	(0.383)	–0.209	(0.203)
Race:Other	0.021	(0.200)	0.275	(0.369)	0.089	(0.171)
Marriage status:Married	0.028	(0.188)	–0.379	(0.277)	–0.158	(0.153)
Marriage status:Divorced	0.214	(0.509)	1.601**	(0.670)	0.402	(0.384)
Number of chidren	–0.139*	(0.082)	0.237**	(0.115)	–0.003	(0.068)
Edu attainment:Bachelor	0.075	(0.159)	0.418	(0.315)	0.137	(0.131)
Edu attainment:Master	0.214	(0.289)	–0.619	(0.699)	0.131	(0.295)
Edu attainment:Doctoral	0.400	(0.362)	–0.949	(0.891)	0.003	(0.387)
Drop out of school	–0.249	(0.168)	–0.150	(0.404)	–0.207	(0.152)
Channel Basic Characteristics						
Channel age	–0.083***	(0.018)	–0.144***	(0.048)	–0.096***	(0.016)
Total subscribers	0.920***	(0.115)	0.825***	(0.310)	0.893***	(0.104)
Language(english)	0.443**	(0.206)	–0.120	(0.333)	0.323*	(0.177)
Face showing	–0.235	(0.214)	–0.828*	(0.490)	–0.292	(0.189)
Category:Entertainment	–0.161	(0.543)	0.330	(0.367)	0.371	(0.286)
Category:Education	–0.501	(0.595)	2.260**	(1.093)	0.489	(0.377)
Category:Animation_Film	–0.179	(0.635)	0.483	(0.673)	0.455	(0.385)
Category:Beauty& Fashion	–1.958***	(0.703)	0.877*	(0.503)	–0.072	(0.439)
Category:Cars	–0.447	(0.584)	*(omitted)*		0.215	(0.431)
Category:DIY&Arts	–0.053	(0.784)	–0.537	(0.919)	–0.013	(0.609)
Category:Finance	–0.701	(0.899)	1.785***	(0.526)	0.641	(0.771)
Category:Food	0.074	(0.587)	0.342	(0.550)	0.669*	(0.342)
Category:Gaming	–0.957*	(0.555)	–0.028	(0.581)	–0.248	(0.312)
Category:Interviews	–1.153	(0.781)	*(omitted)*		–0.506	(0.632)
Category:Music	0.430	(0.657)	0.733	(0.496)	0.988***	(0.373)
Category:News& Commentary	–1.235	(1.081)	*(omitted)*		–0.734	(0.954)
Category:Pets&Animals	0.285	(0.664)	1.027**	(0.450)	1.028**	(0.458)
Category:Sports&Fitness	–0.068	(0.620)	0.462	(0.488)	0.558	(0.384)
Category:Technology	0.259	(0.592)	0.897**	(0.382)	0.382	(0.340)
Category:Travel&Outdoor	–0.816	(0.580)	1.978**	(0.969)	0.115	(0.416)
Category:Kids&Family	–0.128	(0.680)	0.008	(0.935)	0.425	(0.490)
Yearly Production						
Number of yearly uploads	0.001***	(0.000)	0.001	(0.001)	0.001***	(0.000)
Length of yearly uploads	0.116**	(0.058)	0.094	(0.116)	0.101**	(0.051)
Content Frequency						
Average upload interval(in days)	–0.055***	(0.010)	–0.051***	(0.017)	–0.056***	(0.009)
Inactive	–0.876	(0.598)	–0.515	(1.251)	–0.913*	(0.528)
Audience Engagement						
Average likes	0.257***	(0.042)	0.254**	(0.100)	0.268***	(0.039)
Average comments	–0.098*	(0.051)	–0.028	(0.136)	–0.081	(0.050)
constant	11.298***	(0.921)	13.037***	(1.297)	11.266***	(0.694)

Note: Channel categories *Cars*, *Interviews*, and *News & Commentary* are omitted in the female subsample regressions because no female YouTubers are present in these categories. Heteroskedasticity-robust (HC2) standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Heteroskedasticity-robust standard errors in parentheses.

To further examine whether the gender income gap persists across the entire income distribution, we estimate conditional quantile regressions[7] that provide a more comprehensive view beyond the mean association captured by OLS. Figure 2.6 plots the estimated coefficients on Female from conditional quantile regressions at different points of the income distribution (τ = 0.10–0.90), together with the OLS estimate and the corresponding 95% confidence intervals for reference. The quantile estimates remain consistently negative across the distribution, indicating the gender income gap persists at all levels of YouTuber income. The gap is most pronounced at the lowest quantile ($\beta_{0.10} = -0.31$), narrows toward the middle ($\beta_{0.50} = -0.19$), and reaches its smallest magnitude around the 75th percentile, before slightly widening again at the top end.

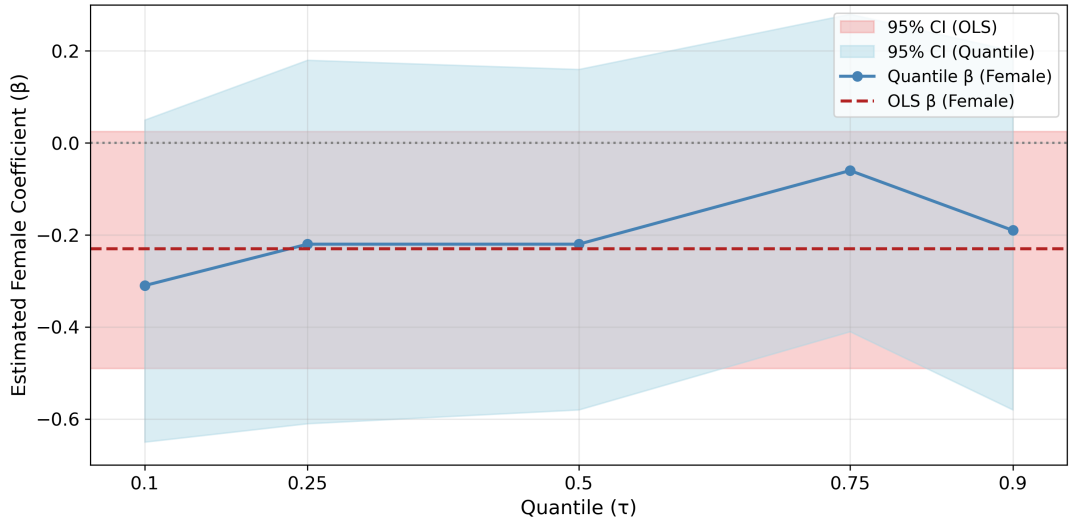


Figure 2.6. Gender income gap across the distribution: Quantile and OLS estimates

Notes: This figure plots the estimated *Female* coefficient (β) from quantile regressions at the 0.10, 0.25, 0.50, 0.75, and 0.90 quantiles of the conditional income distribution, along with the corresponding OLS estimate (red dashed line). Shaded areas indicate 95% confidence intervals.

This pattern suggests that female YouTubers face larger income disadvantages at the lower end of the distribution, where channels typically have limited reach and more fluctuating viewership. At higher quantiles, the smaller gap may reflect convergence in ad-revenue returns once channels attain large and stable view volumes. Overall, both OLS and quantile regression results indicate that gender differences in predicted advertising

[7]Detailed results are summarized in Appendix 2.8 (Table B3).

income persist even after controlling for demographic, educational, and platform-related heterogeneity.

2.7.2 Explaining the gap

Sequential model specification

To identify which observable characteristics account for the estimated gender income gap, we begin by examining how the coefficient on *Female* evolves as additional groups of covariates are introduced sequentially. The results are summarized in Table 2.5. This stepwise specification traces the incremental explanatory power of each covariate block defined in Section 6, allowing us to assess which dimensions of creator heterogeneity most attenuate the raw gender difference in log income. The analysis begins with a baseline specification including only the gender dummy, in which the coefficient on *Female* is negative but not statistically significant (-0.193, $p > 0.1$). This initial model captures the unconditional income difference between male and female YouTubers before accounting for any observable characteristics.

In Column (2), adding demographic controls (*age, race, marital status, number of children, and education*) increases the magnitude of the coefficient to -0.441 and renders it statistically significant at the 5% level. This indicates that once observable demographic differences are taken into account, the adjusted gender income gap becomes even larger, implying that female creators' lower earnings cannot be explained by simple demographic composition effects. After incorporating channel-level characteristics (Column 3), the gap further widens to -0.466, suggesting that channel fundamentals such as channel age, total subscribers, and language do not account for the observed disparity, and may even amplify it.

As additional groups of covariates are introduced sequentially, namely content niche (Column 4), yearly content production (Column 5), uploading frequency (Column 6), and audience engagement (Column 7), the estimated gap gradually narrows from -0.466

Table 2.5. Sequential inclusion of covariate groups and the evolution of the gender income gap

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	−0.193	−0.442**	−0.466***	−0.291*	−0.284**	−0.341**	−0.228*
	[0.187]	[0.194]	[0.163]	[0.168]	[0.145]	[0.140]	[0.132]
Covariate groups included:							
Demographics (G1)		Yes	Yes	Yes	Yes	Yes	Yes
Channel Basics (G2)			Yes	Yes	Yes	Yes	Yes
Content Category (G3)				Yes	Yes	Yes	Yes
Yearly Content Production (G4)					Yes	Yes	Yes
Uploading Frequency (G5)						Yes	Yes
Audience Engagement (G6)							Yes
Observations	518	518	518	518	518	518	518
R^2	0.002	0.088	0.338	0.366	0.550	0.595	0.644

Note: The dependent variable is the logarithm of annual income. The estimation begins with a baseline model that includes only the gender dummy (*Female*), and subsequently adds covariate groups sequentially across columns. Specifically, Columns (2)–(7) incrementally introduce demographic, channel, content, production, activity, and engagement controls as described in Section 2.6. Heteroskedasticity-robust (HC2) standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

to −0.228. This systematic reduction indicates that a substantial portion of the gender income gap is associated with differences in production behavior and engagement metrics. In other words, part of the observed disparity reflects differences between male and female YouTubers in content production intensity and in the ways audiences respond to their content. Nevertheless, even in the fully specified model, the coefficient on *Female* remains negative and statistically significant ($p < 0.10$), implying that observable factors alone do not fully explain women's earnings disadvantage.

Taken together, these stepwise estimates reveal two key insights. First, demographic and basic channel characteristics do not explain the gender income gap and, in some cases, even amplify it. Second, although performance-related factors, particularly posting activity and audience engagement, help narrow the gap, they are not sufficient to close it entirely. The persistence of a negative coefficient after controlling for all observable characteristics points to the presence of unobserved mechanisms such as gendered audience biases, unequal sponsorship opportunities, or algorithmic visibility asymmetries.

Nevertheless, the stepwise approach has an inherent limitation: the contribution at-

tributed to each covariate group is path-dependent. In other words, the observed changes in the Female coefficient depend on the order in which variables are introduced, and alternative inclusion sequences could yield different magnitudes of attenuation. To address this issue, we next employ the Gelbach decomposition, which provides an exact, order-invariant decomposition of the change in the gender coefficient when moving from the baseline to the fully specified model. This method allows us to quantify how much each block of variables contributes to the explained reduction in the estimated gender gap. We then complement this analysis with the Blinder–Oaxaca decomposition, which directly partitions the overall income gap into explained and unexplained components.

Decomposition and extended analyses

Gelbach decomposition: Table 2.6 reports the Gelbach decomposition of the gender income gap. This method decomposes the change in the gender coefficient between the baseline and the fully specified models into additive contributions from each group of covariates, while remaining invariant to the order in which variables are introduced. It thereby identifies how much of the observed reduction (or increase) in the estimated gender gap can be attributed to specific sets of explanatory variables.

Table 2.6. Gelbach decomposition of the gender income gap

Group	Coefficient	Std. Err.	p-value	[95% CI]
G1: Demographics	0.113**	(0.050)	0.025	[0.014, 0.211]
G2: Channel Basics	0.066	(0.072)	0.359	[-0.075, 0.208]
G3: Content Category	-0.003	(0.078)	0.972	[-0.156, 0.150]
G4: Yearly Content Production	-0.067	(0.053)	0.206	[-0.171, 0.037]
G5: Uploading Frequency	0.068	(0.076)	0.374	[-0.082, 0.217]
G6: Audience Engagement	-0.142**	(0.060)	0.018	[-0.260, -0.024]
Total change ($\sum$G1–G6)	0.034	(0.165)	0.835	[-0.289, 0.358]

Note: The dependent variable is the logarithm of annual income. Each row reports the contribution of a covariate group to the change in the gender coefficient when the group is added to the baseline model. Positive coefficients indicate that the group increases the estimated gender gap, while negative coefficients indicate that it reduces the gap. Heteroskedasticity-robust (HC1) standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Consistent with the stepwise results, demographic factors (G1) contribute positively

and significantly to the explained share of the gap (0.1126, $p < 0.05$), suggesting that observable demographic differences widen rather than narrow the estimated disparity. In contrast, audience engagement (G6) exhibits a significant negative contribution (−0.1421, $p < 0.05$), indicating that differences in audience interaction metrics, such as likes and comments, help reduce the gender gap. Other covariate groups, including channel basic characteristics, content niche, and yearly content production (G2–G5), show statistically insignificant effects, implying that these structural or behavioral factors play a limited independent role when demographic and engagement variables are accounted for. Overall, the Gelbach decomposition results reinforce the earlier findings: demographic composition amplifies the gender gap, while engagement-related variables partially offset it, although the total explained change remains statistically insignificant.

Blinder–Oaxaca decomposition: While the Gelbach decomposition attributes the change in the gender coefficient to specific groups of covariates in an order-invariant manner, it mainly captures how each set of variables statistically explains the attenuation of the gender effect as additional controls are introduced. However, this approach remains descriptive: it does not distinguish whether the observed gap arises because men and women differ in their characteristics or because these characteristics yield different economic returns. To address this issue, we further perform the Blinder–Oaxaca decomposition results, which explicitly separates the total gender income gap into an explained component (differences in observable endowments) and an unexplained component (differences in returns to those endowments).

As shown in Table 2.7, the explained portion (−0.037, $p > 0.8$) is negative and statistically insignificant, indicating that differences in observable endowments such as demographics, channel characteristics, or behavioral factors, jointly explain little of the gender income gap. In contrast, the unexplained component (0.231, $p > 0.2$) is substantially larger, suggesting that most of the observed disparity arises from differential returns to these characteristics as well as from unobserved or structural mechanisms such as biased audience preferences, unequal sponsorship opportunities, or algorithmic visibility

that systematically favor male YouTubers.

By covariate group, content niche (0.1448) and audience engagement (0.1624) make the largest positive contributions among the explained components, implying that gender differences in these dimensions slightly favor male YouTubers and help explain part of the income gap. In contrast, demographic characteristics (−0.1881) and channel basic characteristics (−0.1599) exhibit substantial negative contributions, suggesting that if males and females had identical characteristics in these groups, the gender income gap would be even larger. The effects of yearly content production (0.0650) and uploading frequency (−0.0613) are comparatively minor. However, the overall explained component remains statistically insignificant, implying that observable characteristics between male and female YouTubers explain little of the income disparity. Most of the gap arises from the unexplained component, particularly differences in channel basic and demographic characteristics, pointing to potential structural or behavioral factors beyond measured characteristics.

Overall, compared with the Gelbach decomposition results, the Blinder–Oaxaca decomposition reinforces two main conclusions. First, demographic and basic channel characteristics contribute little to explaining the gap and, in fact, appear to generate asymmetric returns that benefit men (as indicated by the large positive coefficients in the unexplained component for demographics and channel basics, 1.0506 and 1.8962, respectively). Second, content niche and performance-related factors help narrow the disparity but cannot eliminate it. Together, these findings confirm that the gender income gap among YouTubers is not primarily driven by compositional differences, but rather by asymmetric economic returns to similar observable characteristics and by unobserved structural or behavioral factors that systematically favor male YouTubers.

Explained contributions across analytical frameworks

Figure 2.7 synthesizes the explained contributions across the three analytical frameworks: sequential inclusion, Gelbach decomposition, and the explained component of the Blinder–

Table 2.7. Detailed Blinder–Oaxaca decomposition of log income by gender

Component	Coefficient	Std. Err.	p-value	[95% CI]
Explained (Endowments)				
Demographics	-0.188			
Channel Basic	-0.160			
Content Category	0.145			
Yearly Content Production	0.065			
Uploading Frequency	-0.061			
Audience Engagement	0.162			
Subtotal Explained	**-0.037**	**0.220**	**0.866**	[-0.469, 0.394]
Unexplained (Coefficients)				
Demographics	1.051			
Channel Basic	1.896			
Content Category	-0.807			
Yearly Content Production	0.198			
Uploading Frequency	-0.057			
Audience Engagement	-0.311			
Subtotal Unexplained	**0.231**	**0.192**	**0.230**	[-0.146, 0.606]
Total Gap (Male – Female)	**0.193**	**0.159**	**0.225**	[-0.119, 0.505]

Note: The dependent variable is the logarithm of annual income. Group-level coefficients are aggregated contributions obtained by summing the corresponding covariate effects from the decomposition output. Bootstrap standard errors are reported for subtotal and total terms. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Oaxaca decomposition. The results exhibit a high degree of consistency in signs of group contributions, although the magnitudes vary across methods. The most notable difference concerns content niche, where the Gelbach decomposition yields a substantially smaller contribution than the other two approaches. This difference likely arises from the distinct logic of the Gelbach decomposition. Unlike the sequential inclusion approach, Gelbach method eliminates path dependence by providing an exact, order-invariant decomposition of the change in the gender coefficient. At the same time, it attributes each covariate group's effect to its incremental contribution conditional on the full specification, rather than to the group's total explanatory share, as in the Blinder–Oaxaca decomposition. Consequently, the Gelbach method often yields more conservative estimates for groups that are highly collinear, or whose explanatory power overlaps substantially with other covariates.

Among the six covariate groups, content niche and audience engagement consistently emerge as the strongest positive contributors to the gender income gap, suggesting that

male YouTubers' concentration in more profitable content domains and stronger audience interactions account for a substantial share of the observed disparity. In contrast, demographic and channel basic characteristics display large negative contributions across all frameworks, suggesting that female YouTubers tend to possess relatively more favorable demographic or channel-related characteristics, which would otherwise narrow the gender gap if men and women were rewarded equally for similar traits.

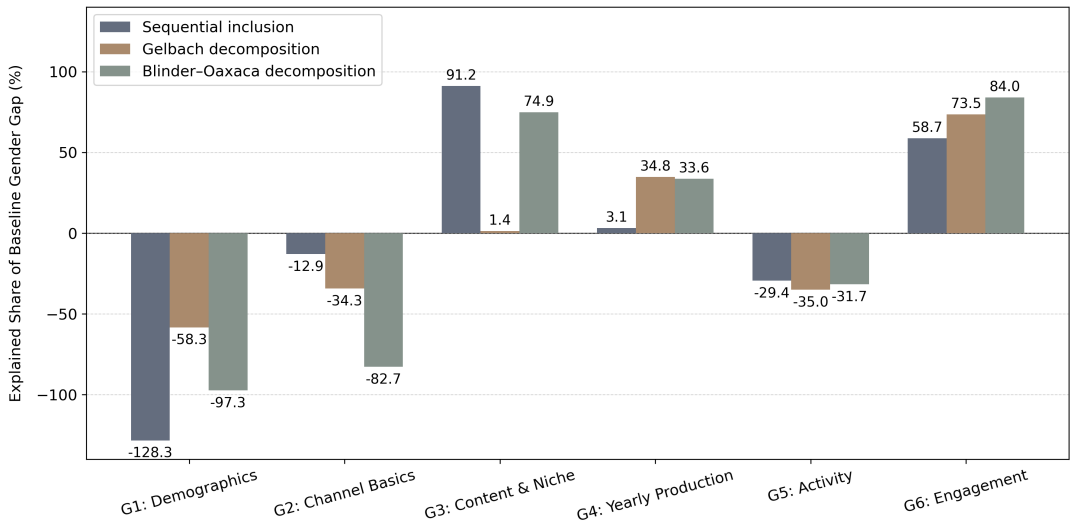


Figure 2.7. Explained contributions to the gender income gap across analytical frameworks

Note: The vertical axis reports the proportion of the baseline gender income gap explained by each covariate group (%), where positive values indicate that the inclusion of a covariate group reduces the gender gap, while negative values indicate that it widens the gap. **Stepwise share:** calculated as $100 \times (|\beta_{k-1}| - |\beta_k|)/|\beta_0|$, where β_k denotes the gender coefficient at model k from sequential inclusion. **Gelbach share:** derived from $100 \times g_k/\beta_0$, where g_k is the Gelbach contribution of covariate group k. **Oaxaca (Explained, Adjusted) share:** computed as $+100 \times E_k/|\beta_0|$, where E_k represents the group's explained (endowment) component from the detailed Oaxaca output. G1–G6 correspond to demographic, channel, content, production, activity, and engagement characteristics, respectively.

Building upon the group-level comparison, Figure 2.8 further disaggregates the explained contributions to the gender income gap at the variable level across the three analytical frameworks. In contrast to broadly consistent patterns observed at the group level, several individual variables display opposite signs across methods, reflecting differences in how each framework attributes overlapping explanatory power among correlated covariates. Nonetheless, focusing on variables that exhibit relatively large and consistent effects provides a clearer picture of robust contributors to the observed disparity. Specifi-

cally, *age, education attainment, channel age, uploading frequency*, and *average comments per upload* all exhibit both stable and significant negative effects across methods, which appear to widen the gap. This suggests that women, despite tending to be younger, better educated, and operating new-established or more regularly updated channels, do not translate these advantages into comparable income gains. In contrast, *content niche* and, particularly *average likes per upload*, stand out as strong positive contributors across frameworks, underscoring that men's higher income is closely linked to concentration in more monetizable niches and their ability to generate stronger audience engagement. Overall, the variable-level analysis reinforces the view that gender disparities among YouTubers are driven less by observable human-capital or structural channel attributes, but more by asymmetries in content niche and how engagement-related outcomes are rewarded.

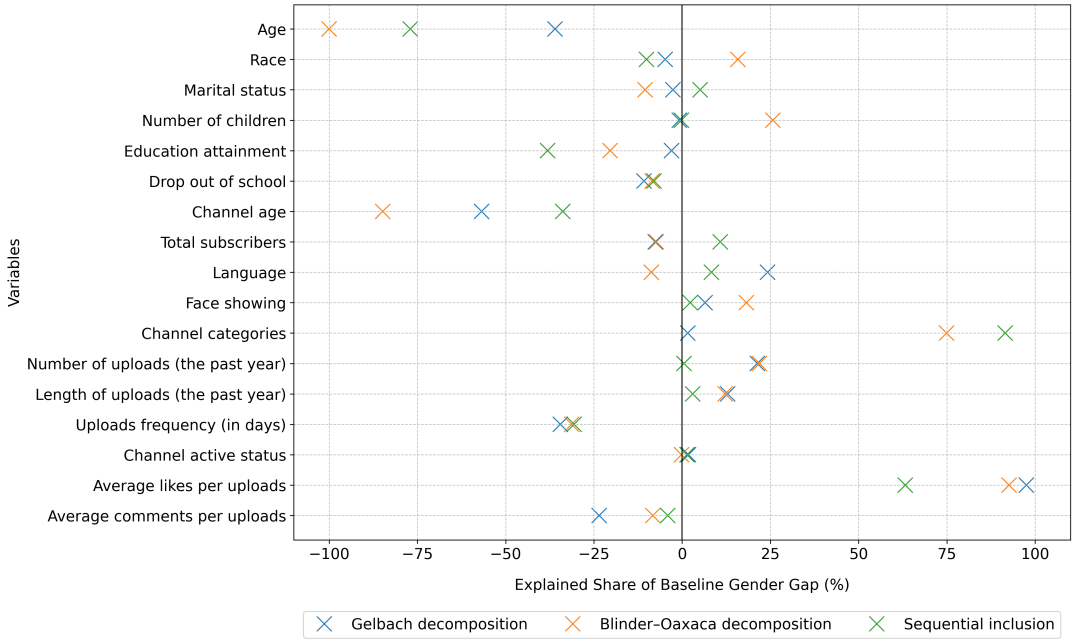


Figure 2.8. Explained contributions to the gender income gap across analytical frameworks

Note: Each point represents the standardized contribution of an individual variable to the explained component of the baseline gender income gap (in percent). Positive values indicate that the variable helps reduce the gender income gap, while negative values indicate that it widens the gap.

2.7.3 Robustness and validation

Beyond the sequential inclusion framework, the two decomposition methods discussed above (the Gelbach and Blinder–Oaxaca decompositions) can be regarded as forms of model structure robustness tests, as they validate the consistency of our results under alternative formulations of how the gender gap is decomposed and interpreted. To further ensure that our main findings are not driven by the choice of income measures, sample composition, or estimation assumptions, we conduct a series of complementary robustness checks. These results are summarized in Table 2.8, which groups robustness tests into three dimensions: alternative income measurement (Panel A), sample structure robustness (Panel B), and estimation robustness (Panel C). Propensity score matching (PSM), which specifically addresses potential sample selection bias, is reported separately later in Table 2.9.

Alternative income measurement

Panel A of Table 2.8 evaluates the sensitivity of the baseline result to alternative income definitions. In addition to log income, we consider two alternative outcome measures: (i) raw income in levels and (ii) CPM-weighted income, which adjusts for heterogeneity in advertising rates across content categories.

Across specifications, the estimated Female coefficient remains negative, although statistical significance varies. Using log income yields a coefficient of -0.228 ($p < 0.10$), consistent with the baseline. When using raw income in levels, the gender coefficient remains negative ($-282,386$ USD) but becomes statistically insignificant, reflecting substantial noise associated with extreme right–tail income values. Under the CPM–weighted specification, the gender coefficient remains negative and becomes more pronounced in magnitude (-0.316), representing an increase of approximately 39% relative to the log–income estimate[8]. This finding indicates that adjusting for systematic differences in monetization rates across content categories leads to an even larger estimated disadvantage

[8]Computed as $(0.316 - 0.228)/0.228$ approximately 0.386 approximately 39%.

for female creators. One interpretation is that female creators are disproportionately concentrated in content categories with lower CPMs, once CPM pricing is equalized across categories, the remaining gap reflects an amplified underlying income disparity rather than improved parity.

Taken together, these results indicate that the gender gap is not driven by a particular income transformation, and that adjusting for heterogeneous monetization conditions reinforces the qualitative conclusion.

Table 2.8. Robustness summary: Alternative income measures, sample structure, and estimation methods

	Obs.	**Female Coef.**	**Std. Error**	**Sig.**	R^2**/Fit**
Panel A. Alternative Income Measurement					
(1) Log Income	518	−0.228	(0.132)	*	0.644
(2) Level Income (USD)	518	-282386.1	(225494.6)		0.447
(3) CPM-Weighted Income	518	−0.316	(0.356)		0.948
Panel B. Sample Structure Robustness					
B1. Alternative Sample Definitions					
(4) Baseline	518	−0.228	(0.132)	*	0.644
(5) + Celebrities	699	−0.087	(0.109)		0.586
(6) + Outliers	531	−0.273	(0.134)	**	0.665
(7) Full Sample	712	−0.128	(0.112)		0.607
B2. By Creator Type					
(8) Entertainment (Male-Dominated)	196	−0.325	(0.174)	**	0.181
(9) People & Blogs (Female-Dominated)	30	−0.165	(0.234)		0.614
(10) Large Channels (> Median Subs)	162	−0.315	(0.072)	*	0.383
(11) Small Channels (< Median Subs)	356	−0.243	(0.159)	*	0.075
Panel C. Estimation Robustness					
(12) OLS–HC3	518	−0.228	(0.138)		0.644
(13) *UQR (Unconditional Quantile Regression)*					
Q10	518	−0.312	(0.181)	**	0.514
Q25	518	−0.216	(0.192)		0.476
Q50	518	−0.213	(0.175)		0.438
Q75	518	−0.057	(0.148)		0.422
Q90	518	−0.187	(0.196)		0.422

Note: Female = 1 for female creators. Panel A compares log income, raw income (USD), and CPM-weighted income. Panel B evaluates robustness to sample structure, including alternative inclusion rules, content type, and channel size. Panel C reports OLS with HC3 standard errors and UQR estimates with bootstrap standard errors. All specifications include demographic and channel-level controls (G1–G6). Standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Sample structure robustness

Panel B of Table 2.8 examines whether the main results are sensitive to sample definition and creator heterogeneity.

Inclusion of celebrities and outliers (Panel B1) We sequentially expand the sample by including (i) celebrity creators, (ii) outliers, and (iii) both groups. Across all specifications, the Female coefficient remains negative. When outliers are included, the coefficient becomes more significantly negative (-0.273, $p < 0.05$), suggesting that including outliers strengthens the estimated gender gap. Overall, changes in sample definition do not alter the baseline conclusion.

Subgroup analysis: By channel type and size (Panel B2) We then estimate regressions across narrower subsamples. The gender gap is economically and statistically meaningful in entertainment channels (-0.325, $p < 0.05$), where male creators are dominant. A weaker and statistically insignificant gap is observed in the People & Blogs category, a female-dominant content niche, consistent with within-category audience preferences. When splitting by channel size, the gap is larger among large creators (-0.315, $p < 0.10$), while smaller creators face a weaker, but still negative penalty. These patterns suggest that monetization asymmetries are more pronounced in competitive or male-dominated content environments.

Sensitivity to estimation procedures

Panel C of Table 2.8 examines whether the coefficient estimates are robust to alternative estimation procedures. First, we re-estimate the baseline specification using HC3 (heteroskedasticity-consistent) standard errors, which are robust to influential observations. The Female coefficient remains nearly unchanged (-0.228).

Then, we implement unconditional quantile regressions (UQR) at the 10th, 25th, 50th, 75th, and 90th percentiles of the income distribution. The estimates remain negative

across all quantiles, with the largest magnitude appearing at the lower tail (−0.312 at Q10, $p < 0.05$), suggesting that relatively lower–earning female creators face the strongest income penalty. The estimated gap remains economically relevant but becomes statistically weaker toward the upper tail of the distribution, where income is highly dispersed. It is worth noting that, because the analytical sample focuses on the top U.S. YouTubers, "lower–earning" creators in this context should be interpreted in a relative sense, that is, within an already high–income subset of the platform. Therefore, the results indicate that even within the upper tier of the creator economy, women in the lower portion of the earnings distribution experience a comparatively larger disadvantage.

Sample selection robustness

To further address potential sample selection bias, we implement a Propensity Score Matching (PSM) analysis that pairs female and male YouTubers with similar observable characteristics (e.g., subscriber size, activity, and channel age). Mean differences in log income before and after matching are reported in Table 2.9.

Before matching, female creators earn on average 0.193 log points less than comparable male creators. After nearest–neighbor matching, the estimated average treatment effect on the treated (ATT) becomes more negative (−0.334), suggesting that, conditional on observables, female creators earn substantially less than otherwise similar males. This increase in magnitude relative to the unmatched estimate indicates that observable characteristics mask part of the underlying gender disparity, once these factors are equalized, the gender gap becomes more pronounced.

The estimated average treatment effect (ATE) is mildly positive (+0.140), reflecting that, on average, female creators would earn slightly more if they were male. Meanwhile, the estimated average treatment effect on the untreated (ATU) is also positive, suggesting that comparable male creators (control group) would earn more if they were female, though these latter estimates should be interpreted cautiously given the limited support structure for counterfactual assignment.

Balance statistics confirm successful matching: mean standardized bias falls to 8.7% (median: 5.3%), and the post-matching pseudo-R^2 drops to 0.110, while the associated likelihood ratio test fails to reject joint insignificance ($p = 0.168$), indicating that pre-treatment covariates no longer predict treatment assignment. The matched sample remains substantial, consisting of 144 female creators and 374 male creators.

Table 2.9. Sample selection robustness: Propensity score matching (PSM)

	Treated Mean	Control Mean	Difference	Std. Error
Unmatched	12.864	13.058	−0.193	0.178
ATT (Matched)	12.864	13.198	−0.334	0.298
ATU (Matched)	13.058	13.380	+0.323	
ATE	–	–	+0.140	
Covariate Balance After Matching				
Mean Bias (%)		8.7		
Median Bias (%)		5.3		
Pseudo-R^2		0.110		
LR χ^2		44.03		
$p > \chi^2$		0.168		
Matched Sample Size				
Treated (Female)		144		
Control (Male)		374		
Total (On Support)		518		

Note: This table reports robustness results using propensity score matching (PSM) to address potential sample selection bias. Propensity scores are estimated using a logit model with all observed characteristics. One-to-one nearest-neighbor matching without replacement is used. The outcome variable is log income. ATT refers to the treatment effect for female creators, ATU to the effect for male creators, and ATE to the population-average treatment effect. Standard errors are matching-based and do not adjust for first-stage estimation.

Taken together, the PSM results suggest that the gender income gap persists, and even widens, after balancing on detailed demographic, educational, channel, content, and engagement characteristics. This pattern implies that observable differences between male and female creators do not explain the gap. Instead, the fact that the ATT becomes more negative after matching indicates that sample selection bias is unlikely to be the source of the estimated disparity. If anything, suppressing observable heterogeneity reveals a larger underlying income penalty faced by female creators. These results support the conclusion that female YouTubers experience systematic income disadvantages that cannot

be explained solely by observable characteristics.

2.8 Conclusion

This paper examines the gender income gap among top YouTubers in the United States, with the aim of identifying the key drivers of this disparity. Based on a empirical specification, we find a persistent gender income gap: female YouTubers earn approximately 22.8% less than male YouTubers, even after controlling for a set of demographic, channel, content, and engagement-related variables.

To understand the explanatory factors underlying this gap, we draw on a set of complementary empirical approaches that yield broadly consistent results. Sequential inclusion and Gelbach decomposition consistently show that demographic attributes such as age, marital status, and education, do not explain the income gap. Instead, they widen the gap. In contrast, three dimensions of characteristics account for a meaningful share of the difference: (i) content niche, (ii) yearly content production, and (iii) audience engagement. Among these, audience engagement, particularly *average likes per upload*, explains the largest share. The Blinder–Oaxaca results further confirm that endowment differences alone cannot account for the overall gap, but again highlight the explanatory power of content niche, production, and engagement. Taken together, these findings suggest that while observable characteristics can partially explain the gender income gap, most of the disparity remains unexplained, consistent with unequal access to monetization opportunities and other unobserved constraints faced by female creators.

We then implemented a series of robustness checks, which reinforce the main finding that female YouTubers consistently earn less than their male counterparts. First, alternative income measures including log income, raw income, and CPM-weighted income, all produce negative gender coefficients, indicating that the estimated gap is not driven by income transformation. Notably, CPM weighting increases the magnitude of the gap by approximately 39%, implying that differential monetization intensifies underlying dis-

parities. Second, various sample structure tests show that results remain stable when expanding the sample to include celebrities and outliers, as well as when examining subgroup heterogeneity by creator type or channel size. Third, estimation robustness checks using HC3 standard errors and unconditional quantile regression confirm that the gender gap persists throughout the distribution and is relatively more pronounced among lower-earning creators within the top-tier sample we focused on. Finally, Propensity Score Matching yields even a more negative treatment effect for female creators, demonstrating that the gap cannot be attributed to sample selection on observables.

Our findings challenge the conventional emphasis on demographic factors as one of the primary explanations for gender income inequality. In traditional labor markets, demographic characteristics such as gender, age, and marital status can partly explain income differences, but in the creator economy, these variables have almost no explanatory power. At the same time, our results also reveal notable continuities with traditional labor markets. In both settings, human capital variables such as education and work experience can hardly explain gender income disparities, while occupational and industry segregation (reflected on YouTube by channel category) remains a major determinant of earnings. Beyond these continuities, however, the creator economy introduces distinctive mechanisms shaped by digital and platform dynamics. Audience engagement, particularly *average likes per upload*, emerges as the second most important factor explaining income differences, underscoring the growing role of algorithmic visibility, user interaction, and attention-based rewards in shaping earnings outcomes. Taken together, these findings suggest that while emerging digital labor markets differ from conventional ones in form, they may reproduce many of the same structural inequalities in substance, now reinforced by platform-specific dynamics. By quantifying the gender income gap within this new form of labor market, this study fills an important empirical research gap and provides one of the first systematic analyses of inequality in the creator economy.

While this study provides new insights into gender inequality in the creator economy, several limitations should be acknowledged. First, the estimated income data from

Social Blade may not perfectly reflect actual earnings, although we verified its strong correlation with self-reported figures whenever possible. Likewise, personal demographic information collected from public sources may contain inaccuracies. Second, potential endogeneity remains a concern. Differences in preferences, personality traits, or gender-role attitudes may lead women to sort into less monetized content niches, while men may be more likely to enter higher-paying fields. Such unobserved self-selection could bias the estimated gender income gap if not properly addressed. Third, the relatively limited sample size reduces the precision of some regression and decomposition estimates and may weaken statistical significance in certain specifications. Despite these limitations, this paper provides an empirical foundation for understanding gendered income dynamics in the digital labor market and opens new avenues for interdisciplinary research at the intersection of digital platforms, algorithmic governance, and gender economics.

Appendix A: Supplementary figures

Figure A1. Timeline of the data collection and sample construction process

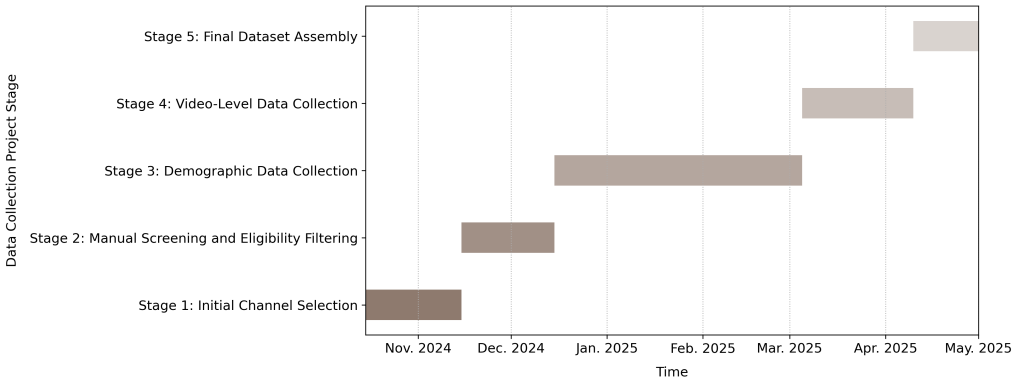
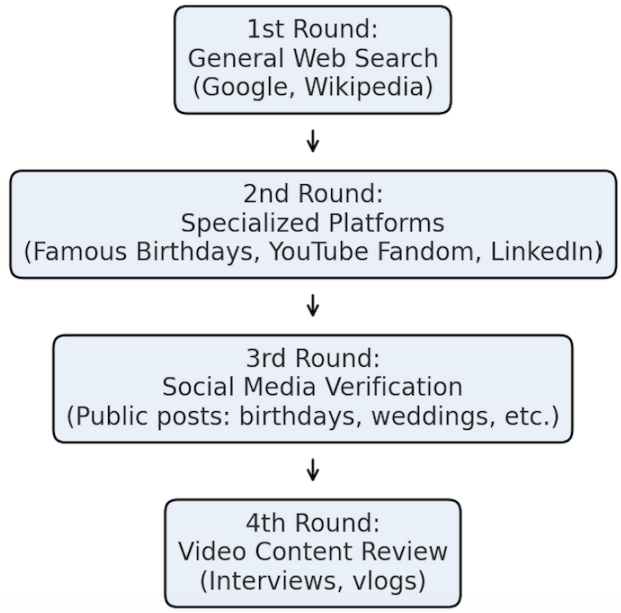


Figure A2. Four-stage demographic data collection process



Appendix B: Supplementary tables

Table B1. Comparison between YouTube videos and shorts

	Videos	**Shorts**
Duration	Long-form	Under 3 min
Ad Revenue	Tied to specific video	Shared from collective pool
Music Usage	May disable monetization	Allowed with adjusted revenue share
Ad Types	Various ad formats	Post-view and skippable ads
CPM	Generally higher	Lower (pooled system)
Impact on Income Studies	Reflects ad engagement	Challenges in income estimation

Table B2. Online sources for average CPM by channel category

No.	**Website Name**	**URL**
1	Tasty Edits	https://www.tastyedits.com/most-profitable-youtube-niches/
2	First Growth Agency	https://firstgrowthagency.com/blog/youtube-cpm-rates-by-niche/
3	Uppbeat	https://uppbeat.io/blog/youtube/most-profitable-youtube-niches
4	Yoola	https://yoola.com/blog/12-best-and-most-profitable-youtube-niches-for-2024
5	Is This Channel Monetized	https://isthischannelmonetized.com/data/youtube-cpm/
6	StartUp Growth Guide	https://startupgrowthguide.com/10-profitable-youtube-niches-with-high-cpm-for-maximum-earnings/
7	TubeBuddy	https://www.tubebuddy.com/blog/profitable-youtube-niches/
8	Teachable	https://teachable.com/blog/most-profitable-youtube-niches
9	Soundstripe	https://www.soundstripe.com/blogs/top-10-most-profitable-youtube-niches-ranked-by-cpm

Table B3. Full-specification conditional quantile regression (CQR) results

	Conditional Quantile Regressions (CQR)									
	10th		25th		50th		75th		90th	
Demographic Characteristics:										
Female	−0.312	(0.212)	−0.216	(0.184)	−0.213	(0.171)	−0.057	(0.170)	−0.187	(0.167)
Age	−0.011	(0.019)	−0.019	(0.014)	−0.020	(0.013)	−0.020	(0.013)	−0.023	(0.015)
Race:Black	−0.191	(0.307)	−0.363	(0.285)	−0.148	(0.408)	0.030	(0.282)	−0.145	(0.289)
Race:Asian	−0.164	(0.346)	0.382	(0.293)	0.379*	(0.204)	0.158	(0.198)	0.138	(0.210)
Race:Latino	−0.233	(0.347)	−0.166	(0.323)	−0.128	(0.229)	0.009	(0.251)	−0.085	(0.326)
Race:Other	0.210	(0.231)	0.198	(0.208)	−0.093	(0.196)	−0.192	(0.245)	0.055	(0.334)
Marriage status:Married	0.067	(0.219)	−0.095	(0.204)	−0.004	(0.189)	−0.065	(0.185)	−0.177	(0.204)
Marriage status:Divorced	0.682	(0.494)	0.385	(0.422)	−0.093	(0.475)	−0.418	(0.828)	0.893	(0.900)
Number of children	−0.062	(0.106)	−0.042	(0.095)	−0.052	(0.099)	0.100	(0.097)	0.049	(0.104)
Edu attainment:Bachelor	0.195	(0.251)	0.166	(0.242)	0.057	(0.169)	0.010	(0.155)	0.094	(0.200)
Edu attainment:Master	0.320	(0.458)	0.197	(0.452)	−0.003	(0.458)	0.238	(0.460)	0.576	(0.635)
Edu attainment:Doctoral	0.532	(0.869)	0.537	(0.434)	0.203	(0.481)	−0.094	(0.651)	−0.296	(0.660)
Drop out of school	0.048	(0.252)	−0.204	(0.238)	−0.244	(0.198)	−0.196	(0.176)	−0.162	(0.217)
Channel Basic Characteristics:										
Channel age	−0.139***	(0.031)	−0.115***	(0.027)	−0.074***	(0.020)	−0.059***	(0.018)	−0.055**	(0.021)
Total subscribers	0.931***	(0.195)	0.750***	(0.152)	0.871***	(0.119)	0.965***	(0.129)	0.998***	(0.139)
Language(english)	0.188	(0.254)	0.267	(0.247)	0.340*	(0.200)	0.431**	(0.199)	0.383	(0.283)
Face showing	−0.316	(0.315)	−0.319	(0.261)	−0.248	(0.240)	−0.209	(0.245)	−0.429	(0.362)
Category:Entertainment	0.469	(0.361)	0.775**	(0.375)	0.439	(0.456)	−0.221	(0.421)	−0.090	(0.341)
Category:Education	0.369	(0.437)	0.555	(0.427)	0.273	(0.606)	−0.019	(0.740)	0.490	(0.766)
Category:Animation_Film	0.956**	(0.485)	0.943*	(0.503)	0.259	(0.596)	−0.488	(0.610)	−0.902	(0.588)
Category:Beauty& Fashion	−0.297	(0.531)	−0.119	(0.566)	−0.206	(0.645)	−0.256	(0.780)	0.063	(0.525)
Category:Cars	1.004*	(0.603)	0.853	(0.585)	0.194	(0.699)	−1.034	(0.652)	−0.978	(0.688)
Category:DIY&Arts	−0.606	(0.760)	0.028	(0.960)	−0.064	(0.868)	0.364	(0.716)	0.034	(0.519)
Category:Finance	0.647	(1.020)	0.103	(1.035)	0.699	(0.992)	0.144	(0.812)	−0.112	(0.772)
Category:Food	0.826*	(0.433)	0.929**	(0.467)	0.497	(0.518)	0.058	(0.526)	0.411	(0.531)
Category:Gaming	−0.381	(0.398)	0.116	(0.404)	0.008	(0.494)	−0.547	(0.452)	−0.331	(0.431)
Category:Interviews	−0.269	(1.000)	−0.550	(1.084)	0.452	(1.303)	−0.122	(1.198)	−0.530	(1.111)
Category:Music	1.329**	(0.626)	1.955***	(0.560)	0.689	(0.566)	0.286	(0.601)	0.617	(0.550)
Category:News& Commentary	−1.464	(1.274)	−0.724	(1.569)	0.454	(1.742)	−0.450	(1.598)	−0.136	(1.579)
Category:Pets&Animals	0.532	(0.633)	1.138	(0.746)	0.873	(0.677)	0.157	(0.522)	0.213	(0.465)
Category:Sports&Fitness	0.303	(0.511)	0.910*	(0.553)	0.385	(0.566)	0.005	(0.610)	0.397	(0.564)
Category:Technology	0.404	(0.636)	0.999*	(0.519)	0.443	(0.524)	−0.547	(0.493)	−0.840*	(0.447)
Category:Travel&Outdoor	0.396	(0.508)	0.575	(0.495)	0.096	(0.627)	−0.896	(0.694)	−0.610	(0.714)
Category:Kids&Family	0.783	(0.932)	0.730	(0.623)	0.939	(0.715)	0.129	(0.615)	−0.000	(0.680)
Yearly Production:										
Number of yearly uploads	0.001***	(0.000)	0.001**	(0.000)	0.001**	(0.001)	0.002***	(0.001)	0.002**	(0.001)
Length of yearly uploads	0.259***	(0.067)	0.195***	(0.072)	0.018	(0.066)	−0.157**	(0.067)	−0.125*	(0.076)
Content Frequency:										
Average upload interval(in days)	−0.061***	(0.012)	−0.065***	(0.012)	−0.069***	(0.012)	−0.060***	(0.011)	−0.043***	(0.014)
Inactive	0.627	(0.806)	−0.143	(0.811)	−1.760**	(0.714)	−3.550***	(0.726)	−3.145***	(0.821)
Audience Engagement:										
Average likes	0.374***	(0.071)	0.350***	(0.057)	0.299***	(0.049)	0.270***	(0.052)	0.195***	(0.049)
Average comments	0.007	(0.092)	−0.056	(0.068)	−0.099*	(0.060)	−0.106	(0.067)	−0.090	(0.086)
Constant	7.878***	(1.128)	9.616***	(1.013)	12.017***	(0.961)	14.411***	(0.893)	14.821***	(1.034)
Observations	518		518		518		518		518	
Pseudo-R^2	0.514		0.476		0.438		0.422		0.422	

Notes: Bootstrap standard errors in parentheses. *, **, and *** denote significance at 10%, 5%, and 1% levels, respectively.

Chapter 3

Do Words Matter? Explaining the Gender Income Gap on YouTube through Top Content

3.1 Introduction

The creator economy has become an increasingly important segment of platform-based labor markets. Digital creators' earnings are shaped not only by advertising revenue itself, but also by a broader ecosystem of platform visibility, audience engagement, and monetization strategies (Sanyoura and Anderson, 2022; Rieder et al., 2023; Meisner, 2023). Yet, despite the rapid growth of this sector, earnings remain highly unequal, and there is still a limited understanding of what drives success. In particular, while the broader labor-market literature has long documented persistent gender inequality in pay, evaluation, and career advancement (Dohmen, 2004; Lyness and Heilman, 2006; Blau and Kahn, 2017; Rivera and Tilcsik, 2019), much less is known about how such inequality emerges in the digital platform-based and audience-driven environment of the creator economy.

This question is especially salient in platform-based creator economies such as YouTube, where economic outcomes depend not only on what creators produce, but even more on how their content is discovered, evaluated, and circulated by audiences (Veletsianos et al., 2018; Rieder et al., 2023; Joseph and Bishop, 2024). In this setting, creators' most visible and top-popular uploads are especially relevant: they are the product most likely to attract attention, engagement, and monetizable opportunities, and therefore make them especially useful for understanding which forms of content and audience response matter most for income. Prior research showed that women in public-facing digital work often face gendered vulnerability, harassment, and unequal forms of visibility

(Döring and Mohseni, 2020; McCarthy, 2021; Duffy et al., 2023). Related evidence from YouTube suggests that audience discourse and comment sentiment vary systematically by the presenter's gender (Veletsianos et al., 2018). These findings suggest that inequality in the creator economy may be shaped not only by creators' personal characteristics or observable channel and production features, but also by how audiences respond to their content.

Textual data are particularly well-suited to studying these mechanisms. Recent advances in text-as-data and natural language processing (NLP) have made it increasingly feasible to extract systematic measures of tone, emotion, and evaluation from large volumes of unstructured language (Németh and Koltai, 2023; Bittermann and Fischer, 2024). This is particularly valuable in platform settings such as YouTube, where titles, descriptions, captions, and comments capture both how creators present their content and how audiences respond to it. More specifically, these textural measures reflect not only how creators frame and present their content through style, tone, and self-presentation, but also audience-side reactions such as support, approval, criticism, and hostility. Existing studies already showed that online discourse and comment sentiments can vary systematically by gender (Thelwall and Más-Bleda, 2018; Veletsianos et al., 2018), suggesting that language may be an important channel through which gendered evaluation is shaped online. However, there is still little quantitative evidence on whether such textual and audience-side measures help explain gender differences in creator income.

We address this question by building on our previous income gap analysis, which documented a 22.8% residual gender income gap even after conditioning on observable creator-, channel-, and non-textual video-level characteristics. Because advertising income is mechanically determined by views and CPM, as discussed in detail in Chapter 2, this chapter does not interpret the residual gap as evidence of direct platform discrimination. Instead, it examines which observable factors help explain gender differences in the underlying margins that ultimately determine earnings. We enrich that baseline with newly collected video-level data from YouTube-related APIs for a final sample of 485

U.S. creators, constructing measures from each creator's up to 50 most popular long-form videos and up to 50 most popular shorts[1], including general top content[2] characteristics, creator-generated text, audience comments, and commercial and interaction measures. Using sequential inclusion regressions and a Gelbach decomposition (Gelbach, 2016), we show that the residual gap can be fully explained once these richer video-level measures are taken into account. General top content characteristics explain an important part of the gap, but they do not explain it on their own. Rather, the broader pattern of results indicates that audience-side measures are more informative than creator-side features. In particular, audience comments emerge as the most informative text-source dimension, sentiment-related measures are the most informative linguistic dimension, and two variable-level factors stand out most clearly: unequal audience endorsement of top content, captured by average likes, and the evaluative tone of audience response, captured by comment sentiment. By contrast, commercial signals such as sponsorship, self-promotion, and interaction prompts add little explanatory power.

Our study contributes in three main ways. First, it provides, to our knowledge, one of the first systematic quantitative analyses of gender differences in estimated advertising income among top U.S. YouTube creators, addressing an important gap in labour economics and platform studies. Second, it contributes a uniquely large-scale text-based dataset, combining information from 41,180 most popular videos and 17.4 million comments. To our knowledge, it is among the first studies to use text-based measures at this scale to examine how creator-side language and audience-side discourse relate to gender inequality in creator income. Third, it contributes a substantive empirical result by showing that the residual gender income gap can be fully explained once a much richer set of observable

[1] YouTube distinguishes between *shorts* and longer standard videos not only by format and duration, but also by how they are distributed and monetized. Under YouTube's current rules, vertical or square videos of up to three minutes may be classified as *shorts*, which are monetized through the Shorts Feed revenue-sharing model, whereas longer standard videos are monetized in the long-form model on the watch page. We therefore analyse the two formats separately and use the term *long-form videos* whenever we refer specifically to videos that are not shorts, in order to avoid confusion with the broader term *videos*, which could otherwise include both formats.

[2] Throughout this chapter, top content refers collectively to creators' most popular long-form videos and most popular Shorts.

top content, textual, and audience-response variables is taken into account.

The remainder of the chapter is organized as follows. Section 3.2 reviews the related literature. Section 3.3 describes the data, sample construction, and variable design. Section 3.4 outlines the empirical strategy. Section 3.5 presents the main results, including the sequential specifications, Gelbach decomposition, separate block inclusion analysis, variable-level evidence, and robustness checks. Section 3.6 concludes.

3.2 Literature review

This study relates to four closely connected strands of literature: (1) the economics of the creator economy and platform monetization, (2) gender inequality in platform labor, (3) language as an economically relevant signal in digital markets, and (4) gendered audience discourse on YouTube. For the purposes of this thesis chapter, the broader literature on gender gaps in labor and platform markets is discussed in Chapter 2, the review here therefore focuses only on the strands most directly relevant to this chapter's additional contribution.

A first strand of the literature examines how creators generate income and visibility in platform-based digital markets. Recent work emphasizes that creator income is shaped less by any single revenue source than by the degree of visibility creators achieve within platform ecosystems. On platforms such as YouTube, monetization depends not only on direct advertising revenue, but also on audience development, engagement, and, to some extent, the additional visibility generated through cross-platform linking to channels such as Instagram and Twitter/X. For example, Rieder et al. (2023) showed that YouTube creators use linking practices to strengthen their online presence, while Sanyoura and Anderson (2022) documented how membership platforms such as Patreon can convert follower attention into recurring financial support. More fundamentally, however, this literature stresses that monetization depends heavily on platform visibility and engagement metrics. Duffy et al. (2021) argued that quantified indicators such as likes, favorites, subscribers,

and shares are widely treated as pathways to professional success in the creator economy, even as they deepen insecurity and dependence on platform systems. Taken together, this strand clarifies why creator labor is both economically meaningful and structurally precarious, but it says much less about whether earnings are systematically stratified by gender.

A second strand examines inequality in platform labor more directly. This literature shows that visibility and monetization are not distributed neutrally across creators. Duffy et al. (2023) argued that public visibility in the influencer economy carries distinctive vulnerabilities, especially for women and other marginalized creators, while Glatt (2023) emphasized that creators' careers depend on relational labor with audiences and that this labor is shaped by broader structural inequalities. Related work also shows that unequal opportunities in creator markets need not arise solely through formal employment relationships. For instance, Christin and Lu (2023) documented an influencer pay gap shaped by racialized market processes, and McCarthy (2021) shows how online misogyny in YouTube comment spaces structures women's experiences of public-facing digital work. This literature is highly relevant to the present study, but it focuses mainly on precarity, harassment, branding, and sponsorship markets. Much less is known about whether comparable gaps arise in YouTube advertising income, and about which observable mechanisms account for them once standard creator- and channel-level characteristics are held constant.

A third strand concerns language itself. In economics and related social sciences, recent advances in text-as-data methods have made it increasingly feasible to convert large volumes of unstructured language into systematic measures of sentiment, evaluation, similarity, and topical content (Gentzkow et al., 2019; Ash and Hansen, 2023). As Ash and Hansen (2023) emphasize, textual data often provide rich, and sometimes unique, information about economically important concepts that are difficult to observe in structured variables alone. This insight is particularly relevant in digital markets, where much of the economically meaningful interaction among producers, platforms, and audiences is

recorded in textual form. In creator markets, language is not merely descriptive. It is integral to how creators construct personal brands and cultivate audience relationships, how viewers evaluate content through comments and sentiment-laden feedback, and how attention is coordinated and amplified across platforms (Tur-Viñes and Castelló-Martínez, 2019; Catalina-García and Suárez-Álvarez, 2022; Joshi et al., 2023; Kim, 2024). At the same time, many related empirical works on creators, influencers, and platform participation has tended to measure success and engagement primarily through platform-visible indicators, such as views, likes, subscribers, retention, or linking behavior, rather than through systematic measures of textual tone or audience discourse (Yang et al., 2022; Sanyoura and Anderson, 2022; Rieder et al., 2023). Although recent studies have begun to analyze creator- and audience- generated text directly, including comments, sentiment, and discursive interaction (Tur-Viñes and Castelló-Martínez, 2019; Poecze et al., 2019; Catalina-García and Suárez-Álvarez, 2022), these approaches remain less central than metric-based analyses in the broader empirical literature on creator markets, especially in economics.

A fourth strand focuses more specifically on gendered audience discourse on YouTube. Veletsianos et al. (2018) analyzed sentiment in comments on YouTube TEDx and TED-Ed videos and showed that comment environments vary with presenter gender, video format, and threading structure. Similarly, Thelwall and Más-Bleda (2018) studied YouTube science channels and asked whether the surrounding comment environment is female-friendly or still marked by sexism. These studies are important because they establish that audience comments are a meaningful site of gendered evaluation on the platform. They also suggest that comment sentiment may capture more than noise: it may reflect systematic differences in how audiences react to male and female presenters. However, this literature stops short of connecting those audience-side textual environments to creators' economic outcomes. In other words, we know that online discourse can differ by gender, but we know much less about whether those differences help explain gender differences in income.

The present study brings these strands together. From the creator economy literature, it takes seriously the idea that visibility, engagement, and monetization are jointly determined and that the most economically relevant part of a creator's output may be the part that becomes most visible. This motivates our focus on each creator's most popular content rather than on a random or purely recent sample of videos. From the labor inequality literature and gendered evaluation literature, it draws the insight that audience response itself may be an important mechanism through which inequality is reproduced online. From the text-as-data literature, it adopts the view that language can be measured systematically and linked to economic outcomes. Our contribution is to bring these insights together by examining whether richer video-level measures from creators' top content, especially creator- and audience-side textual measures, help explain the residual gender gap in YouTube advertising income after conditioning on observable creator, channel, and non-text video-level characteristics.

3.3 Data

3.3.1 Data sources

We begin by assembling a multi-level dataset from two primary data sources, *Social Blade* and *YouTube*, supplemented with creator demographic characteristics that we manually collect and cross-validate across multiple online platforms, including public databases, social media profiles, and publicly available interviews. Our dependent variable is the logarithm of estimated annual advertising income for creator, proxied by the midpoint of *Social Blade*'s reported annual income range. This income measure was introduced and discussed in detail in Chapter 2. Together, these data allow us to examine gender disparities in creator income using observable creator-, channel-, and video-related characteristics, including annual upload measures and aggregates based on the 50 most recent videos covering both long-form videos and shorts. Figure 3.1 provides an overview of the data sources, data enrichment process, and construction of the final analytical dataset.

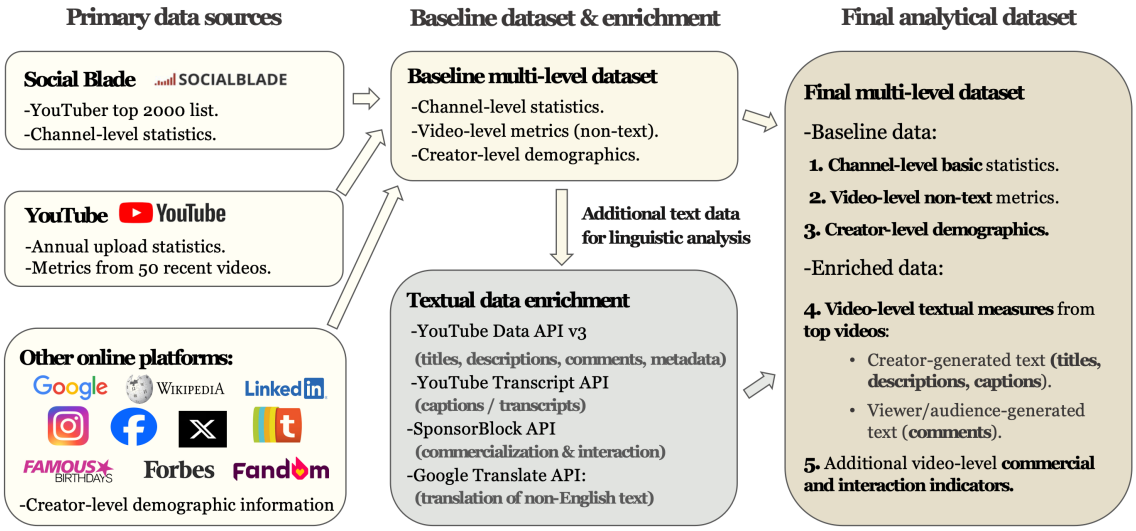


Figure 3.1. Data sources and construction of the analytical dataset

To further enrich the dataset, particularly for the textual analysis, we collect additional video-level textual and contextual information[3] from a bounded set of best performing content, consisting of up to 50 most popular long-form videos and 50 most popular shorts of each creator. These data are retrieved through several data access tools and interfaces. Specifically, we use the *YouTube Data API v3*, the *YouTube Transcript API*, and the *SponsorBlock API* to retrieve video titles, descriptions, captions, comments, and other video-level commercial and audiance- interaction metadata, while the *Google Translate API* is used to automatically translate non-English textual content (although such cases are rare). These enriched measures are then merged with the baseline dataset to support the construction of the final analytical sample. Detailed information on realized data coverage is reported later in Table 3.1.

3.3.2 Sample construction

Figure 3.2 summarizes the sample construction process. We begin with the top 2,000 U.S. YouTube channels ranked by subscribers from *Social Blade* and restrict attention to channels operated by individuals only. Among these channels, we identify those with veri-

[3]Here, "contextual information" refers to non-textual video-level attributes, including metadata such as publication timing, video duration, views, likes, and comment counts, as well as sponsorship, monetization indicators and measures of audience interaction.

fiable gender and demographic information through manual collection and cross-validation across multiple public sources. Following this verification process, the broader creator-level sample contains 518 non-celebrity individual YouTubers with reliable demographic information and valid income estimates. These 518 creators constitute the baseline sample used in the income gap analysis and provide the foundation for the present study.

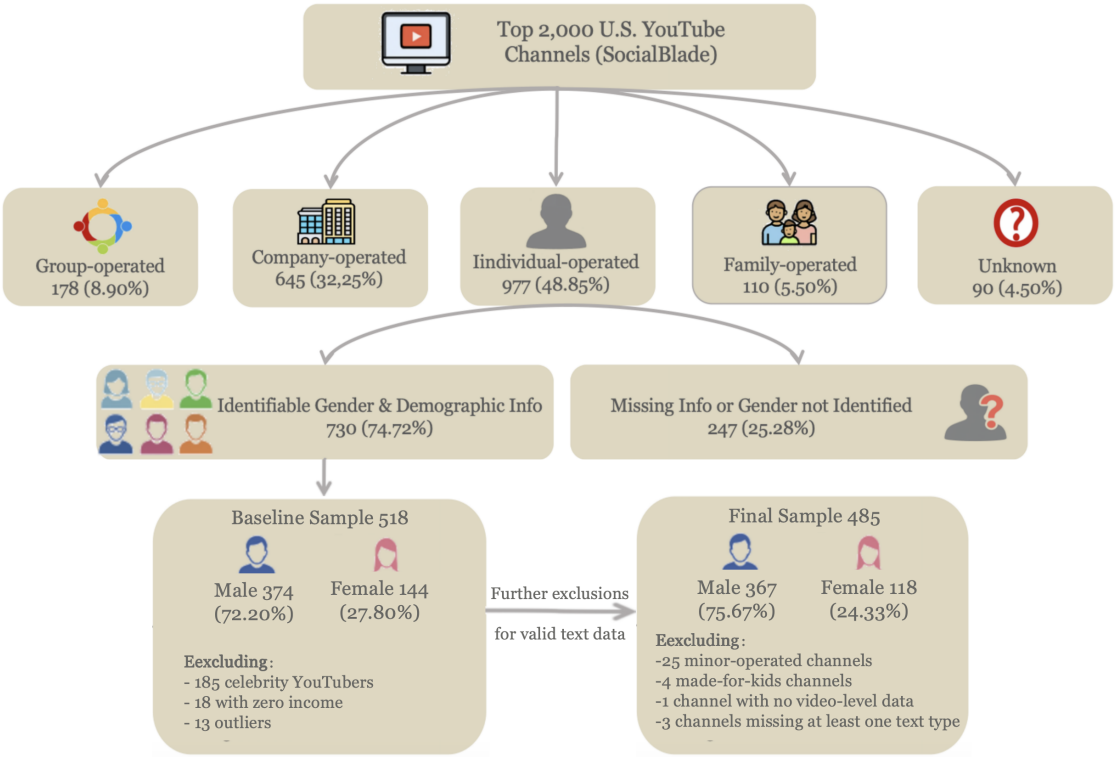


Figure 3.2. Sample construction

To construct the final analytical sample for textual analysis, we impose several additional restrictions on the baseline sample. First, we exclude 25 channels operated by minors and ¨made-for-kids¨ channels[4], because comments are typically disabled for such content, making audience-text analysis infeasible. Second, we exclude 1 channel for which no video-level data were available at the time of collection[5]. Third, we exclude 3 channels for which the sampled long-form videos and shorts contain no valid observations in at least one of the four text categories used in our measures: titles, descriptions, captions, and

[4]Here, "made-for-kids" refers to channels classified under *YouTube*'s official audience setting as primarily directed to children, whether designated by the creator or identified by the platform.

[5]This may occur when a creator has removed or hidden previously public content for personal or strategic reasons, or when videos become unavailable because of platform enforcement actions, including copyright- or policy-related removals.

comments. After these exclusions, the final analytical sample consists of 485 YouTube channels, including 367 male creators and 118 female creators. The next subsection describes how, based on these 485 channels, we identify the top long-form videos and shorts and collect the corresponding video-level textual and contextual information for the empirical analysis.

3.3.3 Text analysis

Building on the final sample of 485 creators, we construct the enriched video-level dataset used for textual analysis. For each creator, we identify a bounded set of most popular content, including both long-form videos and shorts, based on their observed popularity at the time of data collection. This sampling strategy improves comparability across creators while focusing on content that is most representative of audience engagement. Figure 3.3 summarizes the overall data collection and feature construction workflow, which proceeds in three stages.

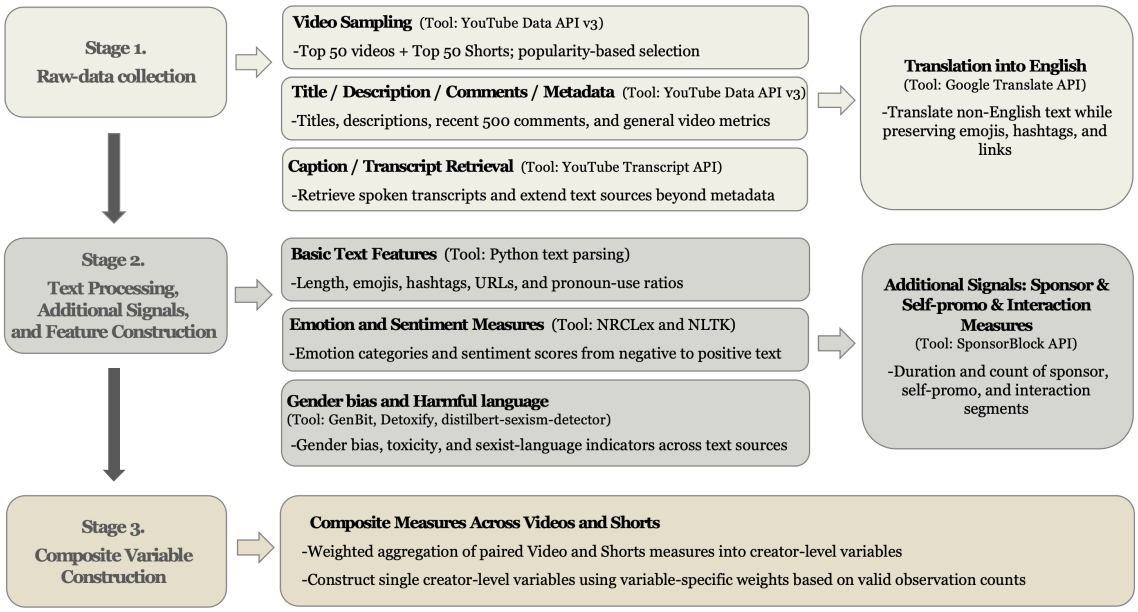


Figure 3.3. Workflow for top-popular video data collection and feature construction

Stage 1: Raw-data collection. In the first stage, we collect up to 50 most popular long-form videos and up to 50 most popular shorts[6] for each creator using the official *YouTube Data API v3*, between November and December 2025. We adopt popularity-based rather than recency-based sampling because the most viewed and most engaged uploads are more likely to capture a creator's economically relevant content style and effective audience reach. For each selected upload, we retrieve the title, description, up to 500 most recent comments[7], and metadata such as publication date, video duration, view count, like count, and comment count. Because captions are not directly available through the official *YouTube Data API*, transcripts are collected separately via the *YouTube Transcript API*. Meanwhile, any non-English text is automatically translated into English using the *Google Translate API* to ensure a consistent measurement framework applied across channels, although such cases are rare. Notably, to ensure comparability of long-form video- and shorts-level measures across creators, time-sensitive metadata such as views, likes, and comments, were extracted within 2 days using multiple API accounts. By contrast, other metadata like captions, are large in volume but fixed once the video is published. Thus, these data were extracted gradually afterwards.

Table 3.1 summarizes the realized coverage of the collected upload sample for the final set of 485 creators. The dataset contains 41,180 uploads in total, including 23,084 (long-form) videos and 18,096 shorts. Title coverage is complete for the full sample, yielding 41,180 valid titles (100.0%). We obtain 34,054 valid descriptions (82.7%) and 37,215 valid transcripts (90.4%). Comments are available for 40,178 uploads (97.6%), from which we collect a total of 17,398,194 comments, averaging 433 comments per upload with comments available. Because data availability differs across uploads and formats, the realized number of observations varies by text source. In particular, while

[6]We use "up to" because not all creators have at least 50 publicly available long-form videos or 50 publicly available shorts at the time of data collection. Some creators specialize in only one format, while others have smaller or incomplete archives.

[7]We use the most recent comments rather than "top comments" because the latter are shaped by *YouTube*'s ranking and moderation mechanisms, which may systematically downweight or filter harmful content such as abusive or gender-targeted comments. As a result, recent comments provide a more natural and less curated view of audience responses.

title coverage is universal, descriptions and transcripts are missing for a subset of uploads, and comment coverage depends on whether comments are enabled and retrievable at the time of collection.

Table 3.1. Coverage of YouTube top content and textual data

	Long-form Videos	Shorts	Total	Presence rate (%)
Panel A. Upload counts				
Total uploads collected	23,084	18,096	41,180	
Panel B. Text-field coverage				
Uploads with valid titles	23,084	18,096	41,180	100.00
Uploads with valid descriptions	22,750	11,304	34,054	82.70
Uploads with valid transcripts	21,574	15,641	37,215	90.37
Panel C. Comment coverage				
Uploads with comments available	22,404	17,774	40,178	97.57
Total comments collected	10,421,204	6,976,990	17,398,194	
Mean comments/upload with comments	465.15	392.54	433.03	

Note: This table reports the coverage of uploads and associated textual data in the collected sample. The presence-rate column reports the share of uploads with the corresponding valid field among all uploads in the full sample of 41,180 uploads.

Stage 2: Text processing, additional signals, and feature construction. Based on the raw inputs collected in Stage 1, we construct measures separately from four text sources: titles, descriptions, captions (transcripts), and comments. The processing of the collected textual data was conducted between December 2025 and January 2026. These textual measures are grouped into three broad categories. The first captures (1) Basic text features, including length, the use of emojis and hashtags as indicators of expressive style and topic engagement, the presence of URLs (primarily in descriptions) as signals of external linking and cross-platform or commercial redirection, and selected pronoun-use ratios derived mainly from captions (transcripts) to characterize speech style and audience orientation. The second captures (2) Emotion and sentiment, which analyzed textual data using the most established and widely used *natural language processing (NLP)* tools. Specifically, *NLTK* supported text processing and *VADER* sentiment (polarity) analysis, together with *NRCLex* for lexicon-based emotion analysis to distinguish more positive, neutral, and negative tones from both the creator and the audience sides. The third captures

(3) Gender bias and harmful language, which analyzed textual data using well-established *NLP* tools available, including *GenBiT* for gender bias, *Detoxify* for toxicity, identity attack, and distilbert-sexism-detector for sexist language. Constructing these measures separately by text source allows us to distinguish creator-generated text from audience-generated text and to identify which textual dimensions are most closely associated with the residual gender income gap.

In parallel, we collect an additional set of video-level commercial and interaction signals using the *SponsorBlock API*. For each upload, we record both the duration and the count of labeled segments in three categories: sponsor content, self-promotion, and explicit interaction prompts, such as reminders to like, subscribe, comment, or follow. Although these signals may not map directly into advertising revenue package, they capture the extent to which creators embed brand-oriented promotion, self-marketing, and audience-activation strategies in their content, all of which may indirectly shape visibility, engagement, and ultimately income.

Stage 3: Composite variable construction. Most measures constructed in the previous stages are obtained separately for long-form videos and shorts. To reduce dimensionality while retaining information from both formats, we aggregate each paired long-form video and shorts measure into a single creator-level composite variable using a weighted mean.[8] The weights are variable-specific and depend on the number of valid observations available for each format. For general content-level measures, the weights are based on the numbers of long-form videos and shorts collected for each creator. For text-based measures, they are based on the numbers of valid observations for the corresponding text source, such as titles, descriptions, captions, or comments. This procedure places greater weight on the format contributing more usable information for a given measure and allows us to retain both long-form video and shorts information without including two highly collinear versions of the same characteristic.

[8]For creator i and characteristic k, the composite measure is computed as $\tilde{x}_{ik} = w_{ik}x^L_{ik} + (1 - w_{ik})x^S_{ik}$, where x^L_{ik} and x^S_{ik} denote the long-form video and shorts measures, respectively, and $w_{ik} = n^L_{ik}/(n^L_{ik} + n^S_{ik})$.

3.3.4 Descriptive statistics

This subsection introduces the main variables used in the empirical analysis and summarizes their distributions in the final analytical sample. As described above, the dataset combines multiple layers of information, which can be grouped into two broad components. The first consists of the baseline controls carried over from our previous income gap analysis: (1) Creator-level demographic and identity characteristics, (2) Channel-level characteristics and content category, and (3) Video-level measures of production, activity, and engagement[9]. The second consists of newly constructed video-level variables based on each creator's top content, including (4) General characteristics, (5) Textual features, and (6) Commercial and interaction signals.

Table 3.2 summarizes these newly constructed variables derived from each creator's top content used in the main analysis. While Table 3.3 further reports descriptive statistics for these variables by creator gender. The table presents sample means for the full sample and separately for male and female creators, together with the corresponding mean differences. Several broad patterns emerge.

From the creator side, male creators tend to have larger values on several general measures of top content. In particular, they exhibit significantly longer average upload duration and higher average likes, whereas gender differences in average comments and upload recency are small and statistically insignificant. Female creators, by contrast, tend to use longer and more positive titles and make greater use of emojis and hashtags in titles. In descriptions, they also display more positive sentiment and greater use of emojis, while gender differences in description length, URLs, and hashtags remain limited. In captions, female creators again appear more positive on average and use a higher first-person pronoun ratio, consistent with a somewhat more personal and affective communication style. At the same time, female creators also score slightly higher on the sexism-related language

[9]These measures combine annual and recent information. Annual production and one activity measure are based on yearly upload data, whereas upload frequency and engagement are based on the creator's 50 most recent videos. Activity is captured both by whether any video was uploaded during the year and by the average gap between consecutive recent uploads.

Table 3.2. Description of newly added top content variables

Variable	Description
General Characteristics (metadata)	
Duration	Duration of long-form videos and shorts.
Likes	Likes on long-form videos and shorts.
Comments	Comments on long-form videos and shorts.
Recency	Days since upload for long-form videos and shorts.
Textual Features (by four text sources)	
1. Title	
Length	Title length.
Sentiment	Title sentiment score.
Emojis and hashtags	Number of emojis and hashtags used in titles.
Bias & harm indicators	Gender bias, identity attack, severe toxicity, and sexualized-content measures.
2. Description	
Length	Description length.
URLs, emojis, and hashtags	Number of URLs, emojis, and hashtags used in descriptions.
Sentiment	Description sentiment score.
Bias & harm indicators	Gender bias, identity attack, severe toxicity, and sexualized-content measures.
3. Caption/Transcript	
Length	Caption/transcript length.
Pronoun use	First-, second-, and third-person pronoun-use ratios.
Sentiment and emotions	Sentiment and emotion-related measures derived from captions/transcripts.
Bias & harm indicators	Gender bias, identity attack, severe toxicity, and sexualized-content measures.
4. Comments	
Length	Comment length.
Sentiment and emotions	Comment sentiment score.
Emojis and hashtags	Number of emojis and hashtags used in comments.
Bias & harm indicators	Gender bias, identity attack, severe toxicity, and sexualized-content measures.
Commercial and Interaction Signals	
Sponsor segment duration	Duration of sponsor segments identified through the SponsorBlock API.
Self-promotion duration	Duration of self-promotional segments identified through the SponsorBlock API.
Interaction segment duration	Duration of segments encouraging likes, subscriptions, comments, follows, or related audience actions.

Note: This table reports newly added video-level variables from creators' top content used in the main analysis. Measures are aggregated to the creator level using each creator's long-form videos and shorts with available text or comment data, following the weighting procedure described in Section 3.3.3. Baseline controls are omitted for brevity. The full set of regression variables, including baseline controls and newly constructed measures, is reported in Appendix B, Table B1.

Table 3.3. Descriptive statistics by gender: Aggregated top content variables

Variables	Mean (N=485)	Male (N=367)	Female (N=118)	Difference Male–Female
General Characteristics				
Average duration (seconds)	596.811	628.458	498.384	130.074**
Average likes	575,000	612,000	462,000	149,000*
Average comments	12,058.241	12,312.099	11,268.701	1,043.398
Average days since upload	1,669.968	1,671.945	1,663.819	8.126
Title Variables				
Title length	7.248	7.028	7.933	−0.905***
Title sentiment	0.008	-0.003	0.042	−0.045***
Title emojis	0.198	0.158	0.320	−0.162***
Title hashtags	0.253	0.205	0.402	−0.197***
Title gender bias ratio	0.010	0.022	-0.027	0.049***
Title identity attack	0.001	0.002	0.001	0.001**
Title severe toxicity	0.000	0.000	0.000	0.000***
Title sexism	0.015	0.014	0.017	−0.003***
Description Variables				
Description length	112.310	111.722	114.142	−2.420
Description sentiment	0.500	0.455	0.642	−0.187***
Description URLs	5.249	5.229	5.312	−0.083
Description emojis	1.457	1.321	1.878	−0.557*
Description hashtags	1.478	1.461	1.530	−0.069
Description gender bias ratio	0.090	0.112	0.021	0.091***
Description identity attack	0.001	0.001	0.000	0.000
Description severe toxicity	0.000	0.000	0.000	0.000
Description sexism	0.017	0.014	0.028	−0.014***
Caption / Transcript Variables				
Transcript length	1,538.062	1,621.766	1,277.727	344.039**
Transcript sentiment	0.613	0.595	0.669	−0.073***
Transcript 1st-person pronoun ratio	0.063	0.062	0.070	−0.008***
Transcript 2nd-person pronoun ratio	0.035	0.035	0.034	0.001
Transcript 3rd-person pronoun ratio	0.037	0.037	0.036	0.001
Transcript NRC positive emotion	0.194	0.190	0.206	−0.016***
Transcript NRC negative emotion	0.142	0.145	0.132	0.013***
Transcript VADER positive sentiment	0.671	0.669	0.677	−0.009
Transcript VADER negative sentiment	0.148	0.159	0.114	0.045***
Transcript gender bias ratio	0.278	0.330	0.118	0.212***
Transcript identity attack	0.005	0.006	0.002	0.004***
Transcript severe toxicity	0.003	0.003	0.002	0.002*
Transcript sexism	0.025	0.024	0.027	−0.003
Comment Variables				
Comment length	12.131	12.339	11.484	0.855*
Comment sentiment	0.108	0.093	0.156	−0.063***
Comment emojis	1.073	0.977	1.371	−0.395***
Comment hashtags	0.003	0.003	0.004	−0.001
Comment gender bias ratio	0.287	0.428	-0.154	0.582***
Comment identity attack	0.004	0.004	0.002	0.002***
Comment severe toxicity	0.003	0.003	0.001	0.002***
Comment sexism	0.985	0.986	0.984	0.002***
Commercial and Interaction Signals				
Sponsor segment duration (seconds)	4.119	5.009	1.350	3.659***
Self-promotion duration (seconds)	1.645	2.004	0.531	1.472***
Interaction segment duration (seconds)	0.998	1.219	0.310	0.910***

Note: This table reports descriptive statistics for creator-level measures obtained by aggregating video-level content variables across each creator's long-form videos and shorts with available underlying data. The last column reports the male–female mean difference. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Baseline-variable statistics are reported in Appendix B, Table B2.

measures in titles, descriptions, and captions. We interpret this pattern cautiously: rather than indicating more hostile gender discourse, it likely reflects a greater use of gender-salient self-presentational language. For example, related research (Duffy and Hund, 2015; Toffoletti and Thorpe, 2018) suggests that language related to femininity, relationships, embodiment, or empowerment, is common in gendered self-branding under platform visibility and attention pressures. Male creators, by contrast, score higher on several bias- and toxicity-related indicators in creator-generated text, especially in captions, and they also devote substantially more time to sponsor, self-promotion, and audience-interaction segments.

From the audience side, the comment measures reveal systematic differences in the discourse surrounding male and female creators. Comments on male creators are longer on average and score higher on the bias, identity-attack, severe-toxicity, and sexism measures, whereas comments on female creators are more positive in sentiment and contain more emojis. One possible interpretation, consistent with prior research, is that different forms of online harm are distributed differently across gendered publics. Existing studies show that women in visible online roles are especially exposed to gendered harassment, including misogynistic, sexualized, or appearance-based attacks, rather than merely higher levels of general toxicity or adversarial language (Chen et al., 2018; Boulianne et al., 2019; Cambronero-Saiz et al., 2023). By contrast, male-dominated or male-centered publics may be more likely to generate openly confrontational, polarized, and profane discussion styles, which are more directly captured by our retained comment measures.

This interpretation is also consistent with research on YouTube and other social platforms showing that highly active commenters tend to be more polarized and more toxic, and that contentious content ecosystems can foster cascades of hostile interaction (Lewis et al., 2021; Locatelli et al., 2022; Miyazaki et al., 2024). In addition, the platform literature suggests that attention-based systems often amplify content that generates controversy, repeated commenting, and sustained engagement, thereby increasing the visibility of adversarial discussion (Tuters and Burton, 2021; Kovacheva et al., 2022; Ringrose et al.,

2024). In our setting, male creators also exhibit higher average likes and longer uploads, which may expose them to larger and potentially more conflict-prone audiences. Taken together, these mechanisms suggest that the higher levels of harmful language observed in comments on male creators may reflect not lower risk for women, but rather a difference in the form of online hostility: women may face more gendered or objectifying abuse, whereas male creators may attract more broadly toxic, polarized, and adversarial comment environments.

3.3.5 Emotion disaggregation

The descriptive table reports broad sentiment differences across titles, descriptions, captions, and comments, and in all four sources the aggregate measures suggest that female creators use or receive more positive language on average. While useful, those summary indicators necessarily compress multiple emotional dimensions into a single polarity score. Recent research in economics emphasizes that text can capture economically meaningful concepts that are difficult to observe in structured variables alone (Gentzkow et al., 2019; Ash and Hansen, 2023). In the present setting, disaggregating sentiment into finer emotional categories is useful because it shows which emotions underlie the broader sentiment differences and whether a more positive average sentiment masks offsetting patterns across specific emotional dimensions.

For this reason, this subsection examines the distribution of emotion specific variables across text sources. Notably, these measures are not included directly in the main regressions. The reason is mainly econometric: within each text source, the emotion scores are highly collinear with one another and overlap substantially with the broader sentiment measures already used in the empirical analysis. Adding the full set of emotion categories would therefore expand the dimensionality of the model while making coefficient interpretation unstable. We therefore treat the emotion variables as supplementary descriptive evidence. Their purpose is not to replace the sentiment measures used in the regressions, but to clarify what those aggregate indicators are summarizing.

The main text focuses on the eight emotion categories that exhibit the most meaningful variation in the current data. The *neutral category* is reported separately in Appendix Figure A1. This separation is substantively useful. In particular, title text is overwhelmingly classified as neutral for both male and female creators, so the title-level emotional profile outside the neutral category is necessarily sparse. Reporting neutral separately therefore prevents the main figures from being dominated by a category that mainly reflects the absence of emotionally distinctive wording rather than variation across specific emotions.

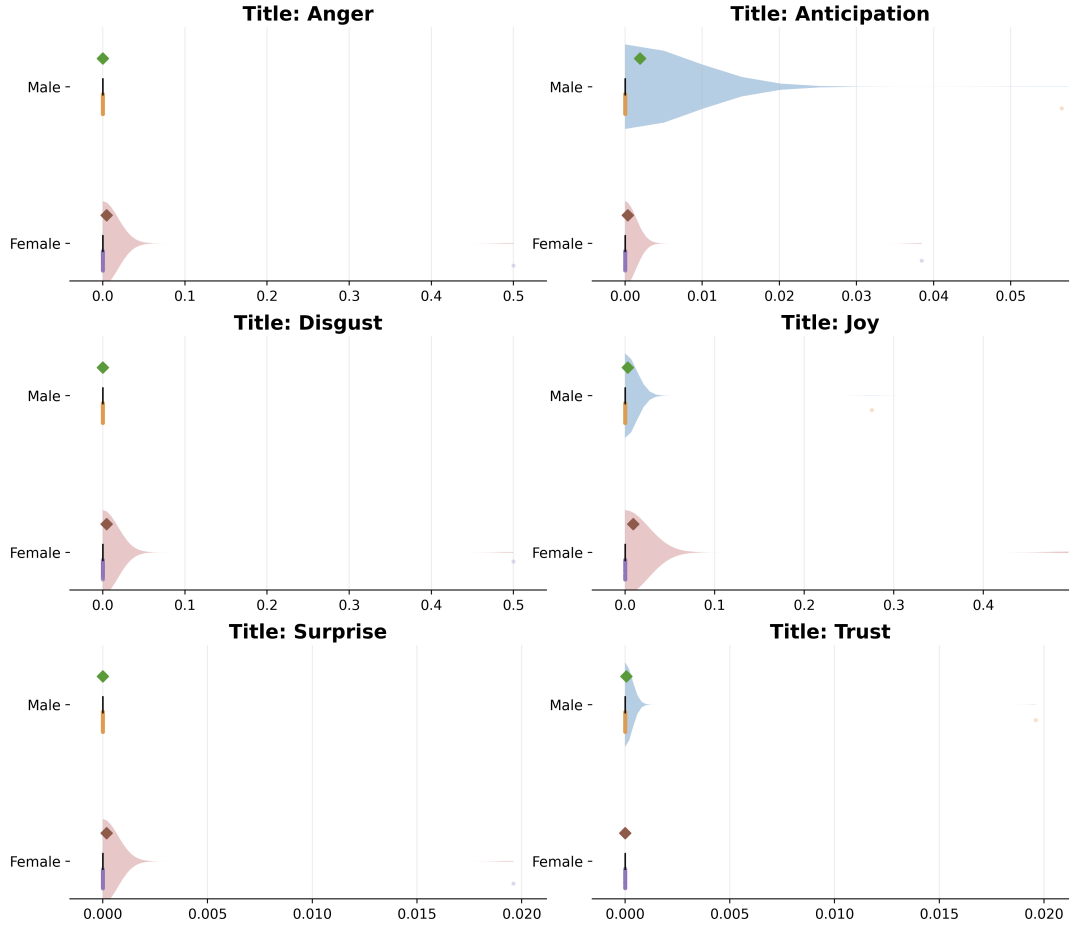


Figure 3.4. Gender differences in emotion profiles: Titles

Figure 3.4 shows that emotional variation in titles is limited overall. Most title observations are concentrated near zero for the *non-neutral categories*, which is consistent with the evidence in Figure A1 that titles are predominantly neutral for both genders. Within that limited variation, female creators show somewhat higher mass in *anger*, *disgust*, and *joy*, while male creators appear slightly more represented in anticipation and *trust*. How-

ever, these differences are small relative to the much larger neutral component of titles, and the distributions overlap heavily. The main implication is therefore not that title emotions differ dramatically by gender, but that title language contains relatively little emotionally distinctive content once neutral wording is set aside. This helps explain why title text is unlikely to be a dominant mechanism in the later regression analysis.

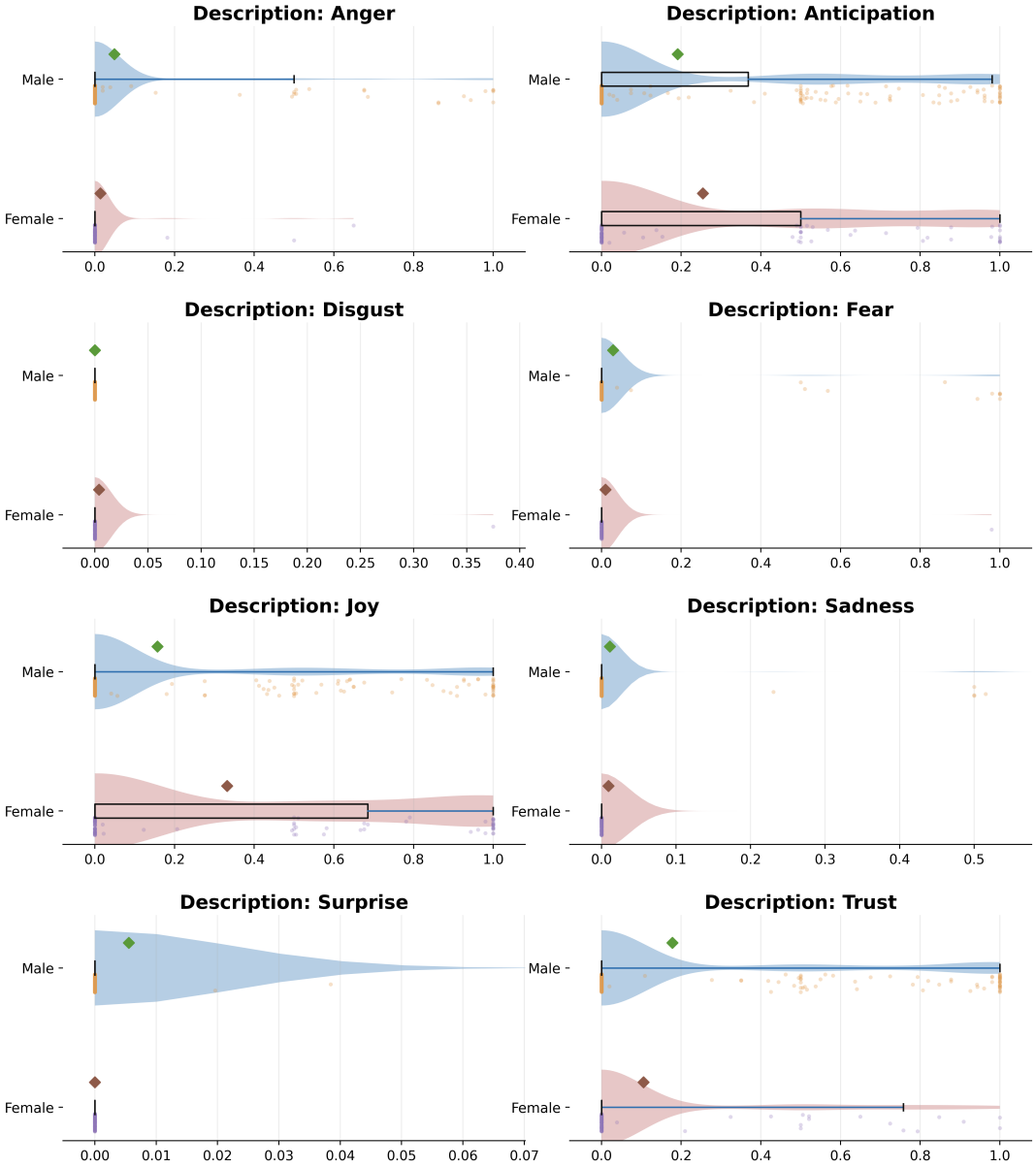


Figure 3.5. Gender differences in emotion profiles: Descriptions

Figure 3.5 reveals a much richer pattern in descriptions. The most pronounced female-skewed category is *joy*, whose distribution is clearly shifted upward for female creators.

By contrast, male creators are more represented in *anger* and *trust*, and they also show somewhat higher values in *fear*. Differences in *anticipation* are more modest, while *disgust*, *sadness*, and *surprise* remain comparatively small and concentrated near zero. This pattern matters because it shows that the more positive average description sentiment among female creators is driven mainly by stronger *joyful* language rather than by a uniform reduction in all negative emotions. At the same time, some seemingly favorable categories, especially *trust*, are actually more prevalent among male creators. The aggregate sentiment gap in descriptions therefore masks meaningful heterogeneity in the underlying emotional composition.

Figure 3.6 is particularly informative because captions contain richer language than titles or descriptions. Here again, *joy* is substantially more prevalent for female creators, which is consistent with the more positive aggregate caption sentiment reported in the descriptive statistics. However, the finer decomposition also shows that male creators are more represented in *fear* and *trust*, while *anticipation* is broadly similar across genders. The remaining categories, such as *anger*, *disgust*, and especially *sadness*, are small in absolute magnitude even when one side appears somewhat more concentrated in the upper tail. This is exactly the type of heterogeneity that a single sentiment score obscures: captions can be more positive on average for female creators while still exhibiting dimensions, most notably *trust* and *fear*, that are more prevalent among male creators.

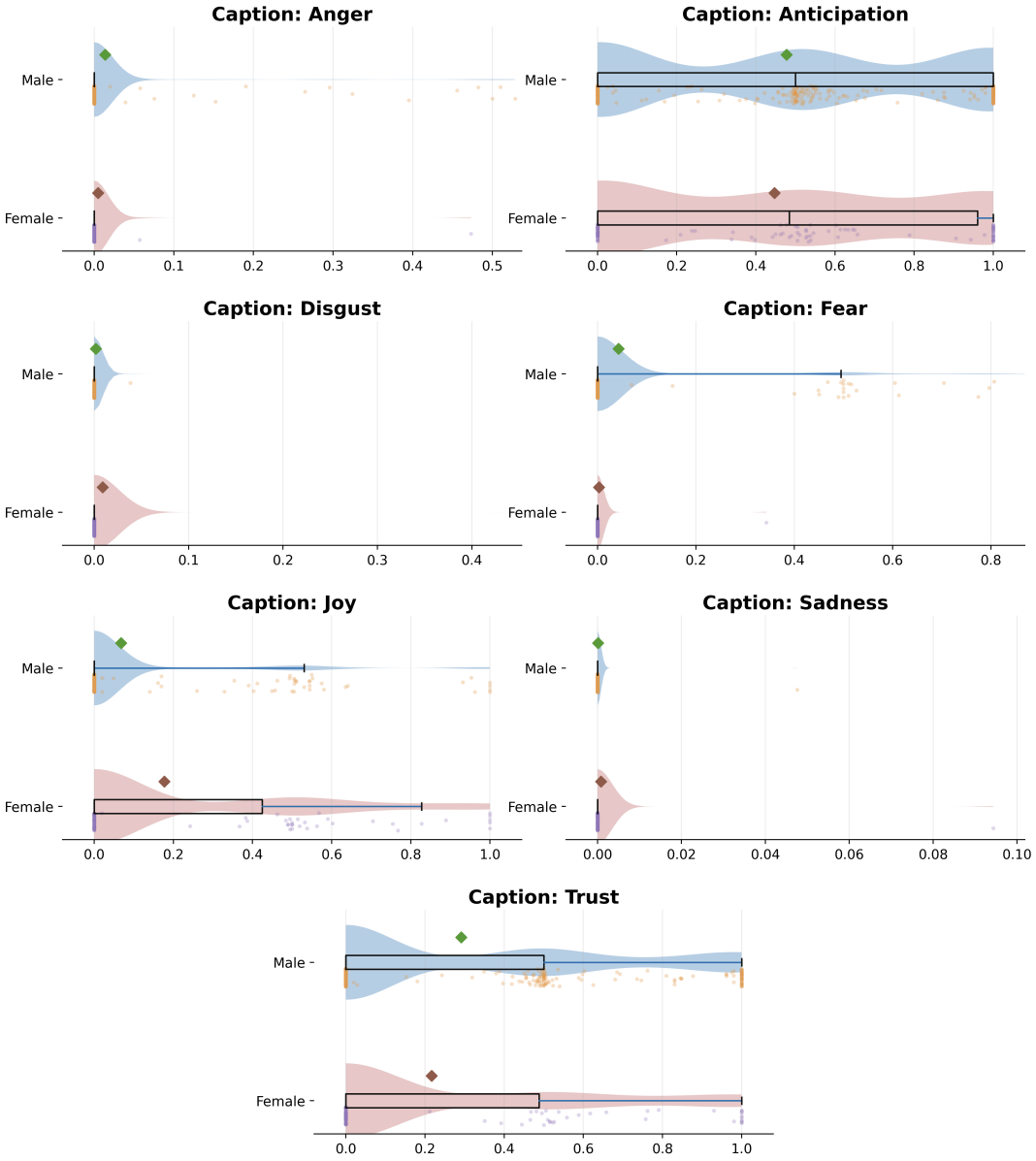


Figure 3.6. Gender differences in emotion profiles: Captions

Figure 3.7 is especially important because comments capture audience-side evaluation rather than creator-side self presentation. The strongest difference is in *joy*: female creators receive comment environments with much higher *joyful* content. At the same time, male creators receive comments with substantially higher *trust*, and also clearly higher *fear*, *anticipation*, and *anger*. By contrast, *sadness* and *disgust* remain small and are much less central to the overall pattern, while *surprise* is also limited in magnitude. This result is substantively useful because it shows that comment sentiment is not summarizing a simple

opposition between "positive" and "negative" reactions. Rather, audience responses differ in their emotional composition: female creators appear to receive more *joyful* comment environments, whereas male creators receive comment environments characterized by more *trust* but also more *fear*, *anticipation*, and *anger*. In this sense, the aggregate sentiment gap masks important heterogeneity in the underlying emotional structure of public response.

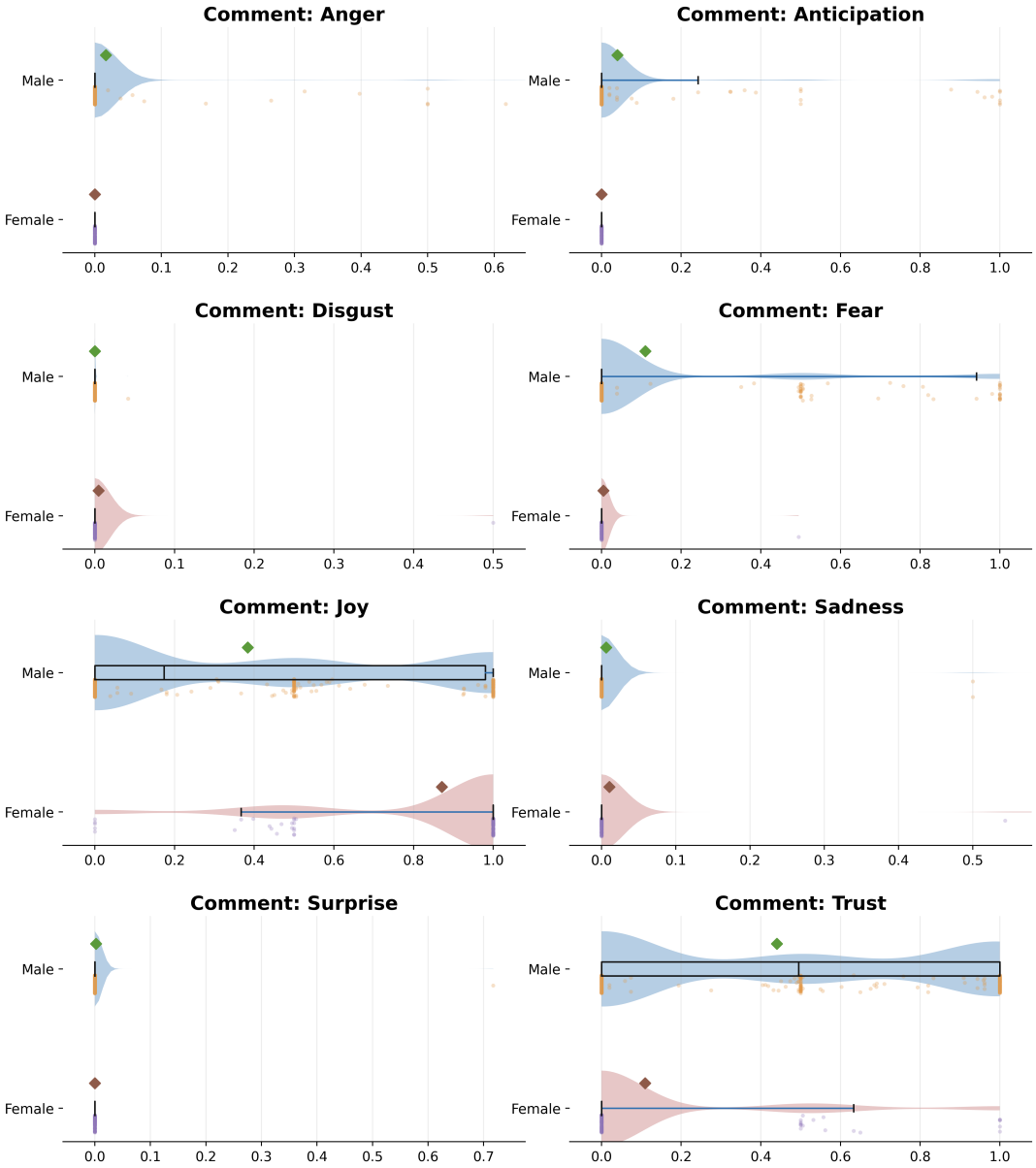


Figure 3.7. Gender differences in emotion profiles: Comments

Taken together, Figures 3.4–3.7 convey three main points. First, the more positive aggregate sentiment measures observed for female creators do not imply that women score

higher on every favorable emotional category or lower on every unfavorable one. Rather, the differences are selective and source-specific. Second, the clearest and most systematic female-skewed category is *joy*, particularly in descriptions, captions, and comments. Third, male creators tend to be more represented in categories such as *trust*, *fear*, *anticipation*, and in some contexts *anger*, particularly on the audience side. This is consistent with the broader interpretation of comments as a gendered evaluative environment rather than just another text field(Veletsianos et al., 2018).

At the same time, these figures remain descriptive. They do not show whether any specific emotional category matters for income once baseline creator-, channel-, and video-level controls are included, nor do they resolve multicollinearity among emotion variables. The analysis therefore relies on more aggregated sentiment measures, while treating these emotion-specific distributions as complementary evidence on what the broader indicators capture.

3.4 Empirical strategy

This study examines the extent to which differences in creator language, audience comments, and related video-level features from top content help account for the gender gap in YouTubers' income. The empirical strategy is organized around two main steps. First, we estimate a sequence of nested regression models that progressively incorporate newly constructed variables derived from each creator's most popular content. Second, we implement a Gelbach decomposition to quantify the contribution of different groups of variables to the change in the estimated gender coefficient in a way that is invariant to the order of inclusion. In addition, we conduct supplementary block-inclusion and alternative-ordering analyses to assess the stand-alone explanatory power of each variable group and the sensitivity of sequential attenuation to specification order. These analyses complement the main evidence from the nested regressions and Gelbach decomposition.

3.4.1 Model specification

We begin with a baseline model that relates creator income to gender and the set of controls used in our previous income gap study. Specifically, we estimate:

$$\ln(\mathrm{Y}_i) = \alpha + \beta\,\mathrm{female}_i + \mathbf{X}_i'\boldsymbol{\gamma} + \varepsilon_i, \tag{3.1}$$

where $\ln(\mathrm{Y}_i)$ is the logarithm of estimated annual income for creator i, proxied by the midpoint of *Social Blade*'s reported annual income range. female_i is an indicator equal to one for female creators and zero for male creators. $\mathbf{X}_i$ is a vector of baseline controls, including creator-level demographic characteristics, channel-level characteristics and content category, and video-level measures of production, activity, and engagement. ε_i is an idiosyncratic error term.

To assess whether newly constructed content-based measures help explain the residual gender gap, we augment the baseline specification with variables derived from each creator's top content. The enriched specification is:

$$\ln(\mathrm{Y}_i) = \alpha + \beta\,\mathrm{female}_i + \mathbf{X}_i'\boldsymbol{\gamma} + \mathbf{Z}_i'\boldsymbol{\delta} + \mathbf{T}_i'\boldsymbol{\theta} + \mathbf{C}_i'\boldsymbol{\lambda} + \varepsilon_i, \tag{3.2}$$

where $\mathbf{Z}_i$ denotes general characteristics of top content, $\mathbf{T}_i$ denotes textual measures constructed from titles, descriptions, captions, and comments, and $\mathbf{C}_i$ denotes commercial and interaction signals, such as sponsor, self-promotion, and audience-engagement prompts.

More specifically, $\mathbf{Z}_i$ includes creator-level aggregates of genral characteristics such as average duration, likes, comments, and recency among top content. $\mathbf{T}_i$ includes textual measures grouped along two dimensions: (1) by *text source* (title, description, caption, and comments) and (2) by *feature type* (text length, emotion and sentiment, text structure and style, and gender-related and harmful-content indicators). $\mathbf{C}_i$ captures additional non-textual signals extracted from top content segments, including sponsor content, self-promotion, and explicit interaction prompts directed at the audience.

The coefficient of primary interest is β, which captures the conditional gender gap in advertising income after accounting for observable characteristics included in the baseline specification. Because estimated advertising income is mechanically determined by views and CPM, the coefficients in this chapter should be interpreted as capturing factors associated with those underlying margins rather than direct platform discrimination. The newly introduced textual and audience-response measures are likely to operate primarily through the visibility, engagement, and circulation of content, and therefore more plausibly through the views margin, whereas some of the broader channel- and category-level characteristics examined earlier may also reflect differences in CPM.

To examine how these additional factors relate to the residual gender gap, we estimate a sequence of nested models that progressively add richer video-level information. We first add general top content characteristics, followed by textual measures and then commercial and interaction signals. Within the textual block, we also consider alternative groupings by text source and linguistic attribute. This approach allows us to track how the estimated gender coefficient changes as richer information on creator content is introduced, and to assess which types of video-level features are most closely associated with advertising income and the residual gender income gap. In addition, we report supplementary exercises, including alternative sequential orderings and separate block inclusion analyses, to assess the robustness of these patterns and the sensitivity of the sequential results to specification order.

3.4.2 Gelbach decomposition

Although sequential inclusion is informative, the estimated contribution of a given block of variables can depend on the order in which regressors are added. To address this issue, we complement the inclusion specifications with a Gelbach decomposition (Gelbach, 2016). In our application, the decomposition compares a reduced specification including only the female indicator with the full model in equation (3.2), thereby providing an exact and order-invariant decomposition of the change in the estimated gender coefficient.

Let $\hat{\beta}^R$ denote the coefficient on Female_i from the reduced specification, and let $\hat{\beta}^F$ denote the corresponding coefficient from the full model. The Gelbach decomposition expresses the difference as:

$$\hat{\beta}^R - \hat{\beta}^F = \sum_k \Delta_k, \tag{3.3}$$

where each Δ_k captures the contribution of a particular group of regressors to the overall change in the estimated gender coefficient.

In principle, the decomposition can be applied to all groups of variables included in the full specification, including both the baseline controls and the newly constructed video-level measures. In the main text, however, we focus on the contribution of the newly added variables, since these are the central objects of interest in this chapter. Accordingly, in the next section we report decomposition results for broad blocks such as general video characteristics, textual measures, and commercial and interaction signals, and, where informative, further decompose the textual measures by text source and feature type.

3.5 Results

This section presents the main empirical results by tracing how the coefficient on *Female* evolves across sequentially augmented specifications and by reporting the corresponding order-invariant Gelbach decompositions. These results speak directly to the paper's central question of whether video-level textual and audience-side patterns help account for the gender gap in creators' estimated advertising income. The analysis therefore begins with the sequential and decomposition results, which most clearly show where the explanatory power is concentrated. A separate-block inclusion analysis is then presented as a supplementary check between the decomposition results and the variable-level analysis. Finally, the full-specification OLS estimates for the male, female, and pooled samples are used more selectively to interpret the main correlates of the gap and the gender heterogeneity in those associations.

3.5.1 A persistent baseline gender income gap

Table 3.4 establishes the baseline empirical benchmark for the analysis. The specification begins with demographic and identity characteristics, then adds channel basics and content category controls, and finally incorporates annual production, activity, and engagement measures. These are exactly the controls used in our previous study, in which the estimated gender gap was reduced but not eliminated, thereby motivating the present extension to richer top content measures. This ordering is deliberate. Together, these variables capture the most immediate observable correlates of creator success in platform labor markets—who the creator is, what type of content the creator produces, how large and visible the channel is, and how actively the creator sustains audience attention and interaction (Christin and Lu, 2023; Glatt, 2023). Related work on YouTube and influencer cultures further suggests that creators are sorted across gendered content niches and modes of self-presentation, implying that channel category and creator profile are not incidental controls but central dimensions of platform inequality (Pattier, 2021; Jorge et al., 2023; Gioia and Morabito, 2025).

Table 3.4. Baseline gender income gap in estimated annual YouTube advertising income

	(1)	(2)	(3)
Female	−0.545**	−0.445**	−0.270**
	[0.227]	[0.184]	[0.137]
Basic control blocks (G0)			
Demographics & identity	Yes	Yes	Yes
Channel basics & content category		Yes	Yes
Annual production, activity & engagement			Yes
Observations	485	485	485
R^2	0.096	0.415	0.689

Note: The dependent variable is estimated annual income. The key variable is an indicator for female creators. Columns (1)–(3) sequentially add the baseline control blocks shown in the table. Robust standard errors are reported in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

The results show that the estimated income gap persistent and remains statistically significant in every column. Although the coefficient on *Female* declines as additional

baseline controls are introduced, it remains around -0.270 in the fullest baseline specification, indicating that those observables combined xplain only part of the gender income gap.

The persistence of this residual gap motivates the next step of further analysis. Research on creator labor emphasizes that platform success depends not only on scale or activity, but also on self-branding, relational labor, and the ability to cultivate and maintain an audience community (Glatt, 2023). In YouTube and influencer settings, these processes are inherently communicative: creators compete through titles, descriptions, and spoken presentation, while audiences respond through visible engagement and comment-based evaluation (Jorge et al., 2023; Glatt, 2023). This suggests that part of the remaining disparity may lie in the textual and interactional environment through which creator content is presented and received. The next step is therefore to examine whether the residual gender gap can be further explained by richer video-level information not captured by the baseline controls, especially textual features derived from creators' titles, descriptions, and captions, as well as from audience comments.

3.5.2 Sequential model specification

The sequential specifications in this subsection are used to trace how the coefficient on *Female* changes as additional blocks of video-level variables are introduced. We organize the same full set of covariates in two complementary ways. The *text-source path* groups variables by where the text comes from, allowing us to distinguish between creator-side and audience-side sources. The *linguistic-attribute path* groups the same variables by what kind of linguistic information they capture, allowing us to assess whether the remaining gender gap is more closely associated with verbosity, sentiment, style, or harmful-content indicators. Taken together, the two paths provide a compact view of how the residual gap responds to alternative but substantively meaningful ways of organizing the same information. At the same time, these sequential paths should be interpreted cautiously. As emphasized in the literature on Gelbach decomposition, sequential attenuation is inherently

order-dependent, so the exact path of coefficient changes cannot by itself establish which block matters most (Gelbach, 2016; Cools et al., 2020). For that reason, the sequential results are best understood as descriptive illustrations, while the main assessment of relative explanatory importance is provided by the Gelbach decomposition and complemented by the separate block inclusion results.

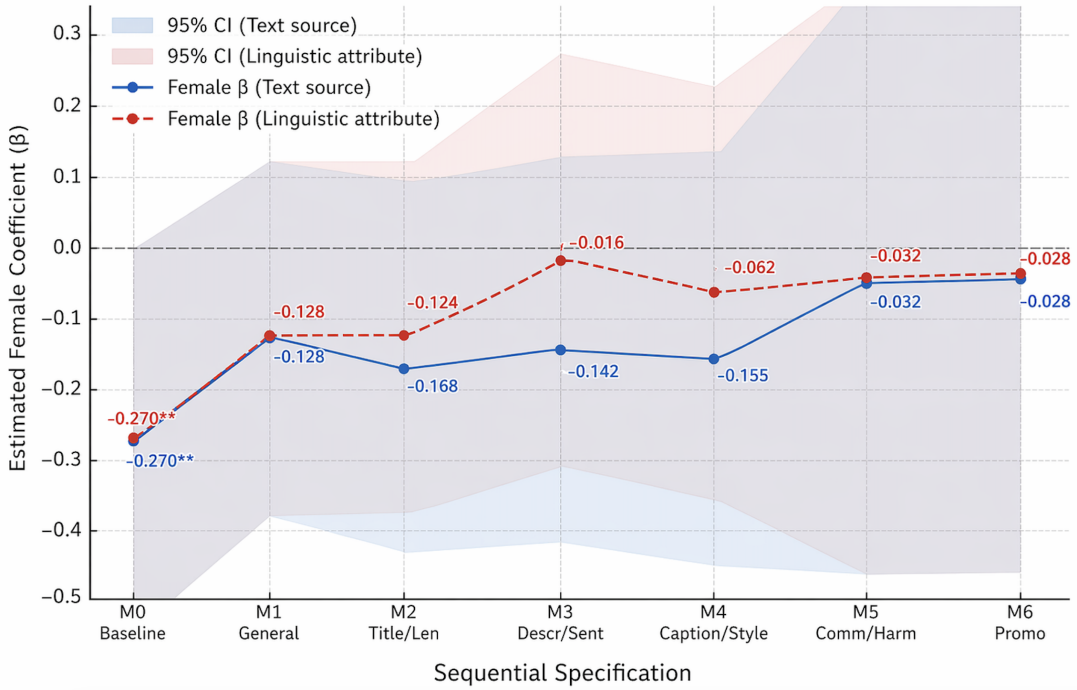


Figure 3.8. Female coefficient paths across sequential specifications

Note: The figure plots the estimated coefficient on *Female* across seven sequential specifications under two grouping schemes: *text source* and *linguistic attribute*. M0 includes baseline controls only; M1 adds general top-video characteristics; M2–M5 add, respectively, title, description, caption/transcript, and comment features in the text-source path, and length and verbosity, sentiment polarity, style and structure, and Bias & harm indicators in the linguistic-attribute path; M6 additionally includes sponsor, self-promotion, and interaction signals. Point estimates are labeled next to each marker, with significance indicated by asterisks; shaded areas denote 95% confidence intervals. Exact estimates and robust standard errors are reported in Appendix B, Table B4 and Table B5.

Figure 3.8 summarizes both coefficient paths and serves as the main reference for this subsection. Three patterns stand out. First, in both paths the largest early compression occurs when general top content characteristics are added in Model 1, indicating that an important part of the baseline female penalty is associated with observable differences

in the performance profile of creators' most visible videos. In the baseline ordering shown here, the coefficient on *Female* is no longer statistically significant from that point onward. This is an important result, but it should not be read too strongly as evidence that general top-content characteristics uniquely explain the remaining gap, because alternative orderings reported in the robustness checks show that audience-side textual variables can also reduce the coefficient to zero before performance metrics are introduced. Second, the two paths diverge in the middle of the sequence. Under the text-source grouping, the largest later shift occurs when comment variables enter in Model 5, whereas under the linguistic-attribute grouping the decisive movement occurs earlier, when sentiment variables are introduced in Model 3. Third, the two paths converge mechanically in the final specification because both ultimately include the same full set of controls. Since Figure 3.8 already reports the estimated coefficients and significance levels, the discussion below focuses only on the main turning points in each path. The exact regression outputs are reported in Appendix B, Tables B4 and B5.

Sequential inclusion by text source

The text-source path shows that the largest single reduction in the coefficient on *Female* occurs between Model 0 and Model 1, where the estimate falls from −0.270 to −0.128. Under the baseline ordering, this indicates that a substantial part of the residual gap is associated with top content general characteristics rather than with gender itself. However, this early compression should be interpreted as the result of one particular ordering rather than as definitive proof that performance metrics are the sole or dominant channel. After that point, creator-generated text contributes relatively little to further attenuation. Moving from Model 1 to Models 2–4, the coefficient remains in a fairly narrow range, indicating that title, description, and caption features are relevant for income overall but do not substantially reduce the remaining gender gap in this ordering.

The main source-specific shift occurs only when audience comments are added in Model 5. At that point, the coefficient declines sharply from −0.155 to −0.032, while

model fit also improves noticeably. This suggests that, conditional on creator and channel characteristics, the most informative textual signal lies less in how creators frame their content than in how audiences respond to it. Adding sponsor, self-promotion, and interaction signals in Model 6 changes the coefficient only minimally, implying that these variables provide little additional explanatory power once top-content characteristics and audience comments are already taken into account.

Sequential inclusion by linguistic attribute

The linguistic-attribute path conveys the same broad message at the beginning and end of the sequence, the key difference lies in the middle of the path. Adding length and verbosity in Model 2 produces almost no additional movement, suggesting that text volume itself is not central to the residual gender difference. The decisive shift occurs in Model 3, when sentiment variables are introduced and the coefficient falls from -0.124 to -0.016. Under this ordering, sentiment-related features are therefore the most informative linguistic margin for explaining the remaining gap.

Subsequent additions are less monotonic. Style and structure variables in Model 4 move the coefficient somewhat away from zero, while bias & harm indicators in Model 5 partially attenuate it again. Even so, the overall pattern remains clear: among linguistic attributes, sentiment contributes the strongest reduction in the residual *Female* coefficient, whereas the other linguistic dimensions are more modest or offsetting in their contribution.

Taken together, the two sequential paths point to a common conclusion. Under the current cumulative ordering of variable groups, the residual female income gap narrows most sharply when general top-content characteristics are introduced. Beyond this initial attenuation, the text-source path shows that audience comments are the most important textual source, while the linguistic-attribute path shows that sentiment is the most important textual dimension. These patterns should be interpreted cautiously, however, because sequential inclusion is inherently order-dependent. The robustness checks using alternative orderings in Section 3.5.6 show that the exact attenuation path changes with the order in

which variable groups are added. The broader conclusion is therefore not that any single block, such as general top-content characteristics, uniquely "kills" the effect, but rather that the remaining disparity is explained by the joint contribution of richer top-content, textual, and audience-response measures. In this sense, the two sequential specifications are complementary rather than definitive: one shows where the most informative textual signal comes from, and the other shows what kind of textual information matters most, while the order-invariant decomposition in the next section provides the main basis for assessing relative explanatory importance.

3.5.3 Gelbach decomposition results

Table 3.5 reports the Gelbach decomposition results, which reallocates the change in the coefficient on *Female* in an order-invariant way. Because the non-text blocks are identical under both grouping schemes (by text source / by linguistic feature), the common components are reported only once in Panel A. The textual variables are then reorganized in two complementary ways, by text source in Panel B and by linguistic feature in Panel C.

The decomposition both confirms and qualifies the sequential evidence. In the sequential specifications, the coefficient on *Female* falls sharply when general top-content characteristics are introduced early in the model. The Gelbach decomposition provides a more cautious interpretation: the contribution of video-level general features is negative, as expected, but smaller and not statistically significant. This suggests that these variables help attenuate the gender income penalty, but their apparent importance in the sequential path partly reflects the order in which variables are added. Sponsor, self-promotion, and interaction signals contribute essentially nothing to the change in the coefficient once the full set of variables is considered.

Panel B shows that, when textual features are grouped by source, comment text has the largest negative contribution among the text blocks. This is consistent with the sequential results, where the main source-specific attenuation occurs when audience comments are added. However, the estimate is imprecise, and the confidence interval is wide. Title,

description, and caption text do not provide a consistent attenuation pattern: title and caption text have positive coefficients, while description text is close to zero. The source-based decomposition therefore supports the idea that audience-side text is more informative than creator-side text, but it does not by itself provide strong evidence that any single text source dominates.

Table 3.5. Gelbach Decomposition of the Change in the Female Coefficient

Group	Coefficient	Std. Err.	p-value	95% CI
Panel A. Common Blocks across Both Decompositions				
G1: Video-level general features	−0.057	(0.069)	0.409	[−0.192, 0.078]
G6: Sponsor / self-promo / interaction	0.000	(0.021)	0.997	[−0.041, 0.041]
Panel B. Textual Features Grouped by Text Source				
G2: Title text	0.022	(0.059)	0.706	[−0.093, 0.138]
G3: Description text	−0.004	(0.051)	0.932	[−0.105, 0.096]
G4: Caption text	0.044	(0.127)	0.730	[−0.205, 0.293]
G5: Comment text	−0.129	(0.236)	0.585	[−0.592, 0.334]
Panel C. Textual Features Grouped by Linguistic Feature				
G2: Length & verbosity	−0.015	(0.035)	0.667	[−0.084, 0.054]
G3: Sentiment polarity	−0.238***	(0.080)	0.003	[−0.396, −0.080]
G4: Style & structure	0.108*	(0.062)	0.081	[−0.013, 0.229]
G5: Bias & harm indicators	0.078	(0.206)	0.706	[−0.326, 0.481]
Total change ($\sum$G1–G6)	−0.324	(0.271)	0.231	[−0.855, 0.207]

Note: The dependent variable is estimated annual income. The Gelbach decomposition is performed relative to a restricted specification including only the indicator for *Female*. Common blocks G1 and G6 are identical across the two decompositions and are reported once. Panel B groups textual variables by text source, while Panel C groups the same variables by linguistic feature type. Each coefficient reports the contribution of the corresponding block to the change in the coefficient on *Female*. Negative coefficients indicate that the block attenuates the female income penalty. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel C provides the clearest refinement. When the same textual variables are grouped by linguistic feature, sentiment polarity emerges as the largest and only statistically significant contributor to the attenuation of the coefficient on *Female*. By contrast, length and verbosity contribute little, while style and structure and bias and harm indicators move the coefficient in the opposite direction. This pattern suggests that the relevant textual margin is not simply text volume, textual structure, or harmful-language markers, but the evaluative and sentiment-related tone of the language associated with creators.

Taken together, the sequential regressions and the Gelbach decomposition point to

a common conclusion, but with different emphases. The sequential paths show where attenuation occurs under a given ordering, while the Gelbach decomposition provides an order-invariant assessment of relative contributions. The results therefore should not be read as showing that any single block, such as general top-content characteristics, uniquely explains the gap. Rather, the remaining female income penalty is accounted for by the joint contribution of richer video-level and textual measures, with the strongest textual evidence coming from audience-side sentiment-related variables.

3.5.4 Separate block inclusion results

Table 3.6 complements the cumulative sequential specifications and the Gelbach decomposition by asking a simpler question: how much does the coefficient on *Female* change when each block is added on its own to the baseline-controls specification? Because these models are not cumulatively nested beyond the common baseline, they should not be interpreted as marginal contributions in the same sense as the sequential paths. Instead, they show the stand-alone explanatory power of each block relative to the residual female income penalty that remains after the baseline controls are included. Appendix Table B8 reports the same exercise relative to the gender-only baseline and helps place the baseline-controls results in a broader perspective.

The clearest common result remains the role of top content genral metrics. Added on their own to the baseline-controls specification, they reduce the coefficient on *Female* from -0.270 to -0.128, corresponding to a 52.6% attenuation. The difference relative to the baseline-controls coefficient is only marginally significant ($p = 0.080$), but the economic reduction is substantial. By contrast, sponsor, self-promotion, and interaction variables leave the coefficient essentially unchanged and show no meaningful explanatory role ($p = 0.944$).

Among the text-source blocks in Panel B, comments again produce the largest attenuation, reducing the coefficient to -0.242, but the change relative to the baseline-controls specification is small and far from statistically significant ($p = 0.884$). Title and caption

blocks move the coefficient further away from zero, while description text generates only a very small attenuation. Taken together, these results suggest that, when considered in isolation, creator-side and audience-side text blocks have limited stand-alone explanatory power once the baseline controls are already included. This is weaker than in the cumulative sequential results, where comments become more informative after other creator- and content-side differences are taken into account.

Table 3.6. Separate block inclusion results relative to the baseline-controls specification

Group	Coefficient (Female)	Robust SE	% change vs. baseline	p-value vs. baseline	R^2
Baseline controls only	−0.270**	(0.137)			0.689
Panel A. Common blocks					
Base + Top content general metrics	−0.128	(0.128)	52.59	0.080	0.758
Base + Sponsor & self-promo & interaction	−0.271**	(0.138)	− 0.37	0.944	0.691
Panel B. Grouped by text source					
Base + Title text features	−0.319**	(0.142)	−18.15	0.308	0.700
Base + Description text features	−0.257*	(0.143)	4.82	0.833	0.705
Base + Caption text features	−0.364**	(0.154)	−34.82	0.295	0.723
Base + Comment text features	−0.242	(0.228)	10.37	0.884	0.731
Panel C. Grouped by linguistic attribute					
Base + Length & verbosity	−0.265*	(0.137)	1.85	0.836	0.693
Base + Sentiment polarity	−0.137	(0.161)	49.26	0.068	0.719
Base + Style & structure	−0.347***	(0.134)	−28.52	0.256	0.728
Base + Bias & harm indicators	−0.028	(0.232)	89.63	0.210	0.720
Observations			485		

Note: The dependent variable is estimated annual income. The main independent variable is an indicator for female creators. All models include baseline controls. Each model adds one block separately to the baseline-controls specification. Panel A reports common blocks that are identical across both grouping schemes. Panel B groups textual variables by text source, while Panel C groups the same variables by linguistic attribute. The column "% change vs. baseline" reports the percentage attenuation in the absolute value of the coefficient on *Female* relative to the baseline-controls specification, computed as $100 \times (|\beta_{\text{base}}| - |\beta_j|)/|\beta_{\text{base}}|$. Positive values indicate attenuation, while negative values indicate amplification. The column "p-value vs. baseline" reports post-estimation tests of equality between the coefficient on *Female* in each specification and that in the baseline-controls specification. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel C sharpens the interpretation by reorganizing the same information by linguistic attribute. Length and verbosity are essentially uninformative. Sentiment polarity, by contrast, reduces the coefficient from −0.270 to −0.137, implying a 49.3% attenuation, and this difference is marginally significant relative to the baseline-controls specification

($p = 0.068$). Style and structure move in the opposite direction, amplifying the female coefficient, while bias-and-harm indicators reduce it sharply to -0.028. However, this large numerical attenuation is not statistically distinguishable from the baseline-controls coefficient at conventional levels ($p = 0.210$). It is therefore better interpreted as strong stand-alone explanatory power than as evidence of a clean and independent channel.

Table B8 in Appendix B helps clarify this pattern. Relative to the gender-only baseline, the results are more polarized. Demographic controls, title text, caption text, style and structure, and bias-and-harm indicators all produce statistically significant changes in the coefficient on *Female*, while comments and sentiment reduce the raw gap more modestly. Most notably, the bias-and-harm block yields a positive coefficient when added on its own to the gender-only specification, indicating over-correction rather than simple attenuation. This suggests that these variables capture a broader set of gender-sorted differences in the raw data. Once the richer baseline controls are introduced, that stronger pattern is substantially muted.

Overall, the separate block inclusion exercise refines the interpretation of the earlier results in two ways. First, it confirms that top-video performance metrics remain the most important common block. Second, it shows that stand-alone explanatory power is weaker and less stable than the cumulative sequential results might suggest. In particular, sentiment emerges as the most informative linguistic block once the baseline controls are included, whereas comments and bias-and-harm indicators appear more sensitive to whether other correlated blocks have already been partialled out. The contrast between the two tables therefore reinforces a simple point: what a block explains on its own need not coincide with what it explains incrementally once other related dimensions are already in the model.

3.5.5 Variable-level contributions

Figure 3.9 reports variable-level contributions to the change in the coefficient on *Female* using two complementary approaches: sequential one-by-one inclusion and Gelbach de-

composition. This helps identify which specific variables drive the block-level patterns documented above. Tables 3.7–3.9 then report the corresponding full-specification ordinary least square (OLS) coefficients, split by mechanism. This sequence is useful because the figure identifies which variables matter most for the *Female* coefficient, while the regression tables show how those same variables are associated with income in the male, female, and pooled samples.

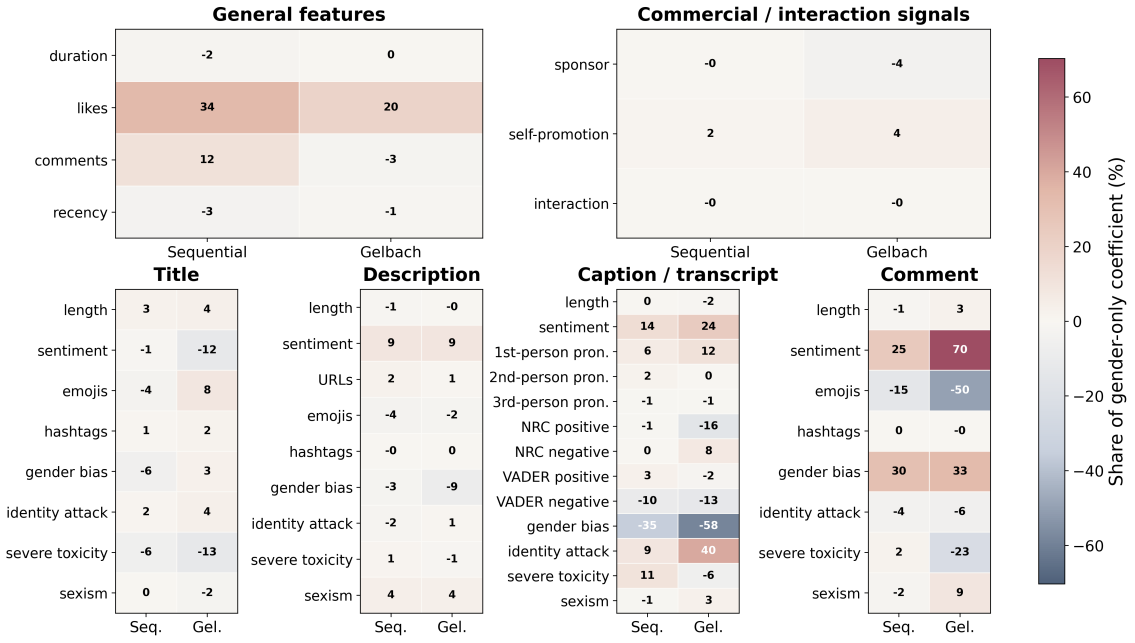


Figure 3.9. Variable-level contributions to the gender income gap

Note: The figure reports variable-level contributions to the coefficient on *Female* under two approaches: sequential one-by-one inclusion and Gelbach decomposition. All values are scaled by the absolute value of the coefficient on *Female* in the gender-only regression. Positive values indicate that the variable attenuates the baseline female income penalty, while negative values indicate that it enlarges it. Variables are grouped into general features, commercial and interaction signals, and textual features derived from titles, descriptions, captions/transcripts, and comments.

Specifically, Figure 3.9 translates the variable-level results into signed percentage contributions to the total change in the *Female* coefficient[10]. The two columns compare the simple sequential metric with the Gelbach-based metric, so the figure should be read as a compact summary of which individual variables account for more, or less, of the attenuation in the residual female income gap. The figure points to a highly concentrated

[10]Variable-level contributions are normalized by the coefficient from the gender-only regression. In the sequential approach, each contribution equals the incremental change in the *Female* coefficient when the variable is added, expressed as a percentage of the absolute gender-only coefficient. In the Gelbach approach, each contribution equals the corresponding decomposition component, scaled in the same way. Positive values indicate attenuation of the female income penalty; negative values indicate amplification.

structure of explanatory power. Two variables stand out above all others: average likes among the best performing content and comment sentiment. The first is the dominant non-textual contributor, while the second is the dominant textual contributor, especially under the Gelbach decomposition. By contrast, most creator-side text variables have small or offsetting effects, and the commercial&interaction signals contribute very little. This pattern suggests a clear economic distinction: the residual female income gap is explained less by how creators package their content in a narrow textual sense, and more by two audience-facing margins: whether top content receives strong endorsement, and what kind of evaluative environment surrounds it in the comment space.

At the same time, the contribution figure and the OLS tables (Tables 3.7–3.9) summarize different quantities and should therefore be interpreted differently. The heatmap tracks how controlling for a variable changes the estimated coefficient on *Female*, while the OLS tables instead show the conditional association between that variable and income within the full specification. A variable can therefore be strongly related to income yet explain little of the gender gap, or vice versa. The most informative mechanism discussion must combine both pieces of evidence with the gender differences in the raw distributions documented earlier.

Non-textual variables: audience endorsement dominates. Table 3.7 shows that among the non-textual variables, the most economically important and most consistent predictor is average likes of creator's most popular videos. Its coefficient is positive and statistically significant in the male, female, and pooled regressions, and Figure 3.9 shows that it is also the single largest non-textual contributor to the reduction in the *Female* coefficient under both sequential inclusion and Gelbach decomposition. This alignment is substantively important. Female creators receive fewer likes on their most popular videos on average than male creators (Table 3.3), yet likes are strongly rewarded in the income regressions. The female-subsample coefficient is, if anything, larger than the male coefficient, suggesting that the issue is not weaker returns to audience approval for women, but lower exposure to that approval on the most visible content.

Other general features are less robust. Average comments and average days since upload enter negatively in the pooled specification, but neither displays the same consistent relevance for the gender gap as likes. Average duration is small and statistically weak throughout. Commercial and interaction variables also play a limited role. Sponsor segment duration, self-promotion duration, and interaction-segment duration are close to zero and statistically insignificant across specifications, and their contributions in Figure 3.9 are negligible or slightly negative. Given that the dependent variable is an advertising income proxy rather than total creator earnings, these variables should be interpreted as signals of monetization oriented content design rather than direct measures of sponsorship revenue. Their weak explanatory role therefore suggests that explicit commercial signaling does not account for much of the advertising-income gap between male and female creators.

Table 3.7. Full-specification OLS Estimates: Non-textual Features

Variables	**Male (N = 367)**		**Female (N = 118)**		**Combined (N = 485)**	
General Features						
Average duration	0.000	(0.000)	0.001	(0.002)	0.000	(0.000)
Average likes	0.000**	(0.000)	0.001**	(0.000)	0.001**	(0.000)
Average comments	−0.007*	(0.004)	−0.035*	(0.018)	−0.009***	(0.004)
Average days since upload	−0.000***	(0.000)	−0.000	(0.000)	−0.001***	(0.000)
Commercial & Interaction Signals						
Sponsor segment duration	−0.005	(0.004)	−0.042	(0.032)	−0.004	(0.004)
Self-promotion duration	0.017	(0.014)	0.043	(0.143)	0.010	(0.014)
Interaction segment duration	−0.006	(0.018)	−0.091	(0.153)	−0.001	(0.016)
Controlled groups						
Baseline controls	Yes		Yes		Yes	
Textual features: title	Yes		Yes		Yes	
Textual features: description	Yes		Yes		Yes	
Textual features: caption	Yes		Yes		Yes	
Textual features: comments	Yes		Yes		Yes	
R^2	0.822		0.958		0.813	

Note: The dependent variable is estimated annual income. The table reports selected rows from the full-specification regressions for the male subsample, female subsample, and pooled sample. All columns include the full set of baseline controls, which are omitted for brevity and reported in Appendix Table B3. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Creator-side text: income-relevant but diffuse. Table 3.8 shows that creator-side textual features are related to income, but the pattern is heterogeneous across text sources, linguistic attributes, and subsamples. In titles, the clearest patterns are that more positive wording is associated with higher income in the male and pooled regressions, whereas severe toxicity is associated with lower income in those same specifications. Most other title features are small and statistically imprecise. In descriptions, the most stable predictor is the number of URLs, which enters negatively and significantly in both the male and pooled regressions. By contrast, other description measures are either weak or sample-specific: for example, description hashtags and description sentiment are negative and significant only in the female subsample, while description emojis are only weakly positive for male creators. These results indicate that description fields contain some income-relevant information, but not a common pattern that maps cleanly onto the pooled gender gap.

Captions contain stronger signals, but these too are internally mixed. In the pooled regression, caption sentiment is negatively associated with income, VADER positive sentiment is positively associated with income, the caption gender-bias ratio is negatively associated with income, and caption identity-attack language is positively associated with income. In the female subsample, second-person pronoun use is negatively associated with income, whereas in the male subsample, VADER positive sentiment and the caption gender-bias ratio remain the clearest significant predictors. Thus, even within captions, which emerged as the most informative creator-generated text source in the earlier block-level results, the relevant mechanisms do not all point in the same direction. Some measures suggest that more positive affect is rewarded, while others indicate that more confrontational or identity-salient language can also be associated with higher income.

Figure 3.9 reinforces this interpretation. A few caption-based variables make significant contributions to the attenuation of the *Female* coefficient, but most title and description variables contribute only modestly, and several of them offset one another in sign. This helps explain why the creator-side text blocks improve model fit without generating a large

or monotonic reduction in the coefficient on *Female*. Creator-controlled language clearly matters for income, but it does not line up with a single dominant creator-side mechanism behind the remaining gender gap. Substantively, this suggests that the residual disparity is not primarily explained by a simple story in which women write systematically less effective titles, descriptions, or spoken text. Rather, creator-side language bundles multiple signals: tone, style, confrontation, and presentation, whose effects partially offset one another. Relative to the strongest non-textual and audience-side variables, creator-controlled text therefore appears to be a secondary channel.

Audience-side text: comment sentiment stands out. Table 3.9 provides the clearest variable-level evidence on the audience side. Among all comment variables, comment sentiment stands out most clearly. In the pooled full specification, it is large, negative, and highly statistically significant; it is likewise strongly negative in the male subsample and remains negative and statistically significant in the female subsample as well. Figure 3.9 shows that comment sentiment is also the largest positive comment-side contributor to the reduction in the coefficient on *Female*, especially under the Gelbach decomposition. Taken together, these results indicate that the evaluative tone of audience comments is the most important audience-side variable associated with attenuation of the residual gender gap.

Other audience-side variables add nuance, but they do not provide the same degree of clarity. Comment emojis are positively associated with income in the pooled and male regressions, yet they make a negative contribution in Figure 3.9, indicating that they do not reduce the coefficient on *Female* once their covariance with other audience-side features is taken into account. Comment gender-bias ratios contribute positively in the figure, especially under the Gelbach decomposition, but their direct OLS coefficients are weak, suggesting that their role comes more from gender-sorted exposure than from a stable within-sample association. Comment severe toxicity and identity attack both enter negatively in the regressions, but neither is estimated precisely, and their contributions are

Table 3.8. Full-specification OLS Estimates: Creator-side Textual Features

Variables	Male (N = 367)		Female (N = 118)		Combined (N = 485)	
Textual Features: Title						
Title length	0.007	(0.033)	0.132	(0.084)	−0.015	(0.023)
Title emojis	0.178	(0.176)	−0.516	(0.511)	−0.180	(0.160)
Title hashtags	−0.195	(0.193)	−0.086	(0.281)	−0.034	(0.111)
Title sentiment	1.231**	(0.584)	−0.280	(1.740)	0.960*	(0.496)
Title gender bias ratio	0.009	(0.007)	−0.002	(0.009)	0.002	(0.005)
Title identity attack	0.148	(0.212)	2.170	(1.611)	0.151	(0.222)
Title severe toxicity	−2.079*	(1.205)	−10.819	(8.363)	−3.008**	(1.390)
Title sexism	−0.015	(0.087)	0.016	(0.267)	0.029	(0.077)
Textual Features: Description						
Description length	0.001	(0.001)	0.003	(0.002)	0.001	(0.001)
Description URLs	−0.028**	(0.014)	0.014	(0.030)	−0.029***	(0.011)
Description emojis	0.030*	(0.015)	−0.025	(0.051)	0.015	(0.016)
Description hashtags	−0.010	(0.012)	−0.131**	(0.062)	−0.005	(0.012)
Description sentiment	−0.135	(0.169)	−2.032**	(0.967)	−0.173	(0.157)
Description gender bias ratio	−0.002	(0.003)	−0.006	(0.007)	−0.003	(0.003)
Description identity attack	−0.289	(0.618)	−6.526	(4.047)	0.074	(0.584)
Description severe toxicity	−0.153	(1.029)	4.583	(8.872)	−0.379	(1.077)
Description sexism	0.156	(0.104)	−0.003	(0.048)	−0.009	(0.023)
Textual Features: Caption						
Caption length	0.000	(0.000)	−0.000	(0.000)	0.000	(0.000)
Caption 1st-person pronouns	−0.034	(0.041)	−0.122	(0.128)	−0.051	(0.035)
Caption 2nd-person pronouns	0.010	(0.053)	−0.316**	(0.141)	0.001	(0.043)
Caption 3rd-person pronouns	−0.059	(0.068)	−0.277	(0.179)	−0.036	(0.056)
Caption sentiment	−0.842	(0.594)	1.007	(1.697)	−1.143**	(0.534)
Caption NRC positive emotion	0.012	(0.025)	0.052	(0.059)	0.034	(0.022)
Caption NRC negative emotion	−0.002	(0.018)	0.004	(0.062)	0.021	(0.014)
Caption VADER positive	0.015***	(0.006)	−0.011	(0.014)	0.009*	(0.005)
Caption VADER negative	0.006	(0.012)	−0.009	(0.046)	−0.010	(0.011)
Caption gender bias ratio	−0.011**	(0.005)	−0.007	(0.010)	−0.010**	(0.004)
Caption identity attack	0.163	(0.112)	1.643	(1.006)	0.349**	(0.135)
Caption severe toxicity	0.016	(0.086)	−0.864	(0.655)	−0.136	(0.121)
Caption sexism	0.002	(0.028)	0.033	(0.094)	−0.036	(0.030)
Controlled groups						
Baseline controls	Yes		Yes		Yes	
General features	Yes		Yes		Yes	
Commercial & interaction signals	Yes		Yes		Yes	
Textual features: comments	Yes		Yes		Yes	
R^2	0.822		0.958		0.813	

Note: The table reports creator-side textual variables from the same full-specification regressions used throughout the analysis. All columns include the full set of baseline controls, which are omitted for brevity and reported in Appendix Table B3. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

not consistent across the two approaches. Relative to these measures, comment sentiment is the only audience-side variable that is simultaneously large, economically meaningful, and central in the decomposition of the gap.

Summary of variable-level evidence. Taken together, the variable-level evidence points to two central correlates of the residual gender income gap. The first is unequal audience endorsement of the top content, captured most clearly by average likes. The second is the evaluative tone of audience response, captured most clearly by comment sentiment. Both are audience-facing rather than purely creator-controlled factors. This suggests that in the platform-based creator labour market, gender inequality is associated not only with what creators produce, but also with how audiences reward and discuss the content that becomes most visible.

Table 3.9. Full-specification OLS Estimates: Audience-side Comment Features

Variables	**Male (N = 367)**		**Female (N = 118)**		**Combined (N = 485)**	
Textual Features: Comment						
Comment length	0.016	(0.022)	0.070	(0.080)	0.011	(0.019)
Comment emojis	0.426***	(0.122)	0.442	(0.280)	0.442***	(0.091)
Comment hashtags	0.032	(0.100)	0.109	(0.239)	0.017	(0.106)
Comment sentiment	−4.895***	(1.353)	−8.619**	(3.745)	−3.911***	(1.208)
Comment gender bias ratio	0.004	(0.007)	0.017	(0.011)	0.002	(0.004)
Comment identity attack	−0.153	(0.214)	−4.156	(2.530)	−0.085	(0.210)
Comment severe toxicity	−0.515	(0.431)	−0.840	(2.900)	−0.484	(0.407)
Comment sexism	0.038	(0.359)	−0.073	(0.932)	0.162	(0.318)
Controlled groups						
Baseline controls	Yes		Yes		Yes	
General features	Yes		Yes		Yes	
Commercial & interaction signals	Yes		Yes		Yes	
Textual features: title	Yes		Yes		Yes	
Textual features: description	Yes		Yes		Yes	
Textual features: caption	Yes		Yes		Yes	
R^2	0.822		0.958		0.813	

Note: The table reports audience-side comment variables from the same full-specification regressions used throughout the analysis. All columns include the full set of baseline controls, which are omitted for brevity and reported in Appendix B, Table B1. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

3.5.6 Robustness check

We conduct seven robustness checks to assess whether the main results depend on measurement choices, sample construction, or the ordering of variable groups. Across these exercises, the central message remains unchanged: the residual female income penalty can be fully explained by richer video-level measures from creators' top content. Audience-side variables remain more informative than creator-side textual features, and sentiment-related measures continue to play an important explanatory role. Detailed results are reported in Tables B6–B14 in Appendix B.

Alternative outcome definitions. First, we re-estimate the sequential specifications using alternative income definitions, including the midpoint, lower bound, upper bound, and winsorized versions of estimated annual income. As shown in Table B9 in Appendix B, the main attenuation pattern is highly stable across all outcome definitions. Under the baseline ordering, the female coefficient declines substantially after top content general features are added, although this step should be interpreted as part of an order-dependent sequence rather than as evidence that this block alone explains the gap. Comment-based variables and sentiment measures continue to provide the strongest additional textual explanatory power.

Heterogeneity by channel type and channel size. Second, we examine heterogeneity by channel type and channel size. Table B10 shows that the main pattern is most pronounced in male-dominated categories and among larger channels, where the baseline *Female* coefficient is larger and declines strongly once audience-side variables are introduced. By contrast, the coefficient is already small among smaller channels, and the female-dominated *People & Blogs* subsample is too small to support stable inference. These results suggest that the main mechanism is concentrated in the more commercially salient parts of the sample.

Alternative top-content constructions: long-form videos only and shorts only. Third, we test whether the results depend on how the best performing content is constructed by comparing the baseline mixed-format measure with long-form-videos-only and shorts-only specifications. Table B11 shows that the main conclusions do not depend on the mixed-format construction. Under the baseline ordering, the *Female* coefficient declines markedly once top-content variables are introduced, although this attenuation should again be interpreted as part of an order-dependent sequence rather than as evidence that this block alone explains the gap. At the same time, the exact path of attenuation differs somewhat by format, with sentiment appearing especially important for Shorts and comment-related or adversarial language playing a stronger role for long-form videos.

Restricting to channels with sufficiently informative comment data. Fourth, we restrict the sample to channels with more informative comment data, using stricter thresholds for comment depth, the number of the most popular videos with comments, and comment coverage. As reported in Table B12, the main pattern remains visible across these restricted samples. Top content general characteristics still account for a large share of the gap, and sentiment-related variables continue to be the most robust audience-side margin. This suggests that the main findings are not driven by sparse or low-quality comment data.

Alternative aggregation of general video-level features. Fifth, we replace mean-based creator-level aggregation of top content general features with median-based aggregation. Table B13 shows that this weakens the explanatory role of the top content somewhat, but it does not overturn the main interpretation. The *Female* coefficient still declines after top content variables are added, and sentiment remains the most informative textual margin. This suggests that the baseline results partly reflect the influence of especially successful videos, but not exclusively so.

Alternative aggregation and measurement of textual tone. Sixth, we test alternative measures of textual tone by replacing mean-based polarity measures with median-based polarity measures and, separately, with a higher-dimensional eight-emotion representation.

Table B14 shows that the results are very similar when polarity is aggregated using medians, confirming that the baseline findings are not driven by the choice between means and medians. The eight-emotion specification produces weaker attenuation, which is consistent with the greater dimensionality and likely multicollinearity of that approach. Overall, the core conclusion remains the same: broad tone measures, especially sentiment polarity, are more informative than simple text volume or packaging variables.

Alternative sequential ordering. Seventh, we test whether the sequential results depend on the order in which variable blocks are introduced. In the baseline ordering, top content general characteristics enter before the textual blocks. Figure 3.10 instead introduces creator- and audience-side text before top content general metrics, with the corresponding estimates reported in Tables B6–B7 in Appendix B. The final result is unchanged: once the full set of top content variables is included, the residual *Female* coefficient is close to zero. The substantive pattern is also similar: creator-side text remains relatively weak, while audience-side text substantially attenuates the coefficient even before general performance metrics are added. At the same time, although adding top content general characteristics at the final step still makes the *Female* coefficient statistically insignificant, their explanatory power is much weaker than in the baseline ordering. This confirms that sequential attenuation is order-dependent, and that no single block should be interpreted as uniquely driving the disappearance of the residual gender gap.

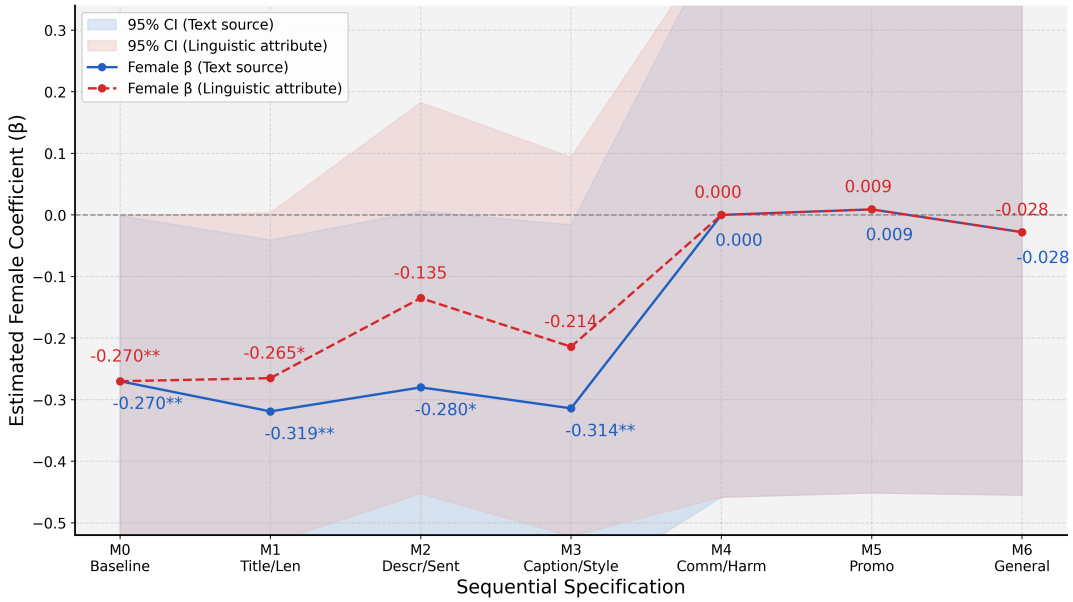


Figure 3.10. Female coefficient paths under alternative sequential orderings

Note: The figure plots the estimated coefficient on *Female* across seven sequential specifications under two grouping schemes: *text source* and *linguistic attribute*, using an alternative ordering in which textual blocks enter before top-video general metrics. M0 includes baseline controls only; M1–M4 add, respectively, title, description, caption/transcript, and comment features in the text-source path, and length and verbosity, sentiment polarity, style and structure, and harmful-content indicators in the linguistic-attribute path; M5 additionally includes sponsor, self-promotion, and interaction signals; M6 finally adds general top-video characteristics. Point estimates are labeled next to each marker, with significance indicated by asterisks; shaded areas denote 95% confidence intervals. Exact estimates and robust standard errors are reported in Tables B6 and B7 in Appendix B.

3.6 Conclusion

This chapter addresses a simple but important question in the creator economy: why do female YouTube creators still earn less advertising income than comparable male creators even after accounting for the observable creator-, channel-, and non-textual video-level characteristics emphasized in our previous analysis? The main finding is clear. Once we account for richer to content variables, the residual gender gap in estimated YouTube advertising income is no longer statistically significant. In other words, the remaining female income penalty identified in Chapter 2 can be fully explained by the additional

variables introduced here.

To reach this result, we build on the baseline dataset from the previous chapter and enrich it with newly collected video-level information from YouTube-related APIs. For the final analytical sample of 485 U.S. creators, we construct measures from each creator's most popular videos, using up to 50 long-form videos and up to 50 shorts. These additional variables include general top-content characteristics, creator-generated text from titles, descriptions, and captions, audience-generated text from comments, and commercial and interaction signals. Empirically, we combine sequential inclusion regressions, Gelbach decomposition, and separate block inclusion exercises to assess how these additional measures account for the remaining coefficient on *Female*.

This main finding is supported by three related results. First, top content general characteristics explain a substantial share of the residual gap, indicating that an important part of the remaining disparity is linked to observable differences in the performance profile of creators' most visible videos. At the same time, the broader set of results shows that this role should not be interpreted too strongly as a unique or independent channel. In the baseline sequential ordering, the *Female* coefficient becomes statistically insignificant once top content general characteristics are introduced, but the robustness checks show that audience-side text can also reduce the coefficient to zero before general performance metrics are added. The order-invariant Gelbach decomposition and the separate block inclusion results therefore suggest a more cautious interpretation: top content general performance is an important part of the explanatory set, but not the sole driver of the result. Second, among the newly added textual variables, audience-side measures are more informative than creator-controlled textual features. Creator-side text explains relatively little on its own, whereas audience-side responses play a much larger role. In particular, the sequential specifications show that the largest attenuation along the text-source path occurs when comment variables are added, while the linguistic-attribute results point to sentiment-related measures as the most informative textual dimension. Third, at the variable and block levels, two factors stand out most clearly: unequal audience

endorsement of the most popular videos, captured most strongly by average likes, and the evaluative tone of audience response, captured most clearly by comment sentiment and, more broadly, sentiment-related measures. The separate block inclusion results further show that top content genral metrics remain the strongest common non-textual block, while sentiment-related variables are the most stable textual dimension when considered on their own. By contrast, explicit commercial signals such as sponsorship and self-promotion add little once richer content and audience measures are taken into account. More broadly, the comparison across the cumulative, alternative-ordering, and separate-block analyses shows that the attenuation of the *Female* coefficient depends partly on how the variables are grouped and ordered. Therefore, a large reduction at one step should not be interpreted as evidence that this variable block alone explains the gap. This is why the sequential results should be interpreted cautiously and read together with the order-invariant Gelbach decomposition.

These findings carry a broader implication for the economics of platform-based creative labour. Because advertising income is mechanically determined by views and CPM, the results should not be interpreted as evidence of direct platform discrimination in monetization. Instead, they show that gender differences in income are associated with differences in the content, audience response, and visibility conditions that shape the inputs to that formula. In this chapter, the evidence points most strongly to views-related channels, especially those linked to audience evaluation and engagement, while some of the broader channel-level differences documented earlier may also reflect CPM-related variation. In this sense, gender inequality in the creator economy is not explained only by what creators produce or how they package their content. It is also shaped by how audiences respond to, reward, and discuss the content that becomes most visible. In a labour market where visibility, engagement, and scalable attention are central inputs into earnings, monetization depends not only on creator-side inputs but also on audience-side evaluation and the performance profile of the content that rises to prominence.

The results also suggest several implications for platform design and policy. Greater

transparency in recommendation, monetization, and engagement systems would make it easier to assess whether platform structures amplify gendered differences in audience response. More robust moderation and healthier comment environments may also matter, not only for creator well-being but for economic opportunity, since audience discourse appears closely tied to monetizable visibility. Likewise, creator-support strategies that focus narrowly on content production or brand strategy may overlook an important part of the problem if unequal audience evaluation remains unaddressed.

At the same time, the study has limitations. The income measure is an advertising-income proxy rather than total creator earnings, and the analysis is observational rather than causal. The results therefore identify explanatory factors and conditional associations rather than definitive causal effects. Even so, the chapter makes the main empirical message considerably clearer: for understanding gender inequality in YouTube advertising income, the key question is not only what creators make, but also how top-content performance, audience response, and platform-mediated visibility are jointly translated into income. That is where a substantial part of the previously unexplained gap is resolved.

Appendix A: Supplementary figures

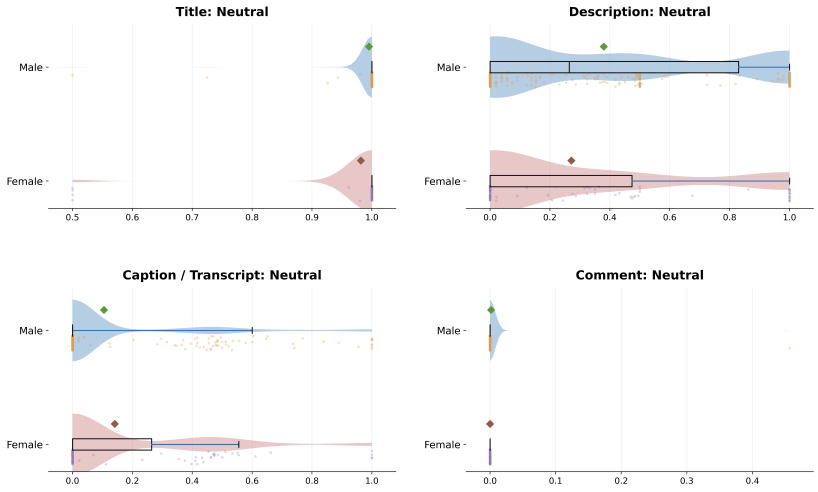


Figure A1. Neutral emotion across text sources

Note: The figure reports raincloud distributions of the *neutral* emotion score across four text sources: titles, descriptions, captions/transcripts, and comments. Each panel compares the distributions for male and female creators.

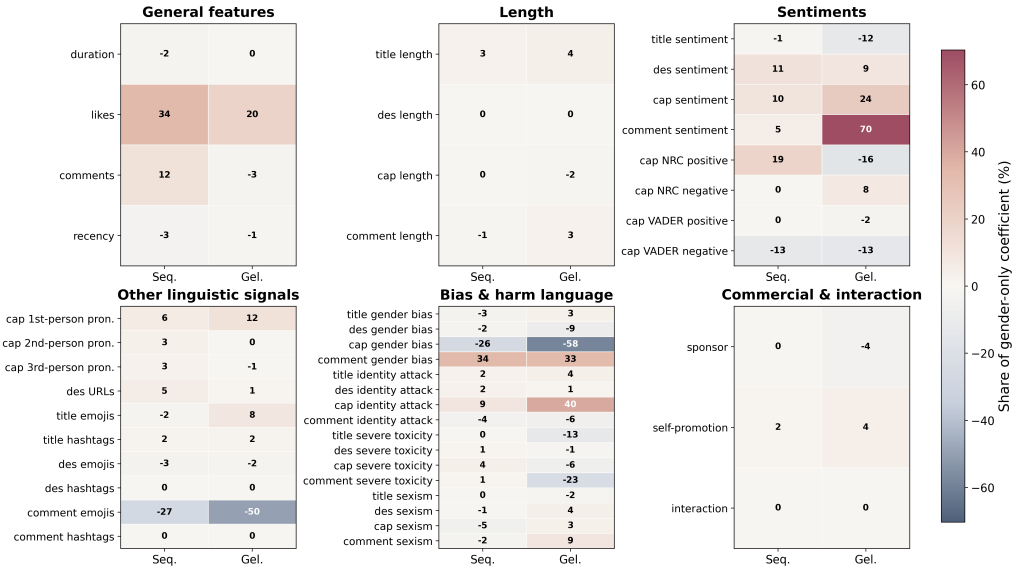


Figure A2. Variable-level contributions to the gender income gap under the alternative feature-based classification

Note: This figure reports variable-level contributions to the gender income gap under an alternative feature-based classification of video-level variables.

Appendix B: Supplementary tables

Table B1. Data description for complete variables

Variable	Description
Demographic and Identity Characteristics	
Female	Indicator equal to one for female creators and zero for male creators.
Age	Age of each creator, measured in years.
Race	Racial identity indicators: Black, Asian, Latino, and Other; White is the omitted reference group.
Marital status	Marital-status indicators: Married and Divorced; unmarried is the omitted reference group.
Number of children	Number of children reported for the creator.
Education	Educational-attainment indicators: Bachelor, Master, and PhD; below-bachelor is the omitted reference group.
Dropout	Indicator for whether the creator dropped out of school.
Language	Indicator for whether the primary channel language is English.
Face showing	Indicator for whether the creator appears visibly on screen.
Channel Basics and Content Category	
Channel age	Years since the YouTube channel was created.
Total subscribers	Total channel subscribers for each creator.
Content category	Channel-category indicators, such as Entertainment, Education, Gaming, Technology, Beauty & Fashion, News & Commentary, Cars, and Interviews.
Production, Activity, and Engagement Controls	
Yearly uploads	Number of uploads posted over the previous 365 days.
Yearly upload length	Total duration of uploads over the previous 365 days.
Average upload interval	Average number of days between recent uploads.

Continued on next page

Table B1. Data description for complete variables (continued)

Variable	Description
Inactive	Indicator for whether the channel had no uploads during the previous 365 days.
Average likes	Average number of likes on the 50 most recent uploads.
Average comments	Average number of comments on the 50 most recent uploads.
Top-Popular Video General Features	
Average duration	Average duration of the creator's top-performing long-form videos and Shorts.
Average likes	Average likes on the creator's top-performing long-form videos and Shorts.
Average comments	Average comments on the creator's top-performing long-form videos and Shorts.
Average days since upload	Average recency of the creator's top-performing long-form videos and Shorts, measured as days since publication.
Creator-Side Textual Features	
Title features	Measures derived from video titles, including length, sentiment score, emojis, hashtags, gender bias, and harmful-content indicators.
Description features	Measures derived from video descriptions, including length, URLs, emojis, hashtags, sentiment score, gender bias, and harmful-content indicators.
Caption/transcript features	Measures derived from captions/transcripts, including length, pronoun-use ratios, sentiment score, emotion composition, gender bias, and harmful-content indicators.
Audience-Side Textual Features	
Comment features	Measures derived from viewer comments, including length, sentiment score, emojis, hashtags, gender bias, and harmful-content indicators.
Commercial and Interaction Signals	
Sponsor segment duration	Total duration of sponsor segments identified through the SponsorBlock API.

Continued on next page

Table B1. Data description for complete variables (continued)

Variable	Description
Self-promotion duration	Total duration of self-promotional segments identified through the SponsorBlock API.
Interaction segment duration	Total duration of segments encouraging likes, subscriptions, comments, follows, or related audience actions.

Note: This table summarizes the definitions of all variables used in the analysis, including baseline creator- and channel-level controls, activity and engagement measures, top-performing video/Short characteristics, creator-side textual features, audience-side comment features, and commercial/intervention signals.

Table B2. Descriptive statistics by gender: Baseline controls

Variables	Mean (N=485)	Male (N=367)	Female (N=118)	Difference Male–Female
Demographic and Identity Characteristics				
Age	33.060	33.270	32.407	0.863
Race: White	0.664	0.695	0.568	0.127**
Race: Black	0.060	0.071	0.025	0.045*
Race: Asian	0.111	0.093	0.169	−0.077**
Race: Latino	0.070	0.057	0.110	−0.053*
Race: Other race	0.095	0.084	0.127	−0.043
Marital status: Unmarried	0.614	0.629	0.568	0.062
Marital status: Married	0.357	0.346	0.390	−0.044
Marital status: Divorced	0.029	0.025	0.042	−0.018
Number of children	0.534	0.572	0.415	0.157
School dropout	0.233	0.253	0.169	0.084*
Educational attainment: Low education	0.489	0.520	0.390	0.131**

Continued on next page

Table B2. Descriptive statistics by gender: Baseline controls (continued)

Variables	**Mean (N=485)**	**Male (N=367)**	**Female (N=118)**	**Difference Male–Female**
Educational attainment: Bachelor's degree	0.425	0.392	0.525	−0.133**
Educational attainment: Master's degree	0.041	0.038	0.051	−0.013
Educational attainment: PhD	0.045	0.049	0.034	0.015
English-language channel	0.895	0.916	0.831	0.085***
Face shown on screen	0.918	0.910	0.941	−0.031
Channel Basics and Content Category				
Channel age	11.590	11.779	11.000	0.779*
Total subscribers (millions)	12.041	13.132	10.009	3.033
Content category: People & Blog	0.064	0.022	0.195	−0.173***
Content category: Entertainment	0.406	0.425	0.347	0.078
Content category: Education	0.060	0.071	0.025	0.045*
Content category: Animation / Film	0.023	0.022	0.025	−0.004
Content category: Beauty & Fashion	0.029	0.011	0.085	−0.074***
Content category: DIY & Arts	0.023	0.016	0.042	−0.026*
Content category: Finance	0.006	0.005	0.008	−0.003
Content category: Food	0.064	0.057	0.085	−0.028
Content category: Gaming	0.167	0.188	0.102	0.086**
Content category: Music	0.029	0.027	0.034	−0.007
Content category: Pets & Animals	0.021	0.025	0.008	0.016
Content category: Sports & Fitness	0.041	0.049	0.017	0.032
Content category: Technology	0.029	0.033	0.017	0.016
Content category: Travel & Outdoor	0.010	0.011	0.008	0.002

Continued on next page

Table B2. Descriptive statistics by gender: Baseline controls (continued)

Variables	**Mean (N=485)**	**Male (N=367)**	**Female (N=118)**	**Difference Male–Female**
Content category: News & Commentary	0.010	0.014	0.000	0.014
Content category: Cars	0.008	0.011	0.000	0.011
Content category: Interviews	0.010	0.014	0.000	0.014
Production, Activity, and Engagement				
Yearly uploads	189.447	194.362	174.161	20.201
Yearly upload duration (hours)	31.670	36.401	16.581	19.820**
Average upload interval (days)	8.890	9.308	7.590	1.718
Inactive	0.056	0.054	0.059	−0.005
Average likes	95.953	109.771	52.951	56.820**
Average comments	170.821	170.049	173.250	−3.201

Note: This table reports descriptive statistics for all baseline control variables used in the regression analysis. The final column reports the mean difference between male and female creators. Asterisks indicate statistical significance from two-sample tests: *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table B3. OLS regression results under the full specification: Male, female, and combined samples

Variables	**Male (N=367)**		**Female (N=118)**		**Combined (N=485)**	
Gender effect						
Female					−0.028	(0.218)
Baseline controls						
Age	0.002	(0.012)	−0.035	(0.031)	0.001	(0.009)

Continued on next page

Table B3. OLS regression results under the full specification: Male, female, and combined samples (continued)

Variables	**Male (N=367)**		**Female (N=118)**		**Combined (N=485)**	
Black	−0.131	(0.212)	0.875	(0.964)	−0.193	(0.206)
Asian	0.179	(0.226)	−0.193	(0.328)	0.029	(0.178)
Latino	0.205	(0.190)	−0.860	(0.599)	−0.001	(0.187)
Other race/ethnicity	−0.195	(0.173)	−0.022	(0.396)	−0.074	(0.151)
Married	0.215	(0.156)	−0.423	(0.407)	−0.011	(0.129)
Divorced	−0.013	(0.409)	0.980	(0.863)	0.233	(0.328)
Number of children	−0.078	(0.066)	0.181	(0.172)	0.008	(0.055)
Dropout	−0.193	(0.152)	0.319	(0.408)	−0.260*	(0.144)
Bachelor's degree	−0.041	(0.142)	−0.150	(0.350)	−0.131	(0.117)
Master's degree	−0.310	(0.230)	−1.045	(0.923)	−0.422**	(0.214)
PhD	−0.058	(0.361)	−0.855	(1.158)	−0.486	(0.336)
English-language channel	0.515**	(0.218)	1.575***	(0.555)	0.617***	(0.180)
Face shown	−0.076	(0.187)	−0.626	(0.552)	−0.011	(0.162)
Channel age	−0.023	(0.017)	−0.033	(0.048)	−0.022	(0.016)
Log total subscribers	0.562***	(0.162)	−0.087	(0.382)	0.447***	(0.145)
Category: Entertainment	−0.007	(0.353)	0.928*	(0.482)	0.346	(0.221)
Category: Education	−0.086	(0.489)	1.042	(1.185)	0.504	(0.349)
Category: Animation/Film	−0.292	(0.466)	0.553	(0.894)	0.160	(0.361)
Category: Beauty/Fashion	−1.010**	(0.510)	1.404**	(0.521)	0.485	(0.355)
Category: DIY/Arts	0.539	(0.587)	−0.711	(0.795)	0.248	(0.512)
Category: Finance	−0.296	(0.601)	1.082	(0.787)	0.611	(0.556)
Category: Food	0.564	(0.407)	0.334	(0.664)	0.808***	(0.276)
Category: Gaming	−0.407	(0.403)	0.324	(0.729)	0.071	(0.278)
Category: Music	0.587	(0.423)	2.189**	(1.058)	1.030***	(0.302)
Category: Pets/Animals	0.721	(0.452)	0.603	(0.902)	1.014***	(0.366)

Continued on next page

Table B3. OLS regression results under the full specification: Male, female, and combined samples (continued)

Variables	**Male (N=367)**		**Female (N=118)**		**Combined (N=485)**	
Category: Sports/Fitness	0.218	(0.451)	1.408	(1.302)	0.639*	(0.337)
Category: Tech	−0.016	(0.407)	−1.079	(0.851)	0.307	(0.286)
Category: Travel/Outdoor	−0.631	(0.388)	0.169	(4.125)	−0.178	(0.325)
Category: News/Commentary	−1.194	(0.975)	omitted		−0.784	(0.920)
Category: Cars	0.032	(0.512)	omitted		0.368	(0.457)
Category: Interviews	−0.798	(0.664)	omitted		−0.435	(0.561)
Number of annual uploads	0.001***	(0.000)	0.001	(0.000)	0.001***	(0.000)
Log total length of annual uploads	0.142**	(0.055)	0.130	(0.145)	0.148***	(0.049)
Average days between uploads	−0.026***	(0.006)	−0.061**	(0.022)	−0.028***	(0.006)
Inactive creator indicator	−0.206	(0.526)	0.121	(1.315)	0.070	(0.475)
Log average likes	0.249***	(0.041)	0.298**	(0.130)	0.288***	(0.037)
Log average comments	−0.123**	(0.056)	−0.182	(0.178)	−0.144***	(0.050)
Video-level general features						
Average duration	0.000	(0.000)	0.001	(0.002)	0.000	(0.000)
Average likes	0.000**	(0.000)	0.001**	(0.000)	0.001**	(0.000)
Average comments	−0.007*	(0.004)	−0.035*	(0.018)	−0.009***	(0.004)
Average days since upload	−0.000***	(0.000)	−0.000	(0.000)	−0.001***	(0.000)
Textual features: title						
Title: length	0.007	(0.033)	0.132	(0.084)	−0.015	(0.023)
Title: number of emoji	0.178	(0.176)	−0.516	(0.511)	−0.180	(0.160)
Title: number of hashtag	−0.195	(0.193)	−0.086	(0.281)	−0.034	(0.111)
Title: sentiment score	1.231**	(0.584)	−0.280	(1.740)	0.960*	(0.496)
Title: bias ratio	0.009	(0.007)	−0.002	(0.009)	0.002	(0.005)
Title: identity attack	0.148	(0.212)	2.170	(1.611)	0.151	(0.222)
Title: severe toxicity	−2.079*	(1.205)	−10.819	(8.363)	−3.008**	(1.390)

Continued on next page

Table B3. OLS regression results under the full specification: Male, female, and combined samples (continued)

Variables	**Male (N=367)**		**Female (N=118)**		**Combined (N=485)**	
Title: sexism	−0.015	(0.087)	0.016	(0.267)	0.029	(0.077)
Textual features: description						
Des.: length	0.001	(0.001)	0.003	(0.002)	0.001	(0.001)
Des.: number of URLs	−0.028**	(0.014)	0.014	(0.030)	−0.029***	(0.011)
Des.: number of emoji	0.030*	(0.015)	−0.025	(0.051)	0.015	(0.016)
Des.: number of hashtag	−0.010	(0.012)	−0.131**	(0.062)	−0.005	(0.012)
Des.: sentiment score	−0.135	(0.169)	−2.032**	(0.967)	−0.173	(0.157)
Des.: bias ratio	−0.002	(0.003)	−0.006	(0.007)	−0.003	(0.003)
Des.: identity attack	−0.289	(0.618)	−6.526	(4.047)	0.074	(0.584)
Des.: severe toxicity	−0.153	(1.029)	4.583	(8.872)	−0.379	(1.077)
Des.: sexism	0.156	(0.104)	−0.003	(0.048)	−0.009	(0.023)
Textual features: caption						
Caption: length	0.000	(0.000)	−0.000	(0.000)	0.000	(0.000)
Caption: 1st-person pronouns	−0.034	(0.041)	−0.122	(0.128)	−0.051	(0.035)
Caption: 2nd-person pronouns	0.010	(0.053)	−0.316**	(0.141)	0.001	(0.043)
Caption: 3rd-person pronouns	−0.059	(0.068)	−0.277	(0.179)	−0.036	(0.056)
Caption: sentiment score	−0.842	(0.594)	1.007	(1.697)	−1.143**	(0.534)
Caption: NRC positive emotion	0.012	(0.025)	0.052	(0.059)	0.034	(0.022)
Caption: NRC negative emotion	−0.002	(0.018)	0.004	(0.062)	0.021	(0.014)
Caption: VADER positive	0.015***	(0.006)	−0.011	(0.014)	0.009*	(0.005)
Caption: VADER negative	0.006	(0.012)	−0.009	(0.046)	−0.010	(0.011)
Caption: bias ratio	−0.011**	(0.005)	−0.007	(0.010)	−0.010**	(0.004)
Caption: identity attack	0.163	(0.112)	1.643	(1.006)	0.349**	(0.135)
Caption: severe toxicity	0.016	(0.086)	−0.864	(0.655)	−0.136	(0.121)

Continued on next page

Table B3. OLS regression results under the full specification: Male, female, and combined samples (continued)

Variables	**Male (N=367)**		**Female (N=118)**		**Combined (N=485)**	
Caption: sexism	0.002	(0.028)	0.033	(0.094)	−0.036	(0.030)
Textual features: comment						
Comment: length	0.016	(0.022)	0.070	(0.080)	0.011	(0.019)
Comment: number of emoji	0.426***	(0.122)	0.442	(0.280)	0.442***	(0.091)
Comment: number of hashtag	0.032	(0.100)	0.109	(0.239)	0.017	(0.106)
Comment: sentiment score	−4.895***	(1.353)	−8.619**	(3.745)	−3.911***	(1.208)
Comment: bias ratio	0.004	(0.007)	0.017	(0.011)	0.002	(0.004)
Comment: identity attack	−0.153	(0.214)	−4.156	(2.530)	−0.085	(0.210)
Comment: severe toxicity	−0.515	(0.431)	−0.840	(2.900)	−0.484	(0.407)
Comment: sexism	0.038	(0.359)	−0.073	(0.932)	0.162	(0.318)
Commercial signals						
Sponsor segment duration	−0.005	(0.004)	−0.042	(0.032)	−0.004	(0.004)
Self-promotion duration	0.017	(0.014)	0.043	(0.143)	0.010	(0.014)
Interaction segment duration	−0.006	(0.018)	−0.091	(0.153)	−0.001	(0.016)
Controls						
Baseline controls included	Yes		Yes		Yes	
Constant	−1.059	(35.403)	25.146	(92.402)	−11.211	(31.553)
R^2	[0.822]		[0.958]		[0.813]	

Note: The dependent variable is the logarithm of estimated annual YouTube advertising income, proxied by the midpoint of *Social Blade*'s estimated annual income range. The table reports OLS estimates under the full specification for the male subsample, the female subsample, and the pooled sample. Baseline controls are reported explicitly in the table. Entries marked "omitted" indicate variables dropped because of collinearity in the female subsample. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table B4. Sequential specifications by text source

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	−0.270**	−0.128	−0.168	−0.142	−0.155	−0.032	−0.028
	[0.137]	[0.128]	[0.133]	[0.140]	[0.149]	[0.218]	[0.218]
Baseline controls							
Baseline controls (G0)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Video-level performance controls							
Top-video general metrics (G1)		Yes	Yes	Yes	Yes	Yes	Yes
Video-level text blocks by source							
Title text features (G2)			Yes	Yes	Yes	Yes	Yes
Description text features (G3)				Yes	Yes	Yes	Yes
Caption / transcript text features (G4)					Yes	Yes	Yes
Comment text features (G5)						Yes	Yes
Video-level commercial/interaction signals							
Sponsor / self-promo / interaction (G6)							Yes
Observations	485	485	485	485	485	485	485
R^2	0.689	0.758	0.764	0.772	0.793	0.812	0.813

Note: The dependent variable is estimated annual income. The main independent variable is an indicator for female creators. All columns include the baseline controls reported in Table 3.4. Column (2) adds video-level general characteristics. Columns (3)–(6) sequentially add text measures from titles, descriptions, captions/transcripts, and comments. Column (7) additionally includes sponsor, self-promotion, and interaction signals. Robust standard errors are reported in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table B5. Sequential specifications by linguistic attribute

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	−0.270**	−0.128	−0.124	−0.016	−0.062	−0.032	−0.028
	[0.137]	[0.128]	[0.126]	[0.149]	[0.148]	[0.218]	[0.218]
Baseline controls							
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Video-level performance controls							
Top-video general metrics (G1)		Yes	Yes	Yes	Yes	Yes	Yes
Video-level text blocks by linguistic attribute							
Length & verbosity (G2)			Yes	Yes	Yes	Yes	Yes
Sentiment polarity (G3)				Yes	Yes	Yes	Yes
Style & structure (G4)					Yes	Yes	Yes
Bias & harm indicators (G5)						Yes	Yes
Video-level commercial/interaction signals							
Sponsor / self-promo / interaction (G6)							Yes
Observations	485	485	485	485	485	485	485
R^2	0.689	0.758	0.760	0.774	0.790	0.812	0.813

Note: The dependent variable is estimated annual income. The main independent variable is an indicator for female creators. All columns include the baseline controls reported in Table 3.4. Column (2) adds video-level general characteristics. Columns (3)–(6) sequentially add text measures by linguistic attribute: length and verbosity, sentiment polarity, style and structure, and bias and harm indicators. Column (7) additionally includes sponsor, self-promotion, and interaction signals. Robust standard errors are reported in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table B6. Alternative ordering of sequential specifications by text source

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	−0.270**	−0.319**	−0.280*	−0.314**	−0.000	0.009	−0.028
	[0.137]	[0.142]	[0.146]	[0.152]	[0.234]	[0.235]	[0.218]
Baseline controls							
Baseline controls (G0)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Video-level text blocks by source							
Title text features (G1)		Yes	Yes	Yes	Yes	Yes	Yes
Description text features (G2)			Yes	Yes	Yes	Yes	Yes
Caption / transcript text features (G3)				Yes	Yes	Yes	Yes
Comment text features (G4)					Yes	Yes	Yes
Video-level commercial/interaction signals							
Sponsor / self-promo / interaction (G5)						Yes	Yes
Video-level performance controls							
Top-video general metrics (G6)							Yes
Observations	485	485	485	485	485	485	485
R^2	0.689	0.700	0.715	0.743	0.773	0.774	0.813

Note: The dependent variable is the logarithm of estimated annual advertising income. The main independent variable is an indicator for female creators. All columns include baseline controls. Columns (2)–(5) sequentially add text measures from titles, descriptions, captions/transcripts, and comments. Column (6) adds sponsor, self-promotion, and interaction signals. Column (7) adds top-content general metrics. Robust standard errors are reported in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table B7. Alternative ordering of sequential specifications by linguistic attribute

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	−0.270**	−0.265*	−0.135	−0.214	−0.000	0.009	−0.028
	[0.137]	[0.137]	[0.162]	[0.157]	[0.234]	[0.235]	[0.218]
Baseline controls							
Baseline controls (G0)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Video-level text blocks by linguistic attribute							
Length & verbosity (G1)		Yes	Yes	Yes	Yes	Yes	Yes
Sentiment polarity (G2)			Yes	Yes	Yes	Yes	Yes
Style & structure (G3)				Yes	Yes	Yes	Yes
Bias & harm indicators (G4)					Yes	Yes	Yes
Video-level commercial/interaction signals							
Sponsor / self-promo / interaction (G5)						Yes	Yes
Video-level performance controls							
Top-video general metrics (G6)							Yes
Observations	485	485	485	485	485	485	485
R^2	0.689	0.693	0.720	0.748	0.773	0.774	0.813

Note: The dependent variable is the logarithm of estimated annual advertising income. The main independent variable is an indicator for female creators. All columns include baseline controls. Columns (2)–(5) sequentially add text measures by linguistic attribute: length and verbosity, sentiment polarity, style and structure, and bias and harm indicators. Column (6) adds sponsor, self-promotion, and interaction signals. Column (7) adds top-content general metrics. Robust standard errors are reported in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table B8. Separate block inclusion results relative to the gender-only baseline

Group	Coefficient (Female)	Robust SE	% change vs. baseline	p-value vs. baseline	R^2
Gender only	−0.352	(0.223)			0.006
Panel A. Baseline control blocks					
Gender + Demographic characteristics	−0.545**	(0.227)	−54.83	0.015	0.096
Gender + Channel feature & content categories	−0.169	(0.177)	51.99	0.255	0.395
Gender + Annual production, activity, & engagement	−0.272*	(0.155)	22.73	0.562	0.528
Panel B. Newly added common content-level blocks					
Gender + Top-video general metrics	−0.282*	(0.144)	19.89	0.637	0.603
Gender + Sponsor & self-promo & interaction	−0.349	(0.226)	0.85	0.909	0.012
Panel C. Newly added text-source blocks					
Gender + Title text features	−0.657***	(0.226)	−86.65	0.002	0.092
Gender + Description text features	−0.322	(0.230)	8.52	0.785	0.118
Gender + Caption text features	−0.795***	(0.230)	−125.85	0.012	0.207
Gender + Comment text features	−0.083	(0.328)	76.42	0.375	0.351
Panel D. Newly added linguistic-attribute blocks					
Gender + Length & verbosity	−0.457**	(0.226)	−29.83	0.056	0.043
Gender + Sentiment polarity	−0.199	(0.228)	43.47	0.191	0.170
Gender + Style & structure	−0.605***	(0.200)	−71.88	0.048	0.282
Gender + Bias & harm indicators	0.366	(0.362)	− 3.98	0.023	0.247
Observations			485		

Note: The dependent variable is estimated annual income. The main independent variable is an indicator for female creators. The first row reports the gender-only specification. Each subsequent model adds one block separately to the gender-only specification. Panel A reports baseline control blocks. Panel B reports newly added common content-level blocks. Panel C reports newly added text-source blocks. Panel D reports newly added linguistic-attribute blocks. The column "% change vs. baseline" reports the percentage attenuation in the absolute value of the coefficient on *Female* relative to the gender-only specification, computed as $100 \times (|\beta_{\text{base}}| - |\beta_j|)/|\beta_{\text{base}}|$. Positive values indicate attenuation, while negative values indicate amplification. The column "p-value vs. baseline" reports post-estimation tests of equality between the coefficient on *Female* in each specification and that in the gender-only specification. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table B9. Robustness check: Mechanism paths under alternative outcome definitions

	Baseline	**Lower / upper bound**	**Winsorized**
	log midpoint	**log income**	**log income**
Panel A. Common specifications			
Baseline controls only (G0)	−0.270**	−0.270**	−0.248*
	(0.137)	(0.137)	(0.132)
+ General video-level features (G1)	−0.128	−0.128	−0.118
	(0.128)	(0.128)	(0.126)
+ Commercial & interaction signals (G6)	−0.028	−0.028	−0.021
	(0.218)	(0.218)	(0.219)
Panel B. Sequential inclusion by text source			
+ Title text features (G2)	−0.168	−0.168	−0.145
	(0.133)	(0.133)	(0.130)
+ Description text features (G3)	−0.142	−0.142	−0.114
	(0.140)	(0.140)	(0.135)
+ Caption / transcript text features (G4)	−0.155	−0.155	−0.160
	(0.149)	(0.149)	(0.152)
+ Comment text features (G5)	−0.032	−0.032	−0.023
	(0.218)	(0.218)	(0.220)
Panel C. Sequential inclusion by linguistic attribute			
+ Length & verbosity (G2)	−0.124	−0.124	−0.114
	(0.126)	(0.126)	(0.125)
+ Sentiment polarity (G3)	−0.016	−0.016	0.018
	(0.149)	(0.149)	(0.134)
+ Style & structure (G4)	−0.062	−0.062	−0.036
	(0.148)	(0.148)	(0.136)
+ Bias & harm indicators (G5)	−0.032	−0.032	−0.023
	(0.218)	(0.218)	(0.220)
Observations	485	485	485
Full-specification R^2	0.813	0.842	0.805

Note: The table reports the coefficient on *Female* from sequential OLS specifications under alternative outcome definitions. The second column combines the lower- and upper-bound income specifications because they yield the same estimates as the baseline midpoint specification. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table B10. Heterogeneity by channel type and channel size

	Channel type		Channel size	
	Entertainment (male-dom.)	**People & Blogs (female-dom.)**	**Below-median subscribers**	**Above-median subscribers**
Panel A. Common specifications				
Baseline controls only (G0)	−0.355***	−0.335	−0.110	−0.389**
	(0.171)	(1.096)	(0.225)	(0.191)
+ General video-level features (G1)	−0.304**	−0.811	0.149	−0.277*
	(0.158)	(1.477)	(0.206)	(0.161)
+ Commercial & interaction signals (G6)	−0.147	−0.241	−0.007	0.008
	(0.326)	(omitted)	(0.417)	(0.270)
Panel B. Sequential inclusion by text source				
+ Title text features (G2)	−0.321*	−0.898	0.088	−0.314*
	(0.180)	(omitted)	(0.201)	(0.166)
+ Description text features (G3)	−0.287	1.383	0.180	−0.342*
	(0.205)	(omitted)	(0.221)	(0.178)
+ Caption / transcript text features (G4)	−0.413*	3.872	0.219	−0.306
	(0.231)	(omitted)	(0.218)	(0.216)
+ Comment text features (G5)	−0.146	−0.265	−0.026	0.024
	(0.319)	(omitted)	(0.414)	(0.269)
Panel C. Sequential inclusion by linguistic attribute				
+ Length & verbosity (G2)	−0.294*	−1.284	0.147	−0.262
	(0.159)	(1.666)	(0.202)	(0.163)
+ Sentiment polarity (G3)	−0.237	−1.653	0.094	−0.148
	(0.216)	(omitted)	(0.262)	(0.184)
+ Style & structure (G4)	−0.300	−0.265	0.077	−0.205
	(0.234)	(omitted)	(0.232)	(0.195)
+ Bias & harm indicators (G5)	−0.146	−0.265	−0.026	0.024
	(0.319)	(omitted)	(0.414)	(0.269)
Observations	197	31	241	244
Full-specification R^2	0.876	1.000	0.842	0.836

Note: The table reports the coefficient on *Female* from sequential OLS specifications estimated separately by channel type and channel size. Robust standard errors are reported in parentheses. "omitted" indicates cases where the corresponding standard error is not available. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table B11. Robustness check: Alternative top content constructions

	Baseline (weighted composite)		**Videos only ($v_$ variables)**		**Shorts only ($s_$ variables)**	
Panel A. Common specifications						
Baseline controls only (G0)	−0.270**	(0.137)	−0.270**	(0.137)	−0.270**	(0.137)
+ General video-level features (G1)	−0.128	(0.128)	−0.209*	(0.115)	−0.192	(0.146)
+ Commercial & interaction signals (G6)	−0.028	(0.218)	0.032	(0.209)	−0.046	(0.229)
Panel B. Sequential inclusion by text source						
+ Title text features (G2)	−0.168	(0.133)	−0.220*	(0.117)	−0.210	(0.148)
+ Description text features (G3)	−0.142	(0.140)	−0.190	(0.126)	−0.217	(0.143)
+ Caption / transcript text features (G4)	−0.155	(0.149)	−0.189	(0.135)	−0.166	(0.152)
+ Comment text features (G5)	−0.032	(0.218)	0.016	(0.210)	−0.052	(0.227)
Panel C. Sequential inclusion by linguistic attribute						
+ Length & verbosity (G2)	−0.124	(0.126)	−0.193*	(0.116)	−0.201	(0.139)
+ Sentiment polarity (G3)	−0.016	(0.149)	−0.192	(0.147)	0.048	(0.141)
+ Style & structure (G4)	−0.062	(0.148)	−0.202	(0.146)	0.034	(0.144)
+ Bias & harm indicators (G5)	−0.032	(0.218)	0.016	(0.210)	−0.052	(0.227)
Observations	485		478		390	
Full-specification R^2	0.813		0.792		0.811	

Note: Column (1) uses baseline creator-level weighted composites that combine videos and shorts using variable-specific valid-count weights. Columns (2) and (3) use long-form video variables only and shorts variables only, respectively. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table B12. Robustness check: Restricting to channels with sufficiently informative comment data

	Full sample	Comment restrictions		
	(baseline)	Depth ≥ 300	Uploads ≥ 50	Coverage ≥ 0.90
Panel A. Common specifications				
Baseline controls only (G0)	−0.270**	−0.266*	−0.252*	−0.230*
	(0.137)	(0.141)	(0.134)	(0.143)
+ General video-level features (G1)	−0.128	−0.148	−0.084	−0.112
	(0.128)	(0.123)	(0.123)	(0.129)
+ Commercial & interaction signals (G6)	−0.028	−0.135	−0.058	−0.071
	(0.218)	(0.221)	(0.223)	(0.225)
Panel B. Sequential inclusion by text source				
+ Title text features (G2)	−0.168	−0.165	−0.118	−0.133
	(0.133)	(0.128)	(0.129)	(0.136)
+ Description text features (G3)	−0.142	−0.137	−0.092	−0.105
	(0.140)	(0.135)	(0.133)	(0.141)
+ Caption / transcript text features (G4)	−0.155	−0.140	−0.146	−0.112
	(0.149)	(0.140)	(0.149)	(0.148)
+ Comment text features (G5)	−0.032	−0.141	−0.064	−0.078
	(0.218)	(0.222)	(0.223)	(0.224)
Panel C. Sequential inclusion by linguistic attribute				
+ Length & verbosity (G2)	−0.124	−0.133	−0.089	−0.101
	(0.126)	(0.122)	(0.123)	(0.128)
+ Sentiment polarity (G3)	−0.016	−0.072	0.054	0.037
	(0.149)	(0.147)	(0.134)	(0.152)
+ Style & structure (G4)	−0.062	−0.117	−0.012	−0.016
	(0.148)	(0.149)	(0.138)	(0.156)
+ Bias & harm indicators (G5)	−0.032	−0.141	−0.064	−0.078
	(0.218)	(0.222)	(0.223)	(0.224)
Observations	485	447	464	460
Full-specification R^2	0.813	0.827	0.809	0.816

Note: Columns (2)–(4) restrict the sample by average comment depth, the number of top-content uploads with comments, and comment coverage, respectively. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table B13. Robustness check: Alternative aggregation of top content features

	Mean-based aggregation (baseline)		Median-based aggregation (alternative)	
Panel A. Common specifications				
Baseline controls only (G0)	−0.270**	(0.137)	−0.270**	(0.137)
+ General video-level features (G1)	−0.128	(0.128)	−0.149	(0.126)
+ Commercial & interaction signals (G6)	−0.028	(0.218)	−0.128	(0.202)
Panel B. Sequential inclusion by text source				
+ Title text features (G2)	−0.168	(0.133)	−0.150	(0.130)
+ Description text features (G3)	−0.142	(0.140)	−0.140	(0.128)
+ Caption / transcript text features (G4)	−0.155	(0.149)	−0.177	(0.137)
+ Comment text features (G5)	−0.032	(0.218)	−0.131	(0.201)
Panel C. Sequential inclusion by linguistic attribute				
+ Length & verbosity (G2)	−0.124	(0.126)	−0.140	(0.125)
+ Sentiment polarity (G3)	−0.016	(0.149)	−0.002	(0.135)
+ Style & structure (G4)	−0.062	(0.148)	−0.094	(0.139)
+ Bias & harm indicators (G5)	−0.032	(0.218)	−0.131	(0.201)
Observations	485		485	
Full-specification R^2	0.813		0.805	

Note: Column (1) uses mean-based creator-level aggregates for duration, likes, comments, and recency; Column (2) replaces them with medians. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table B14. Robustness check: Alternative aggregation and measurement of textual tone

	Baseline polarity (mean aggregates)	**Median-based polarity (median aggregates)**	**Eight-emotion (alternative tone block)**
Panel A. Common specifications			
Baseline controls only (G0)	−0.270**	−0.270**	−0.270**
	(0.137)	(0.137)	(0.137)
+ General video-level features (G1)	−0.128	−0.128	−0.128
	(0.128)	(0.128)	(0.128)
+ Commercial / interaction signals (G6)	−0.028	−0.041	−0.107
	(0.218)	(0.206)	(0.225)
Panel B. Sequential inclusion by text source			
+ Title text features (G2)	−0.168	−0.169	−0.135
	(0.133)	(0.133)	(0.128)
+ Description text features (G3)	−0.142	−0.140	−0.137
	(0.140)	(0.137)	(0.127)
+ Caption / transcript text features (G4)	−0.155	−0.129	−0.151
	(0.149)	(0.138)	(0.135)
+ Comment text features (G5)	−0.032	−0.045	−0.112
	(0.218)	(0.206)	(0.225)
Panel C. Sequential inclusion by linguistic attribute			
+ Length & verbosity (G2)	−0.124	−0.124	−0.124
	(0.126)	(0.126)	(0.126)
+ Sentiment / tone measures (G3)	−0.016	0.007	−0.111
	(0.149)	(0.136)	(0.124)
+ Style & structure (G4)	−0.062	−0.058	−0.136
	(0.148)	(0.138)	(0.126)
+ Bias & harm indicators (G5)	−0.032	−0.045	−0.112
	(0.218)	(0.206)	(0.225)
Observations	485	485	485
Full-specification R^2	0.813	0.815	0.821

Note: Column (1) uses baseline mean-based polarity measures. Column (2) replaces them with median-based polarity measures. Column (3) replaces the baseline polarity variables with an eight-emotion block constructed separately for titles, descriptions, captions/transcripts, and comments. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

General conclusion

This thesis aims to examine how gendered earnings gaps persist and change as labour markets become more flexible, more digitally mediated, and less tied to standard work settings. Across three essays spanning the traditional labour market and the creator economy in the United States, one central conclusion emerges: labour market transformation does not in itself eliminate gender inequality. Rather, it alters the institutional mechanisms through which inequality is generated, sustained, and expressed. In this sense, the thesis both extends and deepens the broader labour-economics literature by showing how the mechanisms underlying gendered earnings gaps evolve as work itself is reorganized.

The first major finding is that changes in work organization can modify the magnitude of gender gaps, but they do not remove the structural conditions under which those gaps arise. Chapter 1 shows that the gender wage gap in the United States narrowed across the wage distribution after the outbreak of the COVID-19 pandemic and that work-from-home played a key role in this process. This finding is important because it suggests that greater flexibility can ease some of the constraints historically associated with rigid schedules, temporal inflexibility, and career discontinuity. At the same time, we found that the effects of labour market change are not uniform. Once occupations are disaggregated into core STEM, STEM-related, and non-STEM fields, substantial heterogeneity becomes visible. The gender wage gap is smallest in core STEM occupations but largest in STEM-related occupations, where a strong and persistent glass-ceiling pattern remains. Traditional human-capital variables explain only a limited share of recent developments in the gap, whereas job-related characteristics, and especially female devaluation, account for a much larger share. Educational attainment is an important determinant only in STEM-related occupations, where educational segregation is stronger, but it does not explain the gap in core STEM and non-STEM occupations, where women tend to have higher educational attainment than men. The broader implication is that flexibility may reduce part of the gender wage gap, but occupational structure, valuation, and field-specific labour-market

organization remain central to understanding why inequality persists.

The second major finding is that even labour markets built around flexibility, low formal entry barriers, and digital intermediation do not automatically deliver more equal outcomes. Chapter 2 turns to the creator economy to ask whether similar or stronger equalizing dynamics emerge in a labour market where flexibility and digital mediation are fundamental rather than incidental. The answer is no. We documented a substantial and persistent gender income gap among top U.S. YouTube creators. Female creators earn significantly less than comparable male creators even after controlling for a rich set of demographic, channel, content, production, and engagement characteristics. As expected, traditional human-capital explanations have little power. Instead, content niche, annual content production, and audience engagement account for much of the explained component of the gap, while a substantial unexplained component remains. These findings challenge the idea that platform-mediated labour naturally leads to fairer outcomes simply because it is more flexible, more transparent, or less tied to conventional organizational structures.

The third major finding is that, in the creator economy, gender inequality is shaped not only by monetization structures and content specialization, but more importantly, by audience engagement, especially the linguistic and emotional forms through which audiences evaluate, endorse, and respond to creators' content. Chapter 3 demonstrated that this additional dimension is crucial for understanding why a residual female income penalty remains after controlling for standard creator and channel characteristics. In digital labour markets such as YouTube, content is not only produced and consumed; it is framed through language, distributed through platform mechanisms, and publicly evaluated by audiences whose responses can affect engagement, visibility, and ultimately monetization. By enriching the analysis with a richer set of video-level measures from each creator's top content, particularly textual features of the videos and audience responses expressed through comments, we found that the residual gender gap can be fully explained with these richer measures. General top content characteristics explain an important part of the gap,

but they do not explain it on their own. Most importantly, the strongest explanatory role is played not merely by creator-controlled textual features, but especially by audience-side measures, in particular audience endorsement as reflected in likes and the evaluative tone of comments. The chapter therefore moves the analysis of creator gender inequality one step further, beyond platform organization and monetization structures to the role of audience-mediated valuation. In doing so, it shows that language is not peripheral to digital labour markets; it could be one of the potential key channels through which economic outcomes are shaped.

Taken together, these findings point to several important conclusions. First, gendered earnings gaps persist across both traditional and emerging labour-market settings. The institutional form of work changes, but the presence of inequality does not disappear. Second, traditional human-capital explanations are consistently limited. Across the thesis, the evidence suggests that gender inequality reflects less a lack of qualifications among women than the gendered ways in which labour markets value, organize, and reward work. Third, what changes across labour-market settings is not the existence of gender inequality, but the mechanisms through which it operates. Chapter 1 shows that in the traditional labour market, gender inequality is tied mainly to occupational heterogeneity, work arrangements, and female devaluation. Chapters 2 and 3 then shift to the creator economy, a distinct platform-based labour market in which earnings depend less on employer-set wages and more on monetization, content specialization, and audience engagement. Within this setting, Chapter 2 identifies the main economic correlates of the gender income gap, whereas Chapter 3 extends the analysis by showing that the large residual female income penalty is further connected to textual framing, audience endorsement, and evaluative discourse. The thesis therefore contributes not only by comparing traditional and emerging labour markets, but also by showing, within the creator economy itself, how the analysis of gender inequality can be pushed one step further, from broad platform-based determinants of earnings to linguistic framing and audience-mediated evaluation.

The thesis also has limitations. All three chapters rely on non-experimental empirical

data, which means that the analyses identify robust patterns and plausible mechanisms, but do not permit the strongest causal claims. In Chapter 1, the IPUMS data make it possible to trace how the gender wage gap evolved before and after the COVID-19 pandemic and to assess the role of work from home, but they cannot directly observe all of the mechanisms through which remote work affects wage-setting. Chapters 2 and 3 rely on estimated advertising income and publicly observable platform data, which do not capture all dimensions of creator earnings, such as sponsorships, affiliate income, or other off-platform revenue streams. Nor do these data allow the analysis to fully observe recommendation systems, advertiser demand, or the causal effects of audience bias.

Despite these limitations, the thesis points to several important policy implications. In the traditional labour market, the results suggest that policies expanding flexibility, including work-from-home and hybrid arrangements, can help narrow gender wage gaps, but only if such arrangements do not generate new penalties in promotion, pay progression, or occupational sorting. At the same time, the strong heterogeneity across occupational fields indicates that one-size-fits-all interventions are unlikely to be sufficient. In particular, STEM-related occupations appear to require more targeted responses to educational segregation, female devaluation, and glass-ceiling dynamics. In the creator economy, the policy challenge is different but equally significant. The evidence suggests that reducing gender inequality in platform labour cannot be approached solely through creator-side skill development or entrepreneurial advice. Greater transparency in monetization systems, closer scrutiny of recommendation and advertising structures, and more active moderation of audience environments may all matter, especially when engagement and evaluative discourse are themselves part of the process through which earnings are generated.

Looking ahead, these findings also open up a broader future research agenda. One promising direction would be to combine platform-internal data with richer longitudinal designs in order to better trace how visibility, engagement, and monetization evolve over time. A second would be to incorporate more direct measures of sponsorship, recommendation exposure, and off-platform revenue in order to understand the full earnings structure

of creator labour. A third would be to use experimental or quasi-experimental approaches to identify the causal effects of audience evaluation, platform response, and algorithmic amplification. Finally, cross-platform and cross-country comparisons would help determine whether the patterns identified here are specific to YouTube and the United States or instead reflect broader transformations in digital labour markets. More broadly, the thesis suggests that as labour markets become more flexible, more platform-mediated, and more dependent on visibility and audience response, future research on gender inequality must also move beyond conventional employment frameworks and examine the new channels through which economic inequality is produced and reproduced.

References

Adams, A., Hara, K., Milland, K., & Callison-Burch, C. (2025). The gender wage gap in an online labor market: The cost of interruptions. *Review of Economics and Statistics*, 107(1), 55–64. doi:10.1162/rest_a_01282.

Albanesi, S., & Kim, J. (2021). Effects of the COVID-19 recession on the US labor market: Occupation, family, and gender. *Journal of Economic Perspectives*, 35(3), 3–24. doi:10.1257/jep.35.3.3.

Analyzify. (2025). YouTube statistics (2025). *StatsUp by Analyzify*. Available at: `https://analyzify.com/statsup/youtube`

Ash, E., and Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, 15(1), 659–688. doi:10.1146/annurev-economics-082222-074352.

Assaf, A., & Scuderi, R. (2020). COVID-19 and the recovery of the tourism industry. *Tourism Economics*, 26(5), 731–733. doi:10.1177/1354816620933712.

Bärtl, M. (2018). YouTube channels, uploads and views: A statistical analysis of the past 10 years. *Convergence: The International Journal of Research into New Media Technologies*, 24(1), 16–32. doi:10.1177/1354856517736979.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70(5), 9–49. Available at: JSTOR.

Berg, J., Furrer, M., Harmon, E., Rani, U., & Silberman, M. S. (2018). *Digital Labour Platforms and the Future of Work: Towards Decent Work in the Online World*. Geneva: International Labour Office. Available at: `https://www.ilo.org/sites/default/files/wcmsp5/groups/public/%40dgreports/%40dcomm/%40publ/documents/publication/wcms_645337.pdf`

Bertaglia, T. F. C. (2024). Decoding digital influence. *Doctoral dissertation, Maastricht University*. doi:10.26481/dis.20241107tb.

Bertrand, M. (2011). New perspectives on gender. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics*, Vol. 4B, pp. 1543–1590. Elsevier.

Bick, A., Blandin, A., & Mertens, K. (2023). Work from home before and after the COVID-19 outbreak. *American Economic Journal: Macroeconomics*, 15(4), 1–39. doi:10.1257/mac.20210061.

Bittermann, A., and Fischer, A. (2024). Natural language processing in psychology. *Zeitschrift für Psychologie*, 232(3), 143–146. doi:10.1027/2151-2604/a000568.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3), 789–865. doi:10.1257/jel.20160995.

Blau, F. D., & Kahn, L. M. (2020). The gender pay gap: Have women gone as far as they can? In J. Brueggemann (Ed.), *Inequality in the United States: A reader* (pp. 345–362). Routledge. doi:10.4324/9781003071709-31.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources*, 8(4), 436. doi:10.2307/144855.

Bolisani, E., Scarso, E., Ipsen, C., Kirchner, K., & Hansen, J. P. (2020). Working from home during COVID-19 pandemic: Lessons learned and issues. *Management & Marketing*, 15(s1), 458–476. doi:10.2478/mmcks-2020-0027.

Boulianne, S., Koç-Michalska, K., and Vedel, T. (2019). Gender and Online Politics: Digital Media as Friend and Foe in Times of Change. *Social Science Computer Review*, 39(2), 175–180. doi:10.1177/0894439319865511.

Cambronero-Saiz, B., Cristófol Rodríguez, C., and Segarra-Saavedra, J. (2023). Content Analysis From a Gender Perspective of Comments Received by Spanish Science YouTubers. *Media and Communication*, 11(1). doi:10.17645/mac.v11i1.5971.

Carlson, D. L., Petts, R. J., & Pepin, J. R. (2021). Changes in US parents' domestic labor during the early days of the COVID-19 pandemic. *Sociological Inquiry*, 92(3), 1217–1244. doi:10.1111/soin.12459.

Catalina-García, B., and Suárez-Álvarez, R. (2022). Twitter interaction between audiences and influencers. Sentiment, polarity, and communicative behaviour analysis methodology. *El Profesional de la Información*, 31(6). doi:10.3145/epi.2022.nov.18.

Chen, G. M., Pain, P., and Chen, V. (2018). 'You really have to have a thick skin': A cross-cultural perspective on how online harassment influences female journalists. *Journalism*, 21(7), 877–895. doi:10.1177/1464884918768500.

Chernozhukov, V., Fernández-Val, I., & Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, 81(6), 2205–2268. doi:10.3982/ECTA10582.

Cho, Y. K. (2021). STEM fields' contexts: Academic diversity and demographic inclusivity, 1995–2018. Doctoral dissertation, University of Wisconsin–Madison. Available at: University of Wisconsin Digital Collections.

Christin, A., and Lu, Y. (2023). The influencer pay gap: Platform labor meets racial capitalism. *New Media & Society*, 26(12), 7212–7235. doi:10.1177/14614448231164995.

Coibion, O., Gorodnichenko, Y., & Weber, M. (2020). Labor markets during the COVID-19 crisis: A preliminary view. NBER Working Paper No. 27017. doi:10.3386/w27017.

Collabstr. (2025). *2025 Influencer Marketing Report: Trends and Statistics for Influencer Marketing, UGC, and the Creator Economy*. Collabstr. Available at: `https://collabstr.com/2025-influencer-marketing-report`

Cook, C., Diamond, R., Hall, J. V., List, J. A., & Oyer, P. (2020). The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. *The Review of Economic Studies*, 88(5), 2210–2238. doi:10.1093/restud/rdaa081.

Cools, S., Finseraas, H., and Rasmussen, M. B. (2020). The immigrant-native gap in union membership: A question of time, sorting, or culture? *Labour*, 35(1), 24–51. doi:10.1111/labr.12186.

Corsi, G. (2024). Evaluating Twitter's algorithmic amplification of low-credibility content: An observational study. *EPJ Data Science*, 13, 18. doi:10.1140/epjds/s13688-024-00456-3.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 448–474. doi:10.1257/jel.47.2.448.

Cunningham, S., & Craig, D. (2019). *Social Media Entertainment: The New Intersection of Hollywood and Silicon Valley*. NYU Press. Available at: `https://www.jstor.org/stable/j.ctv12fw938`

DiNardo, J., Fortin, N. M., & Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973–1992: A semiparametric approach. *Econometrica*, 64(5), 1001–1044. doi:10.2307/2171954.

Djerf-Pierre, M., Lindgren, M., and Budinski, M. A. (2019). The role of journalism on YouTube: Audience engagement with 'superbug' reporting. *Media and Communication*, 7(1), 235–247. doi:10.17645/mac.v7i1.1758.

Dockery, M., & Bawa, S. (2020). Working from home in the COVID-19 lockdown. Bankwest Curtin Economics Centre Research Brief COVID-19 No. 5. Available at: Bankwest Curtin Economics Centre.

Döring, N., and Mohseni, M. R. (2020). Gendered hate speech in YouTube and YouNow comments: Results of two content analyses. *Studies in Communication and Media*, 9(1), 62–88. doi:10.5771/2192-4007-2020-1-62.

Dohmen, T. J. (2004). Performance, seniority, and wages: Formal salary sys-

tems and individual earnings profiles. *Labour Economics*, 11(6), 741–763. doi:10.1016/j.labeco.2004.01.003.

Dube, K., Nhamo, G., & Chikodzi, D. (2020). COVID-19 cripples the global restaurant and hospitality industry. *Current Issues in Tourism*, 24(11), 1487–1490. doi:10.1080/13683500.2020.1773416.

Duffy, B. E., and Hund, E. (2015). "Having it all" on social media: Entrepreneurial femininity and self-branding among fashion bloggers. *Social Media + Society*, 1(2). doi:10.1177/2056305115604337.

Duffy, B. E., Pinch, A., and Sannon, S. (2021). The nested precarities of creative labor on social media. *Social Media + Society*, 7(2). doi:10.1177/20563051211021368.

Duffy, B., Ononye, A., and Sawey, M. (2023). The politics of vulnerability in the influencer economy. *European Journal of Cultural Studies*, 27(3), 352–370. doi:10.1177/13675494231212346.

Eilon, S. (2024). From mommy bloggers to YouTube: The gender pay gap in the creator economy. *ThoughtLeaders*. Available at: `https://www.thoughtleaders.io/blog/from-mommy-bloggers-to-youtube-the-gender-pay-gap-in-the-creator-economy`

Eizmendi-Iraola, M., Peña-Fernández, S., & Morales-i-Gras, J. (2025). The gender gap in science communication on TikTok and YouTube: How platform dynamics shape the visibility of female science communicators. *Journalism and Media*, 6(3), 108. doi:10.3390/journalmedia6030108.

England, P. (Ed.). (1992). *Comparable Worth: Theories and Evidence*. Routledge. doi:10.4324/9781315080857.

Exley, C. L., & Kessler, J. B. (2022). The gender gap in self-promotion. *The Quarterly Journal of Economics*, 137(3), 1345–1381. doi:10.1093/qje/qjac003.

Firpo, S., Fortin, N. M., & Lemieux, T. (2007a). Unconditional quantile regressions. NBER Technical Working Paper No. 339. Available at: NBER.

Firpo, S., Fortin, N. M., & Lemieux, T. (2007b). Decomposing wage distributions using recentered influence function regressions. Unpublished manuscript, University of California. Available at: University of California, Irvine.

Firpo, S., Fortin, N., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3), 953–973. doi:10.3982/ECTA6822.

Fortin, N. M. (2015). Gender role attitudes and women's labor market participation: Opting-out, AIDS, and the persistent appeal of housewifery. *Annals of Economics and Statistics*, 117–118, 379. doi:10.15609/annaeconstat2009.117-118.379.

Frank, R. H., & Cook, P. J. (1996). *The Winner-Take-All Society: Why the Few at the Top Get So Much More Than the Rest of Us*. New York: Free Press.

Fry, R., Kennedy, B., & Funk, C. (2021). STEM jobs see uneven progress in increasing gender, racial and ethnic diversity. *Pew Research Center*. Available at: Pew Research Center.

Gallant, J., Kroft, K., Lange, F., & Notowidigdo, M. J. (2020). Temporary unemployment and labor market dynamics during the COVID-19 recession. NBER Working Paper No. 27924. doi:10.3386/w27924.

Gelbach, J. B. (2016). When do covariates matter? And which ones, and how much? *Journal of Labor Economics*, 34(2), 509–543. doi:10.1086/683668.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3), 535–574. doi:10.1257/jel.20181020.

Gioia, F., and Morabito, L. (2025). Behind the screen: Gender differences in the creator economy. *IZA Discussion Paper No. 17666, IZA Institute of Labor Economics*. Available at: `https://docs.iza.org/dp17666.pdf`

Glatt, Z. (2023). The intimacy triple bind: Structural inequalities and relational labour in the influencer industry. *European Journal of Cultural Studies*, 27(3), 424–440. doi:10.1177/13675494231194156.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, 104(4), 1091–1119. doi:10.1257/aer.104.4.1091.

González-Pérez, S., Mateos de Cabo, R., & Sáinz, M. (2020). Girls in STEM: Is it a female role-model thing? *Frontiers in Psychology*, 11. doi:10.3389/fpsyg.2020.02204.

International Labour Organization. (2021). *The Role of Digital Labour Platforms in Transforming the World of Work*. Geneva: International Labour Organization. Available at: https://www.ilo.org/publications/flagship-reports/role-digital-labour-platforms-transforming-world-work

Ionescu, S., Hannak, A., & Pagan, N. (2023). Group fairness for content creators: The role of human and algorithmic biases under popularity-based recommendations. In *Proceedings of the 17th ACM Conference on Recommender Systems* (pp. 863–870). ACM. doi:10.1145/3604915.3608841.

IPUMS. (2023). U.S. Census data for social, economic, and health research. Available at: IPUMS USA.

IZEA. (2021). *The 2021 State of Influencer Equality*. IZEA Worldwide, Inc. Available at: https://izea.com/press-releases/the-2021-state-of-influencer-equality/

Jiang, X. (2021). Women in STEM: Ability, preference, and value. *Labour Economics*, 70, 101991. doi:10.1016/j.labeco.2021.101991.

Jorge, A., Garcez, B., and Janiques de Carvalho, B. (2023). Parenting on celebrities' and influencers' social media: Revamping traditional gender portrayals. *Journalism and Media*, 4(1), 105–117. doi:10.3390/journalmedia4010008.

Joseph, D., and Bishop, S. (2024). Advertising as governance: The digital commodity audience and platform advertising dependency. *Media, Culture & Society*, 46(6), 1269–1286. doi:10.1177/01634437241237935.

Joshi, Y., Lim, W. M., and Jagani, K. (2023). Social media influencer marketing: foundations, trends, and ways forward. *Electronic Commerce Research*, 25(2), 1199–1253. doi:10.1007/s10660-023-09719-z.

Kahn, S., & Ginther, D. (2017). Women and STEM. NBER Working Paper No. 23525. Available at: NBER.

Kassenboehmer, S. C., & Sinning, M. G. (2014). Distributional changes in the gender wage gap. *ILR Review*, 67(2), 335–361. Available at: JSTOR.

Kim, S. (2024). Understanding political communication and political communicators on Twitch. *PLOS ONE*, 19(11), e0314429. doi:10.1371/journal.pone.0314429.

Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4), 181–209. doi:10.1257/app.20180010.

Kochhar, R. (2023). The enduring grip of the gender pay gap. *Pew Research Center*. Available at: `https://www.pewresearch.org/social-trends/2023/03/01/the-enduring-grip-of-the-gender-pay-gap/`

Koenig, F. (2023). Technical change and superstar effects: Evidence from the rollout of television. *American Economic Review: Insights*, 5(2), 207–223. doi:10.1257/aeri.20210539.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50. doi:10.2307/1913643.

Kovacheva, A., Wiener, H. J. D., and Kareklas, I. (2022). Online Engagement

with Memes and Comments about Climate Change. *Sustainability*, 14(14), 8900. doi:10.3390/su14148900.

Kricheli-Katz, T., & Regev, T. (2016). How many cents on the dollar? Women and men in product markets. *Science Advances*, 2(2). doi:10.1126/sciadv.1500599.

Krueger, A. B. (2005). The economics of real superstars: The market for rock concerts in the material world. *Journal of Labor Economics*, 23(1), 1–30. doi:10.1086/425431.

Kuschel, K., Ettl, K., Díaz-García, C., & Alsos, G. A. (2020). Stemming the gender gap in STEM entrepreneurship – insights into women's entrepreneurship in science, technology, engineering and mathematics. *International Entrepreneurship and Management Journal*, 16(1), 1–15. doi:10.1007/s11365-020-00642-5.

Lazear, E. P., & Oyer, P. (2004). Internal and external labor markets: A personnel economics approach. *Labour Economics*, 11(5), 527–554. doi:10.1016/j.labeco.2004.01.001.

Leighton, H. (2020). Study finds a pay gap between male and female influencers. *Forbes*. Available at: `https://www.forbes.com/sites/heatherleighton/2020/01/16/study-finds-a-pay-gap-between-male-and-female-influencers/`

Levanon, A., England, P., & Allison, P. (2009). Occupational feminization and pay: Assessing causal dynamics using 1950–2000 U.S. census data. *Social Forces*, 88(2), 865–891. doi:10.1353/sof.0.0264.

Lewis, R., and Christin, A. (2022). Platform drama: 'Cancel culture,' celebrity, and the struggle for accountability on YouTube. *New Media & Society*, 24(7), 1632–1656. doi:10.1177/14614448221099235.

Lewis, R., Marwick, A., and Partin, W. C. (2021). 'We Dissect Stupidity and Respond to It': Response Videos and Networked Harassment on YouTube. *American Behavioral Scientist*, 65(5), 735–756. doi:10.1177/0002764221989781.

Lips, H. M. (2012). The gender pay gap: Challenging the rationalizations, perceived equity, discrimination, and the limits of human capital models. *Sex Roles*, 68(3–4), 169–185. doi:10.1007/s11199-012-0165-z.

Litman, L., Robinson, J., Rosen, Z., Rosenzweig, C., Waxman, J., & Bates, L. M. (2020). The persistence of pay inequality: The gender pay gap in an anonymous online labor market. *PLOS ONE*, 15(2), e0229383. doi:10.1371/journal.pone.0229383.

Lobato, R. (2016). The cultural logic of digital intermediaries. *Convergence: The International Journal of Research into New Media Technologies*, 22(4), 348–360. doi:10.1177/1354856516641628.

Locatelli, M. S., Caetano, J. A., and Meira, W. (2022). Characterizing Vaccination Movements on YouTube in the United States and Brazil. *Proceedings of the 33rd ACM Conference on Hypertext and Social Media*, 80–90. doi:10.1145/3511095.3531283.

Lou, S., Adzharuddin, N. A., Syed Zainudin, S. S., & Omar, S. Z. (2025). Discrete emotions shape gender role attitudes: Exploring the impact of gender-stereotyped Douyin urban romantic short dramas on Chinese youth. *Studies in Media and Communication*, 13(3), 282. doi:10.11114/smc.v13i3.7663.

Løvlie, A. S., Ihlebæk, K. A., and Larsson, A. O. (2018). 'Friends call me racist': Experiences of repercussions from writing comments on newspaper websites. *Journalism*, 22(3), 748–766. doi:10.1177/1464884918781794.

Lyness, K. S., and Heilman, M. E. (2006). When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology*, 91(4), 777–785. doi:10.1037/0021-9010.91.4.777.

Mas, A., & Pallais, A. (2020). Alternative work arrangements. *Annual Review of Economics*, 12(1), 631–658. doi:10.1146/annurev-economics-022020-032512.

McCarthy, B. (2021). 'Who unlocked the kitchen?': Online misogyny, YouTube comments and women's professional street skateboarding. *International Review for the Sociology of Sport*, 57(3), 362–380. doi:10.1177/10126902211021509.

Meara, K., Pastore, F., & Webster, A. (2020). The gender pay gap in the USA: A matching study. *Journal of Population Economics*, 33(1), 271–305. doi:10.1007/s00148-019-00743-8.

Meisner, C. (2023). The weaponization of platform governance: Mass reporting and algorithmic punishments in the creator economy. *Policy & Internet*, 15(4), 466–477. doi:10.1002/poi3.359.

Mengel, F. (2021). Gender bias in opinion aggregation. *International Economic Review*, 62(3), 1055–1080. doi:10.1111/iere.12503.

Mincer, J., & Polachek, S. (1974). Family investments in human capital: Earnings of women. *Journal of Political Economy*, 82(2), S76–S108. Available at: JSTOR.

Mitchell, K. M. W., and Martin, J. W. (2018). Gender bias in student evaluations. *PS: Political Science & Politics*, 51(3), 648–652. doi:10.1017/S104909651800001X.

Miyazaki, K., Uchiba, T., Kwak, H., An, J., and Sasahara, K. (2024). The impact of toxic trolling comments on anti-vaccine YouTube videos. *Scientific Reports*, 14(1). doi:10.1038/s41598-024-54925-w.

Nakandala, S., Ciampaglia, G., Su, N., & Ahn, Y.-Y. (2017). Gendered conversation in a social game-streaming platform. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 162–171. doi:10.1609/icwsm.v11i1.14885.

National Science Foundation. (2021). The STEM labor force of today: Scientists, engineers and skilled technical workers. National Center for Science and Engineering Statistics. Available at: NCSES.

Németh, R., and Koltai, J. (2023). Natural language processing. *Intersections*, 9(1), 5–22. doi:10.17356/ieejsp.v9i1.871.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, 122(3), 1067–1101. doi:10.1162/qjec.122.3.1067.

Niederle, M., & Vesterlund, L. (2011). Gender and competition. *Annual Review of Economics*, 3(1), 601–630. doi:10.1146/annurev-economics-111809-125122.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3), 693. doi:10.2307/2525981.

Obakrim, S., Ailliot, P., and Monbet, V. (2024). EM algorithm for generalized Ridge regression with spatial covariates. *Environmetrics*, 35(6). doi:10.1002/env.2871.

OECD. (2023). *Handbook on Measuring Digital Platform Employment and Work*. OECD Publishing. doi:10.1787/0ddcac3b-en.

Ørmen, J., and Gregersen, A. L. (2022). Towards the engagement economy: Interconnected processes of commodification on YouTube. *Media, Culture & Society*, 45(2), 225–245. doi:10.1177/01634437221111951.

Orozco, D. R. (2021). Out on a STEM: Gender wage gap and factors that impact salary in science, technology, engineering, and mathematics careers. Doctoral dissertation, Seattle University. Available at: ScholarWorks at SeattleU.

Pattier, D. (2021). The gender gap among EduTubers and the factors significantly influencing it. *Journal of New Approaches in Educational Research*, 10(2), 313–329. doi:10.7821/naer.2021.7.732.

Peri, G., Shih, K., & Sparber, C. (2015). STEM workers, H-1B visas, and productivity in US cities. *Journal of Labor Economics*, 33(S1), S225–S255. doi:10.1086/679061.

Poecze, F., Ebster, C., and Strauß, C. (2019). Let's play on Facebook: using sentiment analysis and social media metrics to measure the success of YouTube gamers' post types. *Personal and Ubiquitous Computing*, 26(3), 901–910. doi:10.1007/s00779-019-01361-7.

Poggesi, S., Mari, M., De Vita, L., & Foss, L. (2019). Women entrepreneurship in STEM fields: Literature review and future research avenues. *International Entrepreneurship and Management Journal*, 16(1), 17–41. doi:10.1007/s11365-019-00599-0.

Polachek, S. W. (1975). Potential biases in measuring male-female discrimination. *The Journal of Human Resources*, 10(2), 205. doi:10.2307/144827.

Polachek, S. W. (2004). How the human capital model explains why the gender wage gap narrowed. IZA Discussion Paper No. 1102. Available at: IZA Institute of Labor Economics.

Regan, T. L., & Oaxaca, R. L. (2008). Work experience as a source of specification error in earnings models: Implications for gender wage decompositions. *Journal of Population Economics*, 22(2), 463–499. doi:10.1007/s00148-007-0180-5.

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira Jr., W. (2023). The amplification paradox in recommender systems. *Proceedings of the International AAAI Conference on Web and Social Media*, 17(1), 1138–1142. doi:10.1609/icwsm.v17i1.22223.

Rieder, B., Borra, E., Coromina, Ò., & Matamoros-Fernández, A. (2023). Making a living in the creator economy: A large-scale study of linking on YouTube. *Social Media + Society*, 9(2). doi:10.1177/20563051231180628.

Ringrose, J., Milne, B., and Horeck, T. (2024). Postdigital Bodies: Young People's Experiences of Algorithmic, Tech-Facilitated Body Shaming and Image-Based Sexual Abuse during and after the COVID-19 Pandemic in England. *Youth*, 4(3), 1058–1075. doi:10.3390/youth4030066.

Rivera, L. A., and Tilcsik, A. (2019). Scaling down inequality: Rating scales, gender bias, and the architecture of evaluation. *American Sociological Review*, 84(2), 248–274. doi:10.1177/0003122419833601.

Rosen, S. (1976). A theory of life earnings. *Journal of Political Economy*, 84(4), S45–S67. Available at: JSTOR.

Rosen, S. (1981). The economics of superstars. *The American Economic Review*, 71(5), 845–858. Available at: http://www.jstor.org/stable/1803469

Rotman, A., & Mandel, H. (2023). Gender-specific wage structure and the gender wage gap in the U.S. labor market. *Social Indicators Research*, 165, 585–606. doi:10.1007/s11205-022-03030-4.

El Sanyoura, L., and Anderson, A. (2022). Quantifying the creator economy: A large-scale analysis of Patreon. *Proceedings of the International AAAI Conference on Web and Social Media*, 16, 829–840. doi:10.1609/icwsm.v16i1.19338.

Sent, E.-M., & van Staveren, I. (2018). A feminist review of behavioral economic research on gender differences. *Feminist Economics*, 25(2), 1–35. doi:10.1080/13545701.2018.1532595.

Speer, J. D. (2020). STEM occupations and the gender gap: What can we learn from job tasks? IZA Discussion Paper No. 13734. Available at: IZA Institute of Labor Economics.

Teutloff, O., Stenzhorn, E., & Kässi, O. (2025). Skills, job application strategies, and the gender wage gap: Evidence from online freelancing. *ZEW Discussion Paper No. 25-016*. ZEW – Leibniz Centre for European Economic Research. Available at: https://www.econstor.eu/bitstream/10419/314421/1/1920405631.pdf

Thelwall, M., and Más-Bleda, A. (2018). YouTube science channel video presenters

and comments: Female friendly or vestiges of sexism? *Aslib Journal of Information Management*, 70(1), 28–46. doi:10.1108/AJIM-09-2017-0204.

Titus, V., O'Brien, J., and Parker, O. N. (2024). When the spotlight burns: Gender bias in the public perception of entrepreneurs. *Business & Society*, 64(1), 126–162. doi:10.1177/00076503241255512.

Toffoletti, K., and Thorpe, H. (2018). Female athletes' self-representation on social media: A feminist analysis of neoliberal marketing strategies in "economies of visibility." *Feminism & Psychology*, 28(1), 11–31. doi:10.1177/0959353517726705.

Töpfer, M. (2017). Detailed RIF decomposition with selection: The gender pay gap in Italy. In *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen*. ZBW – Leibniz Information Centre for Economics. Available at: EconStor.

Tur-Viñes, V., and Castelló-Martínez, A. (2019). Commenting on Top Spanish YouTubers: "No Comment". *Social Sciences*, 8(10), 266. doi:10.3390/socsci8100266.

Tuters, M., and Burton, A. (2021). The Rebel Yell: On YouTube's Burlesque Traditionalists and Their Alt-Right Audiences. *Canadian Journal of Communication*, 46(4), 757–776. doi:10.22230/cjc.2021v46n4a3937.

U.S. Census Bureau. (2018). Industry and occupation code lists & crosswalks: Occupation – 2018 Census STEM, STEM-related and non-STEM related code list. U.S. Census Bureau. Available at: U.S. Census Bureau.

U.S. Census Bureau. (2023). CPS and COVID-19. U.S. Census Bureau. Available at: U.S. Census Bureau.

United States Congress. (2023). H.R. 382: Pandemic is Over Act. Congress.gov. Available at: Congress.gov.

Vallas, S. P., & Schor, J. B. (2020). What do platforms do? Understanding the gig economy. *Annual Review of Sociology*, 46(1), 273–294. doi:10.1146/annurev-soc-121919-054857.

Veletsianos, G., Kimmons, R., and Larsen, R. (2018). Public comment sentiment on educational videos: Understanding the effects of presenter gender, video format, threading, and moderation on YouTube TED talk comments. *PLOS ONE*, 13(6), e0197331. doi:10.1371/journal.pone.0197331.

Verwiebe, R., Osorio-Krauter, C., Fritsch, N.-S., Weißmann, S., Philipp, A., & Buder, C. (2026). Working on an algorithm-controlled platform as a content creator: What explains unequal earnings on YouTube? *European Societies*, 1–29. doi:10.1162/euso.a.42.

Wotanis, L., & McMillan, L. (2014). Performing gender on YouTube. *Feminist Media Studies*, 14(6), 912–928. doi:10.1080/14680777.2014.882373.

Xie, S., & Connor, D. (2022). Stemming the tide of inequality? Immigration, gender, and spatial income disparities in today's leading technological sectors. *SSRN Electronic Journal*. doi:10.2139/ssrn.4036007.

Xu, Y. (2015). Focusing on women in STEM: A longitudinal examination of gender-based earning gap of college graduates. *The Journal of Higher Education*, 86(4), 489–523. doi:10.1080/00221546.2015.11777373.

Yang, S., Brossard, D., and Scheufele, D. A. (2022). The science of YouTube: What factors influence user engagement with online science videos? *PLOS ONE*, 17(5), e0267697. doi:10.1371/journal.pone.0267697.

YouTube. (2021). 10 ways to monetize on YouTube. *YouTube Official Blog*. Available at: `https://blog.youtube/news-and-events/10-ways-monetize-youtube/`

YouTube. (2025). *A Platform for Growth Through Creativity: Exploring YouTube's U.S.*

Impact in 2024. Available at: https://services.google.com/fh/files/misc/us_full_report.pdf

YouTube Help. (n.d.). YouTube Partner Program overview & eligibility. Google. Available at: https://support.google.com/youtube/answer/72851

Zheng, X., Vastrad, S., and Hu, J.-S. (2023). Contextualizing gender disparities in online teaching evaluations for professors. *PLOS ONE*, 18(3), e0282704. doi:10.1371/journal.pone.0282704.

Zveglich, J. E., Jr., van der Meulen Rodgers, Y., & Laviña, E. A. (2019). Expected work experience and the gender wage gap: A new human capital measure. *Economic Modelling*, 83, 372–383. doi:10.1016/j.econmod.2019.09.003.