Macalester Journal of Economics

Volume 31 May 2022

Department of Economics
Macalester College
Saint Paul, Minnesota
May 2022

FOREWORD

The Macalester College chapter of Omicron Delta Epsilon, the international honors society in economics, proudly edits the Macalester Journal of Economics every year. The editors—Franco Salinas Meza '22 (Arequipa, Peru) and Gracie Ellsworth '22 (Washington, D.C.)—have carefully selected six papers (including two Honors projects) on a variety of important topics. These papers are a sample of the research that our students produced in the last two academic years.

Nicholas (Nick) Di '22 (Saratoga, CA) asks whether the withdrawal of Federal Pandemic Unemployment Compensation (put simply, the unemployment insurance benefits offered during the COVID–19 pandemic) had a differentiated effect conditional on a worker's gender. In theory, this type of insurance can increase the length of an unemployment spell, as it allows for its recipient to choose their new job without financial pressure. As the extra compensation was eliminated, did we see a drop in the unemployment rate? Yes and no. Nick shows that the unemployment rate for males fell more than that of females, and that this effect is especially obvious for female workers who are married, in lower income brackets, or have children. The results from this Honors project provide an opportunity for policymakers to address the asymmetric effects of unemployment insurance payments.

Moving on to development economics, Cheikh Fall '23 (Dakar, Senegal) looks at the relationship between floods and early childhood development in Senegal. As a natural disaster, flooding can disrupt the flow of goods in the short run and agricultural yields in the medium run, which can then influence the well-being and development of children as they grow up. Using Senegal as a case study, Cheikh finds evidence of lower height-for-age as a result of floods. Interestingly, he finds that this impact occurs mostly in urban areas—as rural areas are pretty much unaffected by flooding.

Zefan Qian '23 (Nanjing, China) takes the New Keynesian (NK) model for a ride as he introduces sectoral heterogeneity to the canonical version of the model. The NK model is the bread-and-butter model for macroeconomists, and especially relevant for issues related to monetary policy and inflation. Zefan explores how the results of the baseline model change when we consider several industries (sectors), each with different price-setting constraints. As an exercise, he takes the model to the data by parametrizing it to fit the dynamics of a subset of Chinese industries.

Can gender differences in competitiveness lead to gender differences in education and labor market outcomes? This is the question set forth by Floyd Krom '21 (Naaldwijk, Netherlands) and Aaron Salot '21 (Mumbai, India). There is evidence of gender differences in STEM disciplines at the college level, but where do these differences come from? Using an experimental approach, Floyd and Aaron take Macalester students through a multi-stage trial and conclude that males are more confident than females and more likely to compete in experimental tournaments. Comparing these outcomes with the students' academic background, Floyd and Aaron show that about two thirds of sampled males choose a STEM major, which is more than the (slightly over) half of females who take the same path. They conclude that these results support the hypothesis that differences in overconfidence are likely causes of differences in post-graduation earnings.

The second Honors project, written by Xinyi Wang '22 (Nanjing, China), concerns the gender employment gap—the difference between male and female employment. In particular, Xinyi looks at the effect of the COVID–19 pandemic over this variable, considering the role of K–12 school closures and the primary role of (working) mothers in solving these childcare problems. She finds strong effects of school closures over employment: Other things the same, workers who are also parents of children are less likely to be employed after a school closure. However, this effect is more likely to impact female workers, as they are more likely to return to employment after a school reopening. Once again, these results can be taken as input by policymakers interested in fair outcomes in labor markets.

Franklin Marquette '21 (Little Falls, MN) and Zoraiz Paracha '21 (Lahore, Pakistan) look at the connection between religion and risk preferences, as evidenced by the experiences of Pakistan and the United States. Why does religion have anything to do with an individual's attitude towards risk? Perhaps belief in a higher power is enough to convince them that "everything will be OK in the end." In this line, Frank and Zoraiz perform a multi-dimensional analysis (e.g., differences in risk aversion between Christians and Muslims); using their behavioral economics toolkit, they find that (a) Muslims are less (monetary) risk averse than Christians, (b) frequent prayer outside religious services is linked with lower risk aversion, and (c) an opposition to gambling stems from risk (and not monetary loss) aversion.

Our last paper asks whether tattoos can predict shortsightedness. After all, there's a reason why tattooed people fall prey of stereotyping! To answer this question, Molly Hurley '21 (Stillwater, MN) follows an experimental approach that links impatience, self-reported behavior insights, and cognitive-reflection-test questions. Using a sample of Macalester undergraduates, she concludes that these stereotypes are not representative of our student body; that said, she finds that tattooed participants are more likely to need financial assistance, have lower GPAs, and belong to underrepresented backgrounds.

On behalf of my colleagues in the Economics Department, I am delighted to present the research of these talented students. I am confident that you will find it enlightening and be impressed by the value of a liberal arts education.

Mario Solis-Garcia
Associate Professor of Economics

Macalester Journal of Economics

Volume 31 *Spring 2022*

CONTENTS

Gendered labor market outcomes during COVID-19: Evidence from early withdrawal of Federal Pandemic Unemployment Compensation

Nicholas Di

1. Introduction:

The global economy shrunk 3.5 percent in 2020, a 7 percent loss as opposed to the 3.4 percent growth forecasted in October 2019, primarily due to the emergence of the global COVID-19 pandemic (Yeyati et al. 2020). The labor force participation rate declined from 63.4% in January 2020 to 60.2% in April 2020. As of August 2021, the labor force participation had yet to fully recover, at 61.7% (Center on Budget and Policies, 2021). The economic downturn hit the U.S labor market quickly and forcefully. The drop in the employment rate in post-outbreak months was driven by mass layoffs rather than workers voluntarily quitting their jobs, thus allowing laid-off individuals to qualify for safety net programs (Dias et al. 2020). In the week ending March 14, 2020, there were a total of 250,000 initial unemployment insurance (UI) claims- a jump of 20% from the week before. Just two weeks later, there were over 6 million claims (Bartik et al. 2020).

In March 2020, the CARES Act passed the Federal Pandemic Unemployment Compensation (FPUC), which added a $600 federal supplement to weekly UI benefits until July 2020. The Lost Wages Assistance (LWA) program, a federal-state unemployment benefit program, aimed to provide $300 to $400 in supplemental unemployment insurance benefits from the beginning of August to December 27th 2020. Unfortunately, funding for the program depleted earlier than expected and the LWA ended early.

The CARES Act partially reinstated the supplement at $300 per week in January 2021 until September 6, 2021. Federal payments to unemployment benefits are supplemental to state unemployment benefits, where state UI benefits are pegged to previous salaries. The Pandemic Emergency Unemployment Compensation (PEUC) also made the UI benefits more generous by extending coverage to those who exhausted standard state benefits. In addition to these extra payments, unemployment benefits became exceedingly accessible through the Pandemic

Unemployment Assistance (PUA), which extended benefits to uncovered workers such as the self-employed, freelancers, and part-time workers. All pandemic-related federal unemployment benefits expired on September 6, 2021.

UI benefits provide a temporary wage replacement to workers who become unemployed. The system was set up so individuals may smooth their consumption by creating a balance between spending and saving at times where they may have no working income. As a response to the historical recessions and unemployment, policy makers responded by extending the total eligible length of UI from 13 weeks to a max of 99 weeks— most states had a total eligible length of 26 weeks.[1] This means people who were laid off have more time than usual to remain unemployed and receive payments while looking for work.

Business owners expressed concern that unemployment benefits would deter job re-entry and reduce worker availability (Buchwald 2021). The rationale follows a neoclassical model of the economy where jobs are taken only to generate the income for optimal levels of consumption and leisure. Workers may “price” themselves into a job by lowering wage demands while benefit eligibility rules are not strictly enforced. Therefore any alternative source of unearned income will reduce individual work incentives and increase unemployment. An individual on UI will need a higher wage available, most likely more than historical wages, to rejoin the workforce. (Howell et al. 2011).

Work and life balance during the pandemic meant very different things between women and men, especially for mothers, as they tend to carry an outsize burden and shoulder the load of closed schools and daycare. In general, women are more likely to be part-time workers (Gould et al. 2020). Working women often choose more flexible work because of family responsibilities as

[1] The state decides the logistics as to how it wants to run the unemployment compensation programs and their own qualification guidelines. In 2021, select states withdrew from the FPUC earlier than its planned end date.

they are still often expected to be the primary caretaker for families. (Alon et al. 2020) During COVID-19 the CARES Act made unemployment insurance more accessible through PUA by extending eligibility to self-employed and part-time workers. Therefore, compared to pre-COVID, more women during COVID are able to claim unemployment benefits.

A wide range of literature points to controversial mechanisms as to whether unemployment benefits slow economic recovery as they incentivize certain individuals to remain unemployed. Once the UI benefits end, there will theoretically be an increase in employment- however, the nature of the pandemic may have impacted labor market outcomes differently among females and males (Coombs et al. 2021).

1.1 Research Question:

My paper examines whether the economic recovery differs by gender, particularly focusing on the effect of the expiration of unemployment benefits. During recovery periods of past recessions, women have undergone a slower and weaker recovery compared to men despite lower job losses. From June 2009 through May 2011, men increased their job count by 768,000 jobs and decreased the unemployment rate by 1.1 percentage points down to 9.5%. Women, during the same period, lost 218,000 jobs and increased the unemployment rate by 0.2 percentage points to a total of 8.5% (Kochhar, 2011). Albanesi and Kim (2021) study employment during COVID and categorize the labor force by high or low flexibility and high or low contact. Investigating the employment rates by gender, they found the slowest recovery in employment was represented by women in the flexible and low-contact sectors.

Although there has been literature using the expiration of federal unemployment benefits to gauge economic recovery, there is no current literature regarding how the lapse of benefits leads to different economic recovery between men and women. My hypothesis is that there is heterogeneity in unemployment between males and females once UI benefits are exhausted due to unproportional household production responsibilities and wage differences. To address the question I will use the Current Population Survey (CPS) survey and Bureau of Labor Statistics (BLS) data to run an analysis quantifying differences in unemployment and labor force participation throughout COVID economic recovery. I will then look at heterogeneity through marital status, status of having children, children's age and income. I find that women have a "slower" recovery compared to men similar to past recessions.

2. Literature Review:

Economists debate the effects between UI levels and the labor market. Several studies point to a correlation between increased UI and decreased labor market activity. Meyer (1988) utilizes a Kaplan Meier specification to estimate the relationship between employment and benefit exhaustion (or end). He finds that higher UI benefits reduce the probability of employment, in particular going from 6 weeks to 1 week until exhaustion triples the probability of employment. This means the more individuals receive from safety net programs the higher unemployment rates we will observe. In support of Meyer's study, Hagedorn et al. (2016) utilizes a county border discontinuity design to estimate the effects of UI benefit.[2] They found a 1% drop in benefit duration leads to a statistically significant increase of employment by 0.019 log points. Tao Zhang et al. (2002) uses a flexible hazard rate model to analyze the effect of unemployment compensation on unemployment duration. They conclude that the escape rate[3] rises sharply in months just to benefit exhaustion, meaning the rate at which people find work rises months prior to benefit exhaustion. Furthermore he notes that men are more responsive than women with respect to changes in unemployment compensation– where women are most responsive with

[2] A country border design looks at counties close to each other geographically but in different states, thus different state policies. The two counties will have similar labor markets since they share many geographic properties.
[3] The escape rate, in this context, is the rate at which people go from unemployment to employment.

respect to benefit exhaustion. This implies that women search harder for jobs at the notice of an upcoming benefit termination, which is supported by a higher escape rate for women compared to men.

To the contrary, other studies point toward a null or weak relationship between UI benefits level and labor market participation and unemployment. Gabriel et al. (2019) uses a Diamond Mortensen and Pisarides model to suggest a small and minimal effect of extending benefits. Specifically, Gabriel et al. found that extending benefits increases the unemployment rate by at most 0.3 percentage points. Ammar Farooq et al (2021) finds a controversial effect, where increasing UI has a positive effect on the labor market by improving job match quality in the re-employment job. Furthermore, they find that these effects are greater for women and less educated workers. Similarly, the Congressional Budget Office released a paper in 2012 suggesting that increased UI generosity actually stimulates employment, thus agreeing with Farooq. They theorize that unemployment benefits will stimulate individual’s disposable income, ultimately boosting firms' revenue and job openings for employment (CBO, 2012).

Some studies point to a correlation between UI and unemployment, dependent on group stratification. Raj Chetty et al (2005) examines the effect of UI benefits on unemployment exit hazards between constrained and unconstrained groups in terms of liquidity[4], using nonparametric graphical methods and Cox hazard models. Chetty observes that a 10% increase in UI benefits raises unemployment durations by 6-8% in all the constrained groups but had little to no effect on unemployment duration among unconstrained groups.

Throughout the COVID recession, policy makers extended the length of UI benefits for an additional 39 weeks, loosened eligible requirements, and increased benefits by $600 per week

[4] A constrained group in Raj Chetty et al’s experiment refers to liquidity. For unconstrained group individuals, losing a job will not be too bad on economic health as the individual will be available for debt, savings, and other assets. However, in constrained groups, people will not have access to the above, thus we can look at pure income effects.

in 2020 and $300 per week in 2021, both of which are supplemental to state benefits. UI traditionally has an average income replacement rate at 35-50 percent in most states— however, replacement rates during COVID-19 were estimated to be over 100% for about ⅔ of recipients. The median amount an individual receives amounts to 134 percent[5] of lost wages during COVID-19. Specifically, for the bottom 20% of income distribution, UI benefits were able to more than double their wages (Ganong Et al. 2020).

Several studies examine the $600 dollar federal UI benefits administered from March 2020 to July 2020. Dube (2020) uses a difference-in-difference event study design to estimate the macro employment effects. He finds minimal impact of job gains from the benefit reduction, especially when he focuses on low-income households, who compromise most UI recipients. Altonji et al. (2020) runs a linear probability event study model and found that expanding UI generosity does not depress employment in the aggregate during COVID-19. In fact, workers facing large expansions in UI benefits return to previous jobs at similar rates to those not receiving expansions. A third study by Ganong et al. (2021) finds the negative effects of benefits on employment from discouraged job search were minimal. The job finding rate before and after the $600 supplement led to a reduction in discouraged job search by 0.2-0.4%. There was another round of federal UI benefits at $300 per week taking place at the start of 2021. Vaccines and a healthy labor market were not available in 2020 ; therefore, as unemployment benefits ended, it was exceptionally difficult to become employed or re-employed.

Coombs et al (2021) examine the second round of stimulus in 2021 and find contradicting results to the studies, Dube (2020), Altonji et al. (2020), and Ganong et al (2021), regarding the first round of stimulus. Coombs et al. (2021) studied the $300 dollar supplement round in 2021 and how the early withdrawal of pandemic unemployment insurance affects UI receipt,

[5] Some states saw incredible replacement rates. For example, New Mexico with 177% and Maryland with 129 %

employment and spending. They were able to utilize a natural experiment, as 22 states ended all supplemental pandemic unemployment insurance early. The team found that ending the pandemic UI increased employment by 4.4 percentage points and reduced UI recipiency by 35 percentage points among workers who had UI at the end of April 2021. This paper was published in August 2021, hence did not capture the effects of when UI expired for everyone on September 4th 2021.

2.1 Gender Disparities in the Impact of COVID:

In past American recessions, men typically bear more losses in employment than women. However, in the COVID-19 pandemic, women's unemployment increased by 12.8 percentage points between February and April 2020, opposed to an increase of only 9.9 percentage points for men (Alon et al. 2020). During September 2020, when kids returned to school, the unemployment rate between males and females differed by .3 percentage points. These together imply that having school-aged children placed a disproportionate burden on mothers compared to fathers, especially when schools and day-care centers were closed.

Women have been disproportionately impacted and constitute between 52.2% to 55.8% of unemployment insurance claims (Gould et al. 2020). Women are overrepresented in high-contact and inflexible industries most impacted by the pandemic. A research report conducted by McKinsey estimates that 4.5% of women are at risk of unemployment during COVID, opposed to 3.8% of men given the nature of the industries men and women participate in (Madgavkar et al. 2020). Household production was commonly disproportionately allocated. According to the American Time Use survey, women spent an average of 102 minutes per day while men spent an average of 46 minutes per day caring for and helping household children in 2020 (ATUS, 2021).

Hapkau et al. (2020) conduct a study in Europe and find that women were roughly equally affected in job loss when compared with men, but women in the study provide a larger share of increased unpaid work. Having a greater proportion of home production makes it harder for women to transition back to the workforce as their opportunity cost for work increased throughout COVID-19, providing more incentive to stay unemployed or even leave the workforce. This effect of increased opportunity cost is more for caregivers, particularly married couples with children, and individuals with children under the age of 5 (Heggeness et al 2021). Closures of schools and daycare centers due to COVID-19 have massively increased child care needs, which mostly became the working mother's responsibility despite both parents being home (Landivar et al. 2020).

According to Lee (2021), about one-third of all mothers in the workforce have scaled back or quit their jobs in 2020. This is apparent among heterosexual married couples, where both the mother and father work in telecommuting-capable occupations. Using data from 2017-2018 Current Population Survey, Titan Alon et al. (2020) find only 22 percent of female workers are employed in highly telecommutable occupations as opposed to 28 percent of male workers, thus making it remarkably difficult for women to adapt to new work conditions. To make matters worse, among parents who were able to telecommute to work, mothers saw a greater decrease in labor force compared to fathers. Heggeness and Suri (2021) conclude mothers disproportionately left the labor market at the end of the 2020-2021 virtual school year as opposed to women without children. In fact, mother's labor force participation decreased between 0.1-1.4 percentage points compared to women without dependent children.

The difference in women's and men's labor supply explain why men have an easier time recovering from a recession. Men's labor supply elasticity is lower, especially for married men, compared to women. This suggests that when men face unemployment during a recession, they are likely to stay in the labor force and eventually become employed. However, when women become unemployed in a recession, they are likely to drop out of the labor force or seek part-time work. Together, these patterns suggest during economic recovery for a recession where women are disproportionately affected, women will still face more pressure from a decline in aggregate labor supply (Fukuki et al. 2021).

2.2 Labor Market Conditions Since All States Expired:

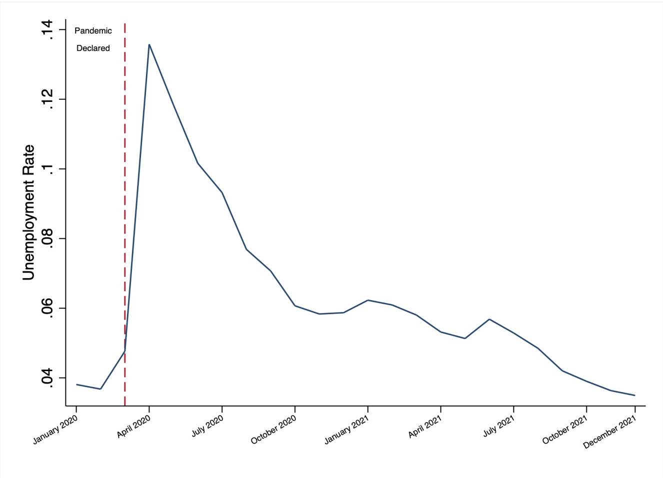
In the 2008 recession, some recipients stayed on UI benefits to improve their careers. At that time, a more generous unemployment insurance program refined the function of the markets by improving efficiency through job matching. Investing in human capital improves the efficiency of the market by boosting productivity, innovation, and general wellbeing. Both workers and employers gain from a generous UI as workers wages and firm productivity both increase (Farooq et al. 2021).

In October 2021, people searched for higher paying and more stable careers, this is reflected by stagnant levels of unemployment throughout the mid months of 2021. Massive layoffs throughout the pandemic prompted individuals, particularly the younger generation, to look for a "stable" job that offers flexibility and external career benefits. Thus, a fraction of the unemployed stayed on UI benefits not because they were incapable of finding a job, but because they would rather take their time to invest in themselves and find a suitable job (Smith, 2021). This effect was more prominent during the $600 dollar round of UI benefit in 2020, as individuals realized the importance of high wages and a stable job (Chalney 2021).

In the month of November 2021, the unemployment rate fell to a pandemic low of 4.2% from 4.6% in October. This drop in unemployment is significant as more than half a million workers, mainly females, returned to the workforce. As mentioned earlier, factors suppressing employment included public health concerns and child-care issues. In November 2021, the labor force participation rose to 61.8 percent, which is the highest level recorded since March 2020. 590,000 workers re-entered the labor force, 304,000 of whom were women.

In December 2021, the unemployment rate declined 0.3 percentage points to 3.9%, approaching the 3.5% unemployment rate in February 2020 before COVID was declared a pandemic. Labor shortages and supply chain worries limited the number of jobs added in December 2021 to 199,000, as opposed to October's 648,000 and November's 249,000. The labor market is still down by 3.6 million jobs from its pre-pandemic levels. The current unemployment rate may be explained by American's unwillingness to settle for lower paying and unstable jobs, as individuals now seek better pay and benefits. Fortunately, the number of initial unemployment claims have fallen below pre-pandemic numbers in recent weeks, implying that companies are holding on to current workers despite the Omicron outbreak (Rosenberg 2022).

Figure 1: Total Unemployment Rate Over Time



3. Theory:

I am going to model the decision of women to stay unemployed once UI expires, among individuals who remain in the labor force. I will do so by explaining the mechanisms and pieces building up to the Duncan model (2003). The decision to be employed will ultimately be a function of reservation wage[6], household production, and personal preferences. The theory behind my paper revolves around the trade-off between goods and time. I split time between leisure, labor, and household production. The foundational mechanics of the theory can be explained by a labor-leisure tradeoff model. Certain individuals face a greater opportunity cost dependent on changes in wage, non-labor income and personal preferences.

I will model the disproportionate impact of unemployment benefit exhaustion through a framework that explains the mechanisms behind market labor supply within a joint household. I do so using the model proposed in Duncan (2003), as this model allows us to examine the interdependence between spouses occupational choices and treat the spouse's characteristics as endogenous, because the spouse’s income is considered as non-work income for their partner.

A fundamental difference between a regular labor-leisure trade off and the Duncan model is the household production curve, where we account for joint maximization between partners– ultimately modeling goods and time spent on household production. This production function stems from the idea that consumers often choose not directly from commodities they purchase, but from commodities they transform into goods through production, which takes time. This is crucial to integrate into models of labor supply especially at a time when individuals are staying at home more than usual, thus consuming more home-produced goods.

[6] Reservation wage is the minimum amount of wage for someone to participate in the labor market.

3.1 Labor-Leisure Trade off in context of Unemployment Benefits:

People budget their time across different activities. An individual decides how much he or she "values" their leisure time versus labor time, which can be converted to goods bought using working wages. The more a person works, the more they will tend to value their remaining leisure time and vice versa. The shape of an individual's indifference curves depends on their preference towards leisure and labor.

When someone is laid off they can apply for unemployment insurance. In the labor-leisure tradeoff, the individual is considered unemployed, therefore all of their time is devoted to leisure and their utility curve sits at the corner solution at point B in figure 2.

Because of the government FPUC program, individuals' consumption levels are not at zero, but are at whatever amount is granted by the federal UI program. This is reflected by person B in figure 3. The budget constraint discontinues drops to the left of B in figure 3, because if someone were to start working, they would lose unemployment benefit.

Figure 2: Labor-Leisure

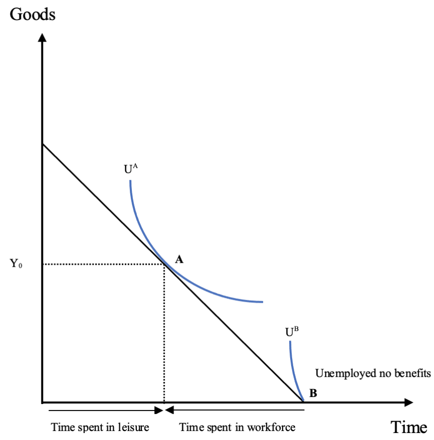
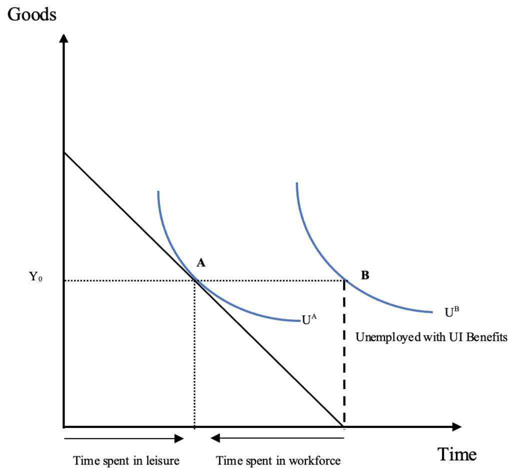


Figure 3: Labor-Leisure with UI Benefits



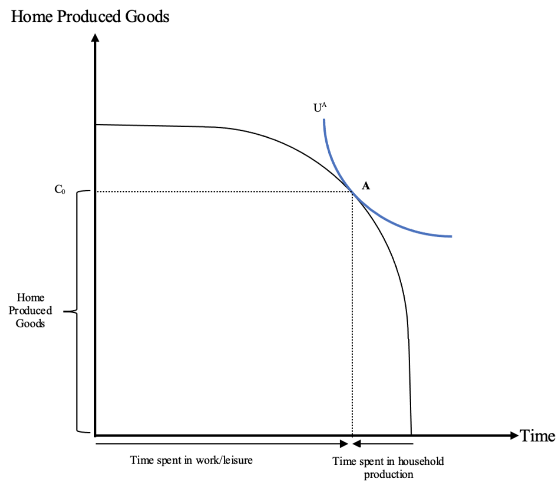
In this particular model, leisure represents all time spent outside of work. This does not necessarily mean the person is unproductive and not producing– as unpaid care is considered under leisure time. This idea is critical to address in the context of COVID as unpaid care was a major burden on households, especially on those with children.

3.2 Duncan Model Household Production Function:

In the Duncan model, there is a household production curve, which models goods and time spent in household production— also referred to as unpaid work. A line with more curvature means the individual tends to gain greater goods from production per unit of time spent on production. A flatter production line with less curvature means the person tends to be less effective in producing household goods, as diminishing returns are more evident.

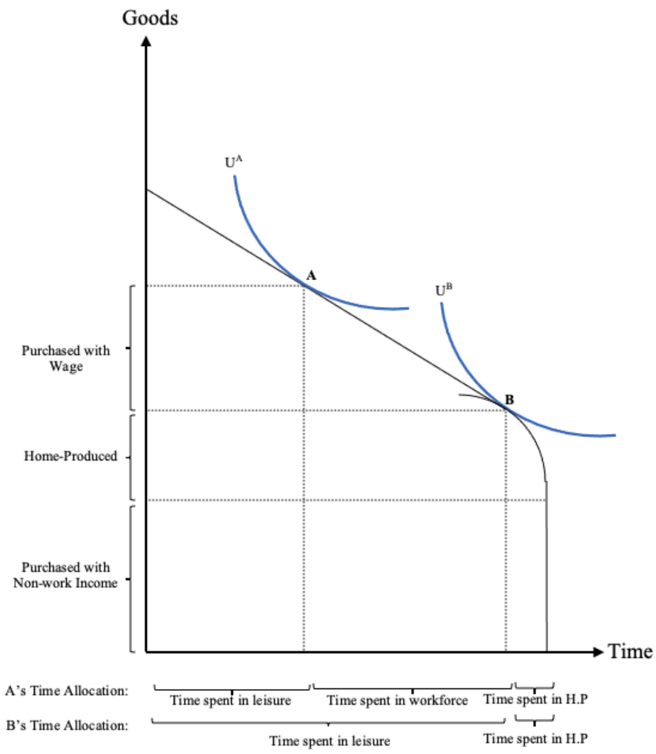
In figure 4, as we go from right to left on the production function, the slope tangent to the curve becomes less steep and approaches 0, this is because we are experiencing diminishing marginal returns to household labor. The x-axis in figure 4 is inverted for time spent in household production. As we spend more time in household production, we gain less and less home produced goods per additional unit increase in time spent on household production

Figure 4: Production Function



Mentioned earlier, the Duncan model is used to model joint maximization within households between the two partners. Goods within the Duncan model are split into 3 categories: goods purchased from wages, household production, and non-work income. The three categories can be seen below on the both X and Y axes in Figure 5. A person's time spent doing household production is dependent on the steepness of the wage slope, represented by the upper part of the "budget constraint", and curvature of the household production curve. An example of someone unemployed and maximizing household production is Person B in graph figure 5. In this particular case, person A and B are spending equal amounts of time in household production, where B is devoting all time outside of household production to leisure. However, since person A is on the wage slope, person A is able to spend time in the workforce by cutting time out of leisure.

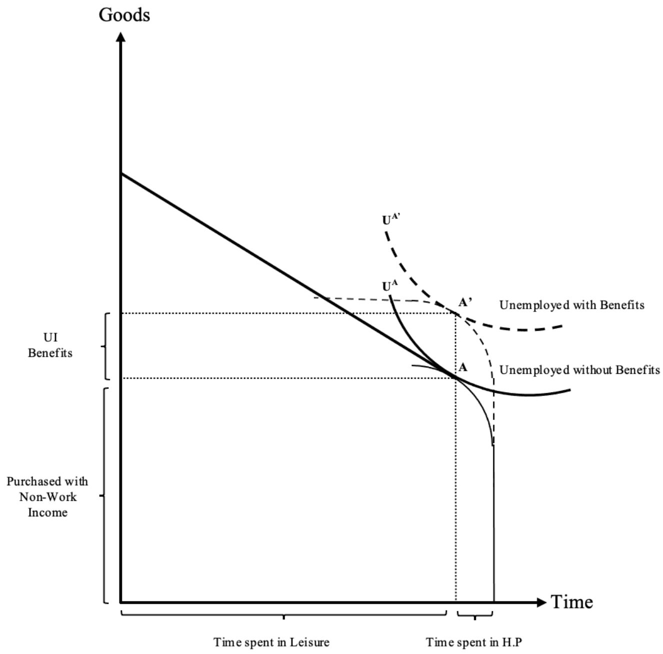
Figure 5: Duncan Model



3.3 Duncan Model During Federal Unemployment Insurance:

Similar to the regular labor-leisure tradeoff, there is a kink when UI is available to the unemployed. However, in the Duncan model, the kink— dotted in Figure 6— includes the household production function, as an individual will still be available for unpaid care while unemployed. Looking at figure 6, Person A is unemployed and splitting time between leisure and household production. However, once unemployment benefits are available, person A can now consume at A'. Person A's indifference curve moves up by the amount of UI benefits offered while keeping the amount of time spent in household production constant.

Figure 6: Duncan Model with Unemployment Benefits

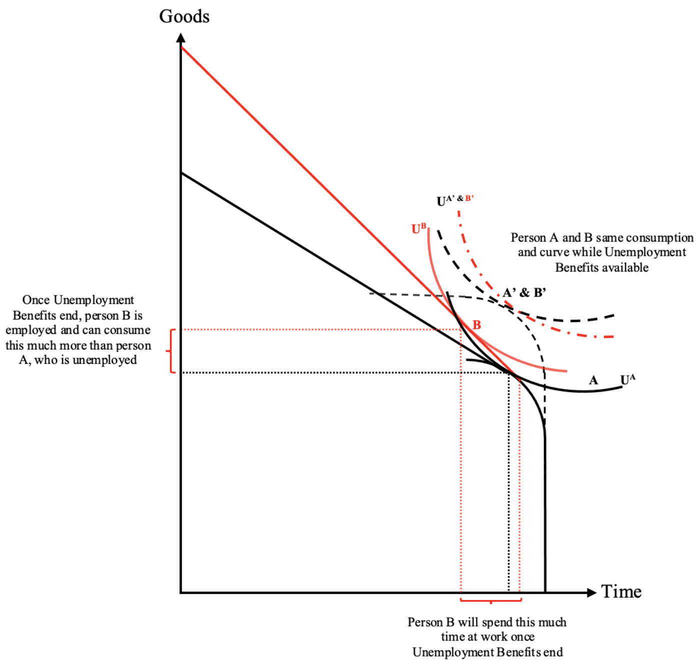


3.4 Duncan Model Different Wages and Production Curves:

An individual's decision to return to work will be dependent on available wages, when off UI, and household production. Throughout COVID-19, household production played an essential role for the working force, especially to parents— as more time spent at home resulted in greater duties and tasks requiring time investment. Once UI benefits are exhausted, we will be able to explain the deciding factors of employment or unemployment.

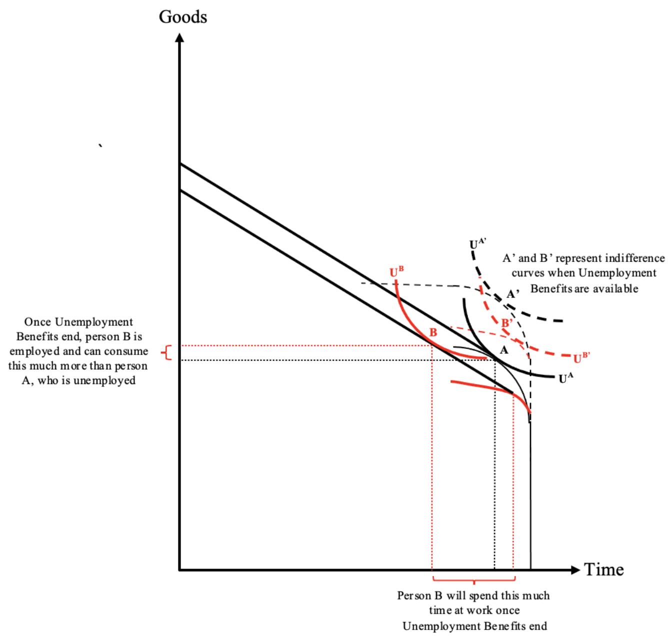
Let us model a situation where A and B have different available wages when off unemployment benefits. In this specific situation, A and B will have the same model except for differentiating wages represented by B having a steeper wage slope. Notice in figure 7, how A and B are consuming the same amount of goods when both are on unemployment benefits at A’ and B’. However, once unemployment benefits are exhausted, B’s red wage slope prompts the individual to work, while A’s wage slope prompts the individual to remain unemployed.

Figure 7: Duncan Model with Different Wage Slope with UI



We proceed by constructing a model dependent on an individual's household productivity. Let A be more productive and face less diminishing returns to home-produced goods than B.[7] When on UI, A' has greater goods compared to B'– as being unemployed will result in more time spent at home and increase home production. However, as unemployment benefits end, B becomes employed while A does not as B's production function is flatter and the wage slope will be able to catch B.

Figure 8: Duncan Model with Different Production Function with UI



In conclusion, using the Duncan model, we will tend to see individuals with lower available wages and more convex production curves to remain unemployed. The nature of

[7] Person A will have a more convex production function compared to person B, meaning that person A will consume at levels yielding greater home produced goods.

different wages and household roles are the main driving mechanisms resulting in a higher unemployment rate for women throughout the pandemic. As states go off federal unemployment benefits, individuals unearned income reduces, and the wage rate or household production income will be the ultimate determinant regarding employment status according to theory and these models (Fry 2022).

Looking at both figure 7 and 8, I hypothesize the ending of UI benefits to impact the change of male unemployment more than the change of female unemployment. For my paper, person A will be a female while person B will be a male, since empirical evidence points towards females experiencing lower wages and more time spent in households. I have a null hypothesis where the difference between male and female unemployment will be zero, thus if my coefficient is statistically significant, we will be rejecting the null hypothesis and conclude the exhaustion of unemployment benefits heterogeneously impacts sexes differently.

4. Data Description:

I use the Current Population Survey (CPS) to identify the change of unemployment and labor participation rates during periods with and without unemployment insurance benefits. I use the UI initial claims data from the Bureau of Labor Statistics (BLS) to provide further state-wide background information regarding total UI claims in each state throughout the pandemic.

4.1 Current Population Survey:

The CPS is a monthly survey of individuals living in households conducted by the Census Bureau. The CPS is currently used to produce estimates on monthly statistics regarding workforce participation, employment, and unemployment that are closely watched by businesses, investors, and policymakers. The dataset uses a rotating panel structure, with households resurveyed for a number of months.

The survey provides detailed economic and demographic data representing everyone age 15 and over in the U.S who is employed, unemployed, or not in the labor force. For my study, I extracted every month from January 2020 to December 2021, which covers both rounds of the FPUC and periods of time before and after benefits. However, for my empirical analysis, I will only look at the year 2021. Each observation has a year, month, and state variable; allowing me to construct a dummy variable indicating if a respondent lived in a state with Federal UI in a given month.

The survey collects observational data from all 50 states, including several other territories. It follows individuals living in a household for 4 months, takes an 8 month break, then interviews them for another 4 months, meaning households follow a 4-8-4 pattern. Therefore, each observation is an individual who is repeated in the survey for a maximum of 8 observations in 8 different months. There are a few rare cases where households only appear once throughout the dataset, hence, I have a combination of both unique and repeated observations. Using the demographic information, I can subcategorize my observations into females, males, low income, race, mothers and fathers. This is critical for heterogeneous analysis on the effects of removing UI.

4.2 Bureau of Labor Statistics UI initial Claims Data:

The Bureau of Labor Statistics is a federal agency that collects data regarding the U.S economy and labor market. Relevant to my research, the BLS collects data on initial UI claims at a state-level and monthly frequency. The report segregates the initial claims by age, industry and sex. Since the CPS data does not include UI data, I will use the BLS data to depict trends in total UI claims throughout the scope of my study to visualize how male and female total UI claims correlate to one another.

4.3 Industry UI Statistics:

Looking at the gendered nature of work across industries, the top five industries by concentration of females were disproportionately impacted by COVID-19 compared to male dominated industries[8].

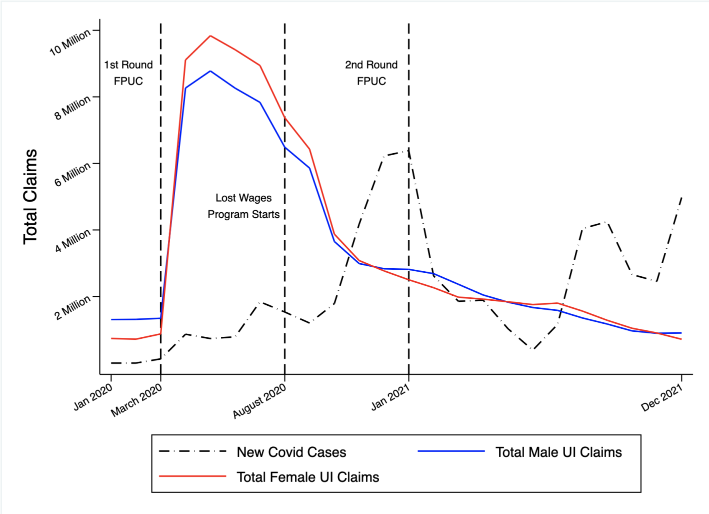
Figure 9: Industries by Percent Female February 2020

Percent Female	Industry
9%	Automotive Repair and Maintenance
11%	Construction
12%	Truck Transportation
24%	Architectural, Engineering, and Related Services
26%	Computer Systems Design and Related Services
76%	General Medical and Surgical Hospitals
76%	Educational Services
77%	Outpatient Care Services
80%	Individual and Famility Services
96%	Child Day Care Services

Child daycare services and schools were immediately shut due to shelter in place orders- most of which had trouble setting up telecommuting work in initial months of the pandemic. Male dominated industries, engineering and computer systems roles were able to transition more smoothly to telecommutable work while construction and truck transportation temporarily halted but gradually continued through the pandemic. This is supported by the trends in UI claims by sex during the first few months of COVID-19 using the BLS data in figure 11.

[8] COVID affected industries differently, as certain industries require more face-to-face interaction. I sort relevant industries by highest and lowest female employment in Table 1 using employment data in February 2020.

Figure 11: Unemployment Claims by Sex

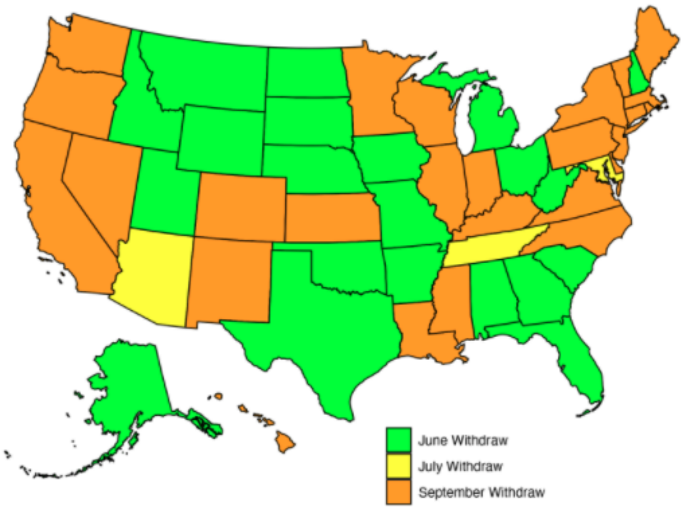


At the start of 2020, total male claims were significantly greater than total female claims, likely due to labor force composition as some industries have more turnover than others. In March 2020, total female UI claims outpaced total male UI claims when the FPUC started, meaning more females were laid off and filed for UI benefits in the early months of the pandemic. When the extra federal payments end in July 2020, we see the number of male and female UI claims become increasingly similar in trend characteristics. A Lost Wages program utilized FEMA funding to smooth consumption between the two rounds of stimulus payments.

I plot the total number of new covid cases within the US along with male and female claims. A spike in December 2020, attributed to the delta variant, led to a gap in male and female claims going into January 2021. The gap suggests that COVID decreased total female UI claims relative to males, which may be a result of many factors, from females dropping out of the labor force to an increased male unemployment rate.

During the second round of federal UI benefits in 2021, contrary to the $600 stimulus round, select states withdrew from the program earlier by cutting off the weekly installments of $300 2-3 months earlier. Below in Figure 12 I map out when each state withdrew from the program.

Figure 12: Federal Pandemic Unemployment Compensation Cutoffs



The June and July states made a decision to withdraw early after making claims that unemployment benefits are encouraging laid-off workers to stay at home instead of looking for jobs. Figure 13 depicts how many people are in UI states each month within the CPS dataset.

Figure 13: Observations in CPS Dataset

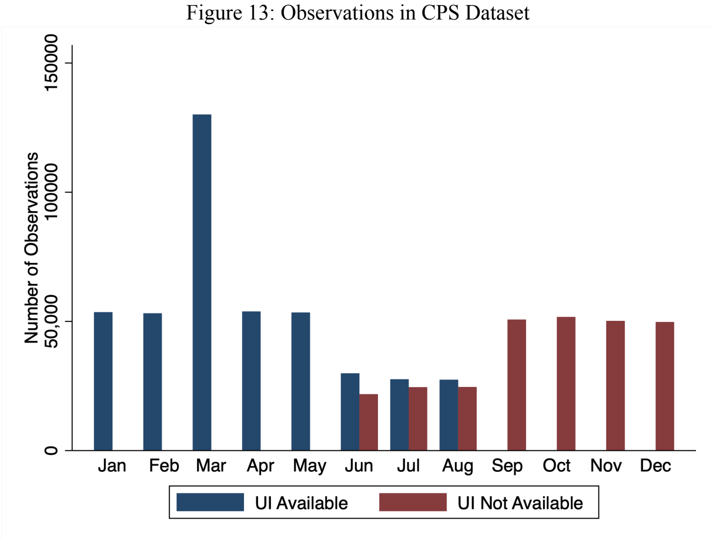
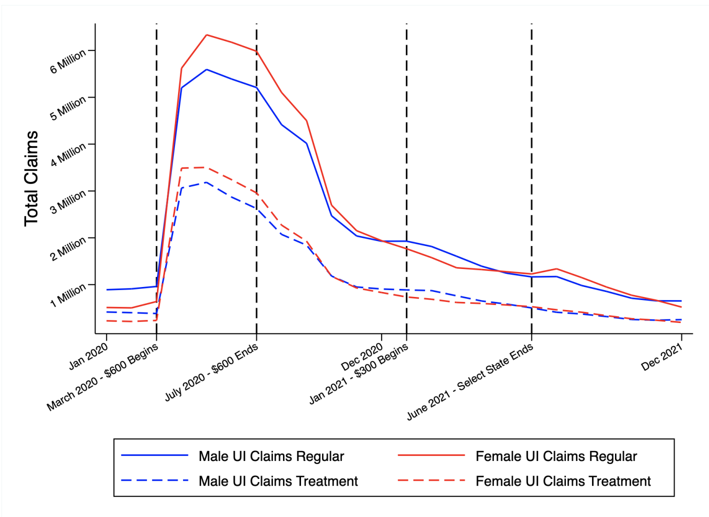


Figure 14: UI claims by Gender and Treatment

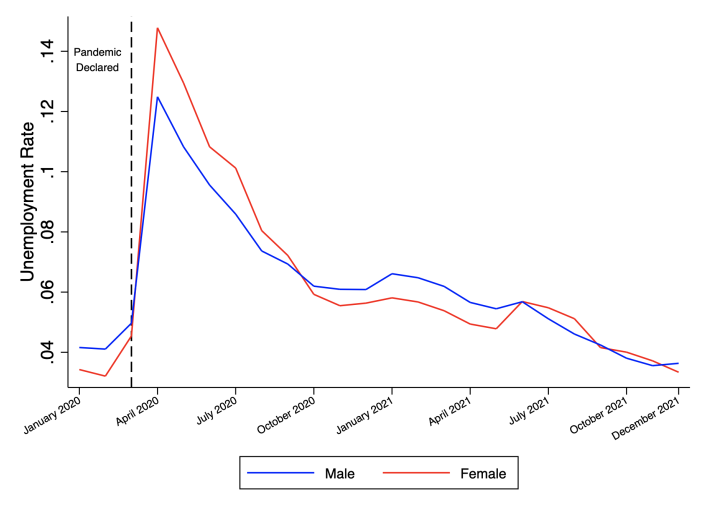


4.4 Unemployment Graphs:

In figure 14, I graph trends in total UI claims by gender and treatment status. “Female control” and “Male control” represent individuals in states that let UI expire at its original date in September rather than withdrawing from the program earlier in June. As treatment states withdrew in June 2021, we note a decrease in both male and female total UI claims, with males claims declining slightly more than females. Meanwhile, states that kept federal UI claims saw a continued growth in claims, more evident among females, the same period treatment group had a drop in claims.

Theoretically there should be a time lag between unemployment and filing for a claim as unemployment should lead to UI claims. Since my data is at a monthly frequency, we may not see this lagged observation, thus unemployment follows a similar trend to total UI claim trends. The overall unemployment trends by gender are displayed in figure 15 below.

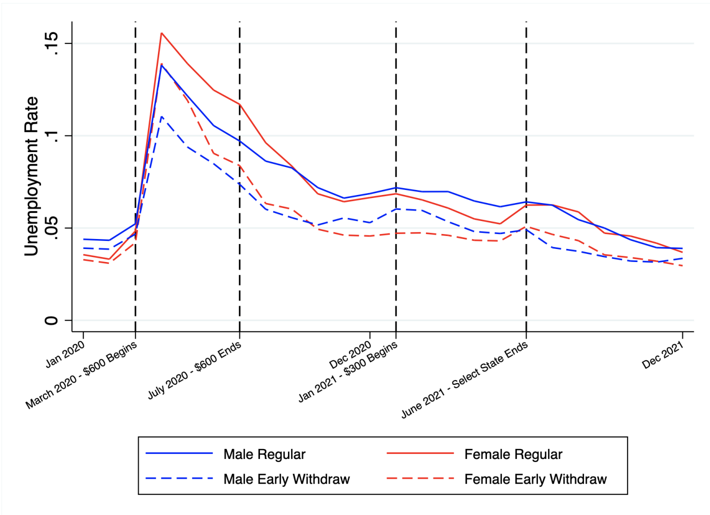
Figure 15: Unemployment by Gender



At the start of 2020, male and female unemployment rates were both below 5%, with male initially being higher than female unemployment. As soon as the pandemic hit, female unemployment jumped to 16% while men's increased to 13.6%. The gender difference disappeared gradually and both rates fell down around 6% in December 2020, this is surprising as some schools stayed closed so I would expect a gap to persist. However, during this time, women labor force participation dropped, leading to a lower unemployment rate.

We can see how gender played a role in determining trends. I plot treatment group vs control group by sex in figure 16.

Figure 16: Unemployment by Gender and Treatment

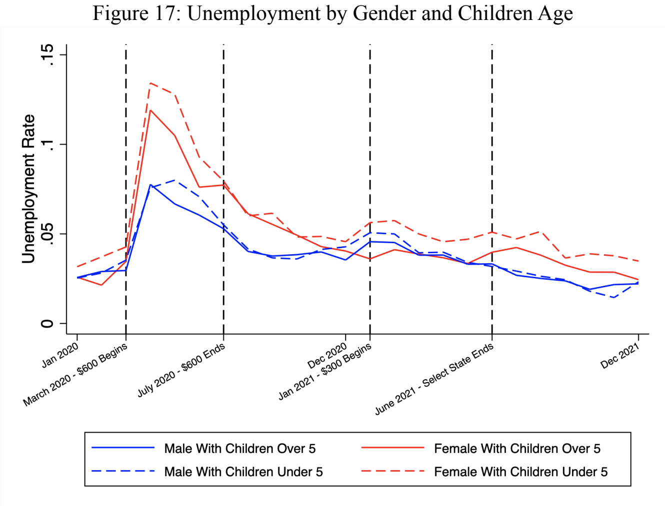


Within early withdrawal states, when UI expires, we note a greater decrease in male unemployment compared to female unemployment. While regular withdrawal states see nearly no significant change in unemployment between male and females, this graph suggests that when individuals have no access to additional federal UI benefits, males see a greater drop in unemployment than females.

We can see the differences in unemployment are small among the four groups at the start of 2020. However, as COVID-19 weakens the economy, female unemployment skyrockets the most in September withdrawal states, which are mostly democratic states by state legislature. Both female and male unemployment in control states are consistently higher than early withdrawal states. An anecdote explaining this mechanism is the harsh and more stringent protocols taken by states under Democratic party control- which ultimately lead to restricted economic activity and higher chance of unemployment among industries that have difficulty going remote (Goolsbee et al. 2020).

In figure 17, it seems that females with children under 5 consistently have higher unemployment rates than females with children over 5. Males seem to have unemployment rates independent of children's age.

Figure 17: Unemployment by Gender and Children Age



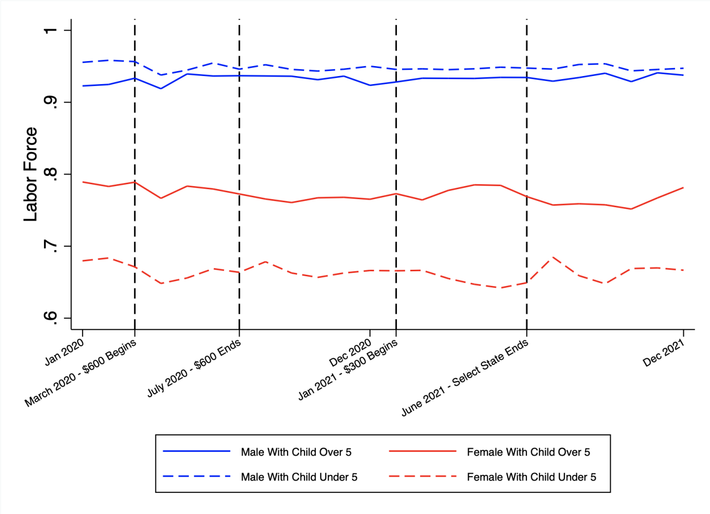
In the U.S, it is standard practice to send children under the age of 5 to daycare. Males are not sensitive to whether they had a child under 5 or not, while females with children under 5 saw greater unemployment rates throughout the pandemic compared to females with children over 5. In fact, males with children under and over 5 behave nearly the same as unemployment benefits end. This is similar to the graph above regarding labor participation by gender and children's age. Before the pandemic, females with children under 5 saw the greatest unemployment out of the four groups. Again, this implies females had an disproportionate amount of unpaid care that ultimately left them voluntarily or obligated to quit their jobs.

4.5 Labor Force Participation Graphs:

It is important to consider labor force participation as employment and unemployment rates are divided by labor force. Throughout the pandemic, women's labor force participation has been declining as a result of additional burdens in the form of needed caregiving and household responsibilities. In order to claim UI, the claimant must be in the labor force actively looking for a job.

Several strong predictive factors in deciding labor participation are children and age of children. Figure 18 analyzes how labor force outcomes sex are dependent on children's age among parents. It is interesting to see how male labor force participation is not influenced by children's age, as both "male with child over 5" and "male with child under 5" hardly differ in labor force participation rate, similar to figure 18 above. Household and childcare responsibilities increased for many during the pandemic, but gender inequalities were most evident among those with children. The gap between "female with child over 5" and "female with child under 5" stays consistent throughout the duration of the pandemic.

Figure 18: Labor Force Participation on Gender and Child Age



4.6 Control Variables:

Control variables are necessary to parse out a casual relationship between individuals and periods on and off of UI. This is because control variables will be able to reduce potential bias caused by omitted variables. The control variables used in my study are age, income, race, and education, all of which may affect employment status and unemployment benefits. Other studies have used similar control variables when measuring labor force outcomes (Aaronson et al. 2021).

4.7 Wage and Work Gap During Pandemic:

Male and females were exposed to different wages throughout the pandemic. Looking at data from 2020 to the end of 2021, females, on average, made 2.31 dollars less in hourly wages. This is important in the context of returning to work after UI programs expire as higher wages will prompt individuals to be more inclined to work. Table 2 below highlights the overall difference in hourly wage between males and females (CPS, 2021).

Table 2: Differences in Wage

	Male	Female	Difference	Standard Error
Average Hourly Wage	21.13	18.82	2.31***	0.0541

Note: *,**,*** mean significance at the 1, 5, and 10 percent level. Standard error is the standard error of the difference.

According to the American Time Use Survey, in terms of providing secondary childcare, which refers to childcare while doing something else, men averaged 4.9 hours per day while women averaged 7.1 hours per day in 2020. The Annual American Time Use Survey also states that men spent an extra 16 minutes per day on housework in 2020, compared to 2019. However, women spent an average of 2.4 hours per day on household work and unpaid care as opposed to 1.6 hours spent by men (MacLellan 2021). Women had to scale back work hours to tend to responsibilities within the household, this effect is more apparent among married individuals. This is shown as we summarize work hours per week below based on marriage status throughout the year 2020 in table 3.

Table 3: Hours Worked

Average Hours Worked per Week	Mean	Standard Error	Total
Married Male	42.74	0.017	365,587
Non Married Male	39.43	0.022	272,772
Married Female	37.78	0.020	305,236
Non Married Female	36.99	0.022	281,577

Note: All Data is from the years 2020-2021 from the CPS.

The gap between married male and females is much greater than between non-married females and males. Furthermore, the standard deviation is greater for married females compared to married males, implying that their hours fluctuate more from the mean throughout the pandemic.

Furthermore, we can see that a disproportionate number of the cases for why women are absent from work are due to "family responsibilities", "child care problems", and "maternity" in table 4. The difference in household production and unpaid care is reflected in the Duncan Model by a flatter production function curve in figure 8.

Table 4: Reasons for Absence from Work

Reason for Absence from Work	Male	Female	Difference	Standard Error	Total
Child Care Problems	17%	83%	65%***	0.0336	508
Maternity/Paternity	13%	87%	74%***	0.0120	2,947
Other Family/Personal	34%	66%	32%***	0.0186	2,585
Vacation Personal Day	44%	56%	12%***	0.0074	17,867
Own Illness/Injury	48%	52%	4%***	0.0088	12,745
School/Training	42%	58%	16%***	0.0312	1,006
Other	48%	52%	5%***	0.0080	15,539
Total	24,239	30,380			54,619

Note: *,**,*** mean significance at the 1, 5, and 10 percent level. Standard error is the standard error of the difference. Standard errors are clustered at the state-level.

5. Empirical Strategy and Results:

We need to understand the mechanisms connecting changes in unemployment rate to the expiration of federal unemployment benefits. Specifically, how gender plays a role between exhausting UI and unemployment. I use two specficiations of statistical models to explain how sex plays a role in unemployment rates once states are off federal programs. The following effects I find are consistent with Albanesi et al (2021). Females have slower "recovery" rates in periods of economic recovery than males. A recovery period defined in this study is defined as periods after UI is terminated.

It is important to reiterate how I do not have individual data on who is on unemployment insurance or not, however, I do have individual data on who is unemployed. My OFFUI300 variable indicates when an individual is in a state with federal unemployment benefits at the specific month rather than when the individual is enrolled in the UI benefits.

5.1 Regression -Two Way Fixed Effect:

I will use a two-way fixed effect model to see how being in a state during periods with federal UI benefits impacts labor market activity. My model takes the general form below:

$$Y_{it} = \beta_0 + \beta_1(OFFUI300_{it}) + \beta_2(Female_i) + \beta_3(OFFUI300_{it}\, X\, Female_i) + \gamma_t + \rho_i + \sigma X_{it} + \epsilon_{it}$$

Y_{it} is an individual *i*'s employment outcome at month *t* within the year of 2021. The independent variable will take a value of 1 if unemployed and 0 if employed. β_1 is the difference in unemployment rate between men on and off unemployment insurance. β_2 represents the marginal effect of being female for being in states when UI is offered.

My coefficient of interest is β_3, an interaction of two dummy variables: access to federal unemployment insurance and female. β_3 represents the marginal effect of being female and off UI, or in other words how transitioning off UI will impact females differently than males. A positive value for β_3 implies that females saw a smaller decrease in unemployment compared to males after transitioning out of unemployment insurance.

The month fixed effect, γ_t, is the indicator for each month of the year within 2021. This controls for variations caused by shocks in different months in 2021. The state fixed effect, ρ_i, is an indicator for each of the 50 states in my sample. This controls for different state level policies, other than UI300, that would influence unemployment levels. A rich set of controls, σX_{it}, account for the individuals age, race, education and industry of current or most recent employment. The purpose of these controls is to remove other possible observable explanations for change in labor market activity throughout the pandemic. My data is at the individual level where we are informed of the individual's state of residency, thus we are able to aggregate unemployment up to a state level unemployment rate to use state fixed effects. ϵ_{it} is the stochastic error term, representing variation not explained by my model, most noticeably state-level policies.

The OFFUI300 dummy variable is 1 for all periods when an individual was in a state without extra UI benefits. Female is a dummy variable taking on a value of 1 if the individual classified themselves as a female and 0 if not.

The focus of the model is to isolate how labor market outcomes changed among males compared to females as federal benefits exhausted. The coefficient β_3, should theoretically be zero if females and males were equally affected when UI programs expired. My alternative hypothesis is that the β_3 coefficient is not equal to zero[9].

5.2 Two Way Fixed Effect Results:

In this section, I look at the regression results from the two way fixed effect difference-in-difference model. In table 5, I run the model with and without controls.

[9] In particular I believe the coefficient to take on a value greater than zero as increase in female unemployment is expected to be higher than increase in male unemployment explained by the premises of the Duncan model.

Table 5: Two Way Fixed Effect Results

Female Change in Unemployment Compared to Males	No Controls	Controls
After FPUC Withdraw	0.00711***	0.00715***
	(0.00107)	(0.00149)
Observations	702,317	702,317
R-squared	0.002	0.045

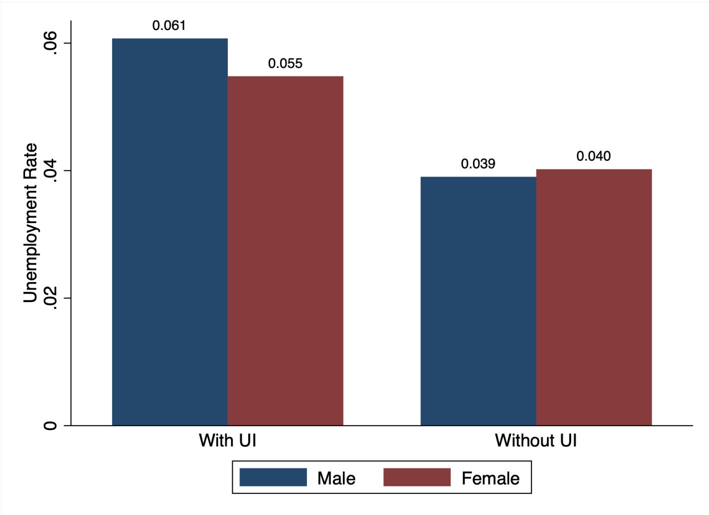
Note: *,**,*** mean significance at the 1, 5, and 10 percent level. Standard Error are clustered at State-level. After FPUC withdraw is the β_3 coefficient from model 1.

When I regress unemployment on an interaction between sex and UI, we can see the expiration of UI impacts female unemployment differently than males. The regression states the difference between female and males off UI and the difference in female and males on UI is 0.715 percentage points, the β_3 coefficient for the regression with controls. This suggests that male unemployment dropped more than the drop in female unemployment by 0.715 percentage points. Although 0.715 percentage points may seem like a small number, relative to the male unemployment of 6.10%, the effect is 11.71% of the male unemployment rate during UI. The effect can be visually represented in Figure 19 and calculated in table 7. With UI, male unemployment was greater than female unemployment, however, without UI, overall unemployment levels dropped, but males dropped more than females to the point where overall male unemployment is lower than overall female unemployment.

Table 6: Calculating Difference in Difference

	Difference in Unemployment (WithOutUI - WithUI)	Difference	Difference in Difference (Female Diff - Male Diff)
Male	0.039-0.061	= -0.022	0.007
Female	0.040-0.055	= -0.015	

Figure 19: Difference in Difference



I combined the effect of coefficients to obtain the overall effect of gender and UI availability status. I tested the three coefficients and tested to see if they were equal to males on UI. Being female, while conditioned on being in a state with UI, generally reduces the chance of being unemployed by 0.706 percentage points compared to males on UI, as seen in table 7 below. However, in our regression, the interaction term assumes going off UI will affect females differently than males. The difference between females and males off UI is 0.005 percentage points. The interaction term's coefficient is 0.711 percentage points, which almost effectively cancels out the female "advantage" during UI. When there is no longer access to federal UI benefits, the gap between men and women closes relative to men. This implies that men may be more motivated to return to work when benefits end, than females are. The overall effects of combining coefficients along with standard errors are displayed below in table 7.

Table 7: Linear Combinations of Coefficient

Overall Effect - Compared to Male & On UI				
	Coefficient	Std. err.	[95% Confidence Interval]	
Male & Off UI	-0.00677***	0.00149	-0.00970	-0.00384
Female & On UI	-0.00706***	0.00086	-0.00876	-0.00537
Female & Off UI	-0.00669***	0.00148	-0.00959	-0.00378

Note: *,**,*** mean significance at the 1, 5, and 10 percent level.

5.3 Two-way Fixed Effect - Heterogeneity Analysis:

I now examine certain characteristics that might influence the magnitude of change in unemployment among females. I do so by including a triple interaction on OFFUI300 x Female with one of the 5 different dummy variables. The coefficient on the triple interaction terms are displayed below in table 8. The regression value of 0.00465 in equation 1 is interpreted as follows: This suggests that female, without children, unemployment dropped more than the drop in female, with children, unemployment by 0.465 percentage points— 11.6% of average female unemployment off UI.

The most notable effect is how interacting income levels affect unemployment change among females, where lower income households saw greater magnitude in change of unemployment than higher income households. More people from low-income households make up UI recipients, as replacement rates will be more attractive to them. Lower wage females experience a 1.18 percentage point change compared to females in high income households, where they only experience a .0912 percent point in change. The difference is consistent with theory as lower wage families will tend to have flatter wage slopes available to them once UI expires, ultimately choosing the household production curve rather than employment in the Duncan Model. Having greater leisure and household production time will be more beneficial for women who face lower wage jobs in the labor market.

Table 8: Heterogeneity Analysis

Change in Unemployment Compared to Females	Children	Children Under 5	Bottom 20th Percentile	Top 20th Percentile	Married
Children	0.00465**				
	(0.00205)				
Children Under 5		0.00530			
		(0.00328)			
Bottom 20th Percentile			.0118***		
			(0.00451)		
Top 20th Percentile				0.000912	
				(0.00212)	
Married					0.00435**
					(0.00162)
Observations	702,317	702,317	702,317	702,317	702,317
R-squared	0.045	0.045	0.045	0.045	0.046

Note: *,**,*** mean significance at the 1, 5, and 10 percent level. Standard errors are clustered at the state-level.

Children's age also impacts female unemployment rates once UI expires- where females with children 5 and under have a greater increase compared to having children in general. There is a great demand for time and resources when families have a child under 5. Higher levels of time investments can be reflected by a more convex household production curve or a flatter wage slope– a reflection of change in opportunity cost. When federal UI benefits expire for everyone, the percentage of unemployment among females with children under 5 decreased by .530 percentage points less than the change in unemployment among females with no children under 5. The standard error for the coefficient is relatively high compared to the estimate, however, this may be a result of the under-represented portion of females with children under 5 in our dataset.

Married individuals have a noticeably lower difference in unemployment rate compared to the average effect among all females. Previous studies point to single adults staying with their parents during Covid-19, contributing to a greater unemployment number (DePaulo, 2018). This effect is prominent among unmarried women, as they have historically dealt with greater unemployment levels and longer unemployment duration (Boushey 2010).

Table 9: Income Sensitivity Analysis

Sensitivity Analysis Difference from Female Average	Bottom 10th Percentile	Bottom 20th Percentile	Top 40th Percentile	Top 20th Percentile
Bottom 10th Percentile	0.0180**			
	(0.00718)			
Bottom 20th Percentile		0.0118***		
		(0.00451)		
Top 40th Percentile			0.00538***	
			(0.00145)	
Top 20th Percentile				0.000912
				(0.00212)
Observations	702,317	702,317	702,317	702,317
R-squared	0.037	0.035	0.024	0.021

Note: *,**,*** mean significance at the 1, 5, and 10 percent level.

Income has a prominent effect on how individuals react when going off UI. We can see that females in the bottom 10th percentile of family income see their unemployment levels increase 1.8 percentage points higher than the average female off UI. This effect decreases in magnitude as we go towards the top 20th percentile. The top 20th percentile of family income has an insignificant coefficient of .09 percentage points at the 10% level. Unfortunately, we cannot model trends prior and after treatment with this specification.

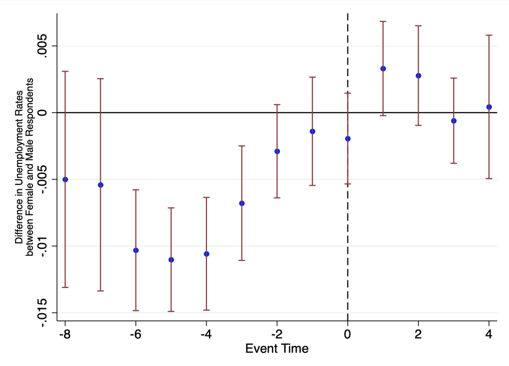
5.4 Event Study Specification:

Mentioned earlier, states experience the withdrawal of unemployment benefits at different times. A majority of the states were treated in June, 3 in July, and the rest in September, therefore we have a staggered treatment effect. An advantage of using an event study regression as opposed to the earlier specification above is that I can disaggregate the beta coefficient effect and can see the magnitude and trend for differences in unemployment prior and after expiration of federal UI benefits.

$$Y_{it} = \alpha + \sum_{\tau=-8}^{4} \beta_{\tau}(FEMALE_{i} X\, MONTH_{t}) + \gamma_{t} + \rho_{i} + \sigma X_{it} + \epsilon_{it}$$

In the equation above, I have information regarding months from 8 months prior to UI expiration to 6 months after expiration. I plot the difference between female and male unemployment along with their standard errors below in my event study results. Similar to my first specification, I have state and month fixed effects. The coefficients are displayed in regression result tables along with their standard errors. The main regression with controls is in table 10 in appendix A as well as graphical representation in figure 20.

Figure 20 : Event Study Main Regression with Controls



Note: Standard Errors are shown above with the 95% confidence interval graphed above. This is the same for all graphs

In the main event study of regression with controls, we do not have a stagnant trend before the event date, rather we can see an upward trend in differences as we approach the event date and onwards. The difference in unemployment becomes significant at the 10% level 1 month after UI exhausts. The results suggest that during periods when states have access to federal unemployment insurance benefits, male unemployment rates were greater than female unemployment rates, the most notable magnitude difference being 6-4 months before UI exhaustion. However, as we approach the event date, the difference becomes smaller and eventually becomes positive once UI expires, suggesting that female unemployment rates gradually increased to a point greater than males once past UI expiration.

5.5 Event Study Heterogeneity Analysis:

Figure 21: Event Study Children

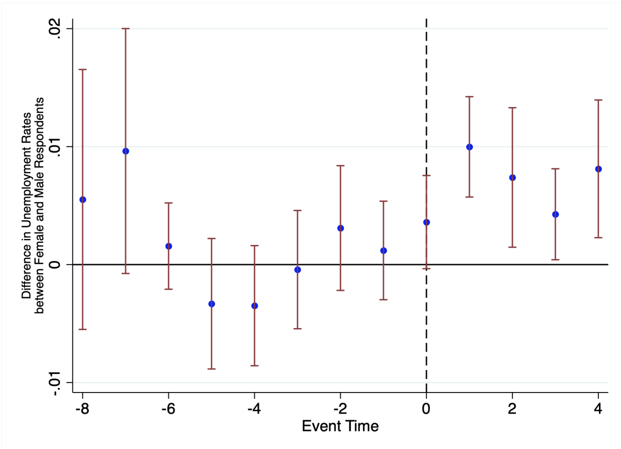


Figure 22: Event Study Children Under 5

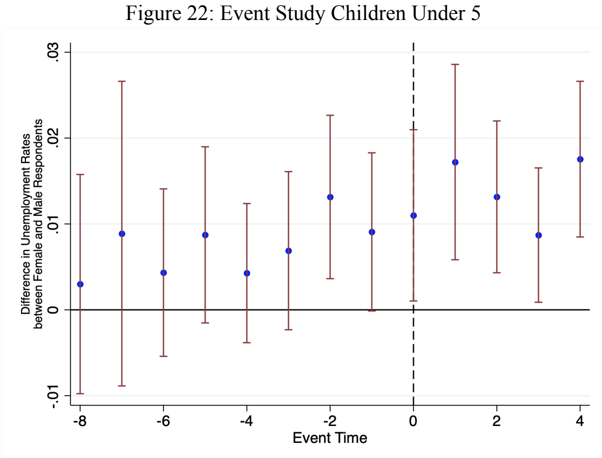


Figure 23: Event Study Bottom Family inc

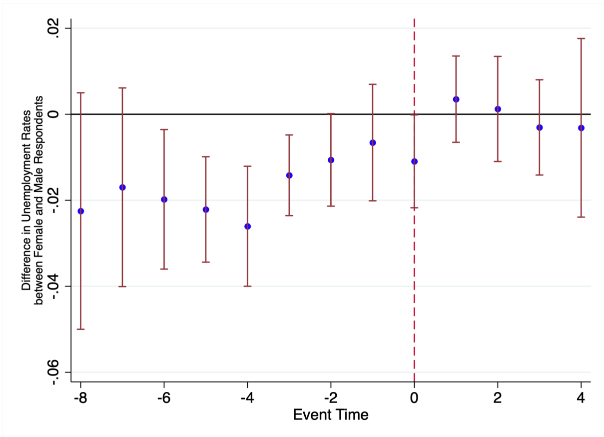


Figure 24: Event Study Top Family inc

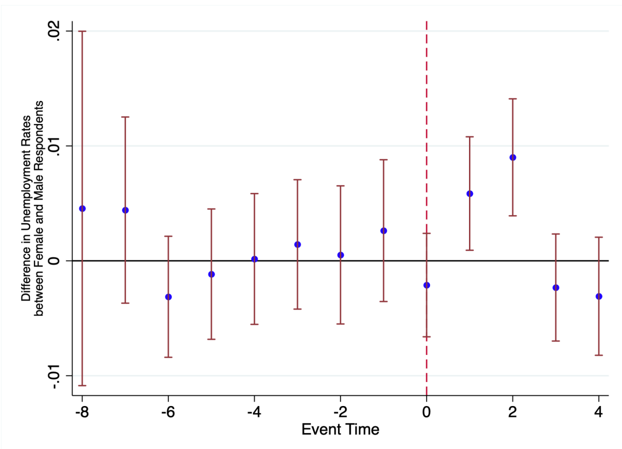
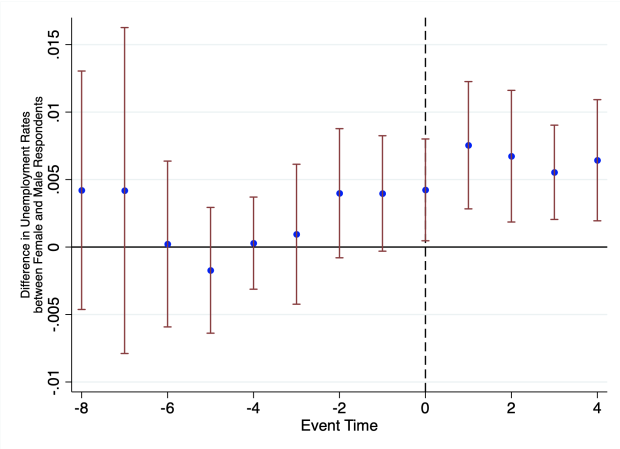


Figure 25: Event Study Married Individuals



Using the same heterogeneous analysis approach as our fixed effect specification, we can break down interesting trends around UI expiration for specific subpopulations. The coefficients for all heterogeneous regression models are displayed in appendix A table 11.

Among the regressions run on children, children under 5, and married individuals, all four months after UI withdrawal lead to a significant difference between female and male unemployment at the 5% level. In the context of the Duncan model, individuals with children, especially under the age of 5, tend to have an unproportional burden of household production and unpaid care placed on females. In married households, stereotypes and roles are more prevalent. Persistent social norms of mothers as primary caregivers contributing to different production functions between males and females. Due to different structural changes in the labor-leisure tradeoff, females will experience a smaller flow of unemployment to employment once UI is unavailable.

Within lower and higher income households, interesting patterns of differences emerge once UI benefits expire. Lower income households have a near 0 difference in unemployment among females and males, however, higher-income households have significantly greater female unemployment. If you are unemployed around expiration date and the other person makes a high level of income, you are more likely to be a female. On the other hand, if you are unemployed and the other person makes a low-level of income, you are equally likely to be female or male.

5.6 Limitations:

Although I have information regarding which states have UI, I do not have information regarding who has UI. Therefore I am capturing the effect of an individual being in a state that offers federal benefits rather than an individual enrolled in the program. Since I do not have a measure of UI at state-level , we can only regress unemployment on an indicator if the state is offering federal benefits. I do not have any data on whether the individuals in my sample are aware of the policy changes and are responding specifically to them. Therefore, it is difficult to infer direct causality between FPUC expiration and unemployment because I assume the majority of people in the state are aware of policy changes and act accordingly because of the policy change.

Using a fixed effect approach, it is difficult to parse out heterogeneous treatment effects due to the difficulty accounting for how states react to the policy differently. If the expiration of federal UI benefits has effects that are constant across states and over months, I will then have an unbiased effect. However, it is unlikely that UI programs impacted all states the same way. For example, without the CARES Act, Massachusetts has weekly unemployment benefits of $823 while Missisispi's weekly benefits are around $235. It is a fair assumption that the extra $300 dollars a week will have a much greater impact on unemployment rates in Mississippi than it would in Massachusetts (Chaisemartin et al. 2019). Future research should limit the empirical approach to states with similar characteristics to control for heterogeneous treatment effects.

Another drawback for both regressions would be endogeneity caused by state specific shocks, as there are always state specific policies that affect unemployment, such as shelter in place orders and minimum wage. States may potentially implement policies related to unemployment during the pandemic that may be a significant driver in variation of unemployment within the state. For example, daycares may have opened up in the state of Michigan in the month of June, the same date as the state's FPUC withdrawal. The event study model will not be able to subtract out the effect of daycares opening on unemployment.

The states that withdrew first also tend to be republican states. Future research should implement a placebo test. What if the government had announced the withdrawal date earlier? We can then question if individuals exhibit the same return to work and spending behavior. A placebo test will often allow us to observe the differential trends in each state of a person who is unemployed. Fixed effects can only get us so far to control for different state governing policies.

6 Discussion:

There has still been a wide-debate with different papers drawing conflicting conclusions as to whether UI programs disincentivize employment. I use the Duncan model to explain how differences in wage and household production can dictate whether an individual returns to work after being on government income. Throughout the pandemic, females have dealt with the longstanding problem regarding wage gap and disproportional unpaid work load. In addition to these disparities, industries highly represented by females were most affected by the nature of the pandemic. COVID-19 evidently negatively impacted females greater than males, as we can measure via labor market outcomes.

How do UI withdrawals impact men and women differently during COVID? From our empirical results above, it seems that during periods of FPUC and UI, both males and females tend to experience high unemployment rates. However, UI was offered during the downturns of the economy, thus may not be the real driver of low unemployment rates— although a body of literature points to controversies in quantifying the extent to which UI motivates unemployment rates. After UI programs expired, both male and female unemployment dropped.

The drop in unemployment was driven by males, as male's drop in unemployment was greater by 0.7 percentage points. This effect is greatest among families with combined income lower than 150,000. Families with lower income may be more motivated to apply for UI. Furthermore, in the context of the Duncan model, households with low income have less luxury in terms of dealing with home production in addition to a flatter wage slope, both of which unproportionately burden females. Therefore as states withdraw from UI, females in lower income households see a substantially smaller decrease in unemployment compared to males. The results of my study are consistent to those of Coombs et al (2021) and Heggeness et al (2021).

Females suffered a greater unemployment rate initially and a slower "recovery" as benchmarked by unemployment compensation expiration. COVID-19 is often referred to as a "SHESESSION" due to the nature of overrepresentation of females in high-contact and low-flexibility jobs. Mentioned throughout my paper, a great deal of this disproportionate decrease in unemployment roots from the fact that females have different wages and unpaid care workloads than males. As necessary when striving closer to gender employment equality, we must support social norms encouraging females to pursue stable and skillfully demanding jobs. We should also encourage households to proportionally split unpaid care, as this will relieve major burden suppressing females. With a higher wage and proportionate household work, we can strive for gender parity.

Appendix A: Event Study Coefficients

Table 10: Event Study Main Analysis

Difference in Unemployment Male - Female	No Control	Controls
6 Months Before Benefits End	-0.00923***	-0.0103***
	(0.00217)	(0.00231)
5 Months Before Benefits End	-0.0105***	-0.0110***
	(0.00204)	(0.00198)
4 Months Before Benefits End	-0.00992***	-0.0106***
	(0.00219)	(0.00216)
3 Months Before Benefits End	-0.00562**	-0.00679***
	(0.00218)	(0.00219)
2 Months Before Benefits End	-0.00178	-0.00289
	(0.00174)	(0.00178)
1 Month Before Benefits End	0.000616	-0.00140
	(0.00212)	(0.00207)
Month of Benefits End	-0.000538	-0.00194
	(0.00182)	(0.00173)
1 Month After Benefits End	0.00468**	0.00331*
	(0.00185)	(0.00180)
2 Months After Benefits End	0.00373*	0.00278
	(0.00203)	(0.00191)
3 Months After Benefits End	-0.000125	-0.000606
	(0.00161)	(0.00163)
4 Months After Benefits End	0.00136	0.000432
	(0.00269)	(0.00274)
Observations	702,317	702,317
R-squared	0.020	0.045

Note: *,**,*** mean significance at the 1, 5, and 10 percent level. Standard Errors are clustered at state-level.

Table 11: Event Study Heterogeneity

Difference in Unemployment Male - Female	(1) Children	(2) Children Under 5	(3) Bottom 20th Percentile	(4) Top 20th Percentile	(5) Married
6 Months Before Benefits End	0.00157	0.00434	-0.0198**	-0.00313	0.000225
	(0.00187)	(0.00498)	(0.00828)	(0.00269)	(0.00313)
5 Months Before Benefits End	-0.00332	0.00873	-0.0221***	-0.00116	-0.00172
	(0.00282)	(0.00523)	(0.00625)	(0.00289)	(0.00238)
4 Months Before Benefits End	-0.00348	0.00428	-0.0260***	0.000160	0.000288
	(0.00260)	(0.00413)	(0.00712)	(0.00290)	(0.00174)
3 Months Before Benefits End	-0.000425	0.00689	-0.0142***	0.00143	0.000949
	(0.00256)	(0.00470)	(0.00479)	(0.00287)	(0.00265)
2 Months Before Benefits End	0.00310	0.0131***	-0.0106*	0.000514	0.00399
	(0.00270)	(0.00485)	(0.00549)	(0.00307)	(0.00244)
1 Month Before Benefits End	0.00120	0.00908*	-0.00659	0.00263	0.00397*
	(0.00213)	(0.00470)	(0.00691)	(0.00315)	(0.00218)
Month of Benefits End	0.00360*	0.0110**	-0.0109*	-0.00211	0.00424**
	(0.00201)	(0.00509)	(0.00551)	(0.00230)	(0.00192)
1 Month After Benefits End	0.00998***	0.0172***	0.00351	0.00586**	0.00754***
	(0.00217)	(0.00580)	(0.00512)	(0.00252)	(0.00241)
2 Months After Benefits End	0.00739**	0.0132***	0.00123	0.00901***	0.00673***
	(0.00302)	(0.00451)	(0.00623)	(0.00260)	(0.00249)
3 Months After Benefits End	0.00427**	0.00870**	-0.00305	-0.00232	0.00554***
	(0.00197)	(0.00399)	(0.00565)	(0.00238)	(0.00178)
4 Months After Benefits End	0.00812***	0.0175***	-0.00315	-0.00308	0.00643***
	(0.00298)	(0.00462)	(0.0106)	(0.00262)	(0.00229)
Observations	295,695	79,970	143,942	140,343	369,261
R-squared	0.050	0.079	0.024	0.016	0.030

Note**:** *,**,*** mean significance at the 1, 5, and 10 percent level. Standard Errors are clustered at state-level.

References

Albanesi, S., & Kim, J. (2021). *The Gendered Impact of the COVID-19 Recession on the US Labor Market.* ().https://10.3386/w28505 https://www.nber.org/papers/w28505

Alon, T., Doepke, M., Olmstead-Rumsey, J., & Tertilt, M. (2020). THE IMPACT OF COVID-19 ON GENDER EQUALITY. *National Bureau of Economic Research,* https://www.nber.org/system/files/working_papers/w26947/w26947.pdf

Alon, T., Doepki, M., Olmstead-Rumsey, J., & Tertilt, M. (2020). THIS TIME IT'S DIFFERENT:

THE ROLE OF WOMEN'S EMPLOYMENT IN A PANDEMIC RECESSION. https://www.nber.org/system/files/working_papers/w27660/w27660.pdf

Altonji, J., Contractor, Z., Finamor, L., Haygood, R., Lindenlaub, I., Meghir, C., O'Dea, C., Scott, D., Wang, L., & Washington, E. (2020). Employment Effects of Unemployment Insurance

Generosity During the Pandemic. https://tobin.yale.edu/sites/default/files/files/C-19%20Articles/CARES-UI_identification_vF(1).pdf

American Time Use Survey Summary. (2021). U.S. Bureau of Labor Statistics. https://www.bls.gov/news.release/atus.nr0.htm

Bartik, A., Bertrand, M., Lin, F., Rothstein, J., & Unrath, M. (2020). Measuring the labor market at the onset of the

COVID-19 crisis. *Brookings,*

Bourne, R., & Partin, E. (2021). *Evidence of the Risks of Elevated Unemployment Insurance Benefits.* CATO Institute. https://www.cato.org/blog/evidence-risks-elevated-unemployment-insurance-benefits

Boushey, H. (2010, May 10,). Unmarried Women Continue to See High Unemployment in April. https://www.americanprogress.org/article/unmarried-women-continue-to-see-high-unemployment-in-april/

Buchwald, E. (2021). *Some states are cutting unemployment payments to push people back to work --- are extra benefits really keeping Americans out of the labor force?* MarketWatch. https://www.marketwatch.com/story/montana-offers-1-200-to-get-people-back-to-work-but-are-extra-unemployment-benefits-really-keeping-americans-out-of-the-labor-force-11620407967

Chalney, R. (2020). *This is how COVID-19 could change the world of work for good.* World Economic Forum. https://www.weforum.org/agenda/2020/04/here-s-how-coronavirus-has-changed-the-world-of-work-covid19-adam-grant/

Chetty, R. (2005). WHY DO UNEMPLOYMENT BENEFITS

RAISE UNEMPLOYMENT DURATIONS?

MORAL HAZARD VS. LIQUIDITY. *Nber,* https://www.nber.org/system/files/working_papers/w11760/w11760.pdf

Chodorow-Reich, G., Coglianese, J., & Karabarbounis, L. (2018). THE MACRO EFFECTS OF UNEMPLOYMENT BENEFIT

EXTENSIONS: A MEASUREMENT ERROR APPROACH.

Coombs, K., Dube, A., Jahnke, C., Kluender, R., Naidu, S., & Stepner, M. (2021). Early Withdrawal of Pandemic Unemployment Insurance:

Effects on Earnings, Employment and Consumption. https://files.michaelstepner.com/pandemicUIexpiration-paper.pdf

Dias, F. A., Chance, J., & Buchanan, A. (2020). The motherhood penalty and The fatherhood premium in employment during covid-19: evidence from The united states. *Research in Social Stratification and Mobility, 69*, 100542. https://10.1016/j.rssm.2020.100542

Dube, A. (2021). *Aggregate Employment Effects of Unemployment Benefits During Deep Downturns: Evidence from the Expiration of the Federal Pandemic Unemployment Compensation.* ().https://10.3386/w28470 https://www.nber.org/papers/w28470

Farooq, A., Kugler, A. & Muratori, U. (2021). *The impacts of unemployment benefits on job match quality and labour market functioning.* Vox eu CEPR. https://voxeu.org/article/unemployment-benefits-job-match-quality-and-labour-market-functioning

Foster, S. (2021). *Survey: 55% Expecting To Search For A New Job Over The Next 12 Months.* Bankrate. https://www.bankrate.com/personal-finance/job-seekers-survey-august-2021/

Fry, R. (2022, January 14,). Some gender disparities widened in the U.S. workforce during the pandemic. https://www.pewresearch.org/fact-tank/2022/01/14/some-gender-disparities-widened-in-the-u-s-workforce-during-the-pandemic/

Fukui, M., Nakamura, E., & Steinsson, J. (2018). *Women, Wealth Effects, and Slow Recoveries.* ().https://10.3386/w25311 https://www.nber.org/papers/w25311

Ganong, P., Greig, F., Liebeskind, M., Noel, P., Sullivan, D., & Vavra, J. (2021). Spending and Job Search Impacts of

Expanded Unemployment Benefits:

Evidence from Administrative Micro Data. https://bfi.uchicago.edu/wp-content/uploads/2021/02/BFI_WP_2021-19.pdf

Ganong, P., Noel J., P., & Vavra S., J. (2020). US Unemployment Insurance Replacement Rates During the Pandemic. *National Bureau of Economic Research,*

Goolsbee, A., & Syverson, C. (2020). *Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020.* ().https://10.3386/w27432 https://www.nber.org/papers/w27432

Gould, E., Zipperer, B., & Kandra, J. (2020, April 15,). Women have been hit hard by the coronavirus labor market: Their story is worse than industry-based data suggest. https://www.epi.org/blog/women-have-been-hit-hard-by-the-coronavirus-labor-market-their-story-is-worse-than-industry-based-data-suggest/

Grossbard, A. S. (2003). A MODEL OF LABOR SUPPLY, HOUSEHOLD PRODUCTION, AND

MARRIAGE. *Advances in Household Economics,* https://econoflove.sdsu.edu/documents/model_for_duncan_w_graphs.pdf

Hagedorn, M., Manovskii, I., & Mitman, K. (2015). THE IMPACT OF UNEMPLOYMENT BENEFIT EXTENSIONS ON EMPLOYMENT:

THE 2014 EMPLOYMENT MIRACLE?

S, M., & Suri, P. (2021). Telework, Childcare, and Mothers' Labor Supply | Opportunity & Inclusive Growth Institute. *Minnesota Federal Reserve Bank,* https://www.minneapolisfed.org:443/research/institute-working-papers/telework-childcare-and-mothers-labor-supply

Howell, D., & Azizoglu, B. (2011). Unemployment Benefits and Work Incentives:

The U.S. Labor Market in the Great Recession.

Hupkau, C., & Petrongolo, B. (2020). Work, Care and Gender during the COVID-19 Crisis*. *Fiscal Studies, 41*(3), 623-651. https://10.1111/1475-5890.12245

Kochhar, R. (2011, July 6,). Two Years of Economic Recovery: Women Lose Jobs, Men Find Them. https://www.pewresearch.org/social-trends/2011/07/06/two-years-of-economic-recovery-women-lose-jobs-men-find-them/

Landivar, L. C., Ruppanner, L., Scarborough, W. J., & Collins, C. (2020). Early Signs Indicate That COVID-19 Is Exacerbating Gender Inequality in the Labor Force. *Socius, 6*, 2378023120947997. https://10.1177/2378023120947997

Lee, D. (2021). *Women put careers on hold during COVID to care for kids. They may never recover.* Los Angeles Times. https://www.latimes.com/politics/story/2021-08-18/pandemic-pushes-moms-to-scale-back-or-quit-their-careers

MacLellan, L.*American men did more housework in 2020, but still not as much as women.* Quartz.

https://qz.com/2036973/women-did-more-housework-during-the-pandemic-according-to-us-data/

Madgavkar, A., White, O., Krishnan, M., Mahajan, D., & Azcue, X. (2020). COVID-19 and gender equality: Countering the regressive effects. *McKinsey,* https://www.mckinsey.com/featured-insights/future-of-work/covid-19-and-gender-equality-countering-the-regressive-effects

Marinescu, I. E., & Skandalis, D. (2019). *Unemployment Insurance and Job Search Behavior.* (). Rochester, NY: https://10.2139/ssrn.3303367 https://papers.ssrn.com/abstract=3303367

Meyer, B. (1988). Unemployment Insurance And Unemployment Spells. *National Bureau of Economic Research,* https://www.nber.org/papers/w2546

Røed, K., & Zhang, T. (2003). Does Unemployment Compensation Affect Unemployment Duration?*. *The Economic Journal, 113*(484), 190-206. https://10.1111/1468-0297.00086

Rosenberg, E. (2022, January 7,). U.S. economy added just 199,000 jobs in December, before labor market confronted omicron surge. *Washington Post* https://www.washingtonpost.com/business/2022/01/07/december-jobs-report-unemployment/

Tracking the COVID-19 Economy's Effects on Food, Housing, and Employment Hardships. Center on Budget and Policy Priorities. https://www.cbpp.org/research/poverty-and-inequality/tracking-the-covid-19-economys-effects-on-food-housing-and

Unemployment Insurance in the Wake of the Recent Recession . (2012). (). https://www.cbo.gov/sites/default/files/cbofiles/attachments/11-28-UnemploymentInsurance_0.pdf

Yeyati, E., & Filippini, F. (2021). Social and Economic Impact of COVID-19. https://www.brookings.edu/wp-content/uploads/2021/06/Social-and-economic-impact-COVID.pdf

Zhang, T., & Roed, K. (2002). Does Unemployment Compensation Affect Unemployment Duration? *The Economic Journal,*

The Impact of Floods on Early Childhood Development:

Evidence from Senegal

Cheikh Fall

1. Introduction

Climate change has caused a rapid increase in the number of natural disasters and extreme weather events occurring around the world. As of 2020, 2.2 billion people (29 percent of the world's population) are vulnerable to a 1-in-100-year flood event (Rentschler & Salhab, 2020). Over the past 25 years, the number of droughts and floods, which are prominent examples of climate change, has doubled (Delesalle, 2019). On top of the increased frequency, the severity of such natural disasters is expected to become more unpredictable (Groppo & Kraehnert, 2015).

Although these weather events affect most of the globe, a disproportionate amount of these shocks occur in countries of the Global South (Hanna & Oliva, 2016). Moreover, it is the poorest people in the Global South countries that are the most exposed to extreme weather events. Rentschler & Salhab (2020) estimate that at least 71 million people in Sub-Saharan Africa live under both extreme poverty and significant flood risk.

Hanna & Oliva (2016) argue that the most impacted by the increased recurrence and severity of these weather shocks are children in Global South countries. As these children are more and more exposed to extreme shocks, concerns about short and long terms effects of climate change arise.

Moreover, extreme weather events often take the form of heavy rainfall. Due to the already fragile state of infrastructure in countries in the Global South, heavy rainfalls result in floods that can affect one village up to a whole region. Kim (2010) claims that these floods can have prolonged impact on human development through two main channels. On one hand, they destroy essential infrastructures such as houses, roads, pipes, schools, hospitals. On the other hand, they also affect households' spending allocation and long-term investment in human capital due to reduced income (Kim 2010). This helps shape my research question as I wonder how the occurrence of floods affects early childhood development, through the use of outcomes of interest such as height-for-age, weight-for-age and weight-for-height.

Senegal is well suited for this analysis. The Centre for Research on the Epidemiology of Disasters (CRED) records 8 floods between 2007 and 2016. These floods can be particularly severe as they can affect up to 264,000 individuals according to CRED's database. The origin of these floods is the occurrence of unexpected heavy rainfall, giving me the opportunity of exploiting the exogenous variation of these weather shocks.

This paper contributes to the literature that investigates the effects of exogenous weather shocks on household and children. Aguilar & Vicarelli (2011) use floods related by the El Niño Southern Oscillation in order to investigate its effects on early child development. The authors find that children that were exposed to the shock have worse physical development (lower height, lower weight) than those who were not exposed to the shock (Aguilar & Vicarelli, 2011). This is in line with the prominent work by Maccini & Young (2009), who exploit variability in early life rainfall to establish the effect this exogenous variable on long term life outcomes for Indonesian women. They also find that negative early life environmental conditions result in lower health, educational and socioeconomic outcomes several decades after these weather events (Maccini & Young, 2009).

The mechanisms uncovered by these studies are similar: weather shocks negatively affect the income of poor households, resulting in contractions of consumption (Aguilar & Vicarelli, 2011). Further Aguilar & Vicarelli (2011) determine that households smoothen consumption by first cutting non-food consumption, then shifting their diet towards cheaper foods. Maccini & Young (2009) uncover a similar effect as they find that infant girls are particularly vulnerable to these weather and income shocks. Like the results of Aguilar & Vicarelli (2011) suggest, consumption smoothing mechanisms were not enough to shield households from reducing their non-food consumption and the quality of their food consumption (Maccini & Young, 2009). The lower health, schooling and nutritional statuses of children impacted by these weather shocks are the drivers of their socioeconomic status during adulthood (Maccini & Young, 2009).

Another study conducted by Duque et al. (2019) uses a difference-in-difference specification paired with a regression discontinuity design in order to investigate how early life weather and income shocks such as conditional cash transfers (CCTs) affect subsequent human capital investment (education investment in this case). CCTs, which are positive income shocks, that happen in early childhood have high returns when they occur with "normal" weather conditions for the children (Duque et al., 2019). However, when they come relatively later into childhood, this effect is much smaller (Duque et al., 2019). These findings suggest that children affected by weather shocks are less likely to enjoy subsequent positive income shocks as much as their unaffected peers because they need to "catch up" to their peers with part of the positive income shock. This confirms previous research (Maccini & Young, 2009; Aguilar & Vicarelli, 2011) because it shows that the weather shocks had a significant blow on children in poor households, which has repercussions on their later life.

So far, the body of literature discussed has mainly used large longitudinal household survey data in order to study the effect of weather shocks on households and small children.

In contrast, Kim (2010) uses a cross sectional data from Mongolia, Cameroon and Burkina Faso to determine the effect of weather shocks on educational attainment. Due to the unavailability of panel data, Kim (2010) choses to use educational attainment by age group in a single cross-section. Although the author admits that this method lacks robustness, he finds evidence that extreme climate events have long-term negative impacts on educational attainment, which aligns with findings of Maccini & Young (2009). The difference-in-difference specification used of Kim (2010) is standard in the literature because it is also used by Groppo & Kraehnert (2015) and Duque et al. (2019). However, this specification allows Kim (2010) to present a benchmark figure of welfare loss by the weather events.

Despite what may seem as a consensus of the effect of weather events on household consumption and welfare, Garbero & Muttarak (2013) bring a more nuanced argument. Their study investigates the differential effects of droughts and floods on village attainment in rural Thailand. Using panel data and a difference-in-difference specification like most of the literature, Garbero & Muttarak (2013) show that, on average, there is no negative effect on consumption, investment in agriculture and education in the short term. The authors argue that the average household in their sample is able to smooth consumption enough in order to keep food and nonfood expenditures unchanged, which includes education spending (Garbero & Muttarak, 2013), which directly contradicts the findings of Aguilar & Vicarelli (2011). Garbero & Muttarak (2013) argue that although non-trivial, their results are also supported by some literature (Dercon & Krishnan, 2000; Davies, 2010). Another notable contribution of this study is that it shows that education decreases vulnerability to weather shocks. Garbero & Muttarak (2013) find that positive externalities of education reduce the negative effect of weather shocks by enabling communities to overcome the shocks. Education also reduces vulnerability to these shocks as it develops human capital, which is essential to increasing income large enough to withstand to smooth consumption post-disaster (Garbero & Muttarak, 2013).

This paper fills gaps in the literature that are left by the study of floods' effect on early childhood development. The empirical research of the impact of floods is often paired with other natural disasters such as droughts or earthquakes (Kim, 2010; Duque et al., 2019; Garbero & Muttarak, 2013) or often studied indirectly through variation in rainfall (Maccini & Young, 2009). I argue that the particular effect of floods on early development needs to be studied in order to get a full picture of the effect of this type of weather shock on the most vulnerable (small children from poor households) in Global South countries. In the context of Senegal, understanding how its most common weather shock (floods) affect early childhood development has important policy implication, motivating this study. I also contribute to the literature by exploring differential impacts of floods on urban and rural areas.

This study shows that floods negatively affect child development by promoting past growth failure (long-term malnutrition) as exposed children have lower height-for-age z-scores than non-exposed children. These results are mostly driven by children in urban areas as there is no evidence of an effect on children living in rural areas. Moreover, I find evidence to confirm that this effect is through poorer nutrition, but the health and employment mechanism channels remain empirically ambiguous. The results of the effect of floods on children's height-for-age are robust to several checks. However, there may be bias in these estimates driven by population changes due to selective migration.

This paper proceeds as follows. Section 2 lays out some economic theory and hypothesizes how floods may affect early childhood development. Section 3 describes the various data sources used for this analysis including the household survey data and the flood data sources. Section 4 outlines my empirical strategy and proposes summary statistics about the data. Section 5 presents the results of my regressions. Then, section 6 discusses the results of this study and outlines its limitations. Section 7 concludes. The references and appendix can be found in Sections 8 and 9, respectively.

2. Theoretical Framework

I use the agricultural household model to describe the effect of a flood on exposed rural households. This model is appropriate to describe floods' impact on Senegalese rural households because of the large share of the rural population engaged in subsistence farming. Indeed, the Food and Agricultural Organization (FAO) estimates that 60% of the rural population of Senegal directly depends on farming for their livelihoods (2021).

Consider a rural agricultural household that has to optimize its production of agricultural products, consumption of goods and how much labor they offer in the labor market. Let's assume that this household's utility function depends on three factors: C, their level of consumption of goods; l, the amount of non-labor activities; and z being some household-specific characteristics. The utility function U can be written as the following:

$$U = U(C, l, z) \tag{1}$$

There are three levels of constraints that apply to this household. The first is a production constraint. This production constraint that is dependent on L, the amount of on-farm labor used, and A, the fixed quality of the land. This constraint can be written as:

$$Q = Q(L, A) \tag{2}$$

Secondly, there is a time constraint *T* facing the household, dependent on the amount of family labor *F* and the amount of non-labor activities *l*. This constraint is described by:

$$T = F + l \tag{3}$$

Finally, there is a budget constraint that is written as:

$$P_A Q + wF - wL = P_A C \tag{4}$$

where P_A represents the price of agricultural products the household buys and *w* is the wage labor earns. I rewrite this budget constraint as follows:

$$P_A(Q - C) + w(F - L) = 0 \tag{5}$$

Let us now combine equations (2), (3) and (5) to get one full income constraint equation that a rural household faces:

$$P_A Q(L, A) - wl + wT - wL = P_A C \tag{6}$$

The optimization problem that this household now faces is described by a maximization of its utility function *U=U(C,l,z)* subject to the full income constraint in equation (6). The Lagrangian equation that follows summarizes this optimization problem:

$$\mathcal{L} = U(C, l, z) - \lambda[P_A C - P_A Q(L, A) + wL - wT + wL] \tag{7}$$

I solve for first order conditions of the Lagrangian to get the equations:

$$\mathcal{L}_C = \frac{\partial U}{\partial C} - \lambda P_A = 0 \tag{8}$$

$$\mathcal{L}_l = \frac{\partial U}{\partial l} - \lambda w = 0 \tag{9}$$

$$\mathcal{L}_L = \frac{\partial Q}{\partial L} \lambda P_A - \lambda w = 0 \tag{10}$$

$$\mathcal{L}_\lambda = P_A C - P_A Q(L, A) + wL - wT + wL = 0 \tag{11}$$

I combine equations (8) and (9) to get equation (12) and I also rewrite equation (10) in equation (13) as follows:

$$\frac{\partial U / \partial l}{\partial U / \partial C} = \frac{w}{P_A} \tag{12}$$

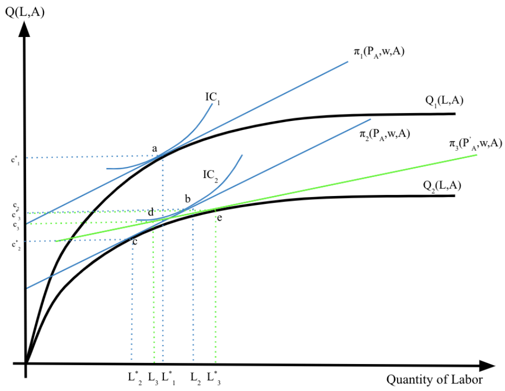
$$\frac{\partial Q}{\partial L} = \frac{w}{P_A} \tag{13}$$

Thus, the condition that needs to hold for the household to optimize its consumption is that the marginal rate of substitution between non-labor activities and consumption and corresponds to the real wage. However, the more useful condition is equation (13), showing that the marginal product of labor should be equal to the real wage for an optimal production for the household.

Let there be a flood affecting this household. On one hand, this shock affects the production function as floods render crop land useless, may sicken livestock and the labor working on it due to stagnant waters and the disease agents they may host (FAO, 2021). Thus, the shock decreases the production function of the household. On the other hand, a flood may also increase local prices (e.g. P_A), leading to a rotation of the profit function (budget constraint).

Figure 1 shows the effect of a flood on a subsistence farming household. Initially, at point a, the farm optimizes production by using only using family labor as on-farm labor (L^*_1) and thus optimizes consumption at c^*_1. Due to the flood, the production function shrinks from Q_1 to Q_2. This leads to a lower optimal on-farm labor (L^*_2). Thus, to optimize utility (consumption at c_2), the family has to offer some off-farm labor ($L_2 - L^*_2$) and buy the difference between their production and consumption ($c_2 - c^*_2$) with the wage from the off-farm labor. This household thus becomes a net buyer of agricultural goods and consumes at a lower consumption level at c_2.

Figure 1: Agricultural household model with weather shock



On top of a lower consumption level, another adverse effect of floods stems from the fact that floods are often covariate shocks and other farms across region may also been affected in the same fashion. Thus, the demand for labor cannot meet off-farm labor supply that affected households offer, leading to unemployment. The optimization condition equation (13) would not hold, resulting in market disequilibrium and suboptimal levels of consumption and production by the agricultural farm.

Let's now explore what occurs when prices increase due to the shock. The profit line π_3 shows the scenario where there is no price stickiness. In this case, this household needs to hire labor outside of family labor in order to produce at point e. Thus, they become a net seller of agricultural goods. This scenario occurs because real wages decrease due to the increase in price levels, making it worthwhile for the household to hire more labor. However, this causes the labor market to be in disequilibrium since the demand for labor exceeds the supply for labor, assuming households in the rural labor market are similar. The market could revert to the profit function π_2 due to real wages increases after this disequilibrium. However, this scenario is highly dependent of labor market functioning well in rural areas, which is often not the case (Kydd & Dorward, 2004). Therefore, the impact of changes in price due to the flood on unemployment is unclear. However, the flood unambiguously causes a decrease in productive capacity, income and/or consumption of goods, as confirmed by the literature (Aguilar & Vicarelli, 2011).

Reduced household consumption could lead to a decrease in the consumption of goods that are essential to child growth. The resulting malnutrition or undernutrition has an especially severe effect on health outcomes when it happens in early childhood. Therefore, I expect the flood to worsen childhood development in local areas.

On the other hand, the agricultural household cannot be applied to urban areas because the average household is not engaged in subsistence farming. A classic optimization of their utility function subject to a budget constraint. Thus, in contrast with the rural household, the urban household only needs to optimize its utility function. I expect the same effect of floods as incomes shrink (shift of budget constraint), resulting in decreased levels consumption and leading to worse levels of nutritional and health outcomes. In summary, I expect an adverse effect of floods on both rural and urban areas.

3. Data Description

The data analyzed in this paper comes from two major sources. The first source is the Demographic and Health Survey (DHS). The DHS data includes 7 continuous surveys conducted in Senegal from 2012 to 2019. These surveys constitute repeated cross-sectional data, that sample the population of Senegal in order to infer characteristics of the greater population. I use the children recode version of the DHS data, which records the survey responses of children aged 0 to 5 years old at the time of the interview. The DHS data is the source of the dependent variable (health and nutrition outcomes) investigated in this analysis and control variables, whether they are characteristics of the children or the households. In total, the combined dataset contains approximately 60,000 observations (children) across the 8 survey years.

Summary statistics can be found in table 1. Most variables have similar means over the years, suggesting little to no change over time. However, I note that the number of years of education for the mother fluctuates more than other variables. Nevertheless, this is not a concern since the standard error of the mother's number of years of education is very large compared to the mean.

Table 1: Summary Statistics by Survey Year

Variables	2012	2013	2014	2015	2016	2017	2018	2019	All
Child-specific Characteristics									
Male	0.507	0.499	0.501	0.498	0.514	0.512	0.508	0.497	0.505
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Age in months	28.14	28.01	28.78	28.36	28.75	28.68	28.81	28.06	28.47
	(16.99)	(17.11)	(17.09)	(17.13)	(17.24)	(17.32)	(17.44)	(17.32)	(17.22)
Height in cm	83.74	83.61	84.11	83.58	84.20	84.64	84.47	83.99	84.09
	(14.76)	(14.37)	(14.28)	(14.37)	(14.63)	(14.38)	(14.38)	(14.66)	(14.46)
Weight in kg	13.09	12.03	12.16	11.63	11.88	11.02	11.03	10.91	11.64
	(14.61)	(10.95)	(10.59)	(9.170)	(9.677)	(3.328)	(3.350)	(3.361)	(8.623)
Birth weight in kg	3.124	3.065	3.085	3.107	3.100	3.090	3.094	3.042	3.086
	(0.745)	(0.711)	(0.700)	(0.732)	(0.675)	(0.694)	(0.659)	(0.631)	(0.693)
# years of education mother	1.898	1.531	1.678	1.800	1.891	2.257	2.235	2.388	1.970
	(3.391)	(2.976)	(3.124)	(3.248)	(3.378)	(4.065)	(3.709)	(4.375)	(3.600)
Household-specific Characteristics									
Male head of household	0.851	0.804	0.796	0.782	0.768	0.772	0.785	0.777	0.789
	(0.256)	(0.397)	(0.403)	(0.413)	(0.422)	(0.420)	(0.411)	(0.416)	(0.408)
# of children under 5	3.787	3.788	3.682	3.722	3.556	3.600	3.582	3.427	3.642
	(2.622)	(2.356)	(2.458)	(2.576)	(2.353)	(2.522)	(2.458)	(2.525)	(2.483)
# household members	15.01	14.85	14.50	14.57	14.33	14.68	14.76	14.03	14.60
	(8.970)	(7.993)	(8.287)	(9.145)	(8.433)	(8.641)	(8.645)	(8.566)	(8.573)
Urban	0.356	0.260	0.308	0.271	0.273	0.336	0.289	0.298	0.298
	(0.479)	(0.439)	(0.460)	(0.447)	(0.445)	(0.472)	(0.453)	(0.452)	(0.457)
Sample size	5132	8592	6842	6935	6725	12185	6719	6125	59255

Note: The estimates presented above are the means of the variables per survey year and standard deviations are shown in parentheses.

My second data source is the Emergency Events Database (EM-DAT). The EM-DAT data was collected by the World Health Organization (WHO) in collaboration with the Centre for Research on the Epidemiology of Disasters (CRED). This database records natural disasters that happen around the world and characteristics of them such as where and when they occur, the type of disaster, how many people were affected etc.

Using the EM-DAT dataset, I find that individuals (children 0-5) in the DHS survey data have been exposed to a total number of 6 floods. I define exposure to a flood using two conditions: being born less than 280 days (pregnancy duration) after the end of the flood and living in a region exposed by the flood. I use children that satisfy the contemporary condition but who do not live in the exposed region as a control group. Table 6 (in appendix) provides details on the floods investigated. I note that some floods affect many regions simultaneously while others may only affect one region. However, defining exposure to floods in this manner assumes that the intensity of the flood is the same across a treated region, which might not be the case.

The primary outcome variables of interest are height-for-age z-scores, weight-for-age z-scores and weight-for-age z-scores. These z-scores are calculated using the WHO standards of child growth. These outcomes allow for an estimation of the development of the child, as they are often used to define malnutrition, including undernutrition (WHO, 2021).

Thus, the child health outcomes are accurate measures of the physical development of a child. Indeed, low height-for-age measures is considered "stunting", which denotes chronic malnutrition and is an indicator for past growth failure. Moreover, "underweight" is defined by a child weighing less than expected given their age. Weight-for-age is an indicator of both current and chronic malnutrition. Finally, the term "wasting" is associated with a child not being able to have enough weight for height. Thus, weight-for-height may be an indicator of current nutritional status (INDEPTH, 2008).

Additionally, I explore the data with the figures below. Figure 2 and 3 show how height-for-age z-scores per age cohort compare between children that have and have not been exposed to a flood in-utero, in the rural and urban area of the Dakar region respectively.

Figure 2: Height-for-age z-score vs age, over in-utero exposure in the rural area of the Dakar region in the entire sample

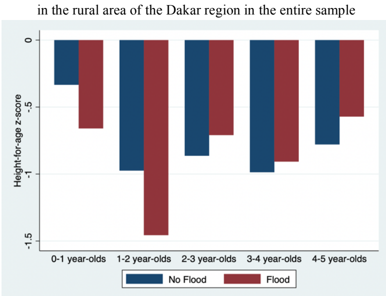
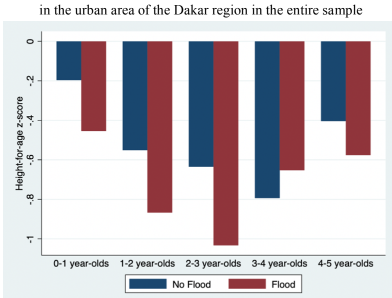


Figure 3: Height-for-age z-score vs age, over in-utero exposure in the urban area of the Dakar region in the entire sample

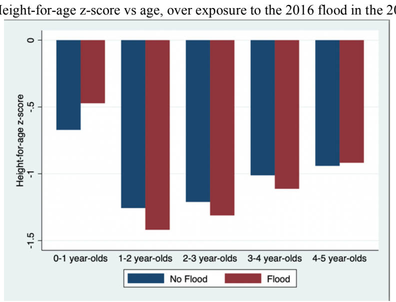


In both bar plots, all z-scores are negative, irrespective of whether the children have been exposed to a flood or not. This suggests that small children in Senegal are on average less physically developed than the average healthy child. Moreover, I note that height-for-age z-score are comparatively worse (more negative) in rural areas than in urban areas.

Figure 2 shows that 0 and 1-year old children exposed to floods in rural areas have worse nutritional outcomes compared to non-affected children. However, this observation does not hold true for 2, 3 and 4-year old children. This contrasts figure 3 which shows that, in urban areas, exposed children have worse height-for-age z-scores than non-exposed children, for all age cohorts expect 3-year old children. I conclude that, for the Dakar region, there is more evidence of a negative effect of floods on height-for-age z-scores in urban areas compared to rural areas.

Furthermore, figure 4 attempts to visualize the impact of only the 2016 flood on height-for-age z-scores:

Figure 4: Height-for-age z-score vs age, over exposure to the 2016 flood in the 2019 survey



This graphic only considers the 2019 survey data and thus compares cohorts that were exposed to the flood in-utero (2 and 3-year-old children), cohorts that were born after the flood (0 and 1-year old children) and a cohort that was born before the flood (4-year-old children). I find that there is some evidence of the adverse impact of floods on height-for-age z-scores before cohorts that were exposed in-utero or born a year after the flood have lower height-for-age than non-exposed children of the same age. Additionally, there is barely any noticeable difference between these two groups for children that were born before the flood. This suggests that the lower height-for-age z-scores for the exposed children compared to non-exposed children can solely be attributed to the flood.

4. Empirical Strategy

4.1. Fixed Effects

Firstly, I use a specification that employs a range of fixed effects in order to determine a causality between flood occurrence and childhood development.

This first identification strategy slightly modifies the one of Agüero (2014). This specification aims to quantify the in-utero effect of each of the 6 floods of interest (table 6 in appendix). The model is specified in the following equation:

$$Y_{ijctrm} = \beta_0 + \beta_1 F_{i,j}^{in\ utero} + \mathbf{X}_i \phi + \gamma_c + \delta_r + \varphi_t + \theta_{r*m} + \epsilon_{ijctrm} \quad (14)$$

In equation (14), *Y* is a child outcome (e.g. height-for-age z-score) of child i, $F^{in\ utero}$ whether they are exposed in-utero to a flood j, **X** a vector of control variables, γ_c a birth cohort fixed effect, δ_r a region fixed effect, φ_t a survey year fixed effect and θ_{r*m} a region-month fixed effect to account for seasonality in the outcomes of interest.

The control variables used in equation (14) are the sex of the child, the size of the household, the sex and age of the household head, the number of years of education of the mother, whether the child lives in an urban or rural setting, whether the child lives with their mother and the level of wealth of the family. The level wealth of the family is measured using an index present in the DHS dataset, classifying the families into 5 categories: poorest, poorer, middle and richer. As theorized in section 2, I expect β_1 to be negative for each flood investigated.

I argue that causal identification is possible with OLS and fixed effects because floods are random and unexpected. The unpredictability of floods randomly assigns exposure to floods to different individuals living in different regions. Thus, I can recover estimates of the effect of flood j on childhood development.

4.2. Difference-in-difference

I also use a difference-in-difference specification to estimate the effect of the floods on childhood development. I now define exposure to the flood by considering children living in regions affected by the flood, not just in-utero exposure. Moreover, to use a difference-in-difference estimation, I only consider the 2016 flood to recover estimates. I make this choice because other floods are too close to each other. Since the last flood before 2016 is the 2013 flood, I can take advantage of two years of pre-treatment and thus isolate the effect of the 2016 flood. The following equation describes this specification:

$$Y_{itcrm} = \beta_0 + \beta_1 F_{it} + \beta_2 after_{it} + \beta_3 D_{it} + \mathbf{X}_i \phi + \gamma_c + \delta_r + \varphi_t + \theta_{r*m} + \epsilon_{itcrm} \quad (15)$$

In equation (15), *F* denotes exposure to the 2016 flood, *after* is an indicator variable for the period after the flood, *D* is the interaction term between *F* and *after* denoting difference-in-difference variable of interest while the control variables and the fixed effects are defined the same as in equation (14).

A difference-in-difference estimation is appropriate to measure the effect of the 2016 flood on the outcome variables of interest because it differences out the fixed effects, accounting for the uniqueness of the individuals/regions, and common time effects, accounting for exogenous shocks or effects that may have affected both treatment and control groups during the period of interest.

There might be concerns about using a difference-in-difference specification, given that I use a repeated cross-sectional dataset rather than panel dataset. However, I argue that a difference-in-difference is still suitable because I assign treatment to regions of Senegal, which are themselves longitudinal. This is a similar method used by numerous empirical studies when dealing with "pseudo-panel" DHS data (Kotsadam & Tolonen, 2015; Benshaul-Tolonen, 2018; Anti & Salemi, 2021).

Similar to OLS, I expect that β_3 in equation (15) has a significantly negative sign, showing an adverse effect of flood exposure on child health and nutritional outcomes.

4.3. Heterogeneity Analysis

As discussed in the theoretical framework of this problem, there are different models that could explained the effect of a flood in rural and urban areas. Thus, I test this heterogenous effect hypothesis by carrying out a triple-difference as follows:

$$\begin{aligned} Y_{itcrm} &= \beta_0 + \beta_1 F_{it} + \beta_2 after_{it} + \beta_3 D_{it} + \beta_4 urban_{it} + \beta_5 (after \times urban)_{it} \\ &+ \beta_6 (F \times urban)_{it} + \beta_7 (D \times urban)_{it} + \mathbf{X}_i \phi + \gamma_c + \delta_r + \varphi_s + \theta_{r*m} + \epsilon_{itcrm} \quad (16) \end{aligned}$$

In equation (16), I build on equation (15) by interacting *D*, the interaction variable between the flood exposure variable and the time variable, with a variable *urban* that denotes whether the child lives in an urban or rural area. This interaction yields *D*×*urban,* which represents the triple difference variable of interest.

It should also be noted that the vector of control variables **X** does no longer contain the *urban* variable like in equation (15), as it is now used in the triple difference.

4.4. Mechanism Analysis

In addition to the main difference-in-difference regression, I test whether there are certain channels through which the effect of floods on child development occurs. With the DHS data available, I identify three main channels of interest: the nutrition channel, the health channel and the employment channel.

Firstly, I use three outcome variables to test the nutrition channel: whether the child has recently been given milk products (e.g. milk, yoghurt), whether the child has recently been given fruits abundant in vitamins and whether the child has recently been given vegetables. Secondly, the health channel is investigated by looking at the effect of the flood on two outcome variables: whether the child has had fever and whether the child coughed in the two weeks prior to the interview. Lastly, I define the employment channel with 2 binary variables of whether the father or mother are employed at the time of the interview.

I note that all the dependent variables used to test these three channels are dichotomous. Thus, one could argue that a logit regression is most appropriate rather than the linear model used in previous specifications. Indeed, a simple idea would be to establish causality by combining the logit estimator with the difference-in-difference specification in equation (15). The result is shown in equation (17) below:

$$log\, odds(Y_{itcrms}) = \beta_0 + \beta_1 F_{it} + \beta_2 after_{it} + \beta_3 D_{it} + \mathbf{X}_i \phi + \gamma_c + \delta_r + \varphi_s + \theta_{r*m} + \epsilon_{itcrms} \quad (17)$$

The logit model restricts predictions of the outcome variables between 0 and 1, which contrasts the linear estimation as predictions might exceed 1 or be lower than 0.

However, the econometric literature points out that the interpretation of interaction terms, such as the variable of interest *D*, in non-linear regressions is often misleading (Ai & Norton, 2003). Moreover, with a difference-in-difference specification, a logit regression model would violate the parallel trends (Lechner, 2010).

Thus, I use a linear probability model with a difference-in-difference method (same specification as equation (15)). This model is not as accurate as the logit model. However, since I am only concerned with causal inference and the difference-in-difference variable of interest, it is still appropriate. Indeed, it allows us to retain interpretability of the coefficients estimates and does not compromise the parallel trends assumption of the difference-in-difference method (Lechner, 2010).

I expect there to be a negative effect of floods on all three mechanisms. The theory (section 2) shows that there should be an adverse effect on employment and/or income of at least one of the heads of the household. This would lead to lower purchasing power in nutritious food, essential to the child's growth. Moreover, the flood may also lead to worse child health due to the diseases that it may convey. Therefore, I also expect an increase in the occurrence of diarrhea and/or fever for exposed children compared to unexposed children.

5. Econometric Results

5.1. Fixed Effects

Fixed effect results are presented in Table 1. Each panel presents regressions ran for each anthropometric outcome. The regressions presented follow the specification as laid out in equation (14). In the first 8 columns, I estimate the effect of in-utero exposure to each flood by comparing the treatment group (in-utero during the flood and in affected regions) and the control group (in-utero during the flood, but in non-affected regions).

Table 2: Fixed effects regressions results of the Effect of In-Utero Exposure to Floods on Child Development

	Dependent Variables: Height-for-age, Weight-for-age and Weight-for-height z-scores								
Variables	2007 Flood	2008 Flood	2009 Flood	2010 Floods	2011 Flood	2012 Flood	2013 Flood	2016 Flood	Any Flood
A. Height-for-age z-score									
In-utero exposure	-0.319	0.774	0.888	0.221*	-0.189	-0.604	-0.573	0.746	0.027*
to flood	(0.586)	(0.662)	(1.394)	(0.128)	(0.484)	(0.592)	(0.466)	(0.523)	(0.160)
Birth Cohort FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Survey Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	529	2041	2650	4243	4677	5765	5016	4557	50472
B. Weight-for-age z-score									
In-utero exposure	-0.195	-0.432	0.102	0.232**	-0.230	-0.777	-0.219	-0.623	-0.018
to flood	(0.452)	(0.523)	(0.111)	(0.104)	(0.412)	(0.517)	(0.408)	(0.466)	(0.138)
Birth Cohort FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Survey Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	529	2041	2650	4243	4677	5772	5024	4581	50558
C. Weight-for-height z-score									
In-utero exposure	-0.010	-0.150***	0.090	0.155	-0.195	-0.432	0.147	0.257	-0.056***
to flood	(0.451)	(0.525)	(1.094)	(0.106)	(0.636)	(0.539)	(0.409)	(0.475)	(0.014)
Birth Cohort FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Survey Year FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	529	2041	2650	4243	4678	5780	5037	4569	50524

Notes: All regressions include extended controls as outlined in equation (14). Their estimated coefficients are omitted. *** p<0.01, ** p<0.05, * p<0.1. Some fixed effects (birth cohort and survey year) are omitted for the 2007 flood because of a lack of groups in the sample along those fixed effects.

I find mixed results on how in-utero exposure to floods has an effect on childhood development. I note that the different floods studied differ drastically in the direction of the relationship: some floods increase the mean child height-for-age, weight-for age and/or weight-for-height z-scores while others decrease these outcome variables, holding all other variables constant. Interestingly, I find a cyclical pattern in this effect. For instance, the 2007 and 2008 flood show that floods have a negative effect on child development whereas the 2009 and 2010 floods suggest the other direction. The same observation can be made by looking at individual outcomes. I note that, for the weight-for-age outcomes, the direction of the relationship is negative in 2007 and 2008, positive in 2009 and 2010, negative again in 2011, 2012, 2013 and 2016. Similar patterns can be seen in the two other outcomes.

A possible explanation for these results is that OLS with fixed effect may not an appropriate specification to study floods that occur in back-to-back years. For example, 6 control regions for the 2008 flood become treated for the 2009 flood. If I assume that the 2008 flood had a negative effect on the outcomes of interest and that effect persists for a year, it is plausible to see that the treatment regions (that were control regions in 2008) have higher child development z-scores. This scenario also depends on the 2009 flood having a smaller negative impact than the 2008 flood, but might be the reason why I note these patterns.

A second reason for the cyclical nature of the direction of the relationship is that OLS only considers the differences between treatment groups in the same year, without accounting for their levels in the years before. Thus, causal identification of the flood's effect on childhood development is hard to make if I am only comparing the groups in one year.

This latter argument is the main reason why I get these mixed results because, although flood treatment is random, the groups that I am comparing are different more often than not. Thus, an empirical strategy that allows for controlling for pre-treatment levels would allow stronger causal identification.

5.2. Difference-in-difference

I address the concerns in the OLS specification by employing a difference-in-difference identification strategy. I solve the first concern of successive floods by only considering the 2016 flood in the difference-in-difference. The second concern is solved by the difference-in-difference method because it accounts for the trends because the weather event and thus, only infers causality if there are changes to the treatment group trend, relative to the control group.

An in-utero treatment framework is not compatible with a difference-in-difference specification because I need to be able to measure the mean z-scores before the weather event. I expand treatment to signify an exposure to the flood between the ages 0 to 5 years old. This allows me to use a difference-in-difference method, even if the available data is of repeated cross-sectional nature. Table 3 shows the results of this specification.

Table 3: Difference-in-difference Regression Results of the Effects of General Exposure to Flood

	Dependent Variable		
Variables	Height-for-age z-score	Weight-for-age z-score	Weight-for-height z-score
After 2016	0.094***	-0.039	-0.134
	(0.029)	(0.025)	(0.025)
Region Treat	0.112	-0.213	-0.389
	(0.029)	(0.171)	(0.172)
DID	-0.100***	-0.030	0.042
	(0.034)	(0.030)	(0.030)
Extended Controls	Yes	Yes	Yes
Birth Cohort FE	Yes	Yes	Yes
Survey Year FE	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Region-month FE	Yes	Yes	Yes
Observations	39094	39180	39146

Notes: All regressions include extended controls as outlined in equation (15). *** p<0.01, ** p<0.05, * p<0.1.

I find that the 2016 flood caused a decrease of the height-for-age z-score by 0.1 standard deviations from the mean healthy height-for-age according to WHO standards. This result is significant at the 1% level. However, I do not find any significant effect of the flood on weight-for-age or weight-for-height z-scores. This suggests that this flood has an effect on long-term nutrition for young children, while there is no evidence that there is no effect on the current nutritional status (weight-for-height z-scores) of children at the time of the survey. Thus, I form the hypothesis that the flood may have had a sustained effect immediately after it occurs. However, this sustained effect may have decreased over time, explaining why there isn't evidence of current malnutrition due to the flood at the time of the interview for the survey.

In other words, I find that children exposed to the flood are more severely stunted than they were before the flood compared to non-exposed children, while there is no evidence of the same effect on the wasting or under/overweight status of children. These findings hold water because it is known that stunting is largely irreversible (WHO, 2015). On top of that, previous studies such as del Ninno & Lundberg (2005) and Dimitrova & Muttarak (2020) support the idea that a flood may cause this negative effect on height-for-age z-scores. These findings emphasize the importance of policy interventions immediately after the flood and/or the importance of policies designed to maintain/improve child development before floods to improve development outcomes in the long run. Del Ninno & Lundberg (2005) preconize the latter policy recommendation by showing that ex-post interventions are not as effective as ex-ante programs.

Moreover, I can contextualize the magnitude of the effect depicted by my results by comparing them to previous studies. Del Ninno & Lundberg (2005) find that children exposed to the 1998 flood in Bangladesh are about 0.2 standard deviations smaller, in terms of height-for-age z-scores, than their non-exposed counterparts. This is twice the effect that I find with the 2016 flood in Senegal, but it is hard to compare the floods due to their different levels of intensity. Another perspective is given by Dimitrova & Muttarak (2020), who find that, in the years 2015 and 2016, a standard deviation increase in rainfall anomalies in India is associated with a decrease of 0.1 standard deviations of children's height-for-age z-scores. Thus, the case can be made that my results are of moderate economic significance, as it is comparable to a moderate increase in rainfall anomalies.

5.3. Heterogeneity Analysis

To estimate the differential effects of floods on urban and rural areas, I employ a triple difference method, as specified in equation (16). Table 4 shows the regression results.

I find that, for the height-for-age outcome, I cannot reject the hypothesis that children in rural areas are not affected by the flood, even if there is a decrease in height-for-age z-score of 0.051 standard deviations on average.

However, there is evidence showing that the significant results in the difference-in-difference specification are driven by children living in urban areas. I note that there is a significant (at the 5% level) decrease of height-for-age z-scores by 0.211 standard deviations from the average healthy child. Similar to the difference-in-difference results, I find no significant effect of the flood on children using the other anthropometric indicators.

These results are not clearly explained by the theory laid out in section 2 of this paper. The theory suggested that there would be a significant negative effect of floods on childhood development indicators in both urban and rural areas. I only find these results in urban areas.

The literature that studies the differential effects on urban and rural areas of floods on children's physical development is very thin, which makes it hard to compare my findings with previous empirical studies. Nevertheless, Dimitrova & Muttarak (2020) find the opposite of my results: the risk of stunting for exposed children living in rural areas is greater due to the flood and this effect is not present for children living in urban areas.

Table 4: Triple Difference Regression Results of the Effects of General Exposure to Flood

	Dependent Variable		
Variables	Height-for-age z-score	Weight-for-age z-score	Weight-for-height z-score
After 2016	0.079**	-0.043	-0.132***
	(0.031)	(0.027)	(0.027)
Region Treat	0.099	-0.226	-0.405**
	(0.200)	(0.174)	(0.175)
After*Treat	-0.051	-0.009	0.029
	(0.039)	(0.034)	(0.039)
Urban	0.072**	0.011	-0.046
	(0.033)	(0.284)	(0.029)
After*Urban	0.044	0.012	-0.003
	(0.370)	(0.032)	(0.032)
Treat*Urban	-0.007	0.025	0.042
	(0.560)	(0.049)	(0.049)
After*Treat*Urban	-0.160**	-0.071	0.036
	(0.066)	(0.058)	(0.058)
Extended Controls	Yes	Yes	Yes
Birth Cohort FE	Yes	Yes	Yes
Survey Year FE	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Region-month FE	Yes	Yes	Yes
Observations	39094	39180	39146

Notes: All regressions include extended controls as outlined in equation (16). *** p<0.01, ** p<0.05, * p<0.1.

A potential scenario in which these regression results are plausible is that rural families have more control on knowing how to mitigate floods effects. This could be through planting more water-intensive crops in areas that are more susceptible of their farm to flooding or just to mitigate risk if other non-water-intensive are decimated by floods. This is not an unlikely scenario because, even if floods are random and unpredictable, it is quite recurrent in Senegal and farmers are aware of the effect of a possible flood on their cultures. Another factor that goes in the favor of rural areas is that the water is can penetrate the soil more easily compared to urban areas that are mostly covered by tarmac. Thus, urban households are vulnerable to the actions of the government such as urban planning, water removal system etc. This is because, in the context of countries of the Global South, only the government has enough resources to fund such flood prevention or treatment actions. However, this scenario is not proved by the data analyzed in this study but only speculation about what could explain the results observed.

5.4. Mechanism Analysis

Using a difference-in-difference specification, I explore different mechanisms through which the negative impact of floods on child development may occur. I identify three main channels: a nutrition channel, a health channel and an employment channel. The results of these regressions are shown in table 5.

Table 5: Mechanism Analysis of the Effect of Exposure to 2016 Flood on Child Development

	Nutrition			Health		Employment	
Variables	Gave Child Milk Prod.	Gave Child Fruits	Gave Child Vegetables	Fever in last 2 weeks	Cough in last 2 weeks	Mom working	Dad working
After 2016	0.0169*	-0.004	0.027**	0.041***	0.052	-0.049***	-0.069***
	(0.093)	(0.010)	(0.011)	(0.009)	(0.008)	(0.110)	(0.006)
Region Treat	0.029	0.044	0.037	0.059	0.121**	0.140*	-0.007
	(0.060)	(0.067)	(0.074)	(0.060)	(0.054)	(0.073)	(0.039)
DID	-0.035***	-0.039***	-0.012	-0.008	0.033***	-0.036***	0.041***
	(0.108)	(0.119)	(0.013)	(0.011)	(0.009)	(0.130)	(0.007)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Survey Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26859	26859	26859	40686	40730	41975	37862

Notes: All regressions include extended controls as outlined in equation (15). *** p<0.01, ** p<0.05, * p<0.1.

The rationale behind the nutrition channel is laid out in the theory section of this paper. The explanation of this mediation is that the flood decreases the income of households, leading to lower overall consumption, more specifically lower consumption in goods essential for child growth. In table 3, I find that the nutrition channel is plausible because there is a significant (at the 1% level) decrease by 4% in children consuming milk products and fruits by children living in exposed regions compared to those in other regions, all else equal. Thus, I find evidence that this mechanism causes at least some of the effect of the flood on child development in affected regions.

Secondly, there could be a mediation of the impact of floods on child development through a health channel. It is possible that the floods may have increased the likelihood of children contracting vector-borne diseases (malaria), cholera, diarrhea etc. through the frequent contact with flood water, lack of access to clean water and sanitation (Zermoglio et al., 2015). I find that there is a significant (at the 1% level) increase of coughing rates in the 2 weeks prior to interview by 3% in children exposed to the flood, compared to non-exposed children, all other things equal.

However, I do not see any significant change of the rate of fever in the 2 weeks prior to interview, holding all control variables constant. Fever is arguably the better proxy of flood-related disease because most of the disease aforementioned (malaria, cholera…) are often accompanied with fever and not coughing necessarily. Thus, the negative effect of floods through the mediation of worse child health is not strongly supported by the evidence.

Thirdly, I explore the mediation channel of employment as laid out in section 2 of this paper. Indeed, it is plausible that the floods decrease employment opportunities in the affected regions. This mechanism may result in a reduced ability to buy foods essential to a healthy child nutrition. However, I theorized that this effect may be ambiguous due to uncertainties about how well labor markets function in rural areas. In table 3, I find that there is a reduction in the employment of mothers in affected regions by 3.6% compared to non-exposed regions. This is however contrasted by a greater increase in the employment of fathers in affected regions by 4.1%. Thus, the employment channel mediation of the main effects remains ambiguous.

5.5. Robustness Checks

The first robustness check I perform is a parallel trend check for the main difference-in-difference in equation (15). In table 7, I test the pre-treatment parallel trends assumption of the difference-in-difference framework. I find that parallel trends are violated for the weight-for-age and weight-for-age z-scores outcomes. However, for height-for-age z-scores, the parallel trends variable of interest (interaction term between treatment and year) is only significant at the 10% level. However, with a significance threshold of 5%, I reject the null hypothesis that there are non-parallel trends of height-for-age z-scores in the pre-treatment period. I conclude that the parallel trends assumption is met for height-for-age z-scores.

I perform a second robustness check by changing the definitions of height-for-age, weight-for-age and weight-for-height z-scores. In the main regression results (tables 4 and 5), I use the WHO definitions of child growth (mean and standard deviation). Thus, to check if these results are robust, I use the DHS definitions of child growth with the same models. The results are shown in tables 8 and 9 (in appendix). I find that difference-in-difference and triple difference results are robust. The only difference is that the significance of the difference-in-difference variable of interest (interaction term) decreases to the 10% level, but I still reject the null hypothesis that there is no effect of floods on height-for-age z-score.

Thirdly, I test how sensitive the difference-in-difference and triple difference estimates are to the number of fixed effects used in the specifications. Thus, I drop the fixed effects (birth cohort, survey year, region and region-month) one by one to isolate the sensitivity of my main results to each level of fixed effect. The results of these regressions are shown in table 10 (for the difference-in-difference) and table 11 (for the triple difference), both in appendix. I find that, for the difference-in-difference, the average effect of the 2016 flood varies between -0.087 and -0.104 standard deviations decrease in height-for-age z-scores. Since the effect remains significant, my difference-in-difference results are robust to different specifications of fixed effects. As shown in table 11, the triple difference results are also robust to different specifications of fixed effects.

6. Limitations

The main limitation of this study is that it does not consider selective migration. Indeed, after the occurrence of a flood, affected households and families are likely to be displaced. This scenario could bias my estimates towards zero since most affected households may move out of the affected regions and decrease the treatment effect. Additionally, a flood may make affected regions attractive to poorer households when relatively richer households, that have the means to leave, do so. Thus, there is a resulting change in composition of regions affected by a flood to a poorer population. This may bias my estimates away from zero and increasing the treatment. Therefore, there is some ambiguity about the effect of selective migration on my estimates.

Another limitation of this study is that there is a lack of external validity to my estimates. Since my robust results are only measured through the effect of the 2016 flood in Senegal, there is no guarantee of a similar effect in another country/region of the world. Moreover, there is no measure of the intensity of this particular flood, compared to other weather events.

7. Conclusion

In sum, I construct a multiple cross-sectional dataset from 7 iterations of the Demographic and Health Surveys (DHS) to estimate the effect of floods on anthropometric indicators such as height-for-age, weight-for-age and weight-for-height for children under 5 years old. Firstly, I use a quasi-experimental research design to investigate the effect of 8 floods that happened in Senegal between 2007 and 2016. My initial specification of OLS paired with fixed effects did not yield reliable results because it doesn't allow a clean separation of the effects of floods occurring in successive years.

Thus, I use a difference-in-difference specification with fixed effects to recover estimates by only focusing on 2016 flood. I find that this flood decreases height-for-age z-scores by 0.1 standard deviations. These results are robust. However, I do not find the same effects on weight-for-age and weight-for-height z-scores. I conclude that the effect of this flood translates mainly through past growth failure as affected children have experienced long-term malnutrition compared to non-exposed children. I also find that these effects on height-for-age z-scores are driven by children living in urban areas by using a triple difference specification, who experience a decrease in their height-for-age z-scores by 0.211 standard deviations on average. There is however no evidence that the 2016 flood affected the development of children living in rural areas of affected regions.

Moreover, I find that results are mainly explained through poor nutrition caused by the flood through a mechanism analysis. There is no significant evidence supporting the hypotheses that the flood affects the health (through diseases) or employment of the parents of children in this dataset.

My study has limitations such as migration bias and external validity. Further research could use panel data to overcome these obstacles. Panel data would allow researchers to take selective migration into account. Moreover, panel data would also allow a more extensive mechanism analysis as it would easier to identify what particular channels drive the effect of floods.

8. References

Agüero, J. M. (2014). *Long-Term Effect of Climate Change on Health: Evidence from Heat Waves in Mexico.* Inter-American Development Bank.

Aguilar, A., & Vicarelli, M. (2011). *El Niño and Mexican children: Medium-term effects of early-life weather shocks on cognitive and health outcomes*.

Ai, C., & Norton, E. C. (2003). *Interaction terms in logit and probit models*. Economic Letters, 80, 123–129.

Almond, D., Currie, J., & Duque, V. (2017). *Childhood Circumstances and Adult Outcomes: Act II*. NBER Working Paper, 23017.

Anti, S., & Salemi, C. (2021). *Hungry hosts? Refugee camps and host community nutritional outcomes in sub- Saharan Africa*. https://ageconsearch.umn.edu/record/316035/.

Benshaul-Tolonen, A. (2018). *Local Industrial Shocks and Infant Mortality*. The Economic Journal, pp. 1–32. doi: 10.1111/ecoj.12625.

Davies, S. (2010). *Do shocks have a persistent impact on consumption?: the case of rural Malawi*. Progress in Development Studies 10:75-79. http://dx.doi.org/10.1177/14-6499340901000105

del Ninno, C., & Lundberg, M. (2005). *Treading water: The long-term impact of the 1998 flood on nutrition in Bangladesh*. Economics and Human Biology, 67–96.

Delesalle, E. (2019). *Good or bad timing? The effects of productivity shocks on education and on schooling performance*.

Dercon, S. (2005). *Vulnerability: A Micro Perspective*. Oxford: Oxford University Press.

Dercon, S., and P. Krishnan. (2000). *Vulnerability, seasonality and poverty in Ethiopia*. Journal of Development Studies 36:25-53. http://dx.doi.org/10.1080/00220380008422653

Dimitrova, A., & Muttarak, R. (2020). After the floods: Differential impacts of rainfall anomalies on child stunting in India. Global Environmental Change, 64.

Duque, V., Rosales-Rueda, M., & Sanchez, F. (2019). *How Do Early-Life Shocks Interact with Subsequent Human Capital Investments? Evidence from Administrative Data*.

Food and Agriculture Organization. (2021). *Floods*. FAO in Emergencies. https://www.fao.org/emergencies/emergency-types/floods/en/?page=2&ipp=10&tx_dynalist_pi1[par]=YToxOntzOjE6IkwiO3M6MToiMCI7fQ==

Food and Agriculture Organization. (2021, February 4). *Senegal published the Annual Agricultural Survey 2019-2020 report, providing data and information on the country's agricultural sector*. Agricultural Integrated Survey Programme. https://www.fao.org/in-action/agrisurvey/news-and-events/detail-events/en/c/1372696/

Garbero, A., & Muttarak, R. (2013). *Impacts of the 2010 Droughts and Floods on Community Welfare in Rural Thailand: Differential Effects of Village Educational Attainment*. Ecology and Society, 18(4).

Groppo, V., & Kraehnert, K. (2015). *The impact of extreme weather events on education*. German Institute for Economic Research.

Hanna, M., & Oliva (2016). *Implications of climate change for children in developing countries. The Future of Children*, (1), 115-132. Retrieved from https://www.jstor.org/stable/43755233

INDEPTH. (2008, June 3). Anthropometric status indicators. INDEPTH Resource Kit for Demographic Surveillance Systems.

Jentsch, A., & Beierkuhnlein, C. (2008). *Research frontiers in climate change: Effects of extreme meteorological events on ecosystems*.

Kim, N. (2010). *Impact of Extreme Climate Events on Educational Attainment: Evidence from Cross-Section Data and Welfare Projection. In Risk, Shocks, and Human Development: On the Brink* (1st ed., pp. 185–206). United Nations Development Programme.

Kotsadam, A., & Tolonen, A. (2015). *African Mining, Gender, and Local Employment*. World Bank Policy Research Working Paper Series, 7251.

Kydd, J., & Dorward, A. (2004). *Implications of Market and Coordination Failures for Rural Development in Least Developed Countries*. Journal of International Development, 16(7).

Lechner, M. (2011). The Estimation of Causal Effects by Difference-in-Difference Methods. Foundations and Trends in Econometrics, 4(3), 165–224.

Maccini, S., & Yang, D. (2008). *Under the weather: Health, schooling, and economic consequences of early-life rainfall*. NBER Working Paper, 14031. http://www.nber.org/papers/w14031

Rentschler, J., & Salhab, M. (2020). *People in Harm's Way: Flood Exposure and Poverty in 189 Countries. 2020 Poverty and Shared Prosperity Report*.

Senegal—Regional Map. (2021). GIS Geography. Retrieved October 7, 2021, from https://gisgeography.com/senegal-map/

Tucci, C. (2007). *Urban Flood Management*. World Meteorological Organization and Cap-Net International Network for Capacity Building in Integrated Water Resources Management.

WHO. (2021, June 9). *Malnutrition*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/malnutrition

WHO. (2015, November 19). *Stunting in a nutshell*. World Health Organization. https://www.who.int/news/item/19-11-2015-stunting-in-a-nutshell

Zermiglio, F., Steynor, A., & Jack, C. (2015). *Climate Change and Health Risks in Senegal* [Technical]. United States Agency for International Development (USAID). https://www.climatelinks.org/sites/default/files/asset/document/180327_USAID-ATLAS_Senegal%20Climate%20and%20Health_Final_rev.pdf

9. Appendix

Figure 5: Regional Map of Senegal

Source: GIS Geography (2021)

Table 6: Floods Investigated

Year (Period)	Regions Affected	N Children Exposed in the Sample	N Children in Control Group
2007 (August 9th – September 20th)	Dakar, Diourbel, Kaolack, Matam, Saint-Louis, Tambacounda, Thiès	374	343
2008 (August 15th – September)	Dakar, Diourbel, Kaolack, Saint-Louis, Thiès	1174	1415
2009 (August 24th – August 25th)	Dakar, Fatick, Kaffrine, Kaolack, Kédougou, Kolda, Matam, Saint louis, Sedhiou, Tambacounda, Thiès	2437	764
2010 (First flood: August 12th – August 23rd; Second flood: October 17th – October 18th)	Kolda (first flood), Saint-Louis (second flood)	578	4542
2011 (August 1st – August 1st)	Thiès	447	5097
2012 (August 15th – August 31st)	Dakar, Fatick, Diourbel	1502	5329
2013 (September 13th – September 13th)	Dakar, Fatick, Kaolack, Thiès	1678	4212
2016 (July – September 11th)	Fatick, Kaffrine, Kaolack, Saint louis	1513	3613

Source: Centre for Research on the Epidemiology of Disasters (CRED) for flood information

Table 7: Parallel Trends Test

Variables	Dependent Variable		
	Height-for-age z-score	Weight-for-age z-score	Weight-for-height z-score
Region Treat=1	-0.003 (0.072)	-0.020*** (0.062)	-0.276*** (0.063)
2015	-0.064** (0.027)	-0.161*** (0.023)	-0.168*** (0.024)
Region Treat=1 × Year=2015	0.093* (0.048)	0.160*** (0.041)	0.141*** (0.041)
Extended Controls	Yes	Yes	Yes
Birth Cohort FE	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Observations	12074	12074	12074

Notes: *** p<0.01, ** p<0.05, * p<0.1

Table 8: Robustness check for difference-in-difference using DHS z-scores

Variables	Dependent Variable		
	Height-for-age z-score	Weight-for-age z-score	Weight-for-height z-score
After 2016	0.113***	-0.030	-0.127***
	(0.027)	(0.024)	(0.022)
Region Treat	0.056	-0.241	-0.375**
	(0.187)	(0.166)	(0.172)
DID	-0.087***	-0.030	0.031
	(0.033)	(0.029)	(0.026)
Extended Controls	Yes	Yes	Yes
Birth Cohort FE	Yes	Yes	Yes
Survey Year FE	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Region-month FE	Yes	Yes	Yes
Observations	38885	38885	38973

Notes: All regressions include extended controls as outlined in equation (15). *** p<0.01, ** p<0.05, * p<0.1

Table 9: Robustness check for triple difference using DHS z-scores

Variables	Dependent Variable		
	Height-for-age z-score	Weight-for-age z-score	Weight-for-height z-score
After 2016	0.104***	-0.034	-0.124***
	(0.030)	(0.026)	(0.024)
Region Treat	0.060	-0.249	-0.382**
	(0.189)	(0.169)	(0.153)
After*Treat	-0.052	-0.015	0.019
	(0.037)	(0.033)	(0.030)
Urban	0.084***	0.024	-0.029
	(0.031)	(0.028)	(0.025)
After*Urban	0.028	0.011	-0.009
	(0.352)	(0.031)	(0.028)
Treat*Urban	-0.033	0.011	0.026
	(0.053)	(0.047)	(0.043)
After*Treat*Urban	-0.113*	-0.054	0.038
	(0.063)	(0.056)	(0.051)
Extended Controls	Yes	Yes	Yes
Birth Cohort FE	Yes	Yes	Yes
Survey Year FE	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Region-month FE	Yes	Yes	Yes
Observations	38885	38885	38973

Notes: All regressions include extended controls as outlined in equation (16). *** p<0.01, ** p<0.05, * p<0.1

Table 10: Sensitivity analysis for difference-in-difference by relaxing fixed effects

Variables	Dependent Variable			
	Height-for-age z-score	Height-for-age z-score	Height-for-age z-score	Height-for-age z-score
After 2016	0.104***	-0.126***	-0.094***	0.046*
	(0.030)	(0.020)	(0.029)	(0.025)
Region Treat	0.076	0.136	0.112	0.043
	(0.201)	(0.196)	(0.197)	(0.041)
DID	-0.099***	-0.104***	-0.100***	-0.087***
	(0.035)	(0.034)	(0.034)	(0.035)
Extended Controls	Yes	Yes	Yes	Yes
Birth Cohort FE	No	Yes	Yes	Yes
Survey Year FE	Yes	No	Yes	Yes
Region FE	Yes	Yes	No	Yes
Region-month FE	Yes	Yes	Yes	No
Observations	39094	39094	39094	39094

Notes: *** p<0.01, ** p<0.05, * p<0.1

Table 11: Sensitivity analysis for triple difference by relaxing fixed effects

Variables	Dependent Variable			
	Height-for-age z-score	Height-for-age z-score	Height-for-age z-score	Height-for-age z-score
After 2016	0.091***	0.110***	0.079**	0.026
	(0.032)	(0.023)	(0.031)	(0.027)
Region Treat	0.074	0.123	0.099	0.055
	(0.204)	(0.199)	(0.200)	(0.222)
After*Treat	-0.054	-0.055	-0.051	-0.041
	(0.040)	(0.039)	(0.039)	(0.033)
Urban	0.078**	0.073**	0.072**	0.057*
	(0.033)	(0.033)	(0.033)	(0.031)
After*Urban	0.038	0.046	0.044	0.062*
	(0.379)	(0.370)	(0.370)	(0.035)
Treat*Urban	-0.030	-0.008	-0.007	-0.012
	(0.057)	(0.056)	(0.056)	(0.053)
After*Treat*Urban	-0.145**	-0.158**	-0.160**	-0.163***
	(0.068)	(0.066)	(0.066)	(0.063)
Extended Controls	Yes	Yes	Yes	Yes
Birth Cohort FE	No	Yes	Yes	Yes
Survey Year FE	Yes	No	Yes	Yes
Region FE	Yes	Yes	No	Yes
Region-month FE	Yes	Yes	Yes	No
Observations	39094	39094	39094	39094

Notes: *** p<0.01, ** p<0.05, * p<0.1

The Canonical New Keynesian Model: Considering Sectoral Price Stickiness

Zefan Qian

Contents

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are widely used for monetary policy analysis. Normally, DSGE models assume that goods are identical, which ignores sectoral heterogeneities. However, empirical studies show that sectoral heterogeneities play a big role in many aspects. For example, studies such that Carvalho (2006) [2] and Eusepi et al. (2011) [6] improve their evaluation of the effect of monetary policy by considering the sectoral heterogeneities. Thus, it is worthwhile to develop the canonical New Keynesian model in a multi-sector framework.

Quantitatively, many studies such as Klenow and Kryvtsov (2008) [9] and Nakamura and Steinsson (2008) [10] adopt microdata to estimate sectoral price stickiness for the European areas. However, microdata rarely exist in many countries. Thus, it is important to find a method to estimate the sectoral price stickiness using aggregate data.

Carvalho and Dam (2010) [3] multi-sector sticky-price models for twelve countries to obtain the cross-sectional distribution of price stickiness using only aggregate data. They assume the specification of staggered price setting inspired by Taylor (1979, 1980) [13] [11] that only a fraction of prices are negotiated each period because of the presence of multiperiod nominal contracts. If firms set a price during a period, then the price will remain in place for a fixed number of periods. Their inferred distributions of sectoral price stickiness conform well to empirical distributions constructed from empirical studies using microdata.

This article studies how to employ previous work on price stickiness to characterize the equilibrium. Specifically, following Hou et al. (2011) [8], I assume the economy consists of eight sectors which correspond to eight major groups defined by the BLS of the United States.

2 Modeling

Assume there is one representative household, one perfectly competitive final good firm, J perfectly competitive composite intermediate good firms, and J intermediate good firms. Each intermediate good firm corresponds to one major sector, consisting of a continuum (0,1) of monopolistically competitive firms. In each period, the representative household chooses consumption and labor supply. The j^{th} intermediate good firm hires labors from the representative household to produce intermediate good j, and sells it to the j^{th} composite intermediate good firm. The j^{th} composite intermediate good firm inputs intermediate good from j^{th} intermediate good firm, and sell it to the final good firm at a perfectly competitive price. The final good firm will input all J intermediate goods to produce the final good, and sell it to the representative household. This article categorizes the United States economy into eight sectors, such that $J = 8$. The flow diagram of this model is shown in Figure 1.

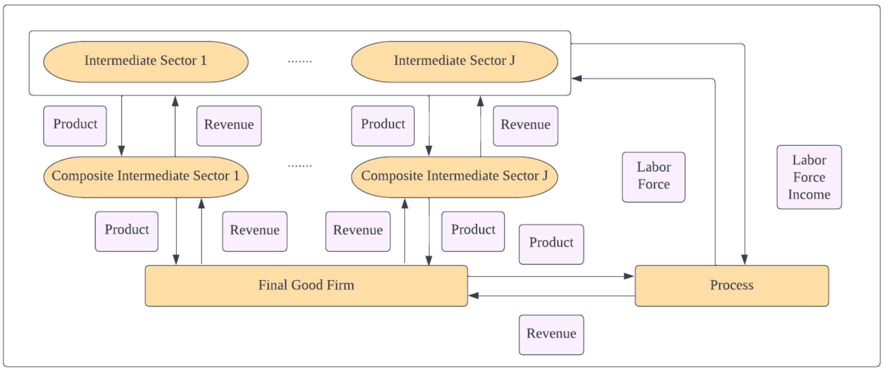


Figure 1: Flow Diagram for Theoretical Model

2.1 Modeling Setup

2.1.1 Representative Household

According to Horvath 2000 [7], since the labor is not perfectly mobile among different sectors, suppose

$$N_t = \left(\sum_{j=1}^{J} \left(\xi^{j,N}\right)^{-1/\zeta} \left(N_t^j\right)^{(\zeta+1)/\zeta} \right)^{\zeta/(\zeta+1)} \tag{1}$$

where N_t^j is the labor supply household provides to j^{th} sector, ζ is the elasticity of substitution of labor force between different sectors, $\xi^{j,N}$ is the percentage of labor income of j^{th} sector. Specifically, since the differentiation of intermediate good's elasticity of substitution and production function are not taken into considerations, $\xi^{j,N} = \xi_j$. This can be log linearized into

$$\hat{N}_t = \sum_{j=1}^{J} \xi^j \hat{N}_t^j \tag{2}$$

Thus, the representative household chooses sequences of consumption C_t, hours worked N_t, and bonds B_t to solve

$$\begin{aligned} \max \quad & E_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\eta}}{1-\eta} - \frac{N_t^{1+\chi}}{1+\chi} \right) \\ \text{subject to} \quad & \sum_{j=1}^{J} N_t^j W_t^j + \frac{R_{t-1} B_{t-1}}{\Pi_t} = C_t + B_t \end{aligned}$$

where the first order condition with respect to C_t is

$$C_t^{-\eta} = \lambda_t, \tag{3}$$

the first order condition with respect to N_t^j is

$$N_t^{\chi} (\xi^j N_t / N_t^j)^{-1/\zeta} = \lambda_t W_t^j, \tag{4}$$

and the first order condition with respect to B_t is

$$\lambda_t = \beta E_t(\lambda_{t+1} R_t / \Pi_{t+1}). \tag{5}$$

Plugging (3) into (4) yields

$$N_t^{\chi} (\xi^j N_t / N_t^j)^{-1/\zeta} = C_t^{-\eta} W_t^j, \tag{6}$$

which can be log linearized into

$$\chi \hat{N}_t + (\hat{N}_t^j - \hat{N}_t)/\zeta = -\eta \hat{C}_t + \hat{W}_t^j, \; j = 1, 2, \ldots, J \tag{7}$$

Also, plugging (3) into (5) yields

$$C_t^{-\eta} = \beta E_t(C_{t+1}^{-\eta} R_t / \Pi_{t+1}) \tag{8}$$

which can be log linearized into

$$-\eta \hat{C}_t + \eta E_t\{\hat{C}_{t+1}\} = \tilde{R}_t - E_t\{\tilde{\pi}_{t+1}\} \tag{9}$$

2.1.2 Final Goods Firm

Intermediate goods are bundled by a final-goods firm that produces the economy's output Y_t. The bundling technology is given by

$$Y_t = \prod_{j=1}^{J} \left(\xi^j\right)^{-\xi^j} \left(Y_t^j\right)^{\xi^j} \tag{10}$$

where $\sum_{j=1}^{J} \xi^j = 1$, which can be log linearized into

$$\hat{Y}_t = \sum_{j=1}^{J} \xi^j \hat{Y}_t^j \tag{11}$$

Plugging composite intermediate good demand function into (10) yields

$$P_t = \prod_{j=1}^{J} \left(P_t^j\right)^{\xi^j} \tag{12}$$

which can be log-linearize as

$$\hat{P}_t = \sum_{j=1}^{J} \xi^j \hat{P}_t^j \tag{13}$$

2.1.3 Composite Intermediate Good Firm

Let Y_t^{ij} denote the output of intermediate good i bought from firm j in period t with the price P_t^{ij}.

$$Y_t^j = \left(\int_0^1 \left(Y_t^{ij}\right)^{\frac{(\psi^j - 1)}{\psi^j}} di\right)^{\psi^j / \left(\psi^j - 1\right)} \tag{14}$$

The profit maximization problem is then,

$$\max_{\left\{Y_t^j\right\}} P_t \prod_{j=1}^{J} \left(\xi^j\right)^{-\xi^j} \left(Y_t^j\right)^{\xi^j} - \sum_{j=1}^{J} P_t^j Y_t^j \tag{15}$$

and

$$\max_{\{Y_t^{ij}\}} P_t \prod_{j=1}^{J} \left(\xi^j\right)^{-\xi^j} \left(\int_0^1 \left(Y_t^{ij}\right)^{\left(\psi^j - 1\right)/\psi^j} di \right)^{\xi^j \psi^j / \left(\psi^j - 1\right)} - \sum_{j=1}^{J} \int_0^1 P_t^{ij} Y_t^{ij} di \tag{16}$$

where we can solve

$$Y_t^j = \xi^j \left(\frac{P_t}{P_t^j} \right) Y_t \tag{17}$$

which can be log linearized into

$$\hat{Y}_t^j = \hat{Y}_t + \hat{P}_t^j \tag{18}$$

and

$$Y_t^{ij} = \left(\frac{P_t^{ij}}{P_t^j} \right)^{-\psi^j} Y_t^j \tag{19}$$

respectively. Plugging (17) into (14) yields

$$P_t^j = \left(\int_0^1 \left(P_t^{ij}\right)^{1-\psi^j} di \right)^{1/\left(1-\psi^j\right)} \tag{20}$$

2.1.4 Intermediate Goods Firm

Using the Calvo (1983) [1] pricing friction, suppose that every period, the probability for j^{th} intermediate goods firm to change their prices is $1 - \theta^j$. Since within the j^{th} intermediate goods firm, the production technology and demand function are the same, there will be an optimal price P_t^{j*} which they would like to change to. Thus, at period t,

$$P_t^j = \left(\theta^j \left(P_{t-1}^j\right)^{1-\psi^j} + \left(1 - \theta^j\right) \left(P_t^{j^*}\right)^{1-\psi^j} \right)^{1/\left(1-\psi^j\right)} \tag{21}$$

Let the inflation of the j^{th} intermediate good sector be $\tilde{\pi}_t^j = \hat{P}_t^j - \hat{P}_{t-1}^j$ such that it can be log linearized into

$$\tilde{\pi}_t^j = (1 - \theta^j)(\hat{P}_t^{j*} - \hat{P}_{t-1}^j) \tag{22}$$

Since intermediate good firm will hire labor from the representative household to produce intermediate goods, suppose that the production function is

$$Y_t^j = N_t^j \tag{23}$$

which can be log linearized into

$$\hat{Y}_t^j = \hat{N}_t^j \tag{24}$$

Let $MC_t^j = W_t^j$, which can be log linearized into

$$\hat{MC}_t^j = \hat{W}_t^j,\ j = 1, 2, \cdots, J \tag{25}$$

the intermediate firm will need to solve

$$\begin{aligned} \max_{P_t^{j*}} \quad & \sum_{k=0}^{\infty} (\theta^j)^k E_t\{Q_{t+k}(P_t^{j*} Y_{t+k}^{ij} - MC_{t+k}^j P_{t+k} Y_{t+k}^{ij})\} \\ \text{subject to} \quad & Y_{t+k}^{ij} = \left(\frac{P_t^{j*}}{P_{t+k}^j}\right)^{\psi^j} Y_{t+k}^j \end{aligned} \tag{26}$$

where $Q_{t+k} = \beta^k (C_{t+k}/C_t)^{-\sigma}(P_t/P_{t+k})$ representing the discount factor of nominal payment, β is the utility discount factor for representative household, σ is the relative risk aversion coefficient. Thus, the first order condition is

$$\sum_{k=0}^{\infty} (\theta^j)^k E_t \left\{ Q_{t+k} Y_{t+k}^{ij} \left(P_t^{j*} - \frac{\psi^j}{\psi^j - 1} MC_{t+k}^j P_{t+k} \right) \right\} = 0 \tag{27}$$

which can be log linearized into

$$\hat{P}_t^{j*} - \hat{P}_{t-1}^j = (1 - \beta\theta^j) \sum_{k=0}^{\infty} (\beta\theta^j)^k E_t\{(\hat{MC}_{t+k}^j - MC_{t+k}^j) + (\hat{P}_{t+k} - \hat{P}_{t-1}^j)\} \tag{28}$$

such that j^{th} intermediate good production sector's NK Phillips curve is

$$\tilde{\pi}_t^j = \beta E_t\{\tilde{\pi}_{t+1}^j\} + \kappa^j \hat{MC}_t^j + \kappa^j \hat{P}_t^j \tag{29}$$

where $\kappa = (1 - \beta\theta^j)(1 - \theta^j)/\theta^j$ which can be further written as

$$\tilde{\pi}_t^j = \beta E_t\{\tilde{\pi}_{t+1}^j\} + \kappa^j \hat{MC}_t^j + \kappa^j \hat{P}_t^j + u_t^j,\ j = 1, 2, \cdots, J \tag{30}$$

and

$$\tilde{\pi}_t = \sum_{j=1}^{J} \xi^j \tilde{\pi}_t^j \tag{31}$$

when we add sectoral inflation stock.

2.1.5 Monetary Policy and Inflation Measure

According to Taylor (1993) [12] and Clarida et al. (2000) [5], the interest rate smoothing is also introduced:

$$\tilde{R}_t = \rho^R \tilde{R}_{t-1} + (1 - \rho^R)(\phi_\pi \pi_t + \phi_Y \hat{Y}_t) + \hat{M}_t \tag{32}$$

2.1.6 Market Clearing

The market clears when

$$Y_t = C_t + G_t \tag{33}$$

which can be log linearized into

$$\hat{Y}_t = \hat{C}_t + \hat{G}_t \tag{34}$$

2.2 Full Model

$$-\eta \hat{C}_t + \eta E_t\{\hat{C}_{t+1}\} = \tilde{R}_t - E_t\{\tilde{\pi}_{t+1}\} \tag{2.2.1}$$

$$\chi \hat{N}_t + (\hat{N}_t^j - \hat{N}_t)/\zeta = -\eta \hat{C}_t + \hat{W}_t^j,\ j = 1, 2, \ldots, J \tag{2.2.2}$$

$$\hat{Y}_t^j = \hat{N}_t^j,\ j = 1, 2, \ldots, J \tag{2.2.3}$$

$$\hat{MC}_t^j = \hat{W}_t^j,\ j = 1, 2, \ldots, J \tag{2.2.4}$$

$$\tilde{\pi}_t^j = \beta E_t\{\tilde{\pi}_{t+1}^j\} + \kappa_j \hat{MC}_t^j + \kappa_j \hat{P}_t^j + u_t^j,\ j = 1, 2, \ldots, J \tag{2.2.5}$$

$$\hat{Y}_t = \sum_{j=1}^{J} \xi^j \hat{Y}_t^j \tag{2.2.6}$$

$$\hat{Y}_t^j = \hat{Y}_t + \hat{P}_t^j \tag{2.2.7}$$

$$\tilde{\pi}_t = \sum_{j=1}^{J} \xi^j \tilde{\pi}_t^j \tag{2.2.8}$$

$$\hat{Y}_t = \hat{C}_t \tag{2.2.9}$$

$$\hat{N}_t = \sum_{j=1}^{J} \xi^j \hat{N}_t^j \tag{2.2.10}$$

$$\tilde{R}_t = \rho^R \tilde{R}_{t-1} + (1-\rho^R)(\phi_\pi \tilde{\pi}_t + \phi_Y \hat{Y}_t) + \hat{M}_t \tag{2.2.11}$$

$$\hat{u}_t^j = \rho^{j,u} \hat{u}_{t-1}^j + \varepsilon_t^{j,u} \tag{2.2.12}$$

$$\hat{M}_t = \rho^M \hat{M}_{t-1} + \varepsilon_t^M \tag{2.2.13}$$

$$\hat{G}_t = \rho^G \hat{G}_{t-1} + \varepsilon_t^G \tag{2.2.14}$$

2.3 Solution Approaches

2.3.1 Mapping to the System of Matrix Equations

First, I present the expressions that go in the first matrix equation

$$0 = \chi \hat{N}_t + (\hat{N}_t^j - \hat{N}_t)/\zeta + \eta \hat{C}_t + \hat{W}_t^j,\ j = 1, 2, \dots, J \tag{2.3.1}$$

$$0 = \hat{Y}_t^j - \hat{N}_t^j,\ j = 1, 2, \dots, J \tag{2.3.2}$$

$$0 = \hat{MC}_t^j - \hat{W}_t^j,\ j = 1, 2, \dots, J \tag{2.3.3}$$

$$0 = \hat{Y}_t - \textstyle\sum_{j=1}^J \xi^j \hat{Y}_t^j \tag{2.3.4}$$

$$0 = \hat{Y}_t^j - \hat{Y}_t + \hat{P}_t^j \tag{2.3.5}$$

$$0 = \hat{Y}_t - \hat{C}_t - \hat{G}_t \tag{2.3.6}$$

$$0 = \textstyle\sum_{j=1}^J \xi^j \tilde{\pi}_t^j - \tilde{\pi}_t \tag{2.3.7}$$

$$0 = \hat{N}_t - \textstyle\sum_{j=1}^J \xi^j \hat{N}_t^j \tag{2.3.8}$$

Second, the expressions that need to be mapped to the second matrix equation,

$$0 = \tilde{R}_t - \rho^R \tilde{R}_{t-1} + (1-\rho^R)(\phi_\pi \tilde{\pi}_t + \phi_Y \hat{Y}_t) + \hat{M}_t \tag{2.3.9}$$

$$0 = \tilde{R}_t - E_t\{\tilde{\pi}_{t+1}\} + \eta \hat{C}_t + \eta E_t\{\hat{C}_{t+1}\} \tag{2.3.10}$$

$$0 = \tilde{\pi}_t^j - \beta E_t\{\tilde{\pi}_{t+1}^j\} + \kappa_j \hat{MC}_t^j + \kappa_j \hat{P}_t^j + u_t^j,\ j = 1, 2, \dots, J \tag{2.3.11}$$

and finally, the equations that need to be mapped to the third matrix equation:

$$\hat{u}_t^j = \rho^{j,u} \hat{u}_{t-1}^j + \varepsilon_t^{j,u} \tag{2.3.12}$$

$$\hat{M}_t = \rho^M \hat{M}_{t-1} + \varepsilon_t^M \tag{2.3.13}$$

$$\hat{G}_t = \rho^G \hat{G}_{t-1} + \varepsilon_t^G \tag{2.3.14}$$

Let x_t denote the set of the endogenous state variables, z_t denote the set of exogenous state variables, and y_t denote the set of other exogenous variables such that

$$\mathbf{x}_t = \begin{bmatrix} \tilde{\pi}_t \\ \tilde{R}_{t-1} \\ \tilde{\pi}_t^j \end{bmatrix} \quad \mathbf{y}_t = \begin{bmatrix} \hat{C}_t & \hat{Y}_t & \hat{N}_t & \hat{Y}_t^j & \hat{N}_t^j & \hat{W}_t^j & \hat{MC}_t^j & \hat{P}_t^j \end{bmatrix}^T \quad \mathbf{z}_t = \begin{bmatrix} \hat{u}_t^j \\ \hat{M}_t \\ \hat{G}_t \end{bmatrix}$$

Plugging (2.3.1) to (2.3.14) in

$$\mathbf{0} = \mathbf{A}\mathbf{x}_{t+1} + \mathbf{B}\mathbf{x}_t + \mathbf{C}\mathbf{y}_t + \mathbf{D}\mathbf{z}_t \tag{2.1}$$

$$\mathbf{0} = E_t\left[\mathbf{F}\mathbf{x}_{t+2} + \mathbf{G}\mathbf{x}_{t+1} + \mathbf{H}\mathbf{x}_t + \mathbf{J}\mathbf{y}_{t+1} + \mathbf{K}\mathbf{y}_t + \mathbf{L}\mathbf{z_{t+1}} + \mathbf{M}\mathbf{z_t}\right] \tag{2.2}$$

$$\mathbf{z}_{t+1} = \mathbf{N}\mathbf{z}_t + \mathbf{e}_{t+1} \tag{2.3}$$

yields matrices $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$

$$\underbrace{\begin{bmatrix} 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ -1 & 0 & \xi^j \\ 0 & 0 & \mathbf{0} \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} \eta & 0 & \chi - 1/\zeta & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ 0 & 0 & 0 & \mathbf{1} & \mathbf{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & 0 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{-1} & \mathbf{1} & \mathbf{0} \\ 0 & 1 & 0 & -\xi^j & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & -1 & 0 & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ -1 & 1 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & 0 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & 0 & 1 & \mathbf{0} & -\xi^j & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\mathbf{C}} \underbrace{\begin{bmatrix} \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & -1 \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \end{bmatrix}}_{\mathbf{D}}$$

the matrices $\{\mathbf{F}, \mathbf{G}, \mathbf{H}, \mathbf{J}, \mathbf{K}\}$

$$\underbrace{\begin{bmatrix} 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} 0 & 1 & \mathbf{0} \\ -1 & 1 & \mathbf{0} \\ 0 & 0 & -\beta \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} (1-\rho^R)\phi_\pi & -\rho^R & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{1} \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} 0 & \eta & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}^T}_{\mathbf{J}} \underbrace{\begin{bmatrix} 0 & \eta & 0 \\ (1-\rho^R)\phi_Y & 0 & 0 \\ 0 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \kappa_j \\ \mathbf{0} & \mathbf{0} & \kappa_j \end{bmatrix}^T}_{\mathbf{K}}$$

and the matrix $\{\mathbf{L}, \mathbf{M}, \mathbf{N}\}$

$$\underbrace{\begin{bmatrix} \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \end{bmatrix}}_{\mathbf{L}} \underbrace{\begin{bmatrix} \mathbf{0} & 1 & 0 \\ \mathbf{0} & 0 & 0 \\ \mathbf{1} & 0 & 0 \end{bmatrix}}_{\mathbf{M}} \quad \underbrace{\begin{bmatrix} \rho^{j,u} & 0 & 0 \\ \mathbf{0} & \rho^M & 0 \\ \mathbf{0} & 0 & \rho^G \end{bmatrix}}_{\mathbf{N}}$$

where we consider $\mathbf{a}$ as a vector or a diagonal matrix of $a \in \mathbb{R}$ corresponding to the equations.

2.3.2 Derivation of Parameters.

Strength of Nominal Interest Rate and Interest Rate Smoothing

I follow methods in Castelnuovo and Efrem (2003) [4] to estimate the strength of nominal interest rate to aggregate inflation and aggregate output gap, and interest rate smoothing (IRS) using data from Federal Reserve Economic Data. FRED Economic Data provides growth, employment, inflation, labor, manufacturing and other US economic statistics from the research department of the Federal Reserve Bank of St. Louis. Specifically, R is derived from the Federal Funds effective rate, π is 10-year breakeven inflation rate, Y is derived from gross domestic product in the United States, and G is government total expenditures. The time frame ranges from the first quarter in 2012 to the fourth quarter in 2021.

Taylor Rule: $R_t = \rho^R R_{t-1} + (1-\rho^R)(\phi_\pi \pi_t + \phi_Y Y_t) + M_t, \quad M_t = \rho^M M_{t-1} + \varepsilon_t^M$			
Variables	Coefficients	Standard Error	95% Confidence interval
IRS of Monetary Policy Shocks	0.087	0.168	$(-.255, 0.428)$
IRS of Nominal Interest Rate	0.413	0.102	$(0.207, 0.619)$
Strength of R to π	5.400		
Strength of R to Y	0.0008		
Government Purchases Shock: $G_t = \rho^G G_{t-1} + \varepsilon_t^G$			
Variable	Coefficients	Standard Error	95% Confidence interval
IRS of Government Purchases	0.826	0.089	$(0.645, 1.008)$
Sectoral Inflation Shock: $u_t^j = \rho^{j,u} u_{t-1}^j + \varepsilon_t^{j,u}$			
Variables	Coefficients	Standard Error	95% Confidence interval
IRS of Inflation Shocks of Sector 1	0.956	0.008	$(0.928, 0.978)$
IRS of Inflation Shocks of Sector 2	0.973	0.078	$(0.813, 1.132)$
IRS of Inflation Shocks of Sector 3	0.923	0.040	$(0.841, 1.004)$
IRS of Inflation Shocks of Sector 4	0.894	0.044	$(0.805, 0.984)$
IRS of Inflation Shocks of Sector 5	0.950	0.008	$(0.931, 0.988)$
IRS of Inflation Shocks of Sector 6	0.930	0.040	$(0.850, 1.011)$
IRS of Inflation Shocks of Sector 7	0.948	0.009	$(0.923, 0.988)$
IRS of Inflation Shocks of Sector 8	0.887	0.080	$(0.725, 1.050)$

Table 1. Regression outputs to derive coefficients

Parameter Summary

These data, in conjunction with the model parameters , were used to calculate the estimated value for each trial as shown in the previous sections following the empirical studies from Hou (2014) [8]. Specifically, Sector 1 is the group “Food and beverages”, Sector 2 is the group “Household furnishings and operations”, Sector 3 is the group “Apparel”, Sector 4 is the group “Transportation”, Sector 5 is the group “Medical care”, Sector 6 is the group “Recreation”, Sector 7 is the group “Education and communication”, Sector 8 is the group “Other goods and services”.

Abbreviation	Symbol	Value
Elasticity of substitution between labor sectors	ζ	1
Relative risk aversion of household	η	1
Frisch elasticity of labor supply	χ	1
Discount factor	β	$0.96^{1/4}$
Strength of nominal interest rate to aggregate inflation	ϕ_π	5.400
Strength of nominal interest rate to aggregate output gap	ϕ_Y	0.0008
Interest rate smoothing of Nominal Interest Rate	ρ^R	0.413
Interest rate smoothing of monetary policy shocks	ρ^M	0.087
Interest rate smoothing of government purchases	ρ^G	0.826
Interest rate smoothing of inflation shocks of sector j	$\rho^{j,u}$	$\{1.06, 0.97, 0.92, 0.89, 1.04, 0.93, 1.05, 0.89\}$
Expenditure share of sector j	ξ^j	$\{0.12, 0.05, 0.05, 0.04, 0.25, 0.11, 0.02, 0.26\}$
Index of price stickiness in sector j	θ^j	$\{0.55, 0.59, 0.41, 0.20, 0.62, 0.46, 0.60, 0.50\}$

Table 2. Major Parameters in Model

2.3.3 Policy Functions

In this new solution method, I am looking for a pair of policy functions of the form

$$\mathbf{x}_{t+1} = \mathbf{P}\mathbf{x}_t + \mathbf{Q}\mathbf{z}_t \tag{2.4}$$

$$\mathbf{y}_t = \mathbf{R}\mathbf{x}_t + \mathbf{S}\mathbf{z}_t \tag{2.5}$$

These equations offers an expression for period-t endogenous/exogenous variables as a function of period $t-1$ endogenous/exogenous variables and period-t shocks. Here, the key ingredient in everything that follows is the concept of matrix vectorization, which is a (matrix) operator that takes any $n \times m$ matrix and produces a new matrix (vector) with dimension $nm \times 1$. The policy functions $\mathbf{x}_{t+1} = \mathbf{P}\mathbf{x}_t + \mathbf{Q}\mathbf{z}_t$ and $\mathbf{y}_t = \mathbf{R}\mathbf{x}_t + \mathbf{S}\mathbf{z}_t$ come from solving the matrix equations

$$\mathbf{0} = \left(\mathbf{F} - \mathbf{J}\mathbf{C}^{-1}\mathbf{A}\right)\mathbf{P}^2 - \left(\mathbf{J}\mathbf{C}^{-1}\mathbf{B} - \mathbf{G} + \mathbf{K}\mathbf{C}^{-1}\mathbf{A}\right)\mathbf{P} - \mathbf{K}\mathbf{C}^{-1}\mathbf{B} + \mathbf{H} \tag{2.6}$$

$$\mathbf{R} = -\mathbf{C}^{-1}(\mathbf{A}\mathbf{P} + \mathbf{B}) \tag{2.7}$$

for $\mathbf{P}$ and $\mathbf{R}$, and solving the matrix equations

$$\mathbf{W}\,\mathrm{vec}(\mathbf{Q}) = \mathrm{vec}\left[\left(\mathbf{J}\mathbf{C}^{-1}\mathbf{D} - \mathbf{L}\right)\mathbf{N} + \mathbf{K}\mathbf{C}^{-1}\mathbf{D} - \mathbf{M}\right] \tag{2.8}$$

$$\mathbf{S} = -\mathbf{C}^{-1}(\mathbf{A}\mathbf{Q} + \mathbf{D}) \tag{2.9}$$

for $\mathbf{Q}$ and $\mathbf{S}$, where

$$\mathbf{W} := \mathbf{N}' \otimes \left(\mathbf{F} - \mathbf{J}\mathbf{C}^{-1}\mathbf{A}\right) + \mathbf{I}_k \otimes \left(\mathbf{J}\mathbf{R} + \mathbf{F}\mathbf{P} + \mathbf{G} - \mathbf{K}\mathbf{C}^{-1}\mathbf{A}\right) \tag{2.10}$$

Proof. See Appendix A. □

3 Results

By solving the policy functions using Python codes shown in Appendix B and C, I derive the matrices $\{\mathbf{P}, \mathbf{Q}, \mathbf{R}, \mathbf{S}\}$ shown in Appendix D. These matrices characterize how endogenous and exogenous variables change over time. Since all the elements in matrix $\mathbf{R}$ are approximately zeros, it is not attached, and it implied that the impacts of endogenous state variables on the exogenous state variables at period t are negligible.

In addition, from matrix $\mathbf{P}$, we can observe that the inflation rate at period t has a big negative effect on the inflation rate at period $t+1$ and the nominal interest rate at period t, and the nominal interest rate at period $t-1$ has a relatively smaller, but a positive effect, on the inflation rate at period $t+1$ and nominal interest rate at period t. Besides, the sectoral inflation rates are independent of each other, and they will always increase by 1% in every period.

From matrix $\mathbf{Q}$ and $\mathbf{S}$, we can learn that the government spending shock always has a big impact on all the state variables, both exogenous and endogenous and the monetary policy shock has a big impact on the aggregate inflation rate. When it comes to sectoral inflation shocks and sectoral inflation rates, we can observe that the sectoral inflation shock has a big positive effect on the corresponding inflation rate for Sector 3 to 8, and the sectoral inflation shock of Sector 2 has no effect on the sectoral inflation rate of Sector 2, and the sectoral inflation shock of Sector 1 has a big negative effect on the inflation rate of Sector 1. Also, sectors are no longer as independent as the observations in matrix **P**. We can observe that there is an interdependence between Sector 1 and 2, and between 4 and 8. Specifically, the sectoral inflation shock of Sector 2 has a big negative impact on the inflation rate in Sector 1, and the sectoral inflation shock of Sector 8 has a strong positive impact on the inflation rate in Sector 4.

4 Conclusion

This article provides a quantitative evaluation of the equilibrium conditions by implementing the canonical New Keynesian Model considering sectoral price stickiness. Specifically, it examines how the interest rate and inflation rate influence each other, and how sectoral inflation rates interact with sectoral inflation shocks among different sectors. The article finds out that there is a level of interdependence between the "Food and beverage" sector and the "Household furnishings and operations" sector, and between the "Transportation" sector and "Other goods and services" sector. More research on the reasoning behind this interdependence will be interesting as further work. Also, one unrealistic finding in this paper is that the impact of endogenous state variables on the exogenous state variables is negligible. There might be a lack of this relationship in this model such that an improvement on this model is needed to better capture the real-world situation.

References

[1] Guillermo A Calvo. "Staggered prices in a utility-maximizing framework". In: *Journal of monetary Economics* 12.3 (1983), pp. 383–398.

[2] Carlos Carvalho. "Heterogeneity in price stickiness and the real effects of monetary shocks". In: *Frontiers in Macroeconomics* 2.1 (2006).

[3] Carlos Carvalho and Niels Arne Dam. "The cross-sectional distribution of price stickiness implied by aggregate data". In: *FRB of New York Staff Report* 419 (2010).

[4] Efrem Castelnuovo. "Describing the Fed's conduct with Taylor rules: is interest rate smoothing important?" In: *Available at SSRN 411595* (2003).

[5] Richard Clarida, Jordi Gali, and Mark Gertler. "Monetary policy rules and macroeconomic stability: evidence and some theory". In: *The Quarterly journal of economics* 115.1 (2000), pp. 147–180.

[6] Stefano Eusepi, Bart Hobijn, and Andrea Tambalotti. "CONDI: A cost-of-nominal-distortions index". In: *American Economic Journal: Macroeconomics* 3.3 (2011), pp. 53–91.

[7] Michael Horvath. "Sectoral shocks and aggregate fluctuations". In: *Journal of Monetary Economics* 45.1 (2000), pp. 69–106.

[8] Cheng-qi Hou, Pin Wang, et al. "AN ESTIMATION OF SECTORAL PRICE STICKINESS USING AGGREGATE DATA1". In: *Romanian Journal of Economic Forecasting* 17.2 (2014), p. 53.

[9] Peter J Klenow and Oleksiy Kryvtsov. "State-dependent or time-dependent pricing: Does it matter for recent US inflation?" In: *The Quarterly Journal of Economics* 123.3 (2008), pp. 863–904.

[10] Emi Nakamura and Jón Steinsson. "Five facts about prices: A reevaluation of menu cost models". In: *The Quarterly Journal of Economics* 123.4 (2008), pp. 1415–1464.

[11] John B Taylor. "Aggregate dynamics and staggered contracts". In: *Journal of political economy* 88.1 (1980), pp. 1–23.

[12] John B Taylor. “Discretion versus policy rules in practice”. In: *Carnegie-Rochester conference series on public policy*. Vol. 39. Elsevier. 1993, pp. 195–214.

[13] John B Taylor. “Staggered wage setting in a macro model”. In: *The American Economic Review* 69.2 (1979), pp. 108–113.

Appendix

Simulation Results

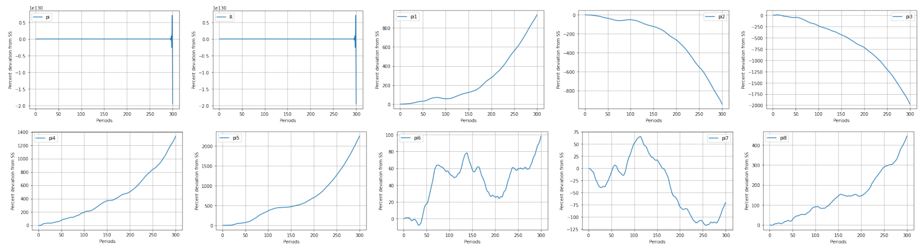
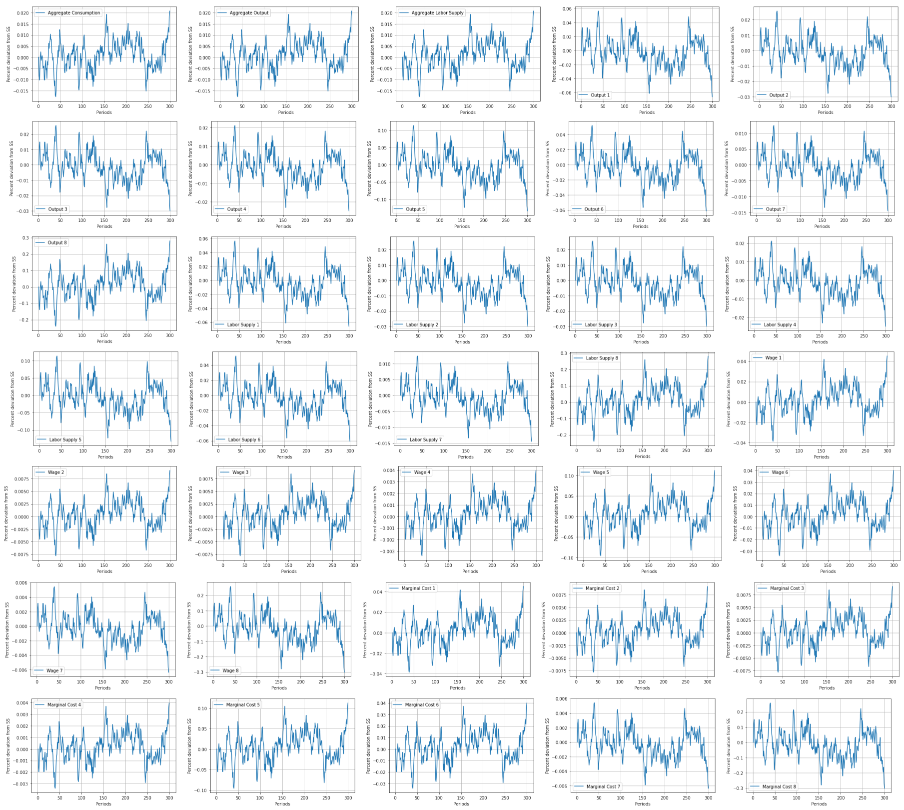


Figure 2: Simulation Results for Exogenous State Variables



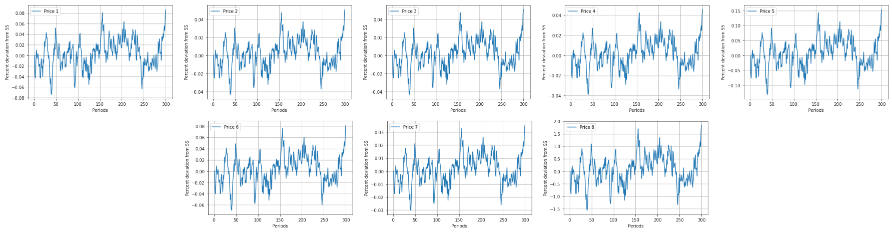


Figure 3: Simulation Results for Other Exogenous Variables

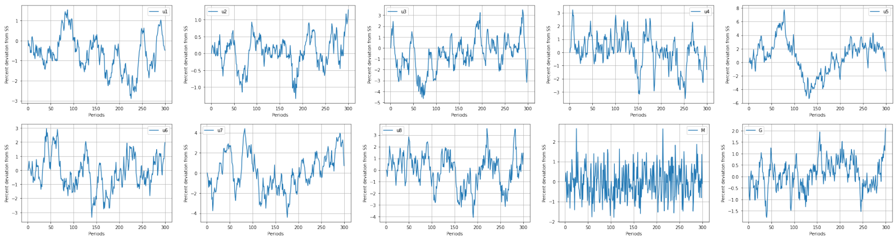


Figure 4: Simulation Results for Other Exogenous Variables

Other files and additional calculations are available from the author upon request.

Gender, Competitiveness, and Academic Track

Aaron Salot & Floyd Krom

I. Introduction

It has been hypothesized that gender differences in competitiveness could be a potential explanation for gender differences in education and labor market outcomes. According to the U.S. Department of Education (2009), gender differences in math and sciences manifests itself at the college level, where women are significantly less likely than men to graduate with a major in science, technology, engineering, or mathematics. Gneezy et al. (2003) provide compelling evidence on the competitiveness differences between men and women, as they conclude that the willingness to enter competitive environments declines for women when the competition group includes men. Gender differences in competitiveness could therefore be a relevant characteristic to explain the disparity between men and women in academic tracks such as science, technology, engineering, and mathematics, which are known to be male dominated and viewed as competitive.

The goal of this paper is to investigate if competitiveness differs among genders at the undergraduate level, and whether that is a potential explanation for gender differences in academic tracks. This is achieved through an experiment that examines the competitiveness of individuals. In order to examine the relationship between competitiveness and academic tracks, participants will complete a survey in which they report their academic track as well as their academic performance. We argue that men pursue academic tracks such as science, technology, engineering, and mathematics on a larger scale, because they are more willing to compete and thus are more predisposed to career paths with greater future expected earnings.

This paper follows the structure of a (I) literature review, (II) experimental design and setup, (III) results, and (IV) conclusion. The literature review examines previous scholars and their theories on gender differences in competitiveness and the relationship between competitiveness and academic track. The next section introduces the experimental design and the intentionality behind certain tasks and what we can extrapolate from them. The results section presents the results of our analysis through various tables and illustrations. The fifth and final section concludes the study and summarizes the results.

II. Literature Review

There are numerous studies that have examined gender differences in competitiveness. A majority of these studies conclude that men are more willing to compete than women as they tend to turn away from competition (Niederle & Vesterlund, 2007). While there are other scholars who claim that due to physiological and cultural factors, women have a more nuanced relationship to competitiveness and are, at times, as competitive as men as they grow older (Flory et al, 2018). Moreover, scholars have argued that one's academic track is a function of their competitive nature and that genders are not equally distributed across academic disciplines (Zhang, 2008). Throughout this section, we will examine previous literature, their research approach and empirical methodology to examine how they arrived at their conclusions.

Buser, et al. (2014) performed a study in which they investigate the relevance of competitiveness for education outcomes and gender differences at secondary schools in the Netherlands. This study uses the measure of competitiveness introduced by Niederle and Vesterlund (2007) in the *Do Women Shy Away from Competition? Do Men Compete Too Much?* study. The scholars performed the experiment with 397 students in grade 9 at the pre-university level of four schools in Amsterdam and surrounding areas. They required participants to solve an arithmetic task thrice, first under a non-competitive piece rate incentive scheme, second under a competitive tournament scheme, and were given the choice between the competitive tournament scheme and the noncompetitive piece rate scheme for the third task.

The results of this study demonstrate that boys are substantially more competitive than girls, even though they have a comparable level of academic strength. When participants were given the choice between the competitive tournament scheme and the noncompetitive piece rate scheme for the third task, 49% of the boys entered the tournament while only 23% of the girls entered the tournament. As boys and girls had very similar performances in the experimental task, the gap in tournament entry between boys and girls cannot be explained by the gender differences in performance. As boys are substantially more competitive than girls, they are also more likely to follow academic tracks which are considered as more prestigious, as they incorporate more math and science intensive courses. This provides evidence that competitiveness is strongly positively correlated with choosing more prestigious academic tracks even conditional on academic ability. The study therefore concludes that the gender difference in competitiveness accounts for 20% of the gender difference in track choice.

Niederle and Vesterlund (2007) performed a study in which they examine whether men and women of the same ability differ in their selection into a competitive environment. The objective of this paper is to investigate whether men and women differ in their preferences for competition and how such differences impact economic outcomes. The experimental design here requires participants, in a laboratory setting, to solve a task twice, first under a noncompetitive piece rate, and second under a competitive incentive scheme. At the end of the tasks, the participants fill out a Belief-Assessment Questionnaire to rank their task preferences and provide background information, allowing the researchers to assess and deduce between the results and the participants' backgrounds.

Their results found that men select the tournament twice as much as women when choosing their compensation scheme for the next performance. While 73% of the men select the tournament, only 35% of the women make this choice. The scholars argue that one potential explanation for this is that men are generally more overconfident - they asserted that women shy away from competition, while men embrace it. Moreover, they claim that this only counts for a share of the gap - another explanation is that women have lower expectations about their relative ability, are more averse to risk, or are more reluctant to be in an environment where they receive feedback on their relative performance.

The scholars lay several reasons as an explanation for the results. First, men simply join more competitions solely because they like to compete more often. Second, men enter such competitions more frequently because they are more overconfident than women. Third, men are less risk-averse than women, making such competitions more appealing. And forth, men enter these tournaments more than women because they are less averse to feedback.

While various studies examined the gender differences in competitiveness, Gneezy et al. (2003) investigated whether men and women differ in their ability or propensity to perform in competitive environments. The experimental design required participants to solve a mazes puzzle in four different incentivized competitive environments: piece rate payment, competitive payment (mixed-sex tournament), random payment, and a competitive payment (single-sex tournament). The participants of this study are engineering students at the Technion (Haifa, Israel), an institution recognized for its competitive nature.

The findings of this study suggest that an increase in the competitiveness of the environment results in a significant increase in performance for men, but not for women. While there is a significant gender gap in the performance in tournaments, there is no gap when participants are paid according to piece rate. This effect is stronger when women compete in a mixed-sex tournament environment compared to a single-sex competitive environment. The study concludes that the increase in performance of women in single-sex tournaments is due to the incentive scheme and not the absence of male participants. This therefore suggests that women may be able to compete in competitive environments. While this paper provides strong experimental evidence, there is uncertainty whether the results provide external validity as the participants of the study are from a selective and competitively recognized engineering program. It is therefore important to examine a multitude of studies and economic experiments performed to better understand how they arrived at their conclusions.

In a review of previous literature on gender differences in economic experiments, Croson and Gneezy (2009) discuss three factors that have been widely examined: risk preferences, social preferences, and reaction to competition. They used economic experiments as well as naturally occurring markets to examine gender differences in these three factors. Their findings identify robust differences in risk preferences, social preferences, and competitive preferences as suggested by past research.

The results of this study suggest that men are more risk-prone than women. This can be explained by the difference in emotional reaction between men and women as well as the overconfidence in men in uncertain situations resulting in different risk taking. Furthermore, the results indicate that the social preferences of women are more situationally specific than those of men. This can be explained by the fact that the social preferences of women are more adaptable to the social environment of their situation. This suggests that women are more sensitive to social clues in determining appropriate behavior which explains the gender difference in social preferences. Lastly, they argue that women are more resistant than men to engage in competitive environments like tournaments, bargaining, and auctions. This attests that women are more averse to competition than men.

While this review of previous literature identifies robust gender differences in risk preferences, social preferences, and competitive preferences, the scholars also identify an important bias in the literature on gender differences. Journals are more likely to publish papers that find gender differences than papers that do not. When looking back, we examined how very few existing studies concluded that there are no gender differences in risk preferences, social preferences, and competitive preferences. Furthermore, the studies discussed in the review of the literature on gender differences are mainly focused on individuals without using age as a variable.

Nevertheless, Flory et al. (2018) assert that previous literature on gender and competition has been singular dimension - since they have only considered two variables: gender and competition. This paper extrapolates from previous literature on gender and competition and argues that it has been singular dimensional - since they have only considered two variables, gender and competition. The scholars, however, sought to introduce age into the metrics and examine whether competitiveness among participants differ as their age changes. Flory et al. (2018) replicate the Niederle & Vesterlund (2007) experimental design and include men & women aged 19 to 90, where they use the age of 50 to divide the participants into two groups - “old” and “young”. The scholars replicated the Neiderle & Vesterlund 2007 study on 700 participants across 12 villages in Malawi.

The scholars assert physiological reasons for using 50 as a dividing factor - they argue that women undergo significant changes in cortisol and progesterone levels after menopause, which has an effect on their competitiveness. In addition to this paper, existing Darwinian theory suggests that the inherited differences in characteristics between men and women stem from the basic differences in brain structure - this also rationalizes how women and men embody different strategies to maximize the fitness of their genes. In particular, genetic or hormonal differences could cause there to be a difference in competitiveness between men and women. Helen S. Bateup et al. (2002) suggest that hormonal levels are correlated with aggression among genders and therefore explain the difference in competitive nature between women and men.

Their results of the study performed by Flory et al. suggest that the age gap between mature women and young women is just as large as the gender gap between young women and young men. They indicate there is at least as much difference in competitiveness between younger women and older women as there is between younger women and men, and that more mature women are no less fond of competition than men. While this paper provides a stronger foundational argument, their experimental setup is at the expense of cultural relativity - since a major proportion of their experiments were conducted in villages in Malawi, there is significant uncertainty as to whether the results provide external validity within different cultural contexts.

After having established that gender and age are variables that influence competitiveness, we decided to examine studies that assess how competitiveness manifests within a college setting, how it influences what students study and potentially their future earnings. Zhang (2008) examines various factors that contribute to a gender and racial gap in earnings among college graduates. The scholar obtained data of over 10,000 baccalaureate recipients and conducted follow-up surveys 1, 5 and 10 years after graduation. The dataset included variables such as demographics, family background, college characteristics, academic track-record and labor market variables to assess the effectiveness of their degrees and its impact on their forward earnings potential. A limitation of the survey was that it only focused on the participants' "primary" job, and, moreover, asked them to include their hourly earnings. While this does not account for individuals with multiple part-time jobs, it also does not encompass individuals who have a disparate hourly and salary income.

The findings suggest that college major remains the most significant factor determining the gender gap in pay. However, the scholar also argues that females lack presence in certain fields such as engineering and science, which further perpetuates a gender gap. For example, the paper highlights how business management is the most popular major for both genders (26% for males and 20% for females), while the 2nd most popular major for females is education (16%)

and engineering for males (12%). Given how genders are not equally distributed among different majors, this further causes an unequal wage distribution among genders.

It is therefore important to further investigate if competitiveness differs among genders at the undergraduate level to extend the currently existing literature. Through our experimental design, we aim to investigate competitiveness among undergraduate students at Macalester College and identify how this correlates with their academic major track.

III. Experimental Design

The basis of our experimental design is to replicate the Niederle and Vesterlund (2007) experimental setup and adapt it for our current environment. Their experiment consists of four assignments in order to investigate whether men and women with equal ability differ in their levels of competitiveness. For our experimental design, we will conduct this experiment with 51 undergraduate students at Macalester College, a competitive liberal arts college with students from every state and 98 countries with academic programs ranked among the top 20 in the nation.

In this section, we discuss the steps of our experimental design and the key differences between our design and that of Niederle and Vesterlund.

The Experiment

After consenting, experiment participants will receive a link to a Google Form which includes the experiment instructions for each task (Appendix A) and the experiment answer sheet for each task (Appendix B). At the end of the experiment, each participant has to complete a questionnaire which includes questions regarding their gender, race, nationality and academic track record at Macalester.

In the experiment, participants will be asked to complete four different tasks. In the first three tasks, they will be asked to calculate the correct sum of a series of four 2-digit numbers. The participants have 90 seconds to complete each task, which encompasses 10 problems to solve. Participants are instructed not to use a calculator, however are allowed to use scratch paper to calculate the sums. In the fourth task, participants are asked to choose which compensation scheme they want to apply to their performance in task 1 with the piece rate compensation scheme. At the end of the entire experiment, we will randomly select one of the four tasks and pay participants based on their performance in that task. The monetary incentive is communicated to all participants before they start the experiment.

Task 1 - Piece Rate

Participants will be asked to calculate the sum of four randomly chosen two-digit numbers. Participants will be given 90 seconds to calculate the sum of a series of these problems. Participants are instructed to set a 90-second timer before starting on task 1 and have to stop providing their responses once time is up.

If task 1 is the task randomly selected for payment, then participants will receive $0.10 for each correctly solved problem in the 90-second time limit. Incorrect answers do not affect the payment, as only correctly solved problems will be considered. Participants are not informed of their performance in this task until after the experiment and payout finds place.

Task 2 - Tournament Compensation Rate

As in task 1, participants will be given 90 seconds to calculate the sum of four randomly chosen two-digit numbers. Once again, participants will be given 90 seconds to calculate the sum of a series of these problems. Participants are instructed to set a 90-second timer before starting on task 2 and have to stop providing their responses once time is up.

In contrast with task 1, the payment for this task depends on the participant's performance in comparison to a group of other participants. These groups consist of four randomly selected participants. The groups are kept unidentified, and participants will be incentivized to compete against each other as the winner of the tournament (the participant with the highest number of correctly solved problems) will receive $0.40 per correctly solved problem if this task is randomly selected for payment. In the situation of a tie, the winner will be randomly selected among the participants with the highest number of correctly solved problems.

Besides the winner of a group, the other participants of the group do not receive any payment. Participants are not informed of their performance in this task until after the experiment and payout finds place.

Task 3 - Choosing between Piece Rate & Tournament Compensation Scheme

Unlike the previous tasks, the participants can select the type of payment they prefer to receive according to the payment scheme - piece rate or tournament rate. As in the previous two tasks, participants will be given 90 seconds to calculate the sum of four randomly chosen two-digit numbers. As in task 1 and 2, participants will be given 90 seconds to calculate the sum of a series of these problems. Participants are instructed to set a 90-second timer before starting on task 2 and have to stop providing their responses once time is up.

Given that participants are allowed to freely select their compensation scheme, this enables us to understand and analyze where participants believe they stand in comparison to their peers. Individuals who are overconfident and competitive may be inclined to select the tournament rate for a greater expected payout, while participants who are more risk-averse may choose the piece rate in order to secure a certain payout.

If task 3 is randomly selected for payment, then participants' payment depends on their selection of compensation scheme. If participants selected the piece rate compensation scheme, then they will receive \$0.10 for each correctly solved problem in the 90-second time limit. If participants selected the tournament compensation scheme, then the number of correctly solved problems will be compared to their group members who were randomly selected in task 2. The groups are kept unidentified, and participants will receive \$0.40 per correctly solved problem if they outperform their group members. To put it into perspective, participants receive five times the payment under the tournament compensation scheme in comparison to the piece rate compensation scheme.

Task 4 - Choice of Compensation Scheme for Past Piece Rate Performance

In this task, participants are asked to choose between the piece rate and tournament compensation scheme which will be applied to their piece rate performance in task 1. Participants therefore get the option to choose between the piece rate compensation scheme ($0.10 for each correctly solved problem) and the tournament compensation scheme (winner of the tournament, the participant with the highest number of correctly solved problems, will receive $0.40 per correctly solved problem).

According to Niederle and Vesterlund, this task will allow us to determine whether general factors such as overconfidence, risk, and feedback aversion by themselves cause a gap in tournament entry. Task 4 therefore gives us the opportunity to demonstrate whether gender differences in competitiveness (choice of compensation scheme) are evident when no tournament performance has been incorporated. Although these factors could potentially have an effect on decision making of the compensation scheme in task 3, they are not isolated to the decision of performing in a competition and thus could have an effect on the decision making of the compensation scheme in task 4. **Text**

Belief Assessment Questionnaire

At the end of the experiment, participants are asked to estimate their task 1 piece rate rank and task 2 tournament rank. In particular, we will ask whether they believe they fit into the "Top 25 Percentile (75-100)", "50-75 Percentile", "25-50 Percentile" or "0-25 Percentile" range. If they guess their rank correctly, they will be rewarded with an additional $0.40.

The goal of this Belief Assessment Questionnaire is to identify and examine whether participants, based on their demographics, are overconfident or have any expectation of their performance without knowing their performance or competition.

Demographics Questionnaire

After the participants have completed all four tasks, they are asked to answer the following questions:

- Age
- Gender - Male, Female, Other
- Race - White, Hispanic or Latino, Black or African American, Native American or Indian American, Asian/Pacific Islander, Other
- Nationality
- Class Year - First Year, Sophomore, Junior, Senior
- Major (or intended major)
- Minor (or intended minor)
- Grade Point Average (GPA)
- SAT/ACT Score
- Athletic Participation: Varsity Athletics - Yes/No

Key Changes to Niederle and Vesterlund Experimental Design (2007)

Our experimental design differs from Niederle and Vesterlund (2007) in two aspects:

First, Niederle and Vesterlund incorporate a lab-experimental setup where they group participants and conduct the experiment in-person. This proves as an advantage since they are able to control the environment and ensure the participants are following the time-guidelines. Our experimental setup, however, will be simulated through an online survey, which makes conducting the experiment in groups and in a timely manner a potential challenge. Moreover, given the online setup, it is easier to deceive participants that they aren't competing against 3 other people in task 2.

Second, whereas Niederle and Versterlund (2007) inform the participants of their results of task 1 - piece rate before choosing a compensation scheme in task 4, we do not inform the participants of their results in task 1. By adding this aspect to task 4 of the experiment, we ensure that participants choose one of the compensation schemes solely based on their performance in task 1.

IV. Results

In this section, we will determine whether women and men differ in their preference for performing under a piece-race compensation scheme or a tournament compensation scheme, conditional on their ability. This experimental setup was selected in order to prevent gender differences in ability that could affect one's tournament entry. We will analyze the participants' performance in task 1 (piece rate) and task 2 (tournament) to determine if women and men perform similarly under the different compensation schemes.

Performance in The Piece Rate (Task 1) and Tournament (Task 2)

TABLE I
Performance Characteristics - Task 1 & 2

	Piece Rate	Tournament
Women	4.571 (0.76)	5.214 (0.65)
Men	5.130 (0.56)	5.696 (0.48)

Standard error in parentheses

Table I illustrates the average performance in the piece rate (task 1) and tournament (task 2) for both genders. Similarly to the results found by Niederle and Vesterlund (2007), we find that there is no significant difference in performance under the piece rate compensation or the tournament compensation across both genders. The average number of correctly solved problems is 4.571 for women and 5.130 for men under the piece rate compensation scheme. The two-sided t-test yields a p-value (p = 0.464) that is statistically insignificant, and therefore indicates that there is no difference in performance between genders. When looking at the tournament compensation scheme, women correctly solve 5.124 problems, while men correctly score 5.696 problems. The t-test results for the tournament compensation scheme are statistically insignificant (p = 0.461) and validate the argument that there is no significant difference in performance across genders.

Table I illustrates both women and men perform better under the tournament compensation scheme than under the piece rate compensation scheme. In particular, women's average scores increase from 4.571 to 5.214 and men's average scores increase from 5.130 to 5.696 when looking at the piece rate and tournament rate, respectively. This can possibly be explained by an initial learning curve, whereby participants are more competent after completing the first task (piece rate) and can perform better through muscle memory. DelleVigna, Malmendier and Vesterlund also found similar results where the participants improved in their performance from task 1 to task 2, however they saw no significant improvement during task 3. Another plausible explanation could be that participants had a greater incentive, a higher expected payout in the tournament scheme, that would have resulted in them improving their performance from task 1 to task 2. While both genders on average scored higher during task 2, it is also important to acknowledge that not all participants scored higher from task 1 to task 2.

In task 3 participants are asked which of the two compensation schemes they would like to apply to their performance - piece rate or tournament. Since we have established no gender difference in performance in task 1 (piece rate) and task 2 (tournament scheme), we hypothesize that there will be no difference in performance for the participants for task 3.

Gender Differences in Tournament Entry (Task 3)

Task 3 provides the participants with the ability to choose how they wish to be compensated for their performance - through either the piece rate or the tournament scheme. Similar to task 2, participants choosing to perform under the tournament compensation scheme in task 3 will be compensated for their correctly solved problems in task 3 if they outperform the number of correctly solved problems by their group members from task 2, the tournament scheme. Since participants are given the choice to select their compensation scheme, we can analyze where participants expect to fare in comparison to their group peers. Participants who are more overconfident and competitive are more likely to select the tournament rate for a greater expected payout, as opposed to participants who are risk-averse and may choose the piece rate to secure a certain payout.

TABLE II
Task 3 - Compensation Scheme

	Men	Women	Total
Piece Rate	5	17	**22**
Tournament	18	11	**29**
Total	**23**	**28**	**51**

Table II illustrates the number of participations of both genders who opted to participate in either the piece-rate or tournament compensation schemes. While women and men had similar performance scores during task 1 and task 2, Table II illustrates their choice of compensation schemes for task 3. As we infer, a greater portion of men choose the tournament compensation scheme in task 3, while a majority of the women chose the piece rate compensation scheme. In other words, we observed 78.2% of men select the tournament compensation, while only 39.3% of women selected this option. After performing Fisher's Exact Test, we arrived at the conclusion that gender differences in the choice of compensation scheme for task 3 is statistically significant (p = 0.010).

Tournament-Entry Decisions Conditional on Performance

TABLE III
Performance Characteristics by Choice of Compensation Scheme (Task 3)

		Average Performance		
	Compensation Scheme	Piece Rate	Tournament	Tournament - Piece Rate
Women	Piece Rate	4.882***	5.118***	0.235
		(0.59)	(0.90)	(0.30)
	Tournament	4.091***	5.364***	1.273**
		(0.90)	(0.72)	(0.43)
Men	Piece Rate	4.200***	4.800***	0.600
		(0.58)	(0.86)	(0.51)
	Tournament	5.389***	5.944***	0.556
		(0.71)	(0.55)	(0.61)

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

We identify the performance characteristics by choice of compensation scheme (task 3) in order to evaluate how performance influences the inclination of women and men to choose the tournament compensation scheme over the piece rate compensation scheme. In Table III, we identify the mean past performance by gender and chosen compensation scheme in task 3. As demonstrated, both women who do and do not choose the tournament compensation scheme increased their performance from piece rate (task 1) to tournament (task 2). However, the increase in performance from piece rate to tournament for women who choose the tournament compensation scheme in task 3 is significantly larger, 1.273 more problems solved correctly, than the increase in performance for women who choose the piece rate compensation scheme in task 3, 0.235 more problems solved correctly. Men who do and do not choose the tournament compensation scheme increased their performance as well from piece rate (task 1) to tournament (task 2). However, we observe a substantial difference in average performance between men who choose the piece rate compensation scheme and the tournament compensation scheme in task 3. In particular, men who choose the tournament compensation scheme in task 3 solve 1.144 more problems correctly in task 2 (tournament) than men who choose the piece rate compensation scheme in task 3. Conditional on the choice of compensation scheme, we observe a gender difference in task 1 and task 2 performance.

TABLE IV
Probit of Tournament Choice in Task 3

Dependent Variable: Task 3 choice of compensation scheme (1-tournament and 0-piece rate)			
	Coefficient	Marginal Effects	P-Value
Female	-1.074	-0.419	0.005
Tournament	0.059	0.023	0.49
Tournament - Piece Rate	0.112	0.044	0.296

Taking from the Niederle and Vesterlund paper, risk-neutral participants became indifferent in choosing between the payment schemes when there is a 25% probability of winning the tournament. For example, if the probability of winning the tournament is greater than 25%, the participants are likely to favor the tournament compensation rate. If the probability is less than 25%, they are more likely to select the piece rate compensation rate. The probit regression as illustrated in Table IV reveals that the participant's gender affects their decision to enter the tournament, while their performance under the two compensation schemes does not significantly affect it. The marginal gender effect of -0.419 reported in Table IV shows that a man with a performance of greater than 6 (in tournament scheme) and 5 (in the piece rate) would have a 41.9% lower probability of entering the tournament if he was a woman. Therefore, after controlling for past performance, we can deduce that women are less likely to select a competitive compensation scheme.

Explanations for the Gender Gap in Tournament Entry

We have concluded thus far that there are no differences in performance between genders under the piece rate or the tournament compensation scheme. However, as demonstrated in Table IV, we deduce that the participant's gender affects their tournament entry decision. A greater share of men choose the tournament compensation scheme in task 3, while a majority of the women choose the piece rate compensation scheme. We can therefore suggest that women with a high level of performance enter the tournament too little while men with a low level of performance enter the tournament at an excessive rate. In the next part of this section, we will further examine the causes of these differences in tournament entry.

There are several explanations for the gender gap in tournament entry. First, women may have lower expectations about their relative ability and therefore are more risk averse. This would naturally make them more hesitant to be present in environments where they receive feedback on their relative performance. Second, men are evidently more overconfident about their abilities and relative performance and have a higher likelihood of entering a tournament. Third, as Niederle and Vesterlund (2007) argue, overconfidence plays a larger role in the difference between genders, while risk and feedback aversion play a negligible role. Based on these multiple factors, we can deduce that a significant portion of the gender differences in tournament entry stem from men and women having different preferences of whether to perform in competitive settings.

Tournament-Entry Decisions Conditional on Performance

After the participants completed all four tasks, they were asked to rank their performance in task 2 against the performance of the three randomly selected group members they competed against in the tournament. Participants were incentivized to guess correctly as they were compensated $0.50 for guessing their correct rank. In Table V, we illustrate how women and men of equal performance have different levels of confidence in their performance in task 2. As earlier established, there is no gender difference in performance in both task 1 and 2. This results in an equal distribution of relative performance ranks between all participants as well as the expected rank within the randomly selected groups for both women and men.

TABLE V
Distribution of Guessed Tournament Rank

	Men		Women	
	Guessed Rank	Incorrect Guess	Guessed Rank	Incorrect Guess
1-Best	6	3	6	4
2	14	8	11	6
3	3	3	11	6
4-Worst	0	0	0	0
Total	**23**	**14**	**28**	**16**

However, participants believe that they are ranked substantially better than their actual performance. Table IV shows the participants' believed rank distributions and the number of incorrect guesses. As observed, there is a higher share of men (26.1%) than women (21.4%) who expected themselves to be the best in the randomly selected group they competed against in the tournament. Furthermore, the number of incorrect guesses shows that both men and women are overconfident in their guessed rank in comparison to their actual rank. However, men are slightly more overconfident, as 60.1% of men overestimated their actual rank while 57.1% of women overestimated their actual rank.

TABLE VI
Ordered Probit of Guessed Tournament Rank

Dependent Variable: Task 4 guessed tournament rank (1-best and 4-worst)			
	Coefficient	Marginal Effects	P-Value
Female	0.285	0.327	0.383
Tournament	-0.214	0.077	0.006
Tournament - Piece Rate	0.053	0.090	0.555

Table VI illustrates an ordered probit of the guessed rank in task 4. In this task, participants are asked to choose between the piece rate and tournament compensation scheme which will be applied to their piece rate performance in task 1. The ordered probit of the guessed tournament rank in task 4 as a function of a female dummy and performance (tournament and tournament - piece rate) demonstrates that conditional on performance, women are less confident about their relative ranking than men and thus less likely to choose the tournament compensation scheme over the piece rate compensation scheme. Furthermore, participants with a better tournament performance believe to have a higher relative performance. We will therefore further investigate the role of overconfidence by men in entering the tournament more frequently than women.

Table VII
Probit of Tournament-Entry Decision (Task 3)

Dependent Variable: Task 3 choice of compensation scheme (1-tournament and 0-piece rate)				
	(1)		(2)	
	Coefficient	Marginal Effects	Coefficient	Marginal Effects
Female	-1.074***	-0.419***	-1.048***	-0.406***
Tournament	0.059	0.023	-0.028	-0.011
Tournament - Piece Rate	0.112	0.044	0.160	0.062
Guessed Tournament Rank			-0.800**	-0.308**
Pseudo R2	0.1428		0.2298	
Sample Size	51		51	

*** p<0.01, ** p<0.05, * p<0.1

Table VII illustrates whether general factors cause gender differences in their compensation scheme choices, when controlling for their expected guessed rank in comparison to their peers. By controlling for their beliefs, we can assess to what extent the gap in compensation scheme accounted for overconfidence as opposed to risk and feedback aversion. Based on the table, we know that the marginal gender effect reported -0.419 shows that a man with a performance of greater than 6 (in tournament scheme) and 5 (in the piece rate) would have a 41.9% lower probability of entering the tournament if he was a woman. However, when accounting for their beliefs and expectations (column 2), we can deduce that this marginal gender effect reported is now -0.406 or has a 1.3% lower probability when accounting for their beliefs and expectations. The p-value is yet statistically significant and the probability is marginally lower. Based on this, we can conclude that a majority of the gender gap originates from the varying levels of overconfidence, while risk and feedback aversion contribute to be a very small portion of this gap.

Do General Factors Cause Gender Differences in Choice of Compensation Scheme?

Table VIII
Probit of Decision to Submit the Piece Rate to a Tournament (Task 4)

Dependent Variable: Task 4 choice of compensation scheme (1-tournament and 0-piece rate)

	(1)		(2)	
	Coefficient	Marginal Effects	Coefficient	Marginal Effects
Female	-0.579	-0.227	-0.571	-0.224
Piece Rate	0.059	0.054	0.116	0.046
Guessed Piece Rate Rank			-0.138	-0.054
Pseudo R2	0.101		0.2298	
Sample Size	51		51	

*** p<0.01, ** p<0.05, * p<0.1

Task 4 is an individual decision that allows us to assess whether gender differences in choice of compensation scheme appear even when no future and past tournament performance is involved. This helps us examine whether overconfidence, risk, and feedback aversion cause a gap in tournament entry. Table VIII illustrates that, conditional on guessed rank and conditional performance, the gender gap in compensation schemes is smaller in comparison to task 3. For example, when controlling for guessed piece rate rank (column 2), we know that the effect of gender on the participant's probability to enter the tournament is only 0.3% lower. We can therefore conclude that when the participants choose a compensation scheme but are not required to actually perform the task, general factors such as risk and feedback aversion account for a greater proportion of the decision making in the task 4 compensation scheme.

The values we deduced for the Female row (column 1 and 2) are statistically insignificant, while Niederle and Vesterlund (2007) obtained a statistically significant p-value for the Female variable in column 1. We do want to highlight, however, that the p-value obtained for Female in column 1 (p-value: 0.117) and column 2 (p-value: 0.123) are very close to the cutoff values to become statistically significant.

Do Preferences for Performing in a Competition Cause Gender Differences in Choice of Compensation Scheme?

Table IX
Probit of Tournament-Entry Decision (Task 3)

Dependent Variable: Task 3 choice of compensation scheme (1-tournament and 0-piece rate)

	(1)		(2)		(3)	
	Coefficient	Marginal Effects	Coefficient	Marginal Effects	Coefficient	Marginal Effects
Female	-1.074***	-0.419***	-1.048***	-0.406***	-0.962**	-0.373**
Tournament	0.059	0.023	-0.028	-0.011	-0.066	-0.026
Tournament - Piece Rate	0.112	0.044	0.160	0.062	0.149	0.058
Guessed Tournament Rank			-0.800**	-0.308**	-0.694**	-0.269**
Piece Rate Submission					0.603	0.234
Pseudo R2	0.1428		0.2298		0.2552	
Sample Size	51		51		51	

*** p<0.01, ** p<0.05, * p<0.1

Table IX illustrates whether gender differences in tournament entry are driven largely by general factors or if there are additional gender differences when it comes to entering a tournament. This section aims to explore whether the tournament-entry gap is explained by a difference in preferences for competitive environments or whether it is accounted for by gender differences in general factors such as confidence and risk and feedback aversion.

Based on the table above, we can conclude that the participant's likelihood to submit past performance to a tournament is not the same as their decision to enter a tournament and then perform. Similar to Table VII, controlling for expectations on relative performance decreased the gender gap for tournament entry from -41.9% to -40.6% (column 1 & 2), while this marginal gender affect effect is further reduced to -37.3% when controlling for the participant's decision to submit a piece rate (column 3). Niederle and Vesterlund assert that this decrease may be explained by control for risk and feedback aversion, since the decision to submit the piece rate is an additional indicator of the participant's degree of confidence. This explanation by Niederle and Vesterlund justifies how the coefficient on guessed tournament rank decreases as we move from column (2) to column (3). After controlling for task 4 and the expected tournament rank (column 3), the marginal gender effect on the participant's decision to enter the tournament is yet -37.3%. This provides enough evidence to conclude that suggests that a gender gap in tournament entry is influenced by a differing preference in women and men to perform in a competitive environment.

After having computed the previous table, we found, similar to Niederle and Vesterlund (2007), that 90.2% of the original gender effects are accounted for by general differences in overconfidence and risk-aversion. We can also deduce that the competitive component contributes to 9.8% of the original gender effect. We can therefore conclude that the gender gap in tournament entry by men and women is accounted for by the difference in preferences of competitive environments.

Do Preferences for Performing in a Competition Cause Gender Differences in Academic Track?

It has been hypothesized that gender differences in competitiveness could be a potential explanation for gender difference in education outcomes (Zhang, 2008). We have established that the gender gap in competitiveness between women and men at the undergraduate level is accounted for by the difference in overconfidence, risk-appetites and general preferences of competitive environments. In this section, we will determine if, when controlling for academic tracks (STEM/Non-STEM), gender continues to play a vital role in accounting for major differences in competitiveness.

Table X Academic Track by Gender			
	Men	Women	**Total**
STEM	15	15	**30**
Non-STEM	8	13	**21**
Total	**23**	**28**	**51**

Table X illustrates the number of participations of both genders and their academic track choice, STEM or non-STEM. As we infer, a greater portion of men choose a STEM major for their academic track. We observed 65.2% of men select a STEM, while only 53.6% of women opted in for STEM academic tracks. After performing Fisher's Exact Test, we arrived at the conclusion that gender differences in the choice of academic track is statistically significant (p = 0.010).

Table XI Probit of Tournament-Entry Decision (Task 3)		
Dependent Variable: Task 3 choice of compensation scheme (1-tournament and 0-piece rate)		
	Coefficient	Marginal Effects
Female	-0.956**	-0.369**
STEM	0.516	0.120
Guessed Tournament Rank	-0.678**	-0.262**
Pseudo R2	0.2273	
Sample Size	51	

*** p<0.01, ** p<0.05, * p<0.1

Table XI illustrates the effects of controlling for academic tracks (STEM/Non-STEM) and assesses whether academic tracks or gender accounts for a significant portion of the difference in preferences. Based on the results above, we can infer that gender is yet statistically significant when controlling for overconfidence, and that females are 36.9% less likely to enter tournaments. Furthermore, we can see that STEM majors are 12% more likely to enter the tournament than Non-STEM majors. Therefore, we can conclude that when controlling for academic tracks (STEM/Non-STEM) and overconfidence, gender continues to play a vital role in accounting for major differences in competitiveness.

V. Conclusion

Our study contributes to the extensive literature on gender and competitiveness and further investigates if competitiveness differs among genders at the undergraduate level, and whether that is a potential explanation for gender differences in academic tracks. While other scholars have contributed to nuanced understandings in the field of gender and competitiveness, we replicated the Niederle & Vesterlund (2007) experimental set up and took it a step further by assessing the impact of differences in overconfidence and risk-aversion on undergraduate's academic tracks and whether they choose STEM or non-STEM majors.

Our experimental setup was intentionally designed to assess and quantify performance, competitive-preferences and risk-aversion. In order to avoid any disparity in the ability to perform, we selected a task in which men and women are able to perform equally well. The task requires participants to complete basic arithmetic calculations, after which they are compensated for their individual and relative performance to others. Our study demonstrates that men are more likely to enter the tournament, despite there being no gender differences in performance. Moreover, we conclude that overconfidence plays a vital role in the gender gap in tournament entry. We found that men, in general, are more confident than women. In particular, we saw that women are 41.9% less likely to enter a tournament than men if they score around the cutoff mark, when controlling for past performance. We also saw that, when given the choice, 78.2% male participants opted for the tournament scheme, while only 39.2% female participants selected the competitive payment scheme. When controlling for the participant's expected performance (a metric of overconfidence), we see that participants are now 40.6% less likely to enter a tournament. Since there was a marginal difference between 41.9% and 40.6%, we can conclude that overconfidence accounts for a greater proportion of the gap, as opposed to risk and feedback aversion. In particular, we saw that 90.2% of the original effects are accounted for by overconfidence as opposed to general preferences.

After establishing that overconfidence differs significantly across genders, and that it plays a role in an individual's choice of tournament-entry, we aim to better understand how this affects undergraduate students, their academic track choices and forward earnings. We found that 65.2% of males select a STEM major, while only 53.6% females select a STEM major. When accounting for STEM in the model, we can deduce that females are 36.9% less likely to enter a tournament, while STEM majors are 12% more likely to participate in a tournament. This helps us establish that there is an unequal distribution of gender across STEM majors. Scholars such as Zhang (2008) also provide evidence to show how genders are not equally distributed across majors and that different majors offer varying levels of competitive salaries. In particular, we know that STEM-related majors such as engineering, computer science, and hard-sciences are more competitive fields that offer graduates a higher expected salary. After having established that women have a lower probability of entering tournaments (competitive environments) than males, we can conclude that they are less likely to pursue academic tracks that are competitive (STEM). In doing so, based on existing literature, we can conclude that women, through their difference in overconfidence, are less likely to participate in competitive academic tracks, and are therefore likely to have a lower expected future earnings.

Although we arrived at this conclusion, we do find it vital to highlight a few shortcomings to our experimental setup and study. First, the type of task we asked participants to do was numerically-focused, and therefore the performance may be influenced by only one aspect of an individual's abilities. Second, the remote setting did not allow us to hold the experiment in a physical space with all participants present. By being physically together in the same space, it is possible that participants may have had increased levels of competitiveness in the Niederle and Vesterlund (2007) study. Since the participants performed our experiment in a remote setting, this could have dampened or increased one's level of competitiveness. Third, as Flory et al (2018) argue, competitiveness among women changes post-menopause. Therefore, while we arrived at certain conclusions, we are particularly interested to examine whether the results would hold when including participants that are over 50 years old - an age cutoff after which women are relatively more competitive. Fourth, and lastly, the participants in the experiment only included undergraduate students from Macalester College. Given that Macalester College is a competitive liberal arts college with a diverse student population, it is plausible that these results may not provide external validity in other college-settings. For future research, it could be compelling to replicate this experimental set up at other competitive liberal arts colleges and state-colleges to see whether the results hold in other college environments. Further analysis is certainly motivated and would be a great continuation of this paper.

References

Bateup, H., Booth, A., Shirtcliff, E., & Granger, D. (2002). Testosterone, cortisol, and women's competition. *Evolution And Human Behavior*, *23*(3), 181-192.

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, Competitiveness, and Career Choices*. *The Quarterly Journal Of Economics*, *129*(3), 1409-1447.

Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal Of Economic Literature*, *47*(2), 448-474.

Flory, J., Gneezy, U., Leonard, K., & List, J. (2018). Gender, age, and competition: A disappearing gap?. *Journal Of Economic Behavior & Organization*, *150*, 256-276.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *The Quarterly Journal Of Economics*, *118*(3), 1049-1074.

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.

U.S. Department of Education. (2009). Students Who Study Science, Technology, Engineering, and Mathematics (STEM) in Postsecondary Education. *National Center for Education Statistics*.

Zhang, L. (2008). Gender and Racial Gaps in Earnings among Recent College Graduates. *The Review Of Higher Education*, *32*(1), 51-72.

Appendix

(A) Instructions for Experiment

In the experiment today you will be asked to complete four different tasks followed by a demographics questionnaire. The experiment will take less than 10 minutes to complete. At the end of the experiment, we will randomly select one of the tasks and pay you based on your performance in that task. To determine which task counts for payment, we will randomly select one out of the four tasks.

The method we use to determine your earnings varies across tasks. Before each task we will describe in detail how your payment is determined. At the end of the experiment, we will randomly select one of the four tasks for compensation. Based on your performance in this task, you will be compensated accordingly. Additionally, you will be able to earn $20 if your name is drawn from the raffle.

Your total payment will be determined after all participants have completed the survey and will be distributed through Venmo. Later in the experiment, you will be requested to provide your Venmo account name in order for us to compensate you based on your performance. If you do not have a Venmo account, you will be able to list another online payment method to be compensated.

This experiment requires a phone to set a timer for each task. Please make sure you have a phone by hand and set a 90 second timer before you start on each task. Once the 90 seconds have passed, participants are requested to stop providing answers and move on to the next section. Any answers provided after the 90 seconds have passed will not be compensated.

Instructions Task 1 - Piece Rate

For Task 1 you will be asked to calculate the sum of four randomly chosen two-digit numbers. You will be given 90 seconds to calculate the correct sum of a series of these problems. You cannot use a calculator to determine these sums, however you are welcome to make use of scratch paper.

If task 1 is the one randomly selected for payment, then you get $0.10 per problem you solve correctly in the 90 seconds. Your payment does not decrease if you provide an incorrect answer to a problem. We refer to this payment as the piece rate payment.

Before you start the task 1, set a 90 second timer and make sure to stop providing answers once the 90 seconds have passed. Any answers provided after the 90 seconds have passed will not be compensated.

Instructions Task 2 - Tournament

As in task 1 you will be given 90 seconds to calculate the correct sum of a series of four 2-digit numbers. However, for this task your payment depends on your performance relative to that of a group of other participants. Each group consists of four people, the three other members of your group are randomly selected members of this experiment. You will not know who is in your group.

If task 2 is the one randomly selected for payment, the individual in the 4-person group who correctly solves the largest number of problems will receive $0.40 per correct problem. The other participants receive no payment. We refer to this as the tournament payment. You will not be informed of how you did in the tournament until later. If there are ties the winner will be randomly determined.

Before you start task 2, set a 90 second timer and make sure to stop providing answers once the 90 seconds have passed. Any answers provided after the 90 seconds have passed will not be compensated. You cannot use a calculator to determine these sums, however you are welcome to make use of scratch paper.

Instructions Task 3 - Piece Rate or Tournament

As in the previous two tasks you will be given 90-seconds to calculate the correct sum of a series of four 2-digit numbers. However, you will now get to choose how you want to be paid: piece rate or tournament.

If task 3 is the one randomly selected for payment, then your earnings for this task are determined as follows:
- If you choose the piece rate you receive $0.10 per problem you solve correctly.
- If you choose the tournament your performance will be compared to the performance of the other three participants of your group in task 2. Task 2 is the one you just completed. The individual in the 4-person group who correctly solves the largest number of problems will receive $0.40 per correct problem.

You will not be informed of how you did in the tournament until later. If there are ties the winner will be randomly determined.

Before you start task 3, set a 90 second timer and make sure to stop providing answers once the 90 seconds have passed. Any answers provided after the 90 seconds have passed will not be compensated. You cannot use a calculator to determine these sums, however you are welcome to make use of scratch paper.

Instructions Task 4

In this task, you will not be requested to add numbers. Instead, you will choose which compensation scheme you want to apply to your performance in task 1 with the piece rate compensation scheme.

If task 4 is the one randomly selected for payment, then your earnings for this task are determined as follows:
- If you choose the piece rate, you receive $0.10 per problem you solve correctly in task 1.
- If you choose the tournament, your performance will be compared to the performance of the other three participants of your group in task 2. The individual in the 4-person group who correctly solves the largest number of problems will receive $0.40 per correct problem.

(B) Answer Sheet for Experiment

Task 1

67 | 87 | 29 | 24 = ____________

75 | 59 | 32 | 32 = ____________

59 | 24 | 95 | 11 = ____________

19 | 10 | 29 | 74 = ____________

80 | 12 | 70 | 56 = ____________

31 | 17 | 21 | 23 = ____________

38 | 91 | 26 | 92 = ____________

21 | 88 | 99 | 46 = ____________

96 | 99 | 76 | 86 = ____________

74 | 56 | 94 | 37 = ____________

Task 2

23 | 73 | 81 | 43 = ____________

67 | 83 | 16 | 31 = ____________

09 | 11 | 41 | 89 = ____________

45 | 67 | 71 | 14 = ____________

37 | 95 | 12 | 28 = ____________

30 | 17 | 85 | 21 = ____________

41 | 76 | 23 | 34 = ____________

12 | 33 | 46 | 81 = ____________

19 | 24 | 83 | 51 = ____________

56 | 31 | 18 | 08 = ____________

Task 3

Compensation Scheme for Task 3:
- ❏ Piece rate payment scheme
- ❏ Tournament payment scheme

12 | 32 | 45 | 73 = ____________

23 | 58 | 31 | 89 = ____________

13 | 47 | 74 | 65 = ____________

10 | 61 | 72 | 39 = ____________

15 | 69 | 42 | 11 = ____________

81 | 17 | 35 | 74 = ____________

88 | 16 | 33 | 41 = ________________

91 | 87 | 34 | 09 = ________________

85 | 64 | 27 | 14 = ________________

73 | 13 | 45 | 68 = ________________

Task 4

Compensation Scheme for Task 4:
- ❏ Piece rate payment scheme
- ❏ Tournament payment scheme

(C) Belief Assessment Questionnaire

How do you think you faired in comparison to your peers in task 1? (the piece rate compensation scheme)
- ❏ 75-100 Percentile (i.e., top 25 percent)
- ❏ 50-75 Percentile
- ❏ 25-50 Percentile
- ❏ 0-25 Percentile

How do you think you faired in comparison to your peers in task 2? (the tournament compensation scheme)
- ❏ 75-100 Percentile (i.e., top 25 percent)
- ❏ 50-75 Percentile
- ❏ 25-50 Percentile
- ❏ 0-25 Percentile

(D) Demographics Questionnaire

Age:
Gender:
Race:
Nationality:
Class Year:
- ❏ First year
- ❏ Sophomore
- ❏ Junior

- ❏ Senior

Major (Intended Major):
Minor (Intended Minor):
Grade Point Average (GPA):

- ❏ 3.5 - 4.0
- ❏ 3.0 - 3.5
- ❏ 2.5 - 3.0
- ❏ 2.0 - 2.5

Did you take the SAT or ACT?

- ❏ SAT
- ❏ ACT

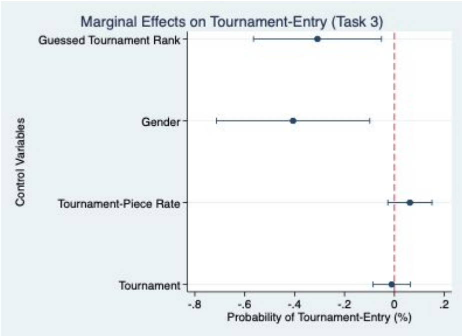
What score did you get on your SAT/ACT test:
Are you a varsity athlete for Macalester College?

- ❏ Yes
- ❏ No

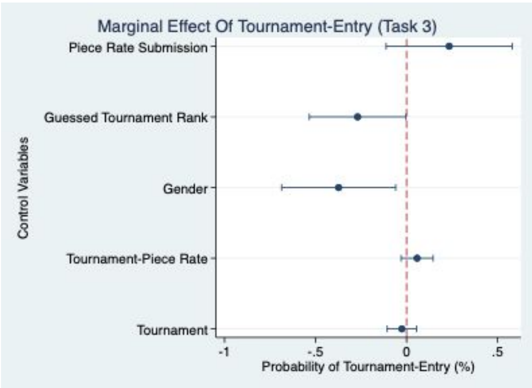
Please provide your Venmo handle for compensation (in case you do not have a Venmo account, please note another online payment method):

(E) Graphic Representation of Results

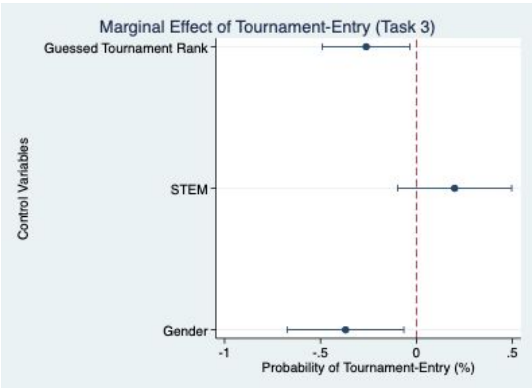
Graph 1.1 illustrates the marginal effects of tournament entry from Table VII



Graph 1.2 illustrates the marginal effects of tournament entry from Table IX



Graph 1.3 illustrates the marginal effects of tournament entry from Table X



Did K-12 school closure and reopening policies in response to COVID-19 enlarge the TBD gender employment gap?

Xinyi Wang

Acknowledgement:

I would like to express my special thanks of gratitude to my advisor Felix Friedt who gave me the golden opportunity to do this honor project on the topic (Did K-12 school closure and reopening policies in response to COVID-19 enlarge the gender employment gap?). I would also like to thank to my committee members Amy Damon and Julie Dolan for their generous suggestions and I came to know about so many new things.

Secondly, I would also like to thank my parents and friends who helped me a lot in finalizing this project within the limited time frame. I would like to thank to Xiaoyan for her emotional support throughout this project.

Thanking you,

Xinyi Wang

April 27, 2022

Abstract:

The COVID-19 pandemic hits female workers the most. This impact on the United States's labor market can be attributed to the limited availability of childcare and schooling options (Stefania and Jiyeon, 2021). With limited resources for childcare and schooling, parents, especially mothers, had to exit the labor force or reduce working hours to stay at home and take care of their children. My study will contribute to understanding the effect of the child penalty, especially under the COVID-19 pandemic and study the impact of school closure and reopening policies. Using data from Current Population Survey (CPS) combined with school closure and reopening data, I conduct both static and dynamic analyses at the extensive (i.e. employed or not) and intensive (i.e. # of hours worked) margins. I find that for a worker, who has at least one child in the household, compared to a worker with the same occupation, in the same industry and similar location is around 76% less likely to be employed when schools are closed while a female worker tends to be 34% more likely to be employed than a male worker after the school has been reopened.

Introduction:

The gender gap index across economic participation, educational attainment, health and survival, and political empowerment subindexes have narrowed over time (World Economic Forum, 2020). However, with the COVID-19 outbreak in Wuhan, China in December 2019, COVID-19 soon grew into a global pandemic within a few weeks. As the impact of the COVID-19 pandemic continues to be felt, the estimated time that is needed to close the whole gender gap has increased from 99.5 years to 135.6 years (World Economic Forum, 2021).

Stefania and Jiyeon (2021) state that unlike other recessions in United States' history, where men usually experience larger employment drops, the COVID-19 recession resulted in larger employment losses for women. This unique impact on the United States's labor market can be attributed to the limited availability of childcare and schooling options (Stefania and Jiyeon, 2021). With such limited resources for childcare and schooling, parents, especially mothers, had to exit the labor force or reduce working hours to stay at home and take care of their children.

My study will contribute to understanding the effect of the child penalty, especially under the COVID-19 pandemic. By examining different phases of school closures and reopening policies and distinguishing between households with and without school-aged children, I will quantify the child employment penalty during the pandemic and differentiate the effects by gender. This study will focus on the different labor outcomes, including the probability of being employed (extensive margin) and the number of hours worked (intensive margin). While the effect of school closure policies has been studied and shown to have an enlarging effect on the gender employment gap, this study will emphasize the effect of school reopening policies and evaluate whether the closure effects are reveresed. This analysis leverages detailed data from the Current Population Survey and attempts to identify the effects at two levels of spatial aggregation. First, at the state-level where I observe school closure/reopening policy data for all states. Second, at the county-level, where I observe school closure/reopening data for California only.

Literature Review:

Gender pay inequality is ubiquitous in society. Hall and Krueger (2012) verified that around one-third of the US working force did not see transparent wages. The gender pay gap is not driven by level of education but by gender discrimination specifically in the labor market. Irimie et. al, (2014) conducted theoretical research which demonstrated gender inequality is a common phenomenon in the labor market, and it is usually caused by discrimination and gender biases. They compared the gender difference between the education period and later career, in which they discovered employers are keen to hire male graduates, regardless of whether a woman would be more appropriate or not. When men are obviously favored in the labor market, women can only work in positions that they are not familiar with, resulting in lower productivity than men, which causes low payment for women.

Although the gender income gap is ubiquitous, COVID-19 made the situation worse. The COVID-19 pandemic shocks the labor market in the US. Based on the experiences during previous recessions in the United States' history, one might have expected men to experience a larger employment drop as usual during COVID-19.

However, the COVID-19 pandemic hit the United States' labor market quite differently, with women being hit the most. Stefania and Jiyeon (2021) studied the effects of the COVID-19 on the labor market of the US, especially from the perspectives of occupation, family, and gender. According to their findings, occupation and the increased childcare factors are two unique factors that may account for the opposite result in the labor market. Since women are mainly employed in the service occupation, which tend to be contact and inflexible jobs, women are more likely to lose their jobs. Plus, the fact that there is a substantial "child penalty" which may reduce women's wages even before they give birth to their first child. This study will focus on the second unique factor, which is the increased childcare factor, and try to explain the gender income and employment gap change during the COVID-19 pandemic.

A new study from the Center for Global Development (2020) suggests that each woman provided up to 173 additional hours in childcare in 2020 through October, compared to only 59 additional hours from men. The reason why I focus on the effect of increased childcare is that it is strongly related to the child penalty. The child penalty has been studied a lot and research suggests that child penalty may contribute a large portion of the gender income gap. Francesca, B, and Helmuth, C, and Chiara, M (2019) constructed a model to study the difference between men and women in informal childcare. They separated informal childcare into two types: basic care (feeding, changing children, baby-sitting) and quality care (activities that stimulate children's social and cognitive skills). Based on their predictions and real data from Italy, mothers tend to devote more time than fathers in both basic and quality informal childcare. More educated mothers also devote more time to quality informal childcare and also spend more time in the labor market than less-educated mothers. Mark Aguiar and Erik Hurst (2007) use five decades of time-use surveys to document trends in the allocation of time within the United States. They find a clear trend that women have carried a heavier load for childcare than men even before the COVID-19. During COVID-19, this gender gap in childcare intensified. Given such disparities in the time allocation on childcare between men and women, the consequence of such differences draws attention to the gender gap in the labor market as women may not be able to work as many hours as men.

Patricia and Jessica (2018) focused on the role of children in explaining the remaining gender gap in the labor market. In their analysis, they discover that over two-thirds of the gender income gap can be explained by the differential impact of children on men and women. Since children play a huge role in explaining the gender income gap, I focus on the school policies to see whether different school policies will affect the gender gap.

School closure policies have been proven to have different impacts on the working styles of men and women. Eiji, Y, and Yoshiro, T (2021) have studied the impact of closing schools on working from home during the COVID-19 pandemic for the period of school closure from mid-March to mid-April 2020. Specifically, they studied the impact of how the presence of the children affects parents' work and how the effect of their children differs between genders. After controlling various variables, they reached the conclusion that mothers are more likely to work at home in order to keep an eye on their children when their children are in primary school while fathers' working styles are less likely to change. However, when their children are in junior high school, both parents' working styles are hardly affected. Such results affirm the heavier burden on mothers of young children compared to fathers. A similar result has been found by Collins C, Ruppanner L, Christin Landivar L, Scarborough WJ (2021). They combined the Elementary School Operating Status database and Current Population Survey to study the effect of remote learning on labor force participation and concluded that the gender gap in labor force participation increased by 5% due to K-12 distance learning.

Acknowledging such differences in childcare for different genders especially with school closure policies, this study will focus on the trade-off between childcare and employment during the COVID-19 pandemic. The study is focused on households with young children. School closure policies have been studied and proven to affect women workers more. However, my study will not only look at the K-12 school closure policies but also the K-12 school reopening policies in order to compare the difference between those with children at home and those without children at home. When comparing the K-12 school reopening policies, my study will try to explain if the gender employment gap caused by the K-12 school closure policy automatically recovers after the K-12 school, or if those K-12 school closures may have a long-term effect on the gender employment gap.

Model:

Theory:

After the CDC confirmed the first laboratory-confirmed case in Washington DC on January 21st 2020, COVID-19 soon spread to all states in the United States within several weeks. COVID-19 spreads when an infected person breathes out the very small particles that contain the virus, and are then breathed in by others or contaminate surfaces. Under such circumstances, people within 6 feet of an infected person are the most likely to be infected (CDC, 2020). Schools, however, fall into the category above of close-contact institutions as most of the activities happen within 6 feet between both students and teachers. Thus, schools have been considered high-risk places where COVID-19 can easily spread. In order to minimize the spread of COVID-19, governors declared states of emergencies and followed the recommendation from the CDC to close the schools.

Under states' orders, K-12 schools had ended in-person instructions in March 2020 and offered remote learning options instead. Students ages 5-18 were forced to stay at home as a result of school closure. This has effects on parents' employment outcomes. In fact, according to Fabrizio, Gomes and Tavares (2021), women with young children, younger than 12 years old, have been disproportionately affected compared with other women and men in terms of employment loss. This can be attributed to the extra childcare that those women have to provide. According to a new study from the Center for Global Development (2020), each woman provided up to 173 additional hours in childcare since COVID-19 has started through October 2020, compared to only 59 additional hours from men. With such a huge difference in the hours for childcare, women are left with shorter time for work. Consequently, they are more likely to be separated from the labor market than men. In addition, COVID-19 hit the service industry the most, which lead to a great reduction in demand for services. Since women are more likely to be employed in a service industry, this compositional effect leads to a greater reduction in demand for female workers.

The employment effects are further complicated by the potential heterogeneity across household income. For example, households with good financial standing will be able to take care of the children if their work is remote. If remote work is unavailable, such household can still afford to pay for a huge amount of money to hire someone to take care of their children or have mom/dad stays at home and take care of the children. On the other hand, these households may be able to afford to take time off work to take care of the children. If it is a mom whose work is not remote and they do not want to hire someone for childcare, then the gender employment gap occurs, while if it is a dad whose work is not remote or they are willing to hire someone for childcare, then there may be no gender employment gap. Although such circumstance rarely happens to a good financial standing household, high-paying jobs are most likely to be remote so that both parents will be able to work at home. In this case, I am expecting a 0 or small effect of school closure on the gender employment gap for those households with good financial standing.

However, things are different for low-income households. They do not have the option to hire others to take care of their children. Moreover, low-paying jobs are less likely to offer a remote work environment when children return home, at least one of the parents needs to reduce working hours or leave the labor market to take care of children. Since females usually work in the service industry and COVID-19 hits the service industry the most, then the demand for female workers decreased a lot. If it is a mom whose work is not remote, then mon has to quit the job for childcare as they cannot leave the child alone at home and the dad has to go to work to support the family. If both parents' works are not remote, then when there is a chance for returning back to work, it is usually the dad who returns back to work and leaves mom at home with children. Thus, it is more likely that the female workers who leave the labor market. In this case, I am expecting a large impact of school closure on the gender employment gap for those households with poor financial standings. Through the above mechanisms, COVID-19 can point to a significant employment penalty for women, especially when schools are closed under the COVID-19 pandemic.

Acknowledging such employment penalty for women under the COVID-19 pandemic, school reopening may counteract such enlarged gender employment gaps. For low-income families, they no longer need to choose between mom and dad who stay at home and take care of the children while the other works. When schools reopen Moms are more likely to be relieved from the extra childcare that usually occurred when their children stay at home and may be able to restart their full-time jobs. However, even if moms are willing to restart their jobs, some of them may be unable to find jobs and remain unemployed. Thus, I am expecting the enlarged gender employment gap to be closed or reduced by the school reopening policies. Similar to the closure effects, school reopening should have a larger effect on the gender employment gap observed for low-income households.

Empirical Model:

To quantify the effect of school closure/reopening on employment, I run two kinds of regression analyses at both the state-level and county-level: one is the static model and the other is the dynamic model. The static model for state from extensive margin perspective for school closure is given as follows:

$$logit(y_{i,k,s,o,t}) = \beta_0 + \beta_1 Female_i + \beta_2 Closure_{s,t} + \beta_3 Closure_{s,t} * Female_{i,t}$$
$$+\gamma * controls + \alpha_k + \alpha_s + \alpha_o \times \alpha_t + \varepsilon_{i,k,s,o,t}$$

Static model for state from extensive margin perspective for school reopening:

$$logit(y_{i,k,s,o,t}) = \beta_0 + \beta_1 Female_i + \beta_2 Open_{s,t} + \beta_3 Open_{s,t} * Female_{i,t}$$
$$+\gamma * controls + \alpha_k + \alpha_s + \alpha_o \times \alpha_t + \varepsilon_{i,k,s,o,t}$$

Where i indexes an individual, k indicates industry, s indicates state, o indicates occupation and t indexes time. The extensive margin is measured via the outcome variable $y_{i,k,s,o,t}$, which indicates the employment status of the individual at time t, in s state and k industry with occupation o. $Closure_{s,t}$ indicates the dummy for the school closure policy within that individual's state at time *t*, with $Closure_{s,t} = 0$ for school has not been ordered closed in that state at time t and $Closure_{s,t} = 1$ for school has been ordered closed in that state at time t. $Open_{s,t}$ indicates the dummy for the school reopening policy within that individual's state at time *t*, with $Open_{s,t} = 0$ for school has been ordered closed in that state at time t and $Open_{s,t} = 1$ for school which is partially opened or fully opened in s state. $Female_i$ indicates the dummy for gender, with $Female_i = 0$ for males and $Female_i = 1$ for females. $controls$ includes a set of controls for age, COVID-19 new cases, vaccine rates, family size, whether the person has difficulty, education attainment, whether the person is white, family income. α_k is the fixed effect term for each industry, α_s is the fixed effect term for each state. $\alpha_o \times \alpha_t$ is the cross fixed effect for occupation and time, which will help control the occupation change throughout the time. By controlling for industry, family income and the cross effect of occupation and time, I will be able to account for most of the difference between different industries and occupations throughout the time. However, this does not take the potential triple-cross fixed effects between time, industry, and occupation into account as this will generates too many levels inside this term, which may not produce statistically insignificant results. And $\varepsilon_{i,k,s,o,t}$ is the error term.

For the school closure effect, this research conducts logistic regression for state level analysis from September 2018 to July 2020[1], which is before the first state that has published any school reopening policy. For the school reopening effect, this research conducts logistic regression from April 2020 to September 2021. By separating the effect of school closure and school reopening policies, this would reduce any potential drawbacks of combining both policies in the same model as the trend caused by school closure or reopening are sperate from each other.

The static model from extensive margin used the difference-in-difference method, and the coefficients of interest are β_2 which captures the effect of school closure or reopening on male workers' employment and β_3 captures the differentiate effect of school policies on female workers' employment other than male workers. Thus, the effect of school closure or reopening policies on female workers' employment would be $\beta_2 + \beta_3$.

The static model for state from intensive margin perspective would be the same model as the extensive perspective model. The key difference is that $y_{i,k,s,o,t}$ indicates the hours worked last month (rather than if an individual is employed or not) and I use $\ln(y_{i,k,s,o,t})$ to replace $\text{logit}(y_{i,k,s,o,t})$ in the equation. The equations for school closure and school reopening policies will be as follows:

$$ln(y_{i,k,s,o,t}) = \beta_0 + \beta_1 Female_i + \beta_2 Closure_{s,t} + \beta_3 Closure_{s,t} * Female_{i,t}$$
$$+\gamma * controls + \alpha_k + \alpha_s + \alpha_o \times \alpha_t + \varepsilon_{i,k,s,o,t}$$

[1] For the county level, to analyze the school closure effect, I include data from September 2018 to August 2020.

Static model for state from intensive margin perspective for school reopening:

$$ln(y_{i,k,s,o,t}) = \beta_0 + \beta_1 Female_i + \beta_2 Open_{s,t} + \beta_3 Open_{s,t} * Female_{i,t}$$

$$+\gamma * controls + \alpha_k + \alpha_s + \alpha_o \times \alpha_t + \varepsilon_{i,k,s,o,t}$$

The static model from intensive margin also uses the difference-in-difference method, and the coefficients of interest are β_2 which captures the effect of school closure or reopening on male workers' working hours and β_3 captures the differentiated effect of school policies on female workers' working hours. Thus, the effect of school closure or reopening policies on female workers' employment would be $\beta_2 + \beta_3$.

Similarly, both models for county-level analysis from an intensive perspective and an extensive perspective will be the same as the models for state-level analysis but instead of having state-level policies, I am having county-level policy here.

One potential drawback of the state-level analysis is that there are other differences between each state that I can hardly control for. These include, for example, the demographic and other labor market policies differences. Different labor market policies in each state may affect the labor outcome of that state and if these effects are correlated with school policies then the conclusion of this research can be overstated. Also, in the state-level dataset that I am using, only 20 states have mandatory school closure or reopening policies while the rest of them will let the districts decide whether to close or reopen schools on their own, which makes it hard to know the real opening status within that state. Therefore, I also consider the county-level analysis to overcome the disadvantages at the state level. A county-level analysis (within the same state) accounts for the difference in state-level labor market policies and alleviates some of the concerns. However, even at the county-level there may be some unobservable county actions during the pandemic that may confound our results. These actions would have to be correlated with school closure/reopening policies.

In addition to the static analysis, I also develop a county-level dynamic model of the school reopening effects. Specifically, I separate monthly effect of reopening policies before and after the announcement of school reopening policy.

A dynamic model for reopening from extensive margin perspective is given as follows:

$$y_{i,k,c,o,t} = \beta_0 + \beta_1 Female_i + \sum_{z=-12}^{12} \beta_z \, Open_{c,t,z} + \sum_{z=-12}^{12} \gamma_z \, Open_{c,t,z} * Female_i + \tau * controls + \alpha_k + \alpha_c + \alpha_o \times \alpha_t + \varepsilon_{i,k,c,o,t}$$

Where i indexes an individual, k indicates industry, c indicates county, o indicates occupation and t indexes time, $y_{i,k,c,o,t}$ is the dependent variable and indicates the employment status of the individual at time t, in c county and k industry with occupation o. $Open_{c,t,z}$ indicates the dynamic dummy for the school reopening policy within that individual's county, with *z<0* represents the time before the school has been reopened and *z>0* represents the time after the school has been reopened, for example, $Open_{c,t,-1}$ represents the dummy for one month before the school ordered open and $Open_{c,t,1}$ represents the dummy for 1 month after the school ordered open. $Open_{c,t,z} = 0$ for school has been ordered closed in that county at time t and $Open_{c,t,z} = 1$ for school which is partially opened or fully opened. $Female_i$ indicates the dummy for gender, with $Female_i = 0$ for males and $Female_i = 1$ for females. $controls$ includes a set of controls for age, COVID-19 new cases, vaccine rates, family size, whether the person has a disability, education attainment, whether the person is white, family income. α_k is the fixed effect term for each industry, α_c is the fixed effect term for each county. $\alpha_o \times \alpha_t$ is the cross fixed effect for occupation and time, which will help control the occupation change throughout the time. And $\varepsilon_{i,k,c,o,t}$ is the error term.

The dynamic model at the county-level from an intensive perspective is similar to the extensive margin model. The key difference is that the outcome variable $y_{i,k,c,o,t}$ captures the hours worked (rather than employment status) and I use $\ln(y_{i,k,c,o,t})$ to replace $y_{i,k,c,o,t}$ in the equation.

Data description

This study uses the Basic Monthly data from Current Population Survey data collected every month during the calendar year from January 2010 to September 2021 and ASEC data from the Annual Social and Economic Supplement of the Current Population Survey from 2010 to 2021. It also combines state-level school closure and reopening data from Education Week and county-level school status data only for California from The Safe Schools For All Hub for California to study the employment status of workers that are in the labor force. I combine these data with information on the cumulative counts of coronavirus cases in the United States from The New York Times as well as the county-level coronavirus infections data from The Los Angeles Times. The closure and reopening policy for schools vary between states in the United States. Since the employment tends to have a delay in reaction to the policy change, I round up the school reopening month. For example, if the school reopened on mid-March 2021, then in my dataset, I count April 2021 as the first month of reopening.

After restricting to individuals that are in the labor force and have at least one child aged from 5-18 years old[2] from March 2019 to Sep 2021, I obtained 650,752 individual-level data for all states in the United States and 38,660 individual-level data for California. The available information includes dependent variables: employment status and hours worked last week; variables of interest: workers' sex (=0 is male, =1 is female), school closure indicator (=0 is open, =1 is closed) and school reopening indicator (=0 is still closed, =1 partially closed, =2 fully opened); and control variables such as: age, state, county, occupation, industry, the number of children they have, the number of own children under age 5, the year and month that the workers reported the data, family income of the household, marital status, worked remotely for pay due to COVID, unable to work due to COVID, COVID-19 cumulative cases, COVID-19 new cases, vaccine rate. I expect someone who has no choice but to keep their children in the household to send their children back to school

[2] Employment status for an individual who is not in labor force will not change with different school policies and school closure or reopening policies will only have very limited effect on the employment status for

if they have such option, which in other words, they will send their children to school when schools are hybrid reopening. So that in our empirical regression, I treat the school reopening indicator as a binary independent variable, with the hybrid reopening to be 1 and closed to be 0.

Table 1 provides summary statistics for our dependent variables and control variables, separating into 3 different time periods. The pre-closure part summarizes the data before any school has been closed (from August, 2018 to February, 2020), the closure part summarizes the data after the first school has been closed and before the first school that has been reopened (from March, 2020 to August,2020) while the reopening part accounts for the data after the first school that has been reopened (from September, 2020 to September, 2021). Control variable "white" is a binary variable that indicates whether the worker's race is white (=0 is not white, =1 is white). Control variable "education" is a variable that indicates the worker's education attainment (=0 high school or below, =1 is some college or above). Table 2 offers these summary statistics at the is a county-level.

Table 1. Summary Statistics at the state-level[3]

	(1)			(2)			(3)				
	pre-closure[4]			closure[5]			reopen[6]			t-test	t-test
	Mean	Std. dev.	Obs	Mean	Std. dev.	Obs	Mean	Std. dev.	Obs	(1)-(2)	(2)-(3)

[3] In this table, I restrict our sample size to individuals that are in the labor force and have at least one child aged from 5-18 years old from March 2019 to Sep 2021 for all states and I also separate this time period into 3 smaller periods and conduct 2 t-tests: (1) between pre-closure period and closure period (2) between closure period and reopen period.

[4] The pre-closure period indicates time from September 2018 to February 2020.

[5] The closure period indicates time from March 2020 to July 2020.

[6] The reopen period indicates time from August 2020 to September 2021.

employed (M)[7]	.978	.148	169,706	.944	.229	52,569	.958	.201	107,192	(+)***	(-)***
employed (F)	.969	.173	165,021	.923	.267	50,983	.951	.216	105,281	(+)***	(-)***
hours worked (M)[8]	175.110	47.309	161,914	170.247	48.204	47,718	171.367	47.02193	99,765	(+)***	(-)***
hours worked (F)	150.785	47.718	154,144	145.283	49.391	43,701	148.174	47.02193	95,579	(+)***	(-)***
female	.493	.500	334,727	.492	.500	103,552	.496	.500	212,473	(0)	(-)**
white[9]	.811	.391	334,727	.8117	.3910	103,552	.800	.400	212,473	(0)	(+)***
education 10	.678	.467	334,727	.706	.456	103,552	.688	.463	212,473	(-)***	(+)***
family income	79060.140	44217.600	334,727	83995.600	44188.04	103,552	82359.910	44804.850	212,473	(-)***	(+)***
family size	1.607	.923	334,727	1.541	.779	103,552	1.534	.774	212,473	(+)***	(+)**

Table 2. Summary Statistics at the county-level[11]

	(1)			(2)			(3)				
	pre-closure[12]			closure[13]			reopen[14]			t-test	t-test
	Mean	Std. dev.	Obs	Mean	Std. dev.	Obs	Mean	Std. dev.	Obs	(1)-(2)	(2)-(3)
employed (M)[15]	.972	.165	11,757	.929	.256	3,472	.940	. 238	7,228	(+)***	(-)**

[7] In this table, employed (M) and employed (F) are the dependent variables from extensive margin. And they indicate employment status for male workers and female workers accordingly.
[8] In this table, hours worked (M) and hours worked (F) are the dependent variables from intensive margin. And they indicate the hours worked last month for male workers and female workers accordingly.
[9] White, education, family income and family size are a set of control variables.
[10] Education indicates education background that the individual has. Education =1 if the individual has a degree of college or above and otherwise it is 0.
[11] In this table, I restrict our sample size to individuals that are in the labor force and have at least one child aged from 5-18 years old from March 2019 to Sep 2021 for California only and I also separate this time period into 3 smaller periods and conduct 2 t-tests: (1) between pre-closure period and closure period (2) between closure period and reopen period.
[12] The pre-closure period indicates time from September 2018 to February 2020.
[13] The closure period indicates time from March 2020 to August 2020.
[14] The reopen period indicates time from September 2020 to September 2021.
[15] In this table, employed (M) and employed (F) are the dependent variables from extensive margin. And they indicate employment status for male workers and female workers accordingly.

employed (F)	.960	.195	10,499	.894	.307	2,908	.922	.268	6,448	(+)***	(-)***
hours worked (M)[16]	167.796	40.829	11,170	160.851	42.250	3,066	163.741	42.275	6,608	(+)***	(-)**
hours worked (F)	147.904	44.124	9,772	141.393	47.078	2,375	145.585	47.277	5,668	(+)***	(-)***
female	.468	.499	22,112	.456	.498	6,380	.471	.499	13,676	(+)**	(-)**
white[17]	.763	.425	22,112	.765	.424	6,380	.751	.432	13,676	(0)	(+)**
education	.607	.488	22,112	.652	.476	6,380	.648	.477	13,676	(-)***	(0)
family income	79585.29	46489.15	22,112	87570.53	47027.23	6,380	83951.45	46838.88	13,676	(-)***	(+)***
family size	1.708	1.183	22,112	1.605	.974	6,380	1.622	1.028	13,676	(+)***	(0)

From the extensive margin, based on the t-test between the pre-closure period and the closure period for the dependent variable: employed, the positive and significant implies that for all workers regardless of their industry or occupations, after school has been closed, their probability of getting employed will be decreased by around 3-4%. The negative sign of the t-test between closure period and reopening period shows that after the school has been reopened, the probability of getting employed will increase. Although the probability of getting employed recovered after, it does not return to the level before the pandemic. There are still plenty workers missing in the labor market due to the COVID-19 pandemic. The number of these missing workers can be even larger as those who lost their jobs may not be accessible

[16] In this table, hours worked (M) and hours worked (F) are the dependent variables from intensive margin. And they indicate the hours worked last month for male workers and female workers accordingly.
[17] White, education, family income and family size are a set of control variables.

for this survey. This will inevitably cause bias to the research that the actual gender employment can be larger, and the effect of school policies can be overstated.

From the intensive margin, based on the t-test between the pre-closure period and the closure period for the dependent variable: hours worked last month, the positive and significant implies that for all workers regardless of their industry or occupations, after school has been closed, their working hours will be reduced by 5-7 hours per month. The negative sign of the t-test between closure period and reopening period shows that after the school has been reopened, the working hours will be increased by 2-4 hours per month. This also suggests that the negative impact on working hours caused by COVID-19 has not been fully removed by the reopening of schools.

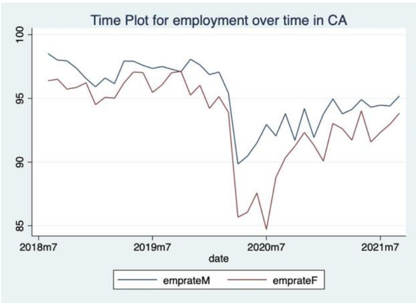
The t-test for female in state-level between pre-closure period and closure period is 0 while it is negative for the t-test between closure and reopening period. This indicates that there are more women in the state-level dataset after the school has been reopened. However, the t-test for female in state-level between pre-closure period and closure period is positive while it is negative for the t-test between closure and reopening period. This shows that, in California, there are fewer female workers after school has been closed and more women after the school has been reopened.

The t-test for our control variables shows that after school has been reopened, there are more white people in the dataset. It also shows that there are more individuals with at least some college education after school has been closed while after school has been closed, this returns to the normal pre-closure level. These two findings can be highly related to the self-selection of labor market in response to COVID-19 that again proves the necessity to control for these variables in the regression.

Figure 1 shows the employment over time in United States for all states and Figure 2 shows the employment over time in California.

Figure 1. Time plot for employment over time in US

Figure 2. Time plot for employment over time in CA

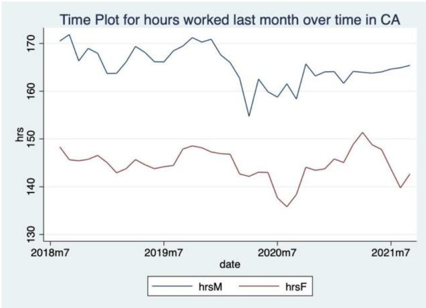


Figures 1 and 2 are visualizations for dependent variable from extensive margin: employment. The blue line represents male workers and red line represents female workers. From September 2018 to September 2021, male workers are more likely to be employed than female workers. Both figures show that when COVID-19 hit the labor market, the probability for both male workers and female workers dropped significantly. However, there is a notable difference in the pandemic effect between male and female workers. The impact is much greater for female workers.

Figure 3 shows the # of hours worked over time in United States for all states and Figure 4 details this information for California only.

Figure 3. Time plot for hours worked last month over time in US

Figure 4. Time plot for hours worked last month over time in CA



Figures 3 and 4 are visualizations for dependent variable from intensive margin: hours worked last month. The blue line is for male workers and red line is for female workers. From September 2018 to September 2021, male workers tend to work more hours each month than female workers while from intensive margin, there is no clear drop in working hours caused by COVID-19 pandemic.

Table 3 below is a detailed timetable of school closure for the states that have a clear state-level closure/reopening policy. It excludes the states that delegate the decision to individual school districts. Input "C" implies that in that month, the schools in that state were ordered to close while Input "O" implies that in that month, the schools were ordered to at least reopen in a hybrid format or fully open that month. For example, for state Texas, all schools were closed from March 2020 until August 2020 when the schools reopened.

Table 3. A timetable of school policies for selected States with exact information[18]

	2020										2021					
State	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
CA	C	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O

[18] Other states that are not in this table do not have state-level closure or reopening policies and each school districts can decide on their own whether to open or close the schools in its area.

DC	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
HI	C	C	C	C	C	C	C	O	O	O	O	O	O	O	O	O
NM	C	C	C	C	C	C	C	O	O	O	O	O	O	O	O	O
VT	C	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O
WV	C	C	C	C	C	C	C	O	C	O	O	O	O	O	O	O
OR	C	C	C	C	C	O	C	O	O	O	O	O	O	O	O	O
KY	C	C	C	C	C	O	O	O	C	O	O	O	O	O	O	O
AR	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O	O
IA	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O	O
MO	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O	O
TX	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O	O
FL	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O	O
AZ	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O	O
NH	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O	O
MA	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O	O
NC	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O	O
SC	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O	O
WA	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O	O
KS	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O	O

From Table 3 I see that many states implemented statewide school closure mandates starting in March 2020. However, it also demonstrates a large degree of variation in the timing of reopening. Texas, for example, was the first to reopen in August 2020, while Washington DC was the last to reopen in February 2021.

Table 4 below is a detailed timetable of school closure for all countries in California. For example, for county Alameda, all schools have been closed from March 2020 until February 2021 when the schools have been reopened.

Table 4. A timetable of school policies for counties in California

	2020										2021					
County	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
Alameda	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Butte	C	C	C	C	C	C	C	C	O	O	O	O	O	O	O	O
Contra Costa	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O	O
El Dorado	C	C	C	C	C	C	C	O	O	O	O	O	O	O	O	O
Fresno	C	C	C	C	C	C	C	C	O	O	O	O	O	O	O	O
Humboldt	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Imperial	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Kern	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
King	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Los Angeles	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Madera	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Marin	C	C	C	C	C	C	C	C	O	O	O	O	O	O	O	O
Merced	C	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O
Monterey	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Napa	C	C	C	C	C	C	C	O	O	O	O	O	O	O	O	O
Orange	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Placer	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Riverside	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Sacramento	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
San Bernardino	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
San Diego	C	C	C	C	C	C	O	O	O	O	O	O	O	O	O	O
San Francisco	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
San Joaquin	C	C	C	C	C	C	C	C	O	O	O	O	O	O	O	O
San Luis Obispo	C	C	C	C	C	C	C	C	O	O	O	O	O	O	O	O
San Mateo	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Santa Barbara	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Santa Cruz	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Shasta	C	C	C	C	C	C	C	O	O	O	O	O	O	O	O	O
Solano	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O

Sonoma	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Stanislaus	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Tulare	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Ventura	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O
Yolo	C	C	C	C	C	C	C	C	C	C	C	O	O	O	O	O

From Table 4, I see less variation in the timing for reopening in California. With county Merced and San Diego being the first two counties that open in September 2020, most other counties reopen around February 2021.

Results:

In this section, I provide and discuss the empirical results. I begin with the static state-level analysis followed by the static county-level analysis and ended with the dynamic county-level analysis and

Table 5: Static State-level logistic regression (hybrid opening[19])[20]

	Extensive		Intensive	
	(1)	(2)	(3)	(4)
VARIABLES	closure	reopen	closure	reopen
female	-0.407***	-0.438***	-0.188***	-0.187***
	(0.0262)	(0.0364)	(0.00182)	(0.00467)
school closure	-1.728***		-0.0747***	
	(0.181)		(0.0199)	
female*closure	0.00753		0.00118	
	(0.0413)		(0.00390)	
school reopening		-0.0203		0.000243
		(0.0697)		(0.00816)
female*reopening		0.216***		0.0193***
		(0.0429)		(0.00514)
age	0.249**	0.115	-0.0130	-0.0310***
	(0.100)	(0.101)	(0.00856)	(0.0112)
new cases	-5.408***	-0.697	-0.220	0.0131
	(1.144)	(0.440)	(0.137)	(0.0498)

[19] Table for static model with full reopening is attached in the appendix.
[20] In this model, I conduct logistic regression from both extensive and intensive margin and include state fixed effect, industry fixed effect and occupation*time fixed effect.

vaccine rate		-0.00268*		-0.000233
		(0.00152)		(0.000155)
family size	-0.0629***	-0.0672***	0.00169**	0.00208*
	(0.00831)	(0.00982)	(0.000795)	(0.00122)
income	15.09***	15.11***	0.544***	0.731***
	(0.269)	(0.266)	(0.0186)	(0.0244)
difficulty	-0.640***	-0.580***	-0.0847***	-0.0681***
	(0.0398)	(0.0427)	(0.00463)	(0.00618)
education	0.0250	-0.0356*	-0.0141***	-0.0155***
	(0.0200)	(0.0202)	(0.00178)	(0.00237)
white	0.355***	0.310***	-0.0186***	-0.0174***
	(0.0207)	(0.0209)	(0.00189)	(0.00244)
Constant	3.008***	2.677***	5.253***	5.246***
	(0.183)	(0.168)	(0.0127)	(0.0155)
Observations	425,051	276,736	395,627	250,406
R-squared			0.078	0.071
State FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Occupation*Time FE	YES	YES	YES	YES

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

Table 5 shows the static state-level logistic regression from both extensive and intensive perspectives, which includes two separate models: one for the effect of school closure, one for the effect of school reopening.

All estimates show a negative relationship between being female and getting employed. For example, when I look at column 2, the coefficient = -0.407 (SE=0.0262) indicates that the female group has $e^{-0.407} = 0.67$ times odds of the male group being employed. Female workers tends to be 33% less likely to be employed than the male workers within the same state and in the same industry and same occupation at the same time. When isolating the school closure effect, the coefficient= -1.728 (SE=0.181) indicates that after the school closure policy has been released, it has $e^{-1.728} = 0.17$ the odds of employment before the school closure policy has been released for male. In other words, it means that a worker tends to be 83% less likely to be employed when the school has been closed than a worker within the same state and in the same industry and same occupation at the same time before the school has been closed. The effect of the school closure policy on female employment will be similar to male while it is not statistically significant. When isolating the school reopening effect, the coefficient= -0.0203 indicates that after the school closure policy has been released, it has $e^{-0.0203} = 98\%$ the odds of employment before the school closure policy has been released for male. And this is not statistically significant which means that the school reopening does not have a significant effect on male employment. While the effect on female employment is different as the coefficient=0.216 (SE=0.0262) suggests that the female group has $e^{0.216} = 1.22$ times the odds of the being employed after school reopened. This means that compared to a male worker in the same state with same industry and occupation under the same time, female workers tend to be 22% more likely to be employed. This result proves that with the reopening policy, women are more likely to get employed and the gender employment gap will be reduced.

At the intensive margin, this study uses linear regression model with log transformation for state-level intensive margin analysis. From Column 3 the coefficient= -0.188 (SE=0.00182) suggests that being a female tends to work around $e^{-0.188} = 82\%$ of the hours that a male worker within the same state and in the same industry and same occupation at the same time. When isolating the effect of school closure, it shows a significant negative effect on males working hours per month. When looking at Column 4 the coefficients show that with school reopening policy has negligible effects on male and female working hours.

Table 6: Static County-level logistic regression (hybrid opening[21])[22]

	Extensive		Intensive	
	(1)	(2)	(3)	(4)
VARIABLES	closure	reopen	closure	reopen
female	-0.661***	-0.626***	-0.166***	-0.163***
	(0.0904)	(0.0957)	(0.00646)	(0.0124)
school closure	-1.413**		-0.0427	
	(0.630)		(0.0838)	
female*closure	-0.130		0.0115	
	(0.146)		(0.0144)	
school reopening		-0.221		-0.0135
		(0.194)		(0.0221)
female*reopening		0.294**		0.00823
		(0.132)		(0.0160)
age	-0.675*	-0.703**	-0.0528*	-0.0108
	(0.348)	(0.342)	(0.0308)	(0.0433)
new cases	-2.559	-0.263	-0.399	0.0232
	(3.551)	(0.748)	(0.410)	(0.0934)
vaccine rate		0.501		0.0309
		(0.978)		(0.120)
family size	-0.108***	-0.0511*	0.000157	0.00414
	(0.0232)	(0.0281)	(0.00231)	(0.00369)
income	13.41***	14.26***	0.528***	0.889***
	(0.935)	(0.902)	(0.0680)	(0.0977)
difficulty	-0.838***	-0.805***	-0.0737***	0.0520
	(0.169)	(0.179)	(0.0199)	(0.0321)
education	0.0109	-0.191***	-0.0245***	-0.0218**
	(0.0792)	(0.0735)	(0.00679)	(0.00977)
white	0.230***	0.376***	0.0168***	-0.00497
	(0.0796)	(0.0719)	(0.00647)	(0.00896)
Constant	2.143***	2.915***	5.108***	5.075***

[21] Table for static model with full reopening is attached in the appendix. I found that with hybrid reopening, it is enough for the female employment to rise that counting full-reopening as 1 is the analysis will eventually mess up the results.
[22] In this model, I conduct logistic regression from both extensive and intensive margin and include county fixed effect, industry fixed effect and occupation*time fixed effect.

	(0.542)	(0.551)	(0.0467)	(0.0606)
Observations	24,995	16,741	25,595	15,414
R-squared			0.079	0.081
County FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Occupation*Time FE	YES	YES	YES	YES

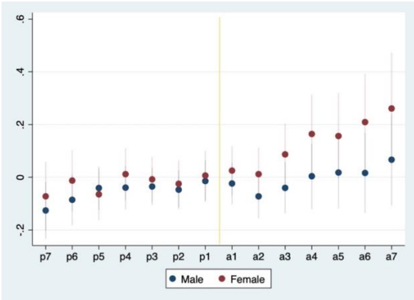
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6 shows similar static results as Table 5 at the county-level. When isolating school closure/reopening effects, both models show a statistically significant negative relationship between being female and getting employed. For example, when I look at column 2, the coefficient =-0.661 (SE=0.0904) indicates that the female group has $e^{-0.661} = 0.53$ times the odds of the male group being employed. Thus, I find that female workers to be 47% less likely to be employed than the male workers in the same state, with same industry and occupation at the same time. The effect of the school reopening policy on male employment is negative with the coefficient from Column 3 =-0.221 (SE=0.194) and the effect will be positive for female employment as the coefficient =0.294 (SE=0.132) and is statistically significant. This means that the log odds of female workers being employed are $e^{0.294} = 1.34$ the odds for male workers. This means that compared to a male worker in the same county with same industry and occupation under the same time, female workers tend to be 34% more likely to be employed after schools reopen.

The linear regression model with log transformation is used for state-level intensive margin analysis. From Column 3, the coefficient = -0.166 (SE=0.00646) suggests that being a female tends to work around $e^{-0.166} = 84\%$ of the hours that a male worker worked per month. When isolating the effect of school closure, it shows a significant negative effect on males working hours per month. When looking at Column 3, the coefficients show that with school closure policy in effect, it generates a negative effect on males working hours and is statistically significant while it generates an almost 0 and non-significant positive for female * school closure on working hours. This means that the school closure policy shows a negative effect on male workers and there is no difference between female and male workers according to this. Similarly, based on the standard errors from the results, the sample mean is an accurate reflection of the actual population mean while the standard errors for the county-level analysis are bigger than the ones from state-level analysis, which can be due to the sample size difference between these two analyses.

All models, regardless of extensive or intensive, state-level or county-level returned similar coefficients for other variables. Age has a positive relationship with employment or working hours and it makes sense that companies tend to hire or use experienced workers for the state-level analysis while for the county-level analysis, it is negative. This can be due to the difference in the industry structure between California and other states. New cases have a negative effect on employment or working hours as people are not able to work or work as many hours as they used to be because of the severe situation of COVID-19. And with a larger family size, the possibility of being employed will decrease. It is also true that with higher family income, people would be less likely to be unemployed. Workers with a disability will be more likely to be unemployed and there is an advantage for white people in finding jobs.

Figure 5: Dynamic county-level linear regression from an extensive perspective (hybrid opening)[23]



Figures 5 and 6 below showed the results from the dynamic county-level analysis. The yellow line in the middle represents the cutting line for reopening policy. p-x represents x months before the event and a-x represents x months after the event. Figure 5 shows that there is no significant difference between males and females before the event (school reopening in this case). While a significant disparity of employment can be seen after the school reopening policy has been in effect: female workers are more likely to be employed than male workers.

Figure 6: Dynamic county-level linear regression from an intensive perspective[24]

---

[23] This uses the dynamic model mentioned earlier in the paper. The dependent variable here is employed. It contains county fixed effect, industry fixed effect and occupation*time fixed effect.

[24] This uses the dynamic model mentioned earlier in the paper. The dependent variable here is hours worked last month. It contains county fixed effect, industry fixed effect and occupation*time fixed effect.

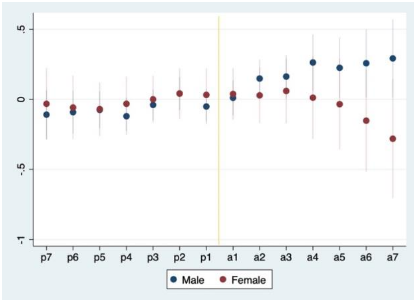


Figure 6 illustrates the dynamic effects at the intensive margin and shows that there is no significant difference between males and females before the event (school reopening in this case). While a significant disparity of working hours can be seen after the school reopening policy has been in effect: female workers work fewer hours than male workers.

Heterogeneity test:

Table 7: A heterogeneity test based on different income levels has been performed

	Extensive Margin-Income levels					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	low[25]	medium[26]	high[27]	low	medium	high
female	-0.244***	-0.312***	-0.417***	-0.199	-0.269***	-0.556***
	(0.0663)	(0.0430)	(0.0395)	(0.151)	(0.0647)	(0.0464)
school closure	-0.464	-1.415***	-2.330***			
	(0.549)	(0.319)	(0.247)			
female*closure	0.130	0.0639	-0.139**			
	(0.140)	(0.0706)	(0.0571)			
school reopening				-0.333	0.0151	-0.000839
				(0.253)	(0.116)	(0.0942)
female*reopening				0.160	0.115	0.350***
				(0.165)	(0.0742)	(0.0571)

25

26

27

Constant	1.108**	3.177***	4.918***	1.054	2.337***	3.986***
	(0.473)	(0.319)	(0.274)	(0.738)	(0.314)	(0.252)
Observations	16,113	91,749	316,667	9,088	56,563	210,710

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

Table 7 shows how different income groups react to the school closure and reopening policies. From an extensive margin, female workers with high-income levels are the most affected group by the school closure while they are also the group which recovered the most after the school has been reopened.

Table 8: A heterogeneity test for different income levels from an intensive margin

	Intensive Margin-Income levels					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	low	medium	high	low	medium	high
female	-0.187***	-0.150***	-0.198***	-0.274***	-0.151***	-0.189***
	(0.0120)	(0.00400)	(0.00208)	(0.0419)	(0.0120)	(0.00509)
school closure	-0.0655	-0.0739*	-0.0805***			
	(0.131)	(0.0426)	(0.0234)			
female*closure	-0.0337	-0.0122	0.00855**			
	(0.0301)	(0.00935)	(0.00436)			
school reopening				-0.0369	0.0172	-0.00407
				(0.0659)	(0.0191)	(0.00906)
female*reopening				0.0961**	0.00484	0.0192***
				(0.0449)	(0.0130)	(0.00562)
Constant	5.134***	5.278***	5.299***	5.376***	5.133***	5.278***
	(0.0871)	(0.0318)	(0.0146)	(0.139)	(0.0433)	(0.0173)
Observations	13,366	83,339	298,922	6,690	48,361	195,355
R-squared	0.128	0.066	0.076	0.140	0.068	0.066

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

Similar results can be found from Table 8, high-income group is still the most negatively affect group, with female workers decreasing their working hours the most. However, after the school has been reopened, female workers from the low-income group increase their working hours even more than the high-income groups. This can be due to the low threshold for entering the low-income jobs than the high-income jobs.

Discussion:

In conclusion, I found that for a worker, who has at least one child in the household, compared to a worker with the same occupation, same county, same industry, he/she is around 76% less likely to be employed when school has been closed while a female worker tends to be 34% more likely to be employed than a male worker after the school has been reopened. Based on our results, both state-level and county-level regression with our assumption that with school reopening policy in effect, women will benefit more than men. More women will return to the labor market and back to work, thus I have seen a significant positive effect on female works employment. This study also confirms the negative effect of school closure, and this effect does not differentiate between female workers and male workers.

However, the school closure or reopening seems to be too large. One potential explanation to this is that this study restricts to individuals with at least one child at home, which would automatically amplify the effect. Also, this study fails to incorporate other Covid-19 issues as it only includes Covid-19 cases and vaccine rate to account for the Covid-19 effects on employment. The effect from other uncontrolled Covid-19 issues will be buried under the effect of school policies thus the conclusion can be overstated for this study.

Interestingly, unlike the employment has been increased significantly for female workers, the working hours per month has been reduced after the school has been reopened. This can be due to more women are being employed after school has been reopened, thus those who remained at the labor market may not be able to work as many hours as they used to be. Also, I sometimes see the small negative effect of school reopening on males, this can be due to the development cost for females. According to Naila Kabeer (2020), there is no free lunch for gender development. Women workers' development can sometimes build up at the cost of male workers.

Although this study qualitatively demonstrates that school reopening policy has a significant positive effect on women's employment from an extensive perspective, it limits its county-level analysis to California. Unlike other states in the United States, California plays a unique role in United States as its advantageous industries include traditional agriculture, cutting-edge high-tech industries, and extremely developed tourism. It is also the technology and cultural center of the United States and the world's film and television center. It is also a state highly engaged in international trade as international trade accounts for 25% of California's GDP, and 45% of U.S. imports pass through California ports. The uniqueness of California made it hardly able to represent United States. So that the conclusion drawn from the county-level analysis should be checked before applying to other states in United States.

Appendix:

Appendix table 1: Static State-level logistic regression (full opening)

	Extensive			Intensive		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	closure only	reopen only	closure+reopen	closure only	reopen only	closure+reopen
female	-0.407***	-0.299***	-0.313***	-0.188***	-0.178***	-0.184***
	(0.0262)	(0.0220)	(0.0187)	(0.00182)	(0.00229)	(0.00148)
sch_reopen		-0.0475	-0.0514		-0.0237***	-0.0221***
		(0.0479)	(0.0418)		(0.00484)	(0.00366)
female_reopen		0.0700	0.0969*		0.0350***	0.0398***
		(0.0503)	(0.0495)		(0.00488)	(0.00452)
age	0.00249**	0.00115	0.00189**	-0.000130	-0.000309***	-0.000181***
	(0.00100)	(0.00101)	(0.000765)	(8.56e-05)	(0.000112)	(6.97e-05)
new_cases	-5.408***	-0.662	-2.149***	-0.220	0.0214	-0.0213
	(1.144)	(0.441)	(0.379)	(0.137)	(0.0499)	(0.0399)
fully_vacc_rate		-0.00276*	-0.000675		-0.000236	-0.000358***
		(0.00152)	(0.00142)		(0.000155)	(0.000130)
pernum	-0.0629***	-0.0674***	-0.0624***	0.00169**	0.00205*	0.00191***
	(0.00831)	(0.00981)	(0.00674)	(0.000795)	(0.00122)	(0.000678)
income	1.51e-05***	1.51e-05***	1.60e-05***	5.44e-07***	7.30e-07***	6.06e-07***
	(2.69e-07)	(2.66e-07)	(2.07e-07)	(1.86e-08)	(2.44e-08)	(1.52e-08)

difficulty	-0.640***	-0.580***	-0.641***	-0.0847***	-0.0684***	-0.0779***
	(0.0398)	(0.0427)	(0.0307)	(0.00463)	(0.00618)	(0.00379)
education	0.0250	-0.0356*	-0.00967	-0.0141***	-0.0155***	-0.0146***
	(0.0200)	(0.0202)	(0.0153)	(0.00178)	(0.00237)	(0.00146)
white	0.355***	0.309***	0.346***	-0.0186***	-0.0174***	-0.0175***
	(0.0207)	(0.0209)	(0.0158)	(0.00189)	(0.00244)	(0.00153)
sch_clo	-1.728***		-0.0850	-0.0747***		-0.0147**
	(0.181)		(0.0661)	(0.0199)		(0.00742)
female_clo	0.00753		-0.0692*	0.00118		-0.00117
	(0.0413)		(0.0360)	(0.00390)		(0.00367)
Constant	3.008***	2.605***	3.252***	5.253***	5.241***	5.258***
	(0.183)	(0.167)	(0.168)	(0.0127)	(0.0154)	(0.0114)
Observations	425,051	276,736	650,752	395,627	250,406	602,688
R-squared				0.078	0.071	0.075

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Appendix table 2: Static County-level logistic regression (full opening)

	Extensive			Intensive		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	closure only	reopen only	closure+reopen	closure only	reopen only	closure+reopen
female	-0.661***	-0.452***	-0.476***	-0.166***	-0.154***	-0.161***

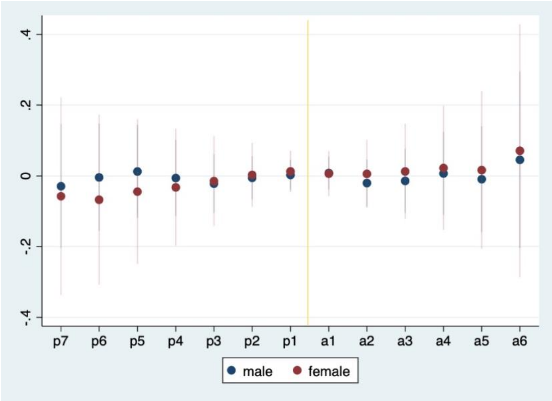
	(0.0904)	(0.0753)	(0.0701)	(0.00646)	(0.00920)	(0.00583)
sch_clo	-1.413**		0.266	-0.0427		0.0146
	(0.630)		(0.195)	(0.0838)		(0.0204)
female_clo	-0.130		-0.189*	0.0115		-0.00153
	(0.146)		(0.109)	(0.0144)		(0.0114)
age	-0.00675*	-0.00698**	-0.00820***	-0.000528*	-0.000118	-0.000228
	(0.00348)	(0.00342)	(0.00257)	(0.000308)	(0.000433)	(0.000258)
new_cases	-2.559	-0.235	-0.186	-0.399	0.0476	-0.0427
	(3.551)	(0.730)	(0.690)	(0.410)	(0.0922)	(0.0780)
fully_vacc_rate	748,497	0.466	-0.248	16,174	0.0217	0.0234
	(1.890e+06)	(0.972)	(0.862)	(41,232)	(0.120)	(0.0937)
pernum	-0.108***	-0.0488*	-0.0769***	0.000157	0.00406	0.00159
	(0.0232)	(0.0280)	(0.0188)	(0.00231)	(0.00369)	(0.00201)
income	1.34e-05***	1.42e-05***	1.44e-05***	5.28e-07***	8.87e-07***	6.45e-07***
	(9.35e-07)	(9.01e-07)	(6.98e-07)	(6.80e-08)	(9.76e-08)	(5.74e-08)
difficulty	-0.838***	-0.809***	-0.829***	-0.0737***	0.0511	-0.0483***
	(0.169)	(0.179)	(0.129)	(0.0199)	(0.0321)	(0.0175)
education	0.0109	-0.188**	-0.102*	-0.0245***	-0.0220**	-0.0247***
	(0.0792)	(0.0735)	(0.0570)	(0.00679)	(0.00977)	(0.00571)

white	0.230***	0.378***	0.278***	0.0168***	-0.00514	0.00771
	(0.0796)	(0.0718)	(0.0569)	(0.00647)	(0.00896)	(0.00539)
sch_reopen		0.240	0.295*		-0.0139	-0.00708
		(0.175)	(0.171)		(0.0185)	(0.0165)
female_reopen		-0.185	-0.143		-0.0205	-0.0157
		(0.179)	(0.177)		(0.0194)	(0.0173)
Constant	2.133***	3.014***	2.627***	5.132***	5.094***	5.111***
	(0.554)	(0.560)	(0.508)	(0.0476)	(0.0623)	(0.0445)
Observations	24,995	16,741	38,660	25,595	15,414	38,495
R-squared				0.079	0.081	0.077

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Appendix Figure 1: Dynamic county-level linear regression from extensive perspective (full opening)



Reference:

1. World Economic Forum. The global gender gap report 2020. (2019, December 16). Retrieved from https://www.weforum.org/reports/gender-gap-2020-report-100-years-pay-equality

2. World Economic Forum. The global gender gap report 2021. (2021, March 31). Retrieved from https://www.weforum.org/reports/ab6795a1-960c-42b2-b3d5-587eccda6023

3. Albanesi, Stefania, and Jiyeon Kim. 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

4. Hall, Robert E., and Alan B. Krueger. 2012. "Evidence on the Incidence of Wage Posting, Wage Bargaining, and On-the-Job Search." *American Economic Journal: Macroeconomics*, 4 (4): 56-67.DOI: 10.1257/mac.4.4.56

5. Sabina Irimie, Roland Moraru, Lucian-Ionel Cioca, & Maria – Elena Boatcă. (2014). Aspects of the gender inequality issue in knowledge society careers. *Polish Journal of Management Studies, 9*, 43. Retrieved from Publicly Available Content Database database. Retrieved from https://search.proquest.com/docview/2505535941

6. Barigozzi, Francesca and Cremer, Helmuth and Monfardini, Chiara, The Gender Gap in Informal Child Care: Theory and Some Evidence from Italy (June 2019). CEPR Discussion Paper No. DP13782, Available at SSRN: https://ssrn.com/abstract=3401869

7. Mark Aguiar, Erik Hurst, Measuring Trends in Leisure: The Allocation of Time Over Five Decades, *The Quarterly Journal of Economics*, Volume 122, Issue 3, August 2007, Pages 969–1006, https://doi.org/10.1162/qjec.122.3.969

8. Center for Global Development, The Global Childcare Workload from School and Preschool Closures During the COVID-19 Pandemic, Retrieved from https://www.cgdev.org/publication/global-childcare-workload-school-and-preschool-closures-during-COVID-19-pandemic

9. Cortes, Patricia, and Jessica Pan. 2018. "Occupation and gender." The Oxford handbook of women and the economy, pp. 425–452.

10. Yamamura, E., Tsustsui, Y. The impact of closing schools on working from home during the COVID-19 pandemic: evidence using panel data from Japan. *Rev Econ Household* 19, 41–60 (2021). https://doi.org/10.1007/s11150-020-09536-5

11. Collins C, Ruppanner L, Christin Landivar L, Scarborough WJ. The Gendered Consequences of a Weak Infrastructure of Care: School Reopening Plans and Parents' Employment During the COVID-19 Pandemic. Gender & Society. 2021;35(2):180-193. doi:10.1177/08912432211001300

12. Fabrizio, S. (2021). COVID-19 she-cession. Washington, DC: International Monetary Fund. doi:10.5089/9781513571157.001

13.Naila Kabeer (2020) Women's Empowerment and Economic Development: A Feminist Critique of Storytelling Practices in "Randomista" Economics, Feminist Economics, 26:2, 1-26, DOI: 10.1080/13545701.2020.1743338

Religion and Risk Preferences: Cross Cultural Evidence from Pakistan TBD and the United States

Franklin Marquette & Zoraiz Paracha

1. Introduction

Religion, that is a “shared set of beliefs, activities, and institutions premised upon faith in supernatural forces” (Iannaccone, 1998), has comforted billions for millennia. It is a unique creation birthed of human’s peculiar psychology and in its name civilizations have been built, wars have been waged, and peoples have been decimated. One can hardly look back upon our history and pick out an era, subject matter, or structure not influenced in some way by religion, from globalization, to science, to secular institutions. Despite its seemingly obvious influences on human behavior in general, and thus on economic behavior in particular, research regarding the link between religion and economics is relatively limited. Hampered by the desire to cram humans into the bounds of rational choice theory and its quintessential figure, homo-economicus, significant inquiry into this fascinating subject has been slow to pick up steam, especially outside of the developed world. Despite this, the rise of behavioral economics in recent decades has provided the academic energy needed to bring the study of the intersection of religion and economics to the level of understanding it deserves.

This intellectual curiosity has fueled countless papers uncovering the many mechanisms linking religion, economic behavior, and the macroeconomy at large; one of the most influential being religion’s impact on risk preferences. In this paper we extend Noussair et. al’s (2012) pathbreaking research and explore four different themes regarding the link between religion and both monetary and psychometric risk preferences: those are 1) the effects of religious adherence on risk preferences, 2) the differences in risk preferences across Christians and Muslims, 3) the effects of belief versus belonging measures of religiosity on risk preferences, and 4) the effects of gambling opinions on risk preferences. Our data, drawn mostly from the United States and Pakistan, show that religious individuals may be less risk averse than non-religious. Further, Muslims are less risk averse on monetary measures compared to Christians, who are themselves possibly less risk averse on psychometric measures. We also discover that greater frequencies of prayer outside of religious services, a measure of both religious belief and religious belonging, may be correlated with lower psychometric risk aversion. However, we could not pinpoint what specific mechanism, whether it be belief measures or social belonging measures, drive risk variances across religious and non-religious individuals. Lastly, we find that gambling aversion is correlated with higher psychometric risk aversion, and uncorrelated with monetary risk aversion.

In the following section, we will analyze the literature landscape surrounding the intersection of religion and economics. Next, we will present our experimental design and data, followed by a discussion of our empirical model. We will discuss our results and the implications of those results for each of our four subjects of study. Finally, we will present the limitations of our research and conclude.

2. Literature Review

Since the work of Weber (1930) scholars have attempted to pin down the mechanisms that relate religion, risk preferences, and economic behavior. Microeconomic research has revealed a strong relationship between religion and economic activity. Research has also shown that religious beliefs have an impact when it comes to corporate decision making in both the United States and China (Hilary and Hui, 2009; Li and Xu, 2020), enrolling in insurance plans (Gill, Mand, Bigger, and Mathur, 2017), and investing principles (Al-Awadhi and Dempsey, 2017). Moreover, the microeconomic interactions lay the groundwork for macroeconomic research around religion and economics. Barrow and McCleary found that religious beliefs of hell and heaven are found to be positively correlated with prosperous institutional structure and a growing macroeconomy caused by the reiteration of reward and punishment for unethical actions in religions. However, there is a more nuanced approach towards understanding the mechanism which links religion and the broader economy.

One potential avenue linking religion and economic activity is through preferences for risk. The social identity theory suggests that self-categorization: a product of membership in a social group such as their religion, forms an individual's self-image. This association influences people's behaviors as they internalize the attitudes, beliefs, and values of their group. Therefore, the moral and ethical teachings of religions might directly influence its followers to behave in a certain way. Both Islam and Christianity generally promote risk aversion in financial and other matters. While Christianity suggests diversification of portfolios (Ecclesiastes 11:2), Islam prohibits investing in risky assets and gambling (Quran, 5:90). Because religious people act on and internalize these risk averse rules, they might be more risk averse than non-religious people. That is, religious figures and community may function as a sort of monitoring system, enforcing rules that limit risk behavior. Related to this is the risk-preference thesis, which argues that in Western societies, where the dominant religions such as Christianity, Islam, and Judaism clearly define non-affiliation as a risky behavior, irreligiousness can be a form of risk-taking behavior that only less risk averse individuals would undertake (Miller and Hoffmann, 1995; Miller, 2000: Miller and Stark, 2002). This does not hold true in Eastern societies where irreligiosity is not an explicit form of risk-taking, such as in Hinduism, Buddhism, and Shintoism (Miller, 2000).

Through risk preferences, religion can channel immense influence over economic behavior and activity. Risk aversion has been linked to behaviors in saving policy (Eeckhoudt and Schlesinger, 2008), healthy policy (Bui, Phuong, Crainich and Eeckhoudt 2005), insurance

(Eeckhoudt, Louis, Meyer, and Ormiston, 1997), and valuation of life (Eeckhoudt and Hammitt, 2001). Therefore, exploring the link between risk aversion and religion can serve paramount importance in understanding its broader economic impact and thus potential policy structure. Further, given that religion is largely affected by culture, exploring the influence that religion has on economic behavior is important in understanding how economic behavior and resulting outcomes vary across cultures (Guiso et. al. 2003, 2003, 2006; McCleary and Barro, 2006; Shu et. al, 2010; Weber, 2014).

Expanding the inferences gleaned from religious texts such as the Bible and Quran, there is vast academic literature discussing the link between religious adherence and risk aversion which predominantly find a positive correlation. One approach to study this link is to correlate county- or regional- religiosity with the financial behavior of companies, firms, or individuals. Hilary and Hui (2009) analyzed the role of religion in corporate decision making in the US. Their research found that firms located in regions with higher levels of religiosity[1] exhibited a lower risk tolerance as shown by their returns on investments and growth rate. Similar to this, Gao, Wang, and Zhao (2017) found that hedge funds operating in religious counties, measured as in Hilary and Hui (2009), tend to be more risk averse and exhibit greater asset diversification. Within religious groups themselves, Shu et. al (2010) find that mutual funds in low-Protestant or high-Catholic areas have higher fund return volatilities, implying lower risk aversion. Kumar et. al (2011) extend this and find that individuals in regions with higher Catholic-Protestant ratios (CPRATIO) have a greater propensity to hold risky individual stocks. They connect this preference to a greater propensity to gamble on the part of Catholics relative to Protestants. These United States-based

[1] Country level religiosity was measured as the ratio of church members in a district relative to the total population of that district.

results suggest both that financial risk preferences may vary across religious affiliation, but also that religious individuals in general share a common aversion to risk.

External to the US, Al-Awadhi and Dempsey (2017) examined Gulf Cooperation Council (GCC) countries which have defined religious rules (a proxy for religiosity). Their study found that Islamic stock markets (governed on Islamic principles of risk aversion) completely avoid high growth value non-Islamic stocks. This has served as an impediment to their global competitiveness. In a similar approach, Alami, Al-Awadhi, Hassan, and Turunen-Red (2018) examined firms in Saudi Arabia, which also has clearly defined religious rules (it is a predominantly Muslim country) and found that firms frequently avoided taking on debt as part of their capital structure, a signal of high risk aversion.

Another approach is to correlate narrower household- or individual-level religiosity data to individual measures of risk taking. Unlike the population-level studies surveyed, which predominantly use macroeconomic financial measures of risk, the individual-level studies use a wide array of risk measurement techniques. Bartke & Schwarze, 2008; Miller, 2000; Liu, 2010; Dohmen et al., 2011; Weber, 2014 use general risk measurements while Miller & Hoffman, 1995; Hong et. al, 2004; Renneboog & Spaenjers, 2011 use specific risk measurement. Moreover, Weber, Blais, & Betz, 2002, use multidimensional and psychometric measures (i.e. "How likely are you to try bungee jumping?"). Barsky et al.,1997, Kahneman and Tversky, 1979, p. 265 leverage hypothetical situation response measures. Other measures include risky monetary decision tasks with real (Benjamin et. al, 2009) or hypothetical stakes (Holt, 2007; Noussair et. al., 2012)[2].

Spanning 11,707 observations, Barsky et. al (1997) find that Protestants are more risk averse than Catholics, as measured by household's responses behavior to risky hypothetical real-

[2] Note that all the individual- or household-level studies reviewed use survey data. Holt (2007) and Noussair et. al (2012)examined both real and hypothetical stakes.

life scenarios. Renneboog and Spaenjers (2011) utilize data of 2,000 Dutch households and find the opposite; Catholics are more risk averse than Protestants, as measured by psychometric financial behavior and economic attitude assessments. Additionally, they find that religious households are more likely to save, an indicator of risk aversion. In contrast, Hong et. al (2004) examine data from 7,465 households and find that those who attend church are more likely to participate in the stock market, contradicting the predominant conclusion that religious individuals are more risk averse.

Beyond households, Bartke and Schwarze (2008) study 22,019 individuals and find a positive correlation between affiliation with a religious group and risk aversion, measured using a general psychometric measure ("On a scale from 1 to 10, how willing are you to take risks, in general"). Moreover, they find that Muslims are less risk tolerant than Catholics, who are themselves less risk tolerant than Protestants. Dohmen et. al (2011) and Weber (2014) corroborate these results, studying the same sample while utilizing the same risk metric. In contrast to Bartke and Schwarze (2008), Weber (2014) finds only weak evidence that Muslims are more risk averse relative to Protestants and Catholics. Utilizing a scale measuring a person's attraction to danger and risk (adventure seeking), Miller and Hoffman (1995) find from 2,408 individuals that those who felt religion was more important in their life exhibited higher levels of risk aversion. Unlike the individual- and household-level studies mentioned thus far, Miller (2000) looked across country lines, using a dataset spanning 1,000 individuals. His study found that higher levels of risk aversion, as measured on a general psychometric scale, were correlated with 1) adherence to a religious denomination, 2) the degree to which religion is comforting in one's life, 3) the important of religion in one's life, and 4) the frequency of attendance to religious services. Interestingly, these correlations only held in Christian and Muslim societies (Western). No significant correlation was found for Buddhist or Hindu societies. Also using a general psychometric risk measure, Liu (2010) studied 2,147 individuals across numerous Asian countries and also found no link between religious adherence and risk preferences. These results are consistent with the aforementioned risk preference thesis mentioned. Contrary to Miller (2000) and the risk preference thesis, Liu found that a greater frequency of religious service attendance was correlated with lower risk appetites.

Some research has been done using religious priming in a lab environment. Benjamin et. al. (2009) asked 817 individuals to choose between a sure versus a lottery-based payment, with real stakes, after priming them with religious ideas. They found that Catholics become less risk averse after priming, possible evidence that Catholicism promotes gambling (Shu et. al, 2010; Kumar et. al, 2011). Jews and Protestants were unaffected by the prime and atheists and agnostics became less risk averse[3].

In this paper, we expand upon Noussair et. al (2012), who use survey data on 2,304 individuals drawn from the Dutch population in the Netherlands. Similar to Benjamin et. al (2009) and work surveyed and done by Holt (2007), they measure risk by asking individuals to choose between a sure and a lottery payoff. 906 of the 2,304 individuals played the game for real stakes. The number of sure payoffs chosen was regressed on numerous measures of religiosity, those being religious adherence, current and past (at age 15) church attendance, prayer outside of services, whether one's parents were church members, and two measures of the strength of religious belief, controlling for exogenous gender and age and other potentially endogenous socioeconomic background variables such as marital, educational, and occupational status, income, health, number of children, and passport type. They find that those adhering to a religion are more risk averse, with some evidence that Protestants are more risk averse than Catholics. In addition, they

[3] Benjamin et. al (2009) claim that this may be an extension of the risk preference thesis. They claim that "when a person takes the risk of turning away from religion, cognitive dissonance causes him to conceive of himself as the kind of person who is less risk averse, so that the initial low-risk aversion that motivated his choice becomes even lower (Festinger, 1957, 1964). Then, whenever religion is made salient, that self-concept of low risk aversion becomes more highly activated"

argue that current *belonging*[4], drives religion's link to risk aversion, *not* a person's beliefs. Further, an individual's attendance and whether an individual's parents were a member of the church *at age 15* is not correlated with risk preferences. Thus, Noussair et. al (2012) conclude that religious upbringing has no permanent effect on risk attitudes.

In summary, the majority of studies, barring Hong et. al (2004), find that general religious adherence is correlated with higher levels of risk aversion across numerous dimensions of risk. Despite this consensus, it should be clear by now that the results with respect to differences in risk aversion between religions is mixed. Barsky et. al (1997), Benjamin et. al (2009), Shu et. al (2010), Kumar et. al (2011), and Noussair et. al (2012) find that Protestants are more risk averse or make safer financial choices than Catholics, while Bartke and Schwarze (2008), Renneboog and Spaenjers (2011), Dohment et. al (2011) and Weber (2014) observe the opposite. There is even less research and consensus regarding Muslims relative risk aversion. Bartke and Schwarze (2008) argue that Muslims are more risk averse than Catholics and Protestants while Weber (2014) finds no significant difference. The results regarding the significance of belonging, that is the social aspect of religion, and belief measures on risk preferences and economic behavior in general are also mixed. Hong et. al (2004) and Noussair et. al (2012) argue that it is belonging while Gebaur et. al (2012) focus on beliefs. Iannaccone (1998) and Barro and McCleary (2003, 2006) do not focus solely on one measure. Rather, they argue that both could be important.

Noussair et al. (2012) noted a lack of academic consensus regarding the effects of belief and belonging measures on risk aversion and regarding relative risk aversion levels *between* religions. They also noted a dearth of cross-cultural studies spanning borders where one religion

[4] A person's social affiliation to a church as measured by attendance and private prayer.

is dominant. Inspired by this, we adjust and expand their design to a cross-cultural study of risk preferences amongst Muslims, Christians, and atheists in Pakistan (predominantly Muslim) and the United States (predominantly Christian). We also explore the effect of gambling preferences on risk aversion, a question posed near the end of Noussair et. al's (2012) paper and which may shed further light on the link between religion and risk. Further, we bolster Noussair et. al's (2012) study by introducing a broader psychometric measure of risk, *as well as* a monetary measure. We discuss these features of our study in detail next.

3. Experimental Design

Noussair et al. (2012) suggested exploring differences in risk aversion across religious populations in countries where those religions were dominant. That is, religions that are practiced by over 50% of the religious population of that country. They also suggested that future researchers investigate whether the results they obtained generalized across borders outside of the Netherlands. In order to add to the growing body of research exploring the relationship between religiosity and risk aversion, we modeled and extended Noussair et al.'s study, instead exploring differences in risk aversion across followers of the Islamic and Christian faiths in both the United States, where Christianity is dominant, and Pakistan, where the Islamic faith is dominant. We obtained data from non-religious individuals as well to explore whether the positive correlation between general religiosity and risk aversion holds in geographical regions outside of the Netherlands, as Noussair et al. (2012) suggested.

We surveyed[5] three population groups across the Islamic, atheist, and Christian affiliations in a 3 x 1, individual-level, correlational[6], between-subject study. One survey form was sent to

[5] Using 'Qualtrics XM' platform.
[6] Note that all studies that we surveyed only implied correlational relationships between risk preferences and religion metrics.

participants in the United States. 121 individuals in and around Minnesota[7] were surveyed, many of which were connected to us in some way. Some participants were from Macalester College[8], which was also randomly sampled (detailed next). 120 *complete* responses were collected from these 121 individuals. The 1 partial response was excluded from our dataset. In addition, 255 randomly selected individuals from Macalester College[9] were surveyed. Of these, 84 complete survey responses and 17 partial survey responses were collected. The partial survey responses were again excluded from our dataset. With the Macalester-based sample, we were able to pick up a wide variety of religious variation, mainly consisting of followers from the Christian, Islamic, and atheist schools of thought. A fairly wide swath of socioeconomic variation was captured as well. Another separate survey form, which was a copy of the first except for the monetary amounts, which were converted from United States dollars (USD) into Pakistani rupees (PKR), was sent to participants in and around Lahore, Pakistan. Many of these individuals were also connected to us in some form[10]. 134 responses were collected, of which 109 were complete. As before. the 25 partial responses were excluded from our dataset. Every survey participant was fluent in English so language barriers across populations were not a concern.

Overall, we collected 313 complete responses over a period spanning from February 16th, 2021 to February 27th, 2021. 43 partial responses were collected and subsequently excluded from our dataset. With these data in hand, we are effectively able to investigate differences in risk preferences between religious and non-religious individuals, and between followers of the Islamic and Christian faiths. Note that we also captured some data on individuals from other faiths. Variations in risk preferences for these individuals was captured in an 'Other Religion' dummy

[7] Note that two individuals, our connections, were located in Europe.
[8] Note that not all individuals from Macalester College are from the United States.
[9] With the help of Macalester College's Institutional Research organization.
[10] Note that for the samples that were effectively self-directed by us, we used a form a pseudo-randomization, pulling on our collective connections to form as diverse and random a sample as possible.

variable, explained in further detail later. Of the 313 individuals surveyed, 90 self-identified as Christian, 118 as Islamic, 53 as "Other", and 52 as atheist[11]. A visual representation of the distribution of religious denomination across samples as well as summary statistics for the participants, for each religious group of focus, can be found in the appendix (Figure 1 and Table 1 respectively).

All participants were informed before the survey period's start that their answers were completely anonymous, to avoid any anonymity bias (Sunstein, 2019, p. 60; Levitt and List, 2007, p. 161). This was especially important given that our Pakistan and Northern Minnesota samples were collected largely from pools of our own connections. To incentivize participation, the surveyed population was told that, upon completion of the survey, they would be entered into a random drawing to receive a reward. For the United States participants, this was a $30 gift card, approximately equal to one hour of average hourly earnings as per the Bureau of Labor Statistics[12], a reasonably meaningful amount of money. The Pakistani group was entered into a drawing to receive PKR 3,000 in cash, approximately $19, a reasonable amount of money for the participants surveyed according to our experience. Note that to preserve anonymity, survey respondents were asked to send their name to us via email at the end of the survey. This allowed anonymity to be maintained while also giving us access to the names needed to conduct the prize drawing.

All participants were asked to read an introductory note which included the purpose of the study[13], explained in vague terms so as to not bias results, as well as confidentiality notices, a note on where to direct questions, rewards details, and the obtainment of consent. The commonality of

---

[11] The majority, 108, of the Islamic individuals were from the Pakistan sample, 7 were from the Minnesota sample, and 3 were from the Macalester sample. The majority, 67, of the Christian individuals were from the Minnesota sample, 23 were from the Macalester sample, and 0 were from the Pakistan sample. Only one individual from the Pakistan sample identified as atheist, 19 from the Minnesota sample, and 32 from the Macalester sample. From the Macalester sample, 26 identified into the "Other" religious category, 0 from the Pakistan sample, and 27 from the Minnesota sample.

[12] https://www.bls.gov/news.release/empsit.t19.htm

[13] We explained the purpose of the study in vague terms to not bias results. In the end of survey message, we did the same. This was done because our respondents completed the surveys at different times and we did not want the purpose of the study conveyed to later respondents by earlier ones, for fear that this would bias results.

this starting sequence should serve to place all participants, regardless of environment, in a similar state of mind[14].

3.1 Measurement of Risk Attitudes

We gauged risk aversion in two distinct ways. The first measure asked participants to choose between a lottery and sure payoff amount. The second asked participants a series of psychometric questions across five domains of risk. Holt (2007) details the use of both lottery versus sure payoff *and* lottery versus lottery choice experiments. We chose a lottery versus sure payoff structure because it is approachable to a wider range of audiences. We wanted to ensure our questions were as cognitively easy as possible, so as to avoid unnecessary noise in our data. Holt (2007) also did not argue that sure versus lottery choice results were any less accurate relative to a lottery versus lottery structure. Also, Benjamin et. al (2009) and Noussair et al. (2012) used a sure versus lottery choice structure, reinforcing it efficacy.

In our sure versus lottery choice structure, participants chose, in five trials, between a lottery that paid $80 (PKR 12,796) or $5 (PKR 800) with equal probability and thus had an expected value of $42.50 (PKR 6,798)[15], and a sure payoff amount that differed by trial. The sure payoff varied from $24 (PKR 3,857) to $48 (PKR 7,714) in steps of $6 (PKR 964)[16]. Each of the five choices was presented on a separate screen and the order of the sure payoffs was counterbalanced among subjects. Half of the subjects was presented with sure payoffs in ascending order and the other half were presented with sure payoffs in descending order. The side of the screen (left or right) on which the lottery and sure payoff amount appeared was also

[14] The only difference in introductions between the US and Pakistan surveys was the rewards structure.
[15] These monetary amounts were based off Noussair et al. (2012) to aid in the comparability of results. They used amounts denominated in Euros. We converted these amounts to USD using an exchange rate of 1 Euro to $1.21 and adjusted the lottery amounts to be in multiples of $5 for ease of understanding. The PKR denominations are based off these adjusted USD amounts, using an exchange rate of 1 USD to PKR 159.95. Exchange rates as of February 8th from Morningstar. Inflation impacts are assumed to be inconsequential as Noussair et al.'s paper is recent. Exchange rates are as of February 8th, 2021.
[16] Sure payoff amounts are directly from Noussair et al. (2012), converted in the same manner as described in footnote (3), rounded to the nearest whole number.

counterbalanced amongst the participants. The lottery choice appeared on the left for half of the participants and on the right for the other half. The participants were not made aware of the outcome of any of the lotteries during the survey.

Unlike Noussair et al. (2012), who presented each lottery in terms of a die roll, we presented each in terms of a flip of a coin (see Figure 2 in the appendix for an example of a screen shot illustrating the format). This is a much more understandable way to exhibit a 50-50 lottery than a die roll and is generalizable across both Pakistani and American cultures. All 313 subjects made this choice for hypothetical stakes. Holt (2007) suggests that at low payoff amounts, the results found using hypothetical payoffs are not significantly different than those found using real payoffs. Noussair et al. (2012) find the same. Our measure of individual-level risk aversion is the number of instances in which a participant chose the sure payoff amount out of the five questions. The higher the number, the more risk averse the participant[17]. This construction is a direct extension of that used in Noussair et al. (2012). Note that given the extreme prohibitions against gambling in the Islamic faith, we took great care to ensure that the word gambling never appeared in association with our lottery versus sure payoff choice structure, so as not to bias our results.

Despite their popularity in measuring risk preferences, lottery-based questions have the disadvantage of only measuring the willingness to assume *financial* risks. They do not tell us much about, for example, the willingness to assume social or health risks. Thus, our second measure of risk aversion involved ten psychometrics questions, selected from a total of 101 as presented by Weber, Blais, & Betz (2002), two from each of *five* domains of risk. Those are financial, health/safety, recreational, ethical, and social. These ten questions were selected because they have a relatively reasonable chance of capturing only risk aversion effects, thereby excluding cultural

[17] Note that by mathematical definition, a risk neutral agent would pick either one or two safe choices. Thus, anything more than 2 choices indicates risk aversion.

and wealth effects. Many of the questions were written with a Western bias and would not have been standardizable across Pakistani and American cultures. The questions all portray risky activities and were presented in the same order for all participants, five questions per screen. Participants were asked to choose the likelihood that they would engage in the activities presented in each question on a scale from 'Very Likely' to "Very Unlikely' (see the appendix for a list of the response prompt, a list of these questions, and their corresponding risk domains [Figure 3], as well as a depiction of the response scale [Figure 4]). The categorical responses were coded into a numerical range between 0 and 4, where 4 corresponds to "Very Unlikely". Our unique measure of individual risk aversion was an average of these coded values across the ten questions. Each question portrays a risky activity. Thus, the higher the average number, the more risk averse the participant is[18].

3.2 Measurement of Religiosity

The literature surrounding the interplay of religion and risk has yet to settle on the exact measure of religiosity that corresponds with risk aversion, whether it be a measure of *belonging* (Hong et. al, 2004; Noussair et. al, 2012), a measure of *belief* (Gebaur et. al, 2012), or a bit of both (Iannaccone, 1998; Barro and McCleary, 2003, 2006). The belonging variables help to capture the social aspects of religious event and service attendance, positing that it may be this social network, *not* religious belief itself, that impacts an individual's risk aversion. Belief variables measure how strongly an individual believes in their faith. Given the literature has yet to settle on which set of

[18] The order of the multiple-choice scale responses was kept constant. We did not randomize or flip the order because we did not want to induce cognitive strain in our participants, as that would add noise to our data. Due to the responses involving personal accounts, we do not expect that respondents answered them through haphazard guessing (as one may do with questions in a survey asking for feedback about an impersonal topic, such as feelings towards a product one has never interacted with). We believe that the cost in cognitive ease to randomize these questions was not economical.

measures is most correlated with risk aversion, we decided to utilize both kinds of measures as was done in Noussair et al. (2012).

We define two specification families, explained in detail in the modelling section below. For one, we define a religious or non-religious adherence dummy variable following the definition of religion from Iannaccone (1998). That is, a "shared set of beliefs, activities, and institutions premised upon faith in supernatural forces". Religious adherents are defined as those who self-identified into the "Islam", "Christian", or "Other" faiths. Non-religious adherents are defined as those who self-identified as atheist. Non-religious adherence is defined as the dummy base. For the other specification family, we define dummies for adherence to the Islamic, Christian, and 'Other' faiths. Other is a general bucket which contains participants of the other myriad faiths which we chose not to concentrate on. Christian individuals were used as the dummy base. Note that in the second specification family, the regressions were ran on a *subsample* of the overall dataset which excludes individuals who identified as atheist. More detail regarding these specifications will be provided in the modelling section that follows.

Beginning with our belief metric, first note that those who self-declared as non-religious were not directed to answer the belief or belonging question measures in the survey as those measures did not apply to them. We used the single-item measure of religious beliefs employed by Gebauer et al. (2012) because the belief measures utilized in Noussair et al. (2012) were specific to the Catholic and Protestants faiths that they investigated. Single-item religiosity measures are common and have been utilized by some of the biggest names in the field (Norenzayan & Hansen, 2006). The measure is the response to a single question: "My personal religious beliefs are important to me" (1 = *not at all*, 7 = *very much*). The higher the numbered response, the higher an individual's "degree of belief". Unlike the measures in Noussair et. al (2012), this metric is standardizable across both the Islamic and Christian faiths.

Continuing to our belonging metrics, informed by Noussair et al. (2012), we define dummy variables for frequency of current religious service attendance [19]. The categories are religious service attendance of 'More than once a week', 'Once a week', 'Once a month', and 'Infrequently', with 'Infrequently' acting as the dummy base. We added the 'Infrequently' category to Noussair et al.'s (2012) construction in order to capture people who consider themselves culturally religious but who may not be too religious in the traditional sense.

We take the same categories of attendance frequency and add a 'Never' category for an additional measure of belonging: religious service attendance at age 15. 'Never' serves as the dummy base. Our survey flow construction has participants who self-identify in the 'atheism' denomination skip the questions regarding religious belief, current religious service attendance, and prayer frequency (described below). We do this because these questions do not apply to those who self-identify in the 'atheism' category. The attendance at age 15 variable, on the other hand, *may* (or may not) apply to them, given that participants could have renounced religious services *after* the age of 15. This is one reason why we include the 'Never' category here. The second is that *currently* religious participants could have started attending religious *after* the age of 15[20]. We also define a dummy variable for whether a participant's parents were a member of their faith at age 15 as an additional measure of belonging. "No" serves as the dummy base.

Dummy variables are constructed with the same attendance frequency levels as with current religious service attendance for the frequency of prayer *outside* of religious services (private prayer). 'Infrequently' is, again, used as the dummy base. Note that because prayer outside

[19] We standardized Noussair et al.'s (2012) measures by swapping 'Church/service attendance' with 'religious service attendance' in the question verbiage.
[20] Given that *only* people who identify with a religious faith are exposed to the two other frequency-based variables [current attendance and prayer outside of religious services], 'Never' would not be a sensible answer as, if this were the case, they would not have identified with a religious faith.

of religious services is done both privately and in social groups it has aspects of both believing and belonging (Noussair et al., 2012).

Noussair et. al. (2012) theorized that the repeated social exposure to sermons reminding Christian individuals of Calvinist and Lutheran prohibitions on gambling could have been a major driving factor behind the higher monetary risk aversion levels that they found on the part of Protestants and Catholics. Inspired by this, we investigated the link between a participant's opinion of gambling and their risk aversion across both samples, both specification families, both sets of controls, and both risk metrics.

Past studies suggest that the link between risk aversion and religion could be driven by the social aspects of religious affiliation rather than belief itself (Iannaccone, 1998; Hong et. al, 2004; Barro and McCleary, 2003, 2006; Gebaur et. al, 2012; Noussair et. al, 2012). One theoretical foundation for this could be that the repeated exposure to religious events stressing the prohibitions on gambling that are present in many of the monotheistic faiths could have an effect on risk aversion with regard to monetary lotteries. This could account for the greater risk aversion on the part of religious individuals discovered in the vast majority of past research. This theoretical framework was presented by Noussair et al. (2012), who theorized that the repeated social exposure to sermons reminding Christian individuals of Calvinist and Lutheran prohibitions on gambling could be a major driving factor behind the higher monetary risk aversion levels that they found on the part of Protestants and Catholics relative to non-religious individuals. To probe this feature, we defined dummy variables for a participant's opinion on gambling: either 'Oppose', 'Support', or 'Neither Oppose nor Support', the final category is defined as the dummy base[21].

[21] Note that like with the psychometric risk measure questions, the order of all of the multiple-choice questions probing religiosity was kept constant. The reasoning for this, as explained earlier, was to not induce unneeded cognitive strain on participants. As before, we do not expect participants to rush through questions do to their non-random configuration because they all probe personal features of one's being.

Note that contrary to this argument, some research argues that Catholicism actually promotes gambling (Benjamin et. al, 2009; Shu et. al, 2010).

3.3 Controls

Our specifications either control for a smaller set of independent variables, Controls A, or a larger set consisting of both Controls A and B. Controls A consist of dummy variables to control for the presentation of the monetary stakes (whether the lottery was presented on the left or right) and the order in which the sure payoffs were given, along with age and dummies for gender. All of these variables are purely exogenous. Controls B consist of socioeconomic variables, some of which in principle are subject to endogeneity. These consist of number of children, gross monthly income[22], health status (1 = *worst*, 5 = *best*), and dummy variables for marital status ('Married', 'Divorced or Widowed', or 'Never Married'), homeownership, educational (completion of higher education or not) and occupational (variables both for whether they are a civil servant and whether they are self-employed) status, citizenship status ('Pakistani', 'American', or 'Other'), and current country of residence. Outside of the last two measures, all of these controls are borrowed from Noussair et al. (2012). Instead of dummy variables considering whether subjects had a Dutch or foreign passport, we used citizenship status. Shu et al. (2010) suggested that cultural differences could have an impact on individual-level risk aversion. We utilized this measure, along with our added measure of current residence, to control for cultural differences. Participants could have lived in a country for which they were not citizens for many years, and this could have a profound impact on their personal culture, an impact that would not solely be captured in the citizenship measure[23].

[22] The United States participants were asked to provide this in USD. The Pakistani group was asked to provide this in PKR. For the purposes of the regression, all amounts were converted to USD using the rates in footnote (3).
[23] As before and for the same reasons, the order of the multiple-choice responses was kept constant.

3.4 Summary Statistics

Table 1 in the appendix provides summary statistics of every variable described in this section for each religious group of focus and for the sample as a whole. Note that given a value of one or two for our monetary risk measure is indicative of risk neutrality (by mathematical definition), it appears our full sample is, on average, risk averse. Overall, people make, on average, 2.72 safe choices in our monetary risk assessment. This is lower than Noussair et al. (2012), whose sample's average was 3.43. We cannot use our psychometric measure to speculate regarding average risk aversion relative to a point of risk neutrality as it is difficult to assess what score in that measure indicates risk neutrality.

4. Empirical Model

We define two "specification families", which both contain two specifications each, one which includes purely exogenous Controls A and one which includes both Controls A and B. Each specification was regressed for both measures of risk aversion, resulting in 8 total specifications.

Specification family one was intended to isolate the overall effect of religious adherence on risk aversion. It helps deduce whether religious people's risk preferences vary significantly from non-religious people, as defined by their self-identification into the religious buckets defined above. This specification family also helps determine whether past belonging, as measured by the attendance at age 15 and parent's membership variables, has a significant effect on current risk aversion levels across all individuals surveyed. Individuals who identify as "atheist" today may have been religious (as measured by belonging) at age 15, and this could impact their current risk preferences and vice versa. It also serves as a robustness test for these two measures, as they are also investigated solely for the subsample of individuals who self-identified into a religion. Lastly, this specification family is used to investigate the effect that participant's opinions of gambling have on their risk preferences across our full sample, which could shed light on the mechanisms driving risk appetite variations. For the same reason as with the prior belonging measures, specification family one serves as a robustness test for the gambling opinion covariate as well. Specifications (1) and (2), the members of specification family one, are given below.

(1) $$R_i = \beta_0 + \sum \beta_1 Religious + \sum \beta_l PB_\sigma + \beta_6 Parents + \sum \beta_\theta Gambling_\rho + \vec{A} + \varepsilon$$

(2) $$R_i = \beta_0 + \sum \beta_1 Religious + \sum \beta_l PB_\sigma + \beta_6 Parents + \sum \beta_\theta Gambling_\rho + \vec{A} + \vec{B} + \varepsilon$$

l = 2, 3, 4, 5 | σ = 'Infrequently', 'Once a month', 'Once a week', 'More than once a week' | θ = 7, 8 | ρ = 'Support, 'Oppose'

R_i is the risk aversion level for individual i, which is measured either using the monetary stakes or the psychometric measure. $Religious$ represents the dummy variable for those who identify as generally religious or non-religious ("atheist") and PB is the dummy set for the frequency of an individual's religious service attendance currently when they were 15. $Parents$ is a dummy which captures whether a respondent's parents were a member of their faith when they were 15. These last two variables measure prior belonging. $Gambling$ is a dummy set which indicates a participant's opinion of the institution of gambling. The corresponding dummy variable bases were explained in section 3.2. $\vec{A}$ is a vector containing the purely exogenous set of Controls A. This includes dummy variables for the presentation of the monetary stakes and the order in which the sure payoffs were given, along with age and dummies for gender. $\vec{B}$ is a vector containing Controls B. This includes controls for socioeconomic factors like number of children, gross monthly income, health status, and dummy variables for marital, homeownership, educational and occupational status, citizenship, and current country of residence. Controls B may introduce some endogeneity. Note that this specification family does *not* include the variables (explained in section 3.2) which capture the potential effects of current belonging and current belief on risk preferences. These effects are bundled together and captured by the religious or non-religious adherence dummy.

Specification family two helps separate these potential effects into the individual current belief and belonging variables. It was applied to a subsample of our dataset which excluded the 52 “atheist” individuals surveyed. This specification family was intended to take a deep dive into the effects underlying the mechanisms that drive risk appetite variation across individuals. It includes the current belief and belonging measures touched upon in the previous section and aims to determine what exact features of an individual’s religiosity, whether it be belief or belonging, affects their risk appetite. “atheist” participants were excluded as they were not directed to answer the current belief and belonging questions, as that would not have been logical (explained in section 3.2). This specification family, by zooming into what exact features affect a *religious* person’s risk appetite, did not need to include those who self-identified as *non*-religious. Specifications (3) and (4), the members of specification family two, are given on the next page.

(3)
$$R_i = \beta_0 + \sum \beta_n Faith_r + \beta_3 Belief + \sum \beta_m CB_\gamma + \sum \beta_l PB_\sigma + \beta_{11} Parents + \sum \beta_\alpha Prayer_\gamma + \sum \beta_\theta Gambling_\rho + \vec{A} + \varepsilon$$

(4)
$$R_i = \beta_0 + \sum \beta_n Faith_r + \beta_3 Belief + \sum \beta_m CB_\gamma + \sum \beta_l PB_\sigma + \beta_{11} Parents + \sum \beta_\alpha Prayer_\gamma + \sum \beta_\theta Gambling_\rho + \vec{A} + \vec{B} + \varepsilon$$

n = 1, 2 | r = ‘Islam’, ‘Other’ | m = 4, 5, 6 | γ = ‘Once a month’, ‘Once a week’, ‘More than once a week’

l = 7, 8, 9, 10 | σ = ‘Infrequently’, ‘Once a month’, ‘Once a week’, ‘More than once a week’ | α = 12, 13, 14

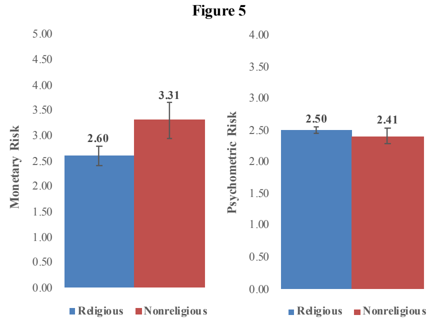
θ = 15, 16 | ρ = ‘Support’, ‘Oppose’

R_i is the risk aversion level for individual i, which is measured either using the monetary stakes or the psychometric measure. $Faith$ represents the dummy set for religious denomination, excluding those who self-identified as "atheist", and $Belief$ indicates an individual's "degree of belief". CB and PB are the dummy sets for the frequency of an individual's religious service attendance currently and when they were 15, respectively. $Parents$ is a dummy which captures whether a respondent's parents were a member of their faith when they were 15, $Prayer$ is a dummy set indicating the frequency of prayer *outside* of religious services (private prayer), and $Gambling$ is a dummy set indicating a participant's onion of the institution of gambling. The corresponding dummy variable bases were explained in section 3.2. $\vec{A}$ and $\vec{B}$ are defined the same as for specification family one.

5.1 Religious Adherence and Risk Aversion

We first consider whether there is an overall correlation between risk aversion and religious affiliation. Looking at the averages exemplified by Figure 5, religious people appear to be less risk averse in our monetary measure of risk aversion. Further, they appear to be more risk averse in our psychometric measure of risk aversion, although this difference is not significant[24]. The raw averages illustrated in Figure 5 fail to control for other influences on risk aversion, which may fall differentially between the religious and nonreligious groups. They are thereby subject to bias. To provide a more robust analysis, we apply specification family one to our full sample

Figure 5



[24] Monetary risk on a scale of most risk averse (5) to least risk averse (0); Psychometric risk on a scale of most risk averse (4) to least risk averse (0); error bars represent 90% confidence intervals; note that if the error bars between comparison groups overlap, the group averages are not significantly different at any level equal to or greater than 10% significance.

of 313 responses and regress for both our monetary and psychometric risk aversion measures. Table 2 on the next page depicts our Ordinary Least Squares regression results for the covariates of interest, for sets of Controls A and B. As measured by our monetary measure of risk aversion, we again find that religious people are less risk-averse compared to non-religious people (the dummy variable base)[25]. However, we only see significance after adding both controls A and B. This result is directionally robust across risk aversion measures, though there are no significant correlations between religious adherence and our psychometric measure of risk aversion. Our result is in direct contrast with the majority of the literature which all found significantly *greater* levels of risk aversion associated with religious affiliation. Only Hong et. al (2004) found a result similar to ours.

Table 2: Risk Aversion between Religious and Nonreligious Respondents

	Monetary	Monetary	Monetary	Psychometric	Psychometric	Psychometric
Controls A	Yes	Yes	Yes	Yes	Yes	Yes
Controls B	No	Yes	Yes	No	Yes	Yes
Macalester	No	No	Yes	No	No	Yes
Religious[a]	-.455 (.313)	-.577 (.331)*	-.407 (.368)	-.058 (.084)	-.105 (.090)	-.072 (.100)
Macalester[b]			.366 (.345)			.071 (.093)
N	313	313	313	313	313	313

Notes: dependent variable is either a monetary (5 = most risk averse, 0 = least risk averse) or psychometric (4 = most risk averse, 0 = least risk averse) measure of risk[24]; OLS regressions, s.e. in parentheses, */**/*** indicate significance at 10%, 5%, and 1% level; a: dummy base category=Nonreligious; b: dummy base category=Not a Macalester student (approximately).

We believe that the contrasting result may be a product of our atheist sample not being representative of the larger population of atheists. Not only is our atheist sample small relative to other studies (Benjamin et. al, 2009 have 269 atheist respondents), almost all atheists in our sample were from the United States (except for one in Pakistan) and were students at Macalester College in the 18-22 age bracket. The demographics and politics of the college might have skewed our results because over time colleges have become centers of restrictive speech and

[25] Note that because our risk metrics are structured such that higher numbers indicate greater risk aversion, negative covariates indicate that risk aversion is *decreasing* relative to increases in that variable. Positive covariates indicate that risk aversion is increasing relative to increases in that variable.

action through political polarization: in essence, one needs to be careful about their behavior to avoid ostracization. This trend could potentially increase risk aversion amongst our atheist sample to non-representative levels. Another reason for higher levels of risk aversion than we would expect amongst the atheist sample might be a product of the uncertainty surrounding COVID-19. There are two approaches to discuss this scenario: the expected utility theory and the evolutionary psychology approach. For an event like COVID-19, we believe that evolutionary psychology fits the narrative because the human reaction to a threat is to avoid it, in essence, they take as little risk as possible. In this case, many college students who had never faced a pandemic before might have had to shell up and preserve through a more risk-averse mentality. This could have spilled over to their risk preferences in our survey. With all things considered, it may not necessarily be true, as our results suggest, that non-religious people are more risk averse relative to religious people. Rather, it could just be that *Macalester students* who are nonreligious are more risk averse.

In an attempt to control for this, we estimated a Macalester dummy variable, defined as participants in the United States samples (as no respondents in the Pakistan sample were from Macalester) who were either surveyed through the random email distribution channel or were atheists, excluding two atheist respondents that did not match the profile of a Macalester college student. Controlling for this variable, we find directionally equivalent but insignificant relationships between religious adherence and both measures of risk aversion. Perhaps the adage "there are no atheists in foxholes" rings true. In a time of global uncertainty and suffering, it could very well be that the differences in risk appetite previously found between religious and non-religious individuals are nullified. After all, no-one has done a study of risk preferences and religion during COVID-19 before. Note that, although insignificant, the covariate on the

Macalester dummy reinforces the argument that Macalester students are, on average, more risk-averse than non-Macalester students, in both risk dimensions.

We also believe wealth effects could have impacted our monetary measures of risk in the religious and non-religious groups. According to Table 1, our sample is wealthier on average compared to the sample presented in Noussair et al. (2012) ($2,657.31)[26] and Pfeifer & Leon (2013) ($3,608). Our atheist sample had two outliers with gross monthly incomes greater than $30,000 but neither of them resided within the United States and they were hence excluded from the Macalester Sample. Removing these outliers and focusing only on the atheists from Macalester College[27] results in a $1,350.12 average monthly salary. This was lower compared to both the average Christian monthly income of $16,999.07 and the Muslim monthly income of $2,271.00. Considering that 87% of Islamic respondents are located in Pakistan (where the national average gross monthly income is $420) and the remainder are predominantly students at Macalester College with limited (if any) earning power, these relative wealth levels are indicative of vast wealth disparity. Note that the vast majority of the Pakistan sample was sourced from Lahore, a large metropolis area, which has much higher levels of average income than the majority of Pakistan. We believe financial concerns might have played a role in the decreased risk aversion of our religious population relative to our poorer atheist population. Weber's (1998) findings showed that greater financial concerns led to increased levels of monetary risk aversion and conversely that an increase in wealth led to lower levels of monetary risk aversion. Considering our religious samples are much wealthier than our non-religious, especially factoring in local average incomes, they may be willing to take on more monetary risk. This wealth effect could have factored into our anomalous correlation. Commenting on the survey, one Christian Minnesotan participant stated,

[26] Euro to Dollar conversation at 1:1.21 as of February 8th, 2021.
[27] Recall that an overwhelming majority of atheists (49/52) came from the Macalester sample.

“I chose to just gamble in every case because the amount meant nothing to me, who cares about $50, or even $80?” One Pakistani respondent was quoted saying “I am willing to gamble an honest half-day of work.” Statements like these exemplify the biasing power of the wealth effect.

Our counter result may have also been caused by regional cultural effects not captured in the broad ‘citizenship’ and ‘country of residence’ variables. For example, individuals in Northern Minnesota tend to be incredibly risk seeking in psychometric measures, despite their religious affiliation, which could perhaps bias the religious sample’s psychometric risk aversion downwards relative to non-religious individuals, potentially explaining the null but directionally negative correlation between psychometric risk aversion and religion. By failing to add finer controls for regional differences in both Pakistan and the United States, we could have biased our results. The relative youth of our sample could also be introducing bias, especially in the atheist sample. Looking at Table 1 in the appendix, the average age of our respondents is 30.91 years and only 21.21 years for the atheist subgroup (Noussair et. al’s (2012) sample has an average age of 49.60).

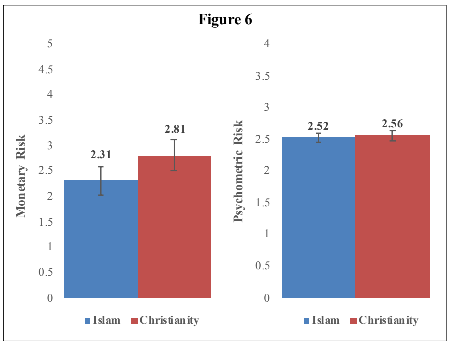
5.2 Muslims and Christians

The previous section establishes a negative but possibly null correlation between overall religiosity, as defined by self-selection into one of three religious categories[28], and risk aversion and attempts to explain this unexpected result. We now consider whether there are differences in average risk attitude between Muslims and Christians, on both the monetary and psychometric measures of risk. From Table 1 in the appendix, it is clear that there are differences between the two denominations in terms of the intensity of religious activities and beliefs. On average, Muslims hold stronger beliefs as measured by our beliefs indicator[29]. Muslims also attend

[28] Islam, Christianity, Other.
[29] Based on the numerical answer to the question “My personal religious beliefs are important to me” (1 = *not at all*, 7 = *very much*)

religious services more frequently, but that can be attributed to the higher frequency of services compared to Christianity inherent to the Islamic faith. Lastly, note that Muslims are shown to pray less frequently *outside* of religious services, but this could be due to the fact that they attend religious services with greater frequency.

Bartke and Schwarze (2008) and Dohmen et. al (2011) suggest that Muslims are slightly more risk-averse than Christians, and due to the stricter doctrines in Islam (For example, The Quran, 5:90), we expected to corroborate their findings. However, as shown in Figure 6, we find the opposite[30]. That is, our data suggests that Muslims are *more* risk-taking than Christians, at least in the monetary measures. The opposite is true in the Psychometric measures. However, neither difference is statistically significant[30].



Like with Figure 5, the raw averages illustrated in Figure 6 fail to control for other influences on risk aversion, which may fall differentially between the two groups. To control for these influences, we applied specification family two to a subset of our overall sample, excluding atheist respondents and instead focusing on our religious respondents. Table 3 on the next page provides the covariates of interest for regressions ran on both measures of risk aversion, including sets of either controls A or controls A and B. For the monetary measure of risk, we again find evidence that Muslims are less risk-averse relative to Christians (the dummy variable base is Christians)[31]. This is robust

[30] Monetary risk on a scale of most risk averse (5) to least risk averse (0); Psychometric risk on a scale of most risk averse (4) to least risk averse (0); error bars represent 90% confidence intervals; note that if the error bars between comparison groups overlap, the group averages are not significantly different at any level equal to or greater than 10% significance.

[31] Note that because our risk metrics are structured such that higher numbers indicate greater risk aversion, negative covariates indicate that risk aversion is *decreasing* relative to increases in that variable. Positive covariates indicate that risk aversion is increasing relative to increases in that variable.

across both specifications, though the result becomes insignificant once controls B are added. Along the psychometric dimension of risk, we find that Muslims could be more risk-averse than Christians, though this can be said with no certainty as the result is significant. This could serve as evidence for the regional cultural factors, discussed in section 5.1, which tend to bias Northern Minnesotans to the risk seeking end of the spectrum, especially when it comes to the psychometric risk domains (such as trying bungee jumping). Overall, our results seem to align with Weber (2014) who found no significant difference in risk aversion amongst Muslims relative to Christians once adequate controls were added.

Table 3: Risk Aversion by Denomination

	Monetary	Monetary	Psychometric	Psychometric
Controls A	Yes	Yes	Yes	Yes
Controls B	No	Yes	No	Yes
Muslims	-.713 (.351)**	-.701 (0.780)	.013 (.090)	.171 (.201)
N	261	261	261	261

Notes: dependent variable is either a monetary (5 = most risk averse, 0 = least risk averse) or psychometric (4 = most risk averse, 0 = least risk averse) measure of risk[30]; OLS regressions, s.e. in parentheses, */**/*** indicate significance at 10%, 5%, and 1% level; a: base category=Christians.

5.3 Belief versus Belonging

An important question regards the mechanisms linking an individual's religiosity to risk preference variation. That is, whether the relationship is driven by religious beliefs, by the social effects of religious participation (Hong et. al, 2005; Noussair et. al, 2012), or by a mixture of both. To probe this question, we applied specification family two to our religious subsample. In doing so, we pulled out some of the variation lumped together in the religious adherence dummy used in specification family one into more specific buckets, attempting to pinpoint the mechanisms driving risk preference variation between religious and non-religious people. Those buckets were, (1) the frequency of current religious service attendance, (2) the frequency of private prayer, and (3) a single item belief measure. Focusing on the belief variables to begin, the covariates of the first two measures for both risk metrics and including both sets of controls A and controls A and B, are provided in Table 4 on the next page.

Table 4: Risk Aversion and Beliefs/Prayer

	Monetary	Monetary	Psychometric	Psychometric
Controls A	Yes	Yes	Yes	Yes
Controls B	No	Yes	No	Yes
Belief	-.042 (.086)	.008 (.090)	-.006 (.022)	.006 (.023)
Praying (private)[a]				
>1 per week	.299 (.317)	.219 (.328)	.0145 (.082)	-.041 (.085)
=1 per week	.020 (.373)	.076 (.378)	.149 (.096)	.125 (.098)
=1 per month	.280 (.503)	.356 (530)	-.213 (.129)*[b]	-.231 (.137)*[c]
N	261	261	261	261

Notes: dependent variable is either a monetary (5 = most risk averse, 0 = least risk averse) or psychometric (4 = most risk averse, 0 = least risk averse) measure of risk[31]; OLS regressions, s.e. in parentheses, */**/*** indicate significance at 10%, 5%, and 1% level; a: base category=Infrequently; b: p > |t| = .101 but approximated to <= 0.10; c: p > |t| = .093.

Corroborating Noussair et. al (2012), we find no significant effect of the strength of religious beliefs (a pure belief measure) on risk aversion across all specifications. When considering the psychometric measure of risk, we find weak evidence (barely 10% significance) that praying more frequently outside of religious services is correlated with lower risk aversion (the dummy variable base is infrequent private prayer)[32]. Note that collapsing the private prayer propensity variable into a dichotomous measure removes all significant correlation. Additionally, this dichotomous measure insignificantly indicates a positive, not negative, correlation between risk aversion and private prayer propensity[33]. No other significant correlations between private prayer propensity and risk aversion levels were found. All but one of the other private prayer propensity covariates were insignificantly positive[34]. Given the insignificance of the majority of

---

[32] Note that because our risk metrics are structured such that higher numbers indicate greater risk aversion, negative covariates indicate that risk aversion is *decreasing* relative to increases in that variable. Positive covariates indicate that risk aversion is increasing relative to increases in that variable.
[33] This was constructed by collapsing the 'More than once a week' and 'Once a week' categories into a 'Belong/Believe' private prayer propensity level and the 'Once a month' and 'Infrequently' categories into a 'Not' private prayer propensity level. As touched upon in the limitations section to follow, this was done to investigate the effects of imperfect multicollinearity.
[34] Despite the significance of these two private prayer covariates, note that of the ten others, none is significant. Hence, the emphasis on *weak* evidence. Further, note that only one of the ten others is directionally equivalent, most of them predict the opposite relationship, with higher frequencies of prayer, relative to the dummy base category, being correlated with higher levels of risk aversion, although all of these effects are insignificant. Due to this, the best we can claim is weak *negative* correlation between psychometric risk aversion and private prayer propensity.

these results, we assume a *weak* negative to null correlation between private prayer propensity and psychometric risk aversion for the remainder of this paper. This result is directly contrary to Noussair et. al (2012) who found that greater frequency of prayer outside of services was correlated with *higher* levels of risk aversion. Recall that the frequency of prayer outside of religious services is presumably correlated with stronger beliefs but could also be tied to greater interaction with other faith members. That is, it carries both the belief *and* the belonging aspects of religiosity (Noussair et. al, 2012). In the absence of significant effects between risk aversion and our pure belief indicator, we cannot yet say whether belief or belonging play a larger role in risk preference variation across religions.

To examine this question further, we provide the covariates on the current and past belonging variables, as measured by current and past religious service attendance and parent's faith membership, in Table 5 on the next page. Contrary to Hong et. al (2004) and Noussair et. al. (2012), we find no significant relationship between current belonging, as measured by the frequency of religious service attendance, and risk aversion. This result is robust across both measures of risk and both control configurations. We also find no significant relationship between religious service attendance at age 15 and risk aversion, nor between a respondent's parent's faith membership at age 15 and risk aversion. This result is robust across risk metrics and controls. Recall that we also asked atheist individuals about their parent's faith membership and their prior religious service attendance. Taking advantage of this fact, we applied specification family one to our full sample of data (denoted as "All" in Table 5) and found the same result. Thus, it does not appear that exposure to religion during a participant's upbringing permanently affects risk appetites. This runs contrary to the notion that parents' religion has an impact on their offspring (Guiso et. al, 2003, 2003, 2006) but parallels Noussair et. al. (2012).

Table 5: Risk Aversion, Attendance, and Religion Membership

	Monetary	Monetary	Monetary	Monetary	Psychometric	Psychometric	Psychometric	Psychometric
	Sub.	Sub.	All	All	Sub.	Sub.	All	All
Controls A	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls B	No	Yes	No	Yes	No	Yes	No	Yes
Current Attendance[a]								
>1 per week	.526 (.410)	.457 (.436)			-.119 (.105)	-.080 (.113)		
=1 per week	.289 (.315)	.341 (.323)			.006 (.081)	.033 (.083)		
=1 per month	-.113 (.381)	-.146 (.393)			-.079 (.098)	-.552 (.102)		
Prior Attendance[b]								
>1 per week	-.589 (.512)	-.570 (.535)	-.385 (.414)	.321 (.424)	-.017 (.132)	-.042 (.138)	-.045 (.111)	-.034 (.115)
=1 per week	-.178 (.481)	-.204 (.450)	.165 (.382)	.045 (.395)	.093 (.124)	.043 (.129)	.067 (.103)	.037 (.107)
=1 per month	-.338 (.513)	.349 (.529)	.336 (.424)	.303 (.435)	.080 (.132)	.053 (.137)	.102 (.114)	.079 (.118)
Infrequently	-.349 (.487)	-.318 (.501)	-.375 (.380)	-.230 (.392)	-.040 (.125)	-.073 (.129)	-.060 (.102)	-.055 (.106)
Religion Membership[c]								
Parents	.185 (.398)	.301 (.409)	.199 (.316)	-.025 (.322)	-.088 (.102)	-.083 (.106)	-.009 (.085)	.001 (.087)
N	261	261	313	313	261	261	313	313

Notes: dependent variable is either a monetary (5 = most risk averse, 0 = least risk averse) or psychometric (4 = most risk averse, 0 = least risk averse) measure of risk[31]; OLS regressions, s.e. in parentheses, */**/*** indicate significance at 10%, 5%, and 1% level; a: base category=Infrequently; b: base category=Never; c: base category=Parents not a member of the participants current religion.

In the absence of significant correlational effects between risk preferences and both our pure belonging *and* belief measures, we cannot conjecture which mechanism underlying the weak to null correlational relationship between private prayer propensity and risk aversion is most at work. We can only say that a mixture of stronger belonging and belief mechanisms, communicated through increased private prayer propensity, are *potentially* correlated with lower psychometric risk aversion[35]. Note that this reinforces our earlier evidence supporting the possible negative correlation between religious adherence and risk aversion, and it shines light on a possible mechanism underlying this relationship.

This result is not only in direct contrast with Hong et. al (2004), which claim that church attendance is positively correlated with risk aversion, but also in direct opposition to Noussair et.

[35] Note that this mixed link corroborates the theoretical underpinnings laid down by Iannaccone (1998) and Barro and McCleary (2003, 2006), which posit that both the belonging and belief aspects of religion could have significant impacts on economic behavior. Note that because we cannot pinpoint any single mechanism, be it belief or belonging, our results also corroborate Gebaur et. al (2012), who implies that belief could significantly effect economic behavior.

al (2012), which claim both that *only* belonging, as measured by current attendance and private prayer propensity, has significant effects on risk preferences and that that private prayer propensity is *positively*, not negatively, correlated with risk aversion.

Our implication that greater belief, as measured by private prayer propensity, may exhibit a weakly negative to null correlation with risk aversion is also directly divergent from Norenzayan's *"Big Gods: How Religion Transformed Conflict"*, who believes that religion serves as a guiding mechanism for individuals to *avoid* risky behaviors, and indirectly divergent from inferences gleaned from religious texts like the Quran and Bible, which tend to promote risk aversion across multiple dimensions. It also directly contradicts the risk preference thesis (Miller and Hoffman, 1995; Miller, 2000). Despite this, we believe that there is a case for the opposite to be true whereby some individuals might engage in risky behavior because they believe their all-powerful God can serve as an effective safety net for them. In essence, their strong belief in God provides an additional safety net that they leverage when making riskier decisions. To make an empirical case for this is an inciteful avenue for future research.

On the belonging side, whereby we find that risk aversion and belonging (measured by private prayer propensity) may also exhibit a weakly negative to null correlation, it may be that risk seeking individuals are naturally drawn to others who are risk seeking via religious gatherings or, relatedly, it may be the case that already prevalent risk seeking behavior is socially transmitted to others via these gatherings. In the latter, if religious individuals act in relatively risk seeking ways due to the perceived presence of a safety net, this risk seeking behavior could be transmitted to others via religious gatherings creating a spillover effect negatively correlating risk aversion and measures of belonging.

On a final note, given there is a chance our results imply a null correlation between private prayer propensity and risk aversion, it may also be the case that we simply did not find any significant relationship between risk preferences and *our* measures of religiosity, whether that be due to imperfect multicollinearity (discussed further in later sections) or the lack of a proper measure of religiosity. Future research could be done exploring other measurement techniques that may have better correlational capabilities.

An overarching reason for the relative dearth of significant results regarding the correlation connecting religious adherence, religious denomination, belief, or belonging to risk preferences, could be due to an implication of the risk preference thesis. Many parts of the world, the United States and large metropolises in particular, are modernizing and secularizing. That is, religions are becoming laxer in some areas, decreasing the risks imposed by non-affiliation and irreligiosity. The risk preference thesis argues that in the presence of strict, non-affiliation prohibiting religions, irreligiousness can be a form of risk-taking behavior that only less risk averse individuals would undertake (Miller and Hoffmann, 1995; Miller, 2000: Miller and Stark, 2002). Just as this does not hold true in Eastern societies where irreligiosity is not an explicit form of risk-taking, it may also have started to not hold true in modernizing and secularizing regions, such as the United States and large Pakistani cities like Lahore[36], where religions are becoming less strict and more open. Just as Miller (2000) and Liu (2010) failed to find a significant correlation between religious adherence and risk preferences, and Miller (2000) additionally failed to find a significant correlation between religious service attendance (*belonging*) and risk preferences, so too may we have failed in many ways to link measures of religiosity to risk preferences.

[36] Recall this is where the majority of our Pakistan sample is sourced from.

5.4 Gambling

Noussair et. al. (2012) theorized that the repeated social exposure to sermons reminding Christian individuals of Calvinist and Lutheran prohibitions on gambling could have been a major driving factor behind the higher monetary risk aversion levels that they found on the part of Protestants and Catholics[37]. Inspired by this, we investigated the link between a participant's opinion of gambling and their risk aversion across both samples, both specification families, both sets of controls, and both risk metrics. Table 6 shows the covariates of interest for this investigation. We find that a participant's opposition of gambling is correlated with higher levels of risk aversion (the dummy variable base is a neutral opinion of gambling)[38], as Noussair et. al. (2012) would have expected. This result is directionally robust across both risk metrics, both control configurations, and both sample sets (the full sample is denoted by "All" and the subsample is denoted by "Sub. in the table), though only significant for the psychometric measure of risk aversion. This is interesting as it would suggest that a person's opposition of gambling may leak into their risk preferences on non-monetary dimensions. Also, given Muslim's greater average gambling opposition relative to Christians (77.97% vs 26.67% as per Table 1), this adds some support to the insignificant conclusion from 5.2 which suggests that Muslims are *more* risk averse

Table 6: Risk Aversion and Gambling

	Monetary	Monetary	Monetary	Monetary	Psychometric	Psychometric	Psychometric	Psychometric
	Sub.	Sub.	All	All	Sub.	Sub.	All.	All.
Controls A	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls B	No	Yes	No	Yes	No	Yes	No	Yes
Opinion of Gambling[a]								
Oppose	.315 (.282)	0.146 (.310)	.155 (.223)	.285 (.264)	.156 (.073)**	.187 (.080)**	.131 (.060)**	.173 (.072)**
Support	-.193 (.353)	-.200 (.362)	-.074 (.299)	-232 (.309)	.105 (.091)	.114 (.093)	-.005 (.081)	-.005 (.084)
N	261	261	313	313	261	261	313	313

Notes: dependent variable is either a monetary (5 = most risk averse, 0 = least risk averse) or psychometric (4 = most risk averse, 0 = least risk averse) measure of risk[37]; OLS regressions, s.e. in parentheses, */**/*** indicate significance at 10%, 5%, and 1% level; a: base category=Neither Oppose nor Support.

[37] Note as well that Benjamin et. al (2009) and Kumar et. al (2011) propose just the opposite.

[38] Note that because our risk metrics are structured such that higher numbers indicate greater risk aversion, negative covariates indicate that risk aversion is *decreasing* relative to increases in that variable. Positive covariates indicate that risk aversion is increasing relative to increases in that variable.

relative to Christians on the psychometric risk dimension. Greater psychometric risk aversion on the part of Muslims could be due to their greater opposition to gambling.

Extending past our own data, indeed it is true that Islamic prohibitions against gambling are stronger than are Christian prohibitions (The Quran, 5:90) and some Christian faiths have even been shown to have a *propensity* to gamble (Benjamin et. al, 2009, Kumar et. al, 2011). Like with psychometric risk aversion, we expected this gambling opposition, amongst other features, to channel into higher *monetary* risk aversion on the part of Muslims relative to Christians but, as shown in section 5.2, this was (significantly) not the case. The null correlation between monetary risk preferences and gambling shown in Table 6 could offer an explanation. Despite the fact that this correlation is between gambling opinion *for all religious denominations* (as well as for non-religious individuals in the full ["All"] sample) and risk preferences, we believe it still sheds some light on two *strong* factors that could be biasing our Muslim risk aversion measurement downwards, especially given that Muslims are the largest religious group in our data. Recall that the word "gamble" was excluded from our monetary risk dimension questions in an effort to ensure that the Islamic monetary risk aversion measurements were not biased upwards due simply to the greater gambling prohibitions inherent in their culture. Despite the seemingly logical underpinnings of this design, we are attempting to measure religion's impact on risk preferences in this study and, it is likely exactly through greater prohibitory rules on gambling that the Islamic religion is related to risk appetites, among other features. By delinking the word "gambling" from our monetary risk measure, we may have biased our Muslim monetary risk aversion measurements downward relative to what they may truly be in the real world, where connotations are *not* delinked from gambling-type activity, and thus where it seems less ethical to choose lottery-type choices over sure payouts. We could have probed this point further by using the word "gambling" as a primer in some of the Pakistani survey groups. Future research could be done in this area. Additionally, anonymity might have played a role. While the majority of Muslims are opposed to gambling in our sample the anonymity of gambling might have induced some individuals to gamble considering they would have never gotten caught. This argument stems from Norenzayan (2013) and Barro & McCleary (2003), both of whom state that humans tend to take riskier or unethical decisions when they believe they are not being watched. This could also serve to unlink the oppositions to gambling from the monetary risk behavior we observed. Both of these effects likely served an important role in biasing our Muslim sample's monetary risk aversion levels downward relative to what we expected.

6. Limitations

Conducting a cross-cultural study within strict time constraints[39] can be prone to many limitations. We took care to control for confounding variables and maintain uniformity between the two surveys and country sample groups[40], but cross-cultural studies may pick up institutional differences that cannot be controlled for (Guiso et. al, 2003) among other things such as regional cultural differences and/or current college attendance (i.e. the issues discussed in section 5.1 and 5.2 and, relatedly, the Macalester College dummy). Given this was an online-based experiment, we had little control over the environment people conducted the survey in. We constructed the introduction such that everyone read the same message to begin with and thus, theoretically, was in the same headspace. Despite this, people's choices are incredibly complex and are affected by a vast set of factors such as relational situations, social norms, frames, and past experiences that

[39] We completed our entire research process in less than two months.
[40] Such as by keeping the surveys identical in terms of payouts and figures, simplicity of interpretation, culturally transferable psychometric questions, and time to fill in the survey, as well as by neutrally priming each individual with the same introduction in an attempt to standardize participant's frame of mind throughout the completion of the survey (see section 2).

have a profound impact and are often *not* fully controllable (Heinrich et al., 2005)[41]. A lack of proper controls may lead to omitted variables bias, which could have erroneously affected our results.

Further, our sample size was much lower compared to *every* other study surveyed in the literature review, an unavoidable feature of our short data collection time frame. In addition, our samples were much more stratified and much less heterogenous than we had hoped. This is likely because the survey was sent to many of our own connections, who tend to be in similar income brackets and have similar cultural backgrounds, experiential toolsets, and demographic features. Due to this, our sample was wealthier than in many other studies (see section 5.1), contained more gender biases (84.75% of the Islamic sample was male), and exemplified an age bias not seen in studies such as Noussair et. al (2012) (average age of 30.91 versus 49.60, see section 5.1). This lack of heterogeneity could have caused our results to lean on the side of err. In addition, the fact that we knew most of the individuals in our samples could have led to erroneous anonymity bias in their responses, even though we took special care to avoid this (see section 3) (Sunstein, 2019, p. 60; Levitt and List, 2007, p. 161).

As it relates to our survey construction, our survey flow in Qualtrics was meant to randomly distribute our monetary risk measurement questions such that 50% of the participants viewed the lottery on the left, 50% on the right, and that 50% viewed the sure payoffs in ascending order, and 50% in descending. In essence, there were four sets of risk aversion questions for each possible combination of presentation and sequence. Thus, each question was supposed to be shown 25% of the time. Qualtrics came close to a random distribution but there were some question setups that

[41] Levitt and List (2007) argue that choices can be influenced by at least five main categories of factors: 1) the presence of moral and ethical considerations; 2) scrutiny of one's actions by others; 3) situational context; 4) self-selection of the participants making the choices; and 5) stakes.

were presented more than others[42]. Moreover, as it relates to the wording of the survey, we believe that a lot of individuals may have entered their *yearly* income instead of *monthly* income as the latter is the standard of report. This may have led to the massive wealth disparity explained in 5.1. We believe that the Pakistani sample might have also intermixed private prayer propensity with religious service attendance, which could have biased one or both variable covariates and/or led to imperfect multicollinearity. To avoid confusion, we could have instead been more specific (i.e. "How often do you go to the mosque for Jummah prayer [Friday prayer]?"). The potentially erroneous data collection with regard to these variables can be seen in Table 1, which shows that Muslim individuals pray less frequently than Christian individuals, even though there are many more required weekly prayers (~35). Lastly, the question regarding parent's faith membership asked if a respondent's parents were a member of *their* faith at age 15. This should have been broadened to a member of *any* faith to capture more data regarding the effect of parent's religions affiliation (to any religion) on their children's risk preferences.

Further, even though Holt (2007) and Noussair et al. (2012) suggest that at low payoff amounts, the monetary risk aversion measure results using hypothetical payoffs are not significantly different from those found using real payoffs, Holt (2007) also mentions that in the absence of a widely accepted theory regarding when real and hypothetical incentive choices coincide, one should be cautious assuming that real incentives are not needed. The absence of real incentives in our study could be biasing our results. That is, our results may change drastically in the presence of real incentives, especially if these real incentives are large, as that tends to increase risk aversion. Additionally, our monetary risk aversion results are based on contrived gambles for small hypothetical stakes and may not be generalizable to actual real-world scenarios. Kahneman

[42] Left lottery presentation with ascending sure payoffs shown 27.48% of the time; Left lottery presentation with descending sure payoffs shown 25.24% of the time; Right lottery presentation with ascending sure payoffs shown 23.96% of the time; Right lottery presentation with descending sure payoffs shown 23.32% of the time.

and Tversky (1979, p. 265) suggest that people often know how they would behave if presented with a hypothetical example of a real-world situational choice. Thus, asking real world questions with high hypothetical stakes may aid in external validity[43].

In statistical matters, we could be experiencing the potential side effects of imperfect multicollinearity whereby at least one of two highly correlated covariates is estimated imprecisely. This effect is especially potent when considering our Macalester college dummy variable in section 5.1, which is highly correlated with the religious adherence dummy due to the prevalence of atheists at the school. It is also important when considering our religiosity measures, especially the frequency-based variables, which are likely highly correlated with each other[44]. These effects could explain the lack of significance found in this study. Similarly, there could be endogeneity introduced into our model as the belief and belonging measures, as well as many of controls B, could be correlated with each other and could be interacting in ways that we did not take into account.

7. Conclusion and Future Research

While the effect of religion on economic activity has been researched for two centuries, a vast majority of the literature fails to reach a consensus regarding, or simply fails to explore, four features that we believe are key in understanding the mechanisms linking religion and risk preferences, and thereby religion and economic activity. Those are 1) whether it be belief or belonging measures which influence cross-religion risk aversion variation, 2) relative risk aversion

[43] For an example of one of these real-world situational questions, imagine that someone's car was not performing correctly. They could take it to the mechanic now and pay for the repair or they could wait and keep driving it There is a chance it will never break, but there is also a chance of catastrophic failure (which is much more costly). A researcher could assess an individual's risk appetite by asking them whether or not they would risk continuing to drive or if they would rather just pay the mechanic bill right away.

[44] To explore this effect, we collapsed the private prayer propensity, current belonging, and prior belonging variables to dichotomous measures [Believe or not, belong or not, etc.] to investigate the within-measure effects of multicollinearity. None of the dichotomous measures were significant, implying that within-measure multicollinearity had a likely muted, not profound, effect. Note that collapsing the private prayer propensity variable made it insignificant even for the psychometric measure. Additionally, it caused all correlations to align in a positive direction.

levels *between* religions, 3) cross-cultural evidence spanning borders *outside* of the Western hemisphere where one religion is dominant, and 4) the effect of gambling opinions on risk preferences. Using survey data collected from the United States, Pakistan, and beyond containing revealed risk attitude measures, as well as detailed information regarding respondent's religious practice and beliefs, we expand Noussair et al. (2012) and explore these features.

First, we find evidence that religious people are less risk averse than non-religious people, significant only for the monetary measure of risk aversion. After introducing a variable to control for Macalester students, risk preferences and religious adherence become effectively unrelated. This could point to the lack of heterogeneity in our results, a lack of proper controls, effects of COVID-19, or an instantiation of the risk preference thesis due to the continued modernization and secularization of the world.

Second, we find evidence that Muslims are less risk averse than Christians on the monetary dimension and show no effective risk preference difference on the psychometric dimension. This could potentially be due to an inadvertent delinking of Islamic gambling prohibitions from our monetary risk measure.

Third, we find that private prayer propensity shows weakly negative to null correlation with risk aversion. Due to the lack of correlative evidence surrounding our pure belief or belonging measures[45], we are unable to posit whether it is specifically a belief or belonging measure that drives risk variations across religious affiliations (or non-affiliations) and in what fashion. This may serve as evidence that *both* belief and belonging aspects of religiosity drive risk preferences.

[45] Those bring our single item belief measure and the frequency of current religious service attendance.

Further, past religiosity[46] seems to have no effect on an individual's *current* risk aversion, implying that a religious upbringing does not permanently affect risk preferences.

Fourth, an opposition to gambling is correlated with higher levels of psychometric risk aversion. Relative to our other results, this is highly significant and robust and implies that an opposition to gambling may leak into non-monetary measures of risk. The lack of a significant relationship between gambling opinions and *monetary* risk preferences is evidence for the delinking problem described earlier.

Religion and the culture that shapes it have a profound impact on human nature and economic behavior. Understanding the link connecting our religious backgrounds with our economic behaviors can have profound policy implications spanning insurance demand, healthcare, financial market regulation, and international consumer exchanges. It continues to be unclear how measures of belonging, that is the social aspect of religion, and belief differentially effect risk aversion, as well as how risk aversion varies across religious groups and across countries and continents. It is also yet unclear how these specific factors could function to influence risky behavior. For example, are risk preferences transmitted through religious gatherings? Do the potential safety nets setup by a strong belief in a protective god cause risk seeking behavior? These questions are obvious avenues for future work. The effects of gambling on risk aversion *across religions* is also not well studied. Further, while correlational studies are prevalent in the literature, no *causal* study has yet been devised to explore risk preferences and religion. In the face of the relentless stride of modernization and secularization, future research should also be directed towards understanding the implications these processes are no doubt inflicting on the link between

[46] As measured by the frequency of religious service attendance when the respondent was 15, and by whether the respondent's parents were a member of their faith when they were 15.

religion and risk preferences, as well as on economic behavior in general. It would also be interesting to see further work completed in *controlled* environments.

Structurally, work should be done to account for the limitations exhibited in our study. Most importantly, the exploration of more pertinent controls and more explanatory measures of religiosity, and the use of both large and small *real* stakes (Holt, 2007), as well as the use of hypothetical real-world scenarios (Barsky et al.,1997; Kahneman and Tversky, 1979, p. 265), in measuring monetary risk aversion. Future research is also needed to explore better models of risk preferences dependent on religiosity measures and other factors, paying close attention to the potential negative effects of imperfect multicollinearity and endogeneity.

As academia continues to wrestle with this fascinating topic, we look forward to the many conversations to come and the insights which will no doubt be undusted and shown the light.

Appendix

Figure 1. The distribution of religious affiliation across samples.

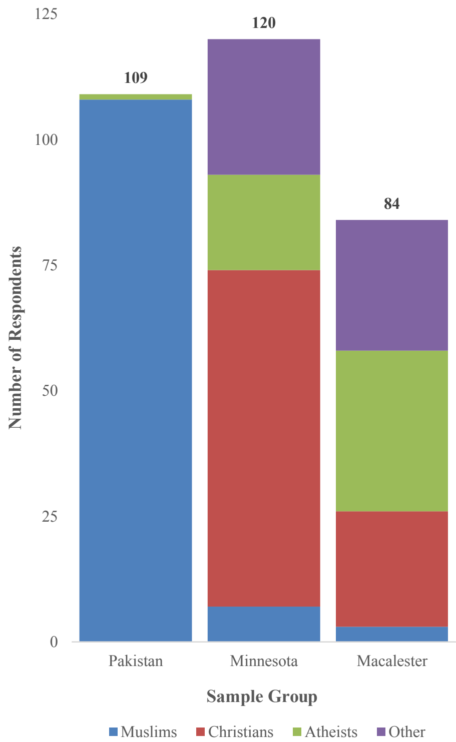


Figure 2. The format for our first risk aversion measure. In this example, the lottery is on the *right* of the screen and the sure payoff is the first in an *ascending* series. This example is one shown in US Dollars to the United States respondents.

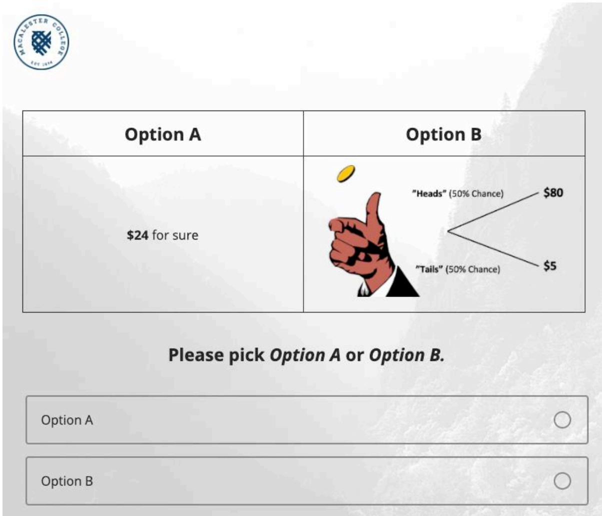


Figure 3. The response prompt and psychometric measures used for our second risk aversion measure, along with their risk domains.

For each of the following statements, please indicate the likelihood that you would engage in each activity.

1. Asking your boss for a raise. (S)
2. Cheating on an exam. (E)
3. Dating someone that you are working with. (S)
4. Going camping in the wild. (R)
5. Taking a medical drug that has a high likelihood of negative side effects. (H)
6. Lending a friend an amount of money equivalent to one month's income. (F)
7. Never wearing a seatbelt. (H)
8. Shoplifting a small item (e.g. a lipstick or a pen). (E)
9. Spending money impulsively without thinking about the consequences. (F)
10. Trying bungee jumping. (R)

Note: E = ethical, F = financial, H = health/safety, R = recreational, and S = social domains.

Figure 4. The scale used for the psychometric measure responses.

Table 1: Summary Statistics

Variable	# obs.	Mean	Islamic	Christian	Atheist
Risk					
Monetary Measure (min 0, max 5)	313	2.72	2.31	2.81	3.31
Psychometric Measure (min 0, max 4)	313	2.49	2.52	2.56	2.41
Religious Affiliation					
Islamic	313	37.70%			
Christian	313	28.75%			
Atheist	313	16.61%			
Other	313	16.93%			
Religiosity					
Belief (min 1, max 7)[a]	261	5.25	6.33	5.28	
Pray >1 per week	261	39.85%	44.92%	55.56%	
Pray =1 per week	261	13.41%	15.25%	12.22%	
Pray =1 per month	261	6.13%	5.93%	7.78%	
Pray = infrequently	261	40.61%	33.90%	24.44%	
Attendance >1 per week	261	13.41%	27.12%	3.33%	
Attendance =1 per week	261	22.22%	22.03%	35.56%	
Attendance =1 per month	261	16.67%	9.32%	16.67%	
Attendance = infrequently	261	53.26%	41.53%	44.44%	
Attendance >1 per week (age 15)	313	19.49%	33.05%	16.67%	5.77%
Attendance =1 per week (age 15)	313	34.50%	27.12%	61.11%	13.46%
Attendance =1 per month (age 15)	313	12.46%	12.71%	8.89%	7.69%
Attendance = infrequently (age 15)	313	15.65%	19.49%	11.11%	15.38%
Attendance = never (age 15)	313	17.89%	7.63%	2.22%	57.69%
Parents Member of Faith[b]	313	77.96%	92.37%	94.44%	38.46%
Gambling					
Oppose	313	42.17%	77.97%	26.67%	19.23%
Neither Oppose nor Support	313	42.49%	20.34%	50.00%	65.38%
Support	313	15.34%	1.69%	23.33%	15.38%

Notes: a: based on one question; b: when respondent was age 15, question posed as "were their parents a member of their faith".

Table 1: Summary Statistics (continued)

Variable	# obs.	Mean	Islamic	Christian	Atheist
Controls A[c]					
Male	313	58.47%	84.75%	44.44%	42.31%
Female	313	38.66%	15.25%	53.33%	51.92%
Other	313	2.88%	0%	2.22%	5.77%
Age	313	30.91	34.98	36.04	21.21
Controls B					
Married	313	40.58%	66.95%	47.78%	1.92%
Divorced or Widowed	313	1.92%	0.00%	3.33%	1.92%
Never Married	313	57.51%	33.05%	48.89%	96.15%
# children	313	1.00	1.54	1.38	0.06
Home owner	313	30.03%	39.83%	47.78%	1.92%
Health status (1=worst, 5=best)	313	4.15	4.18	4.27	3.96
High education (college or more)	313	54.31%	86.44%	54.44%	21.15%
Civil Servant	313	8.31%	5.93%	13.33%	11.54%
Self-employed	313	12.46%	15.25%	16.67%	5.77%
American	313	59.74%	4.24%	95.56%	88.46%
Pakistani	313	35.14%	92.37%	0.00%	1.92%
Other	313	5.11%	3.39%	4.44%	9.62%
United States[d]	313	63.58%	9.32%	100.00%	88.46%
Pakistan[d]	313	33.23%	87.29%	0.00%	1.92%
Other[d]	313	3.19%	3.39%	0.00%	9.62%
Gross monthly income ($)	313	$ 6,473.44	$ 2,271.96	$ 16,999.07	$ 2,528.96

Notes: c: in regression analyses, Controls A also includes controls for presentation and sequence in the monetary risk elicitation task; d: respondent's *current* country of residence.

Bibliography

Alalmai, Somaiyah et al. "The Influence of Religion on The Determinants of Capital Structure: The Case of Saudi Arabia". *Journal of Islamic Accounting and Business Research*, vol 11, no. 2, 2020, pp. 472-497. *Emerald*, doi:10.1108/jiabr-03-2018-0043.

Al-Awadhi, A., & Dempsey, M. (2017). Social norms and market outcomes: The effects of religious beliefs on stock markets. *Journal of International Financial Markets, Institutions and Money*, *50*, 119-134. doi: 10.1016/j.intfin.2017.05.008

Barsky, Robert B., F. Thomas Juster, Miles S. Kimball, & Matthew D. Shapiro (1997). Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study. *Quarterly Journal of Economics* 112, 537–579.

Barro, Robert J. and Rachel M. McCleary (2003) Religion and Economic Growth across Countries. *American Sociological Review* 68, 760–781.

Benjamin, Daniel, James J. Choi, and Geoffrey Fisher (2009). Religious Identity and Economic Behavior. Working paper, Cornell.

Bui, Phuong, David Crainich and Louis Eeckhoudt (2005). Allocating Health Care Resources under Risk: Risk Aversion and Prudence Matter. *Health Economics* 14, 1073-1077.

Bartke, S., & Schwarze, R. (2008). Risk-Averse by Nation or by Religion? Some Insights on the Determinants of Individual Risk Attitudes. SSRN Electronic Journal. doi:10.2139/ssrn.1285520

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner (2011). Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences. *Journal of the European Economic Association* 9, 522–550.

Eeckhoudt, Louis, Jack Meyer and Michael Ormiston (1997). The Interaction between the Demands for Insurance and Insurable Assets. *Journal of Risk and Uncertainty* 14, 25–39.

Festinger, Leon (1954). A Theory of Social Comparison Processes. Human Relations 7, 117-140.

——— A Theory of Cognitive Dissonance (Stanford, CA: Stanford University Press, 1957).

——— Conflict, Decision, and Dissonance (Stanford, CA: Stanford University Press, 1964).

Gao, L., Wang, Y., & Zhao, J. (2016). Does Local Religiosity Affect Organizational Risk-Taking? Evidence from the Hedge Fund Industry. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2868233

Gebauer, J. E., Sedikides, C., & Neberich, W. (2012). Religiosity, Social Self-Esteem, and Psychological Adjustment: On the Cross-Cultural Specificity of the Psychological Benefits of Religiosity. *Association for Psychological Science, 23*(2), 158-160. doi:10.1177/0956797611427045

Guiso, L., Sapienza, P., & Zingales, L. (2003). *People's Opium? Religion and Economic Attitudes*. (NBER Working Paper No. 9237). National Bureau of Economic Research. https://www.nber.org/papers/w9237

Guiso, L., Sapienza, P., & Zingales, L. (2003). Does Culture Affect Economic Outcomes? *Journal of Monetary Economics, 50*, 225-282.

Guiso, L., Sapienza, P., & Zingales, L. (2006). Does Culture Affect Economic Outcomes? *Journal of Economic Perspectives, 20*, 23-48.

Hilary, G., & Hui, K. (2009). Does religion matter in corporate decision making in America?. *Journal of Financial Economics*, *93*(3), 455-473. doi: 10.1016/j.jfineco.2008.10.001

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., . . . Tracer, D. (2005). "Economic Man" in Cross-Cultural Perspective: Ethnography and Experiments from 15 Small-Scale Societies. *Behavioral and Brain Sciences, 28*(6), 795-815. doi:10.1017/s0140525x05000142

Hilary, Gilles and Kai Wai Hui (2009). Does religion matter in corporate decision making in America? *Journal of Financial Economics* 93, 455–473.

Holt, C. A. (2007). Markets, games, and strategic behavior: Recipes for interactive learning. In *Markets, games, and strategic behavior: Recipes for interactive learning* (pp. 51-62). Boston, Massachusetts: Pearson Addison Wesley.

Hong, Harrison, Jeffrey D. Kubik and Jeremy C. Stein (2004). Social Interaction and Stock Market Participation. Journal of Finance 59, 137–163.

Iannaccone, Laurence. "The Economics of Religion". *Journal of Economic Literature*, vol 36, no. 3, 1998, pp. 1465-1495. *JSTOR*, Accessed 6 Mar 2021.

Kumar, Alok, Jeremy K. Page, and Oliver G. Spalt (2011). Religious Beliefs, Gambling Attitudes, and Financial Market Outcomes. *Journal of Financial Economics* 102, 671–708.

León, A. and Pfeifer, C., 2013. *Religious activity, risk-taking preferences and financial behaviour: Empirical evidence from German survey data*.

Levitt, S. D., & List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives, 21*(2), 153-174. doi:10.1257/jep.21.2.153

McCleary, Rachel M., and Robert J. Barro, ‘‘Religion and Economy,’’ Journal of Economic Perspectives 20 (2006), 49–72.

Miller, Alan S., ‘‘Going to Hell in Asia: The Relationship between Risk and Religion in a Cross-Cultural Setting,’’ Review of Religious Research 42 (2000), 5–18.

Miller, A., & Hoffmann, J. (1995). Risk and Religion: An Explanation of Gender Differences in Religiosity. *Journal for The Scientific Study of Religion*, *34*(1), 63. doi: 10.2307/1386523

Mishra, M., & Mishra, S. (2016). Financial Risk Tolerance among Indian Investors: A Multiple Discriminant Modeling of Determinants. *Strategic Change*, *25*(5), 485-500. doi: 10.1002/jsc.2075

Norenzayan, A., & Hansen, I. G. (2006). Belief in Supernatural Agents in the Face of Death. *Personality and Social Psychology Bulletin, 32*, 174-187.

Noussair, C. N., Trautmann, S. T., van de Kuilen, G., & Vellekoop, N. (2012). *Risk Aversion and Religion*. (CentER Discussion Paper; Vol. 2012-073). Economics.

Renneboog, L., and C. Spaenjers. "Religion, Economic Attitudes, And Household Finance". *Oxford Economic Papers*, vol 64, no. 1, 2011, pp. 103-127. *Oxford University Press (OUP)*, doi:10.1093/oep/gpr025.

Shu, Tao, Johan Sulaeman, and P. Eric Yeung (2010). Mutual Fund Risk-Taking and Local Religious Beliefs. Working paper, University of Georgia.

Shu, T., Sulaeman, J., & Yeung, P. (2012). Mutual Fund Risk-Taking and Local Religious Beliefs. *Management Science, 58*(10).

Sunstein, C. R. (2019). *Conformity: The Power of Social Influences* (p. 60). New York City, New York: New York University Press.

Weber, Christopher. "Determinants of Risk Tolerance". *International Journal of Economics, Finance and Management Sciences*, vol 2, no. 2, 2014, p. 143. *Science Publishing Group*, doi:10.11648/j.ijefm.20140202.15.

Weber, E. U., Blais, A., & Betz, N. E. (2002). A Domain-specific Risk-attitude Scale: Measuring Risk Perceptions and Risk Behaviors. *Journal of Behavioral Decision Making, 15*, 263-290. doi:10.1002/bdm.414

Weber, Max. *The Protestant Ethic and The Spirit Of Capitalism*. 2nd ed., 1930, pp. 1-266.

Bikers, Sailors, and Convicts: Can Tattoos Predict Shortsightedness?

Molly Hurley

I. INTRODUCTION

Whether it's dying your hair, the clothes you choose to wear, or a new piercing, people like to find ways to express their individuality. Although a bit more permanent than wearing your favorite t-shirt, tattoos have been a popular form of self-expression for centuries, dating all the way back to 3,200 B.C. Throughout history tattoos have been used to signal status, membership in social circles, and protection during childbirth, and their cultural significance can still be seen today in cultures like that of the indigenous Māori people of New Zealand.[1] In cultures like the U.S., tattooing holds less cultural significance but is still practiced, growing in popularity in recent years. According to a survey conducted by Statista in 2019, 44% of Americans have at least one tattoo and 39% of Americans without tattoos are considering getting one.[2] Despite nearly half of all Americans having at least one tattoo, there is a cultural stigma around tattoos, associating tattooed people with criminals, drug addicts, and generally more reckless individuals. In the same survey by Statista, respondents were most likely to describe people with tattoos as "Rebellious", including respondents with tattoos.[3] Negative stereotypes of people with tattoos may have real life consequences, leading employers to make premature judgements about an applicant's employability based on their choice to get a tattoo. Employers may think having something permanently tattooed on your body is an indication of shortsightedness and impulsive behavior. Yet is there any merit to these claims or are they just lingering sentiments from a time when only the lower class and criminals had tattoos? Through this study, I examined if having tattoos can predict people's time preferences, using responses from a discounting experiment,

[1] Cate Lineberry, "Tattoos: The Ancient and Mysterious History," *Smithsonian Magazine*, January 1, 2007, https://www.smithsonianmag.com/history/tattoos-144038580/

[2] Alexander Kunst, "United States - Americans with Tattoos 2019," *Statista*, December 12, 2019, https://www.statista.com/statistics/719662/americans-with-tattoos/.

[3] Statista report discussed in Mia Taylor, "How Much People Spend on Tattoos and Other Amazing Facts About Ink in America," *Cheapism*, March 05, 2020, https://blog.cheapism.com/tattoo/#slide=5

cognitive-reflection-test (CRT) questions, and self-reported financial, health, social, and ethical behaviors. I conducted regression analysis to test the relationship between and statistical significance of tattoo status and time preferences. A significant relationship would argue that there are differences in time orientation between people with tattoos and without tattoos, suggesting that tattoos are an indication of shortsightedness and impulsivity. Any findings from this study should not be used to justify taste-based discrimination against people with tattoos, but instead should be more telling of their differing work styles and what environment is most compatible with their time preferences.

This paper follows the study conducted by Ruffle and Wilson (2019) and is organized as follows: Section I introduces tattooing, attitudes towards people with tattoos, and the motivation for researching the time preferences of people with and without tattoos. Section II reviews the literature on tattoo stigma, potential employment disadvantages, and differences in time preferences. Section III describes the experimental design and procedures of the study. Section IV presents the initial results of the survey and findings of the regression models. Section V interprets the study's findings, and Section VI summarizes the study, discusses limitations, and proposes areas for future research related to tattoos and time preferences.

II. LITERATURE REVIEW

Without knowing any other characteristics of a person, people often associate tattoos with recklessness, impulsivity, and criminal history. Adams (2009) explored the history and current perceptions of tattoos, discussing their association with deviant behavior and self-expression. Tattoos have cycled in and out of popular fashion but have always been seen as a sign of deviance for lower classes due to its popularity with criminals and the working class (Adams 2009, 270). The medical field contributed to the stigmatization of tattoos, labeling them as dirty and indicators of risky and poor judgement (Adams 2009, 273). During the tattoo renaissance of the 1960s through the 1980s, tattoo interest spread to higher class professionals and women, shifting the perception from deviance to art (Adams 2009, 271). To study current tattoo prevalence and perception, Adams conducted surveys across the US and compared perceptions of tattoos between tattooed and non-tattooed participants. He found strong negative relationships between education and age, and tattoo status, contrary to the demographic shifts seen during the renaissance (Adams 2009, 281). He did find that there were no significant differences in tattoo prevalence between men and women, suggesting that the male norm of tattoos no longer remains (Adams 2009, 282). Looking at the stigmatization of tattoos, Adams found that the strongest predictor of having a tattoo was spending three or more days in jail, followed by having friends or family members with tattoos, and educational attainment was a significant negative predictor (2009, 282-3). He also found that the only significant predictor of having a face, neck, or hand tattoo was spending three or more days in jail (Adams 2009, 283). This suggests that there still is a cultural association of tattoos with criminal history, especially for certain parts of the body, and although the demographics of people with tattoos is widening, tattooing remains prevalent with younger, lower class social deviants.

Supporting Adams' (2009) study, recent research by Broussard and Harton (2018) examined current tattoo perception among college students and the larger community. In their first study, college students were shown pictures of men and women with and without tattoos and were asked to rate them on character attributes. The participants were then evaluated on their own personality traits, drinking behavior, and cognitive ability, and their results were separated between male and female participants and those with and without tattoos. They found that participants rated tattooed pictures more negatively than non-tattooed pictures and tattooed participants engaged in more drinking behavior, yet this finding could be due to age differences (Broussard and Harton 2018, 532). Their second study used the same procedure as the first study but with an older and more diverse sample. They found tattooed pictures were also rated more negatively than non-tattooed pictures but did not find any significant differences between participants with and without tattoos in drinking behavior, cognitive ability, or personality (Broussard and Harton 2018, 535). Their findings argue that there is a significant bias against people with tattoos, despite little to no measured differences in personality, cognitive ability, or drinking behavior.

Researchers have examined the employment effects of tattoo stigma, testing if people with tattoos have a more difficult time finding employment than those without tattoos. Brallier et al. (2011) conducted a study to examine the impact of visible tattoos on employment in the restaurant service industry. They found significant decreases in willingness to hire tattooed applicants compared to non-tattooed applicants (Brallier et al. 2011, 74). Despite reporting the same qualifications across all applications, managers preferred to hire non-tattooed candidates, suggesting that they hold negative biases against applicants with tattoos. In addition, they found a significant difference in willingness to hire between the non-tattooed female and both the tattooed male and female applicants, whereas the willingness to hire the non-tattooed male was not statistically different from the tattooed male and female (Brallier et al. 2011, 75). This argues that visible tattoos pose a greater employment disadvantage for women than men in this industry, possibly due to existing gender stereotypes and judgements about who tattoos are acceptable for.

Although this study suggests that hiring managers hold negative stereotypes about applicants with visible tattoos, studies looking at actual employment and income statistics have found little to no difference in employability between people with and without tattoos. Dillingh, Kooreman, and Potters (2016) conducted a study in The Netherlands, collecting demographic, financial, and tattoo-related information from individuals. They identified tattoo status by visibility level, differentiating between objectively visible tattoos, such as on the face, neck, and hands, and subjectively visible tattoos. Their initial data summary showed significant differences in characteristics of the tattooed and non-tattooed participants. On average, those with tattoos were younger, more often disabled or blue-collar workers, lower educated, more often had poorer mental and physical health, and more often smoked or used drugs (Dillingh, Kooreman, and Potters 2016, 195). Their regression analysis showed no support for age differences but found a significant negative correlation between education level and having a tattoo (Dillingh, Kooreman, and Potters 2016, 197) and a significant positive correlation between substance use and visible tattoos (203). Relating to employment, they found no significant relationship between tattoo status and employment but found a significant negative relationship between objectively visible tattoos and income, yet the relationship was insignificant after controlling for fixed effects (Dillingh, Kooreman, and Potters 2016, 200). Despite the sociodemographic differences between tattooed participants and non-tattooed participants, they found that tattoo status could not significantly predict income or employment.

Although Dillingh, Kooreman, and Potters (2016) could not confidently conclude that tattoo status had a significant impact on employment, their findings on the relationship between education and visible tattoos supports Adams' (2009) finding that lower education level is a strong predictor of tattoo status. Considering their other findings related to criminal history and substance use, research suggests that there are common characteristics of people who choose to get tattoos that differ from those who choose not to get tattoos, and differences in work-specific characteristics may help explain the relationship between tattoos and employment. Researchers

Ruffle and Wilson (2019) tested differences in work-related characteristics of people with and without tattoos, finding significant differences in time preferences between participants without tattoos, participants with tattoos that could be hidden with clothing, and participants with at least one visible tattoo, such as on the face, neck, or hands. They conducted three experiments, one measuring the participant's time discounting and two measuring honesty, followed by an extensive questionnaire with questions about risky behaviors, cognitive-reflection-test (CRT) questions to measure impulsivity, and detailed tattoo-related questions. From the first experiment, they found that participants with visible tattoos were more irrational than participants with no or hidden tattoos, as well as more shortsighted than those with hidden tattoos, who were more shortsighted than those without tattoos (Ruffle and Wilson 2019, 7). In a separate paper, Ruffle and Wilson found no significant difference in honesty between participants with and without tattoos (Ruffle and Wilson 2018, 6). In the self-reported behaviors, they found tattooed participants were more shortsighted in their financial, health, and social behaviors such as saving less for retirement, drinking and smoking more, and posting more personal or controversial comments on social media (Ruffle and Wilson 2019, 9). Tattooed participants were also more shortsighted in their CRT responses. Overall, those without tattoos performed the best and those with hidden tattoos performed better than those with visible tattoos for 3 out of the 4 questions. The participants with tattoos had greater intuitive incorrect answers, suggesting they are more impulsive and answer without proper deliberation (Ruffle and Wilson 2019, 9). With their findings, Ruffle and Wilson conducted a series of regression analyses, examining the impact of hidden and visible tattoos on participants' time preferences. They found that those with visible tattoos were significantly more present-oriented than non-tattooed participants, regardless of the number of tattoos they have, motive for getting a tattoo, the recency of their last tattoo, the age that they got their first tattoo, their likelihood to take risks, or their impulsivity (Ruffle and Wilson 2019, 18). They also found that intention to get a tattoo in the next year predicted time preference, regardless of if they had a tattoo already or not, suggesting that short-sightedness causes people to get tattoos, not the other way around (Ruffle and Wilson 2019, 15). Ruffle and Wilson argue that people's time preferences can accurately be predicted by their likelihood of getting or having a tattoo, therefore people with tattoos are more likely to exhibit shortsighted characteristics. They conclude that their findings should not encourage discrimination in the workplace against people with tattoos but should instead suggest different work style preferences for people who have tattoos (Ruffle and Wilson 2019, 19).

III. METHODS

a. Experimental Design

I surveyed a random sample of Macalester students chosen by the Institutional Research Office as well as a sample of my friends and peers. I used a 3x4 experimental design, observing responses to a time preference experiment, self-reported behaviors, and cognitive-reflection-test (CRT) questions based on tattoo status. I differentiate tattoo status by no tattoos, hidden tattoos, semi-visible tattoos, and visible tattoos. I listed examples of areas that correspond to the different categories, but participants' tattoo identification ultimately depended on the participants' discretion. Hidden tattoos were described to the participants as those on the chest, back, and upper legs. Semi-visible tattoos were described to the participants as those on the wrists, arms, and lower legs. Visible tattoos were described to the participants as those on the face, neck, and hands. Building on Ruffle and Wilson's (2019) study, I further differentiate between tattoo visibility, adding a semi-visible category because I was interested in seeing if there are significant differences between people with more visible tattoos, such as arm tattoos, and people with more hidden tattoos, such as back tattoos. In Ruffle and Wilson's (2019) study, arm and back tattoos would both be considered hidden tattoos, yet in an employment context, this is only true if the workplace allows long sleeves. Required work uniforms vary across industries and there are some where arm tattoos would be visible, such as the service industry (Brailler et al. 2011). Therefore, by further differentiating hidden tattoos into semi-visible and hidden categories, I hoped to find a stronger relationship between people with semi-visible tattoos and shortsightedness.

b. Survey Design

The survey was comprised of a time preference experiment, followed by self-reported financial, health, social, and ethical behaviors, 3 CRT questions, and a sociodemographic questionnaire with tattoo specific questions. The time preference experiment was first developed by Coller and Williams (1999) using a hypothetical payment schedule to determine individual discount rates among their participants. I used the simplified version adapted by Ruffle and Wilson (2019), summarized below.

Table 1: Time Preference Experiment

Scenario	Option A	Option B
1.	$1.00 in 24 hours	$1.00 in 3 weeks
2.	$1.00 in 24 hours	$1.05 in 3 weeks
3.	$1.00 in 24 hours	$1.10 in 3 weeks
4.	$1.00 in 24 hours	$1.20 in 3 weeks
5.	$1.00 in 24 hours	$1.30 in 3 weeks
6.	$1.00 in 24 hours	$1.45 in 3 weeks
7.	$1.00 in 24 hours	$1.65 in 3 weeks
8.	$1.00 in 24 hours	$1.90 in 3 weeks
9.	$1.00 in 24 hours	$2.20 in 3 weeks
10.	$1.00 in 24 hours	$2.50 in 3 weeks

I changed the time of Option A payment to 24 hours to ease participants' strain of imagining the present payment. This way, they will receive Option A exactly one day later, instead of having to mentally calculate 18 hours from now. Participants saw all 10 scenarios and indicated if they would prefer to receive Option A or Option B for each pair. Option A was always $1.00 in 24 hours and represents the present-oriented option. Option B started at $1.00 and gradually increased with each scenario, equaling $2.50 in scenario 10. Option B was scheduled to be paid in 3 weeks, representing the future-oriented option. Because the payments in Option A and Option B are equal for the first scenario, participants were expected to prefer Option A. As the scenarios continue and the payment of Option B increased, participants were expected to switch from preferring Option A to preferring Option B and continue to prefer Option B for the remaining scenarios. Participants' discount rate or time preference was observed by the scenario that they switch from Option A to Option B, where switching sooner indicated future orientation and switching later indicated present orientation. Additionally, switching option preference multiple times (i.e., switching from Option A to Option B, then from Option B to Option A, then back to Option B) indicated irrationality because if a participant preferred a larger future payment over a smaller present payment, their preference should not change when the future payment increased and the present payment stayed the same. I compared the scenario that participants switched from Option A to Option B across the 4 tattoo status categories and noted any irrational behavior. If tattoos predict time preferences, I expected to see later scenario switching for tattooed participants compared to non-tattooed participants, indicating those with tattoos are more shortsighted and present-oriented. Even though Option A is the present-oriented option, payment is delayed by 24 hours to control for Option B's risk of non-payment.

Next, participants rated their likelihood of engaging in 8 behaviors, 2 financial behaviors, 2 health behaviors, 2 social behaviors, and 2 ethical behaviors. Each question identified a present-oriented behavior that could have negative future implications, and regular participation in these behaviors suggests a pattern of shortsightedness (Ruffle and Wilson 2019). Assessing participation in behaviors from each of these categories measured participants' risk preferences (Weber, Blais, and Betz 2002). I used the health and social behaviors adapted by Ruffle and Wilson (2019), ethical behaviors adapted by Weber, Blais, and Betz (2002), and financial behaviors I specifically identified for a college-aged population. The behaviors are summarized below in Table 2.

Table 2: Self-Reported Behavior Assessment	
Financial:	How likely are you to order delivery food when you have groceries at home? How likely are you to not consider your recent expenditures before making a purchase?
Health:	How likely are you to not exercise regularly? How likely are you to smoke or vape more than 3 days a week?
Social:	How likely are you to post personal or private information on social media? How likely are you to post controversial statements or opinions on social media?
Ethical:	How likely are you to copy a classmate's test answers? How likely are you to shoplift a small item from Target?

Participants rated their likelihood of engaging in these behaviors on a scale from 1 to 5, 1 indicating "Not At All Likely", 2 indicating "Slightly Unlikely", 3 indicating "Neither Likely Nor Unlikely", 4 indicating "Somewhat Likely", and 5 indicating "Very Likely". I compared risk preferences across each tattoo status category and expected to see higher risk-taking behaviors and shortsightedness in tattooed participants than non-tattooed participants.

After completing the self-reported behavior assessment, participants answered 3 cognitive-reflection-test (CRT) questions to evaluate their impulsivity. I included 2 of the common CRT questions developed by Frederick (2005) and 1 less common CRT question developed by Thomson and Oppenheimer (2016). Similar to Ruffle and Wilson (2019), I changed specific details from each question to reduce familiarity, as well as asked participants how familiar they are with each of the CRT questions. Participants chose from 3 answers, one that is intuitive but incorrect, one that is correct, and one that is unintuitive and incorrect (wrong). The intuitive answer suggests impulsivity given limited deliberation, the correct answer suggests careful and sufficient deliberation, and the wrong answer suggests impulsivity given no deliberation. The CRT questions are summarized in Table 3 below.

Table 3: CRT Questions

Question	Intuitive	Correct	Wrong
A burger and fries cost $5.90. The burger cost $5 more than the fries. How much do the fries cost?	$0.90	$0.45	$0.75
If you are running a race and you pass the person in 3rd place, what place are you in?	2nd place	3rd place	4th place
It takes 15 printers 15 minutes to print 15 pages. How long would it take 20 printers to print 20 pages?	20 min	15 min	30 min

First developed by Frederick in 2005, the cognitive-reflection-test included 3 questions that required slow, reflective deliberation to answer correctly. Without thoughtful calculation, people often provide an incorrect, yet intuitive, answer, suggesting they are more impulsive. Frederick (2005) also found that CRT responses and time preferences are correlated, finding people who answered more CRT questions incorrectly preferred sooner, smaller payments over larger future payments. For my study, I examined the correlation between impulsivity and tattoo status and expected to find that tattooed participants responded with more incorrect (intuitive and wrong) answers than nontattooed participants.

Finally, participants completed a sociodemographic questionnaire with tattoo specific questions. I included many of the demographics included in Ruffle and Wilson's (2019) study, as well as Macalester College specific questions such as eligibility for a work study through financial aid. The tattoo specific questions asked participants how many tattoos they have and where they are, their likelihood to get a tattoo, and how prevalent tattoos are among their family and friends.

c. *Procedures*

After completing the survey, participants had the option to enter a raffle to win 1 of 3 $10 prizes, distributed through Venmo or electronic Amazon gift card by email. I expected this monetary incentive would encourage participants to complete my survey, ensuring a substantial sample size (*N*). Once all the survey data was collected, I conducted regression analyses to determine if tattoo status can significantly predict shortsightedness, measured by the scenario number that participants switched from Option A to Option B in the time preference experiment, omitting any irrational behavior. I controlled for risk preference, measured by the self-reported behavior assessment, impulsivity, measured by the CRT responses, and sociodemographic characteristics. I was especially interested in gender differences and conducted separate regression analysis by gender.

IV. RESULTS

a. *Survey Results*

i. Sociodemographic Questions

I received a total of 85 complete responses, 48 female students (56%), 28 male students (33%), and 7 non-binary students (8%), with 5 students (6%) identifying as transgender. Most participants were seniors (40%), followed by juniors (25%), sophomores (18%), and first-year students (16%). The majority of participants identified as white (80%) followed by Asian or Asian American (22%).[4] Most participants identified as either heterosexual (51%), bisexual (18%), or queer (11%). About half of the participants (49%) were majoring in a social science field, followed by natural science and mathematics (32%) and fine arts and humanities (28%).[5] 38% of participant were from the Midwest, 21% from the West, 15% from the Northeast, 5% from the Southwest, 6% from the Southeast, and 6% of participants were international students. Participants mostly identified as Democrats (78%), were eligible for work study through financial aid (62%), and had no or weak religious beliefs (49%). 89% of participant GPAs ranged from 3.00 to 3.99 on a 4.00 scale. 60 participants (71%) did not have any tattoos and 25 participants (29%) had at least one tattoo, 4 with visible tattoos, 13 with semi-visible tattoos but no visible tattoos, and 8 with only hidden tattoos. On average, tattooed participants had 2.52 tattoos.[6] 83% of non-tattooed participants were considering getting a tattoo and 35% of all participants were considering getting a tattoo within the next year. Detailed demographic summary statistics by tattoo status are shown in Table 4 below. Values may total less than 100% due to participants who chose not to respond to the question, and values may total greater than 100% due to multiple responses per participant (e.g., double majors, multiracial identities).

[4] Values may total greater than 100% due to biracial and multiracial responses.
[5] Values may total greater than 100% due to double and triple major responses.
[6] This is an underestimate due to the survey's format. Participants with 6 or more tattoos were recorded as having only 6 tattoos.

Table 4: Sociodemographic Summary Statistics, by tattoo status	Tattooed			Non-tattooed
	Visible	Semi	Hidden	
Class Year:				
First-Year	0%	0%	0%	23%
Sophomore	0%	15%	13%	20%
Junior	25%	23%	13%	27%
Senior	75%	62%	75%	28%
Major:				
Social Studies	25%	69%	25%	50%
Natural Sciences and Mathematics	50%	23%	50%	30%
Fine Arts and Humanities	25%	31%	25%	28%
Undeclared	0%	8%	0%	13%
GPA:				
2.00 - 2.69	0%	8%	0%	2%
2.70 - 2.99	0%	0%	0%	2%
3.00 - 3.69	50%	38%	63%	30%
3.70 - 3.99	25%	46%	38%	60%
4.00	25%	0%	0%	7%
Hometown (Region):				
West	25%	15%	38%	20%
Southwest	0%	0%	0%	7%
Midwest	25%	31%	38%	40%
Southeast	0%	8%	0%	7%
Northeast	25%	15%	0%	17%
International	0%	0%	0%	8%
Eligible for Work Study:	75%	54%	88%	60%
Gender:				
Female	75%	54%	75%	53%
Male	0%	23%	25%	38%
Non-binary	25%	15%	0%	7%
Transgender	0%	15%	0%	5%
Race and Ethnicity:				
White	50%	92%	63%	82%
Middle Eastern or North African	25%	0%	0%	2%
Hispanic, Latino, Latina, or Latinx	0%	8%	13%	3%
Black or African American	0%	8%	0%	2%
Asian or Asian American	25%	0%	25%	27%
American Indian or Alaska Native	25%	0%	0%	2%
Native Hawaiian or Other Pacific Islander	25%	0%	0%	0%

Sexual Orientation:				
Heterosexual	50%	38%	63%	52%
Bisexual	25%	31%	25%	13%
Gay	0%	0%	0%	5%
Queer	0%	31%	13%	7%
Lesbian	0%	0%	0%	7%
Asexual	0%	0%	0%	7%
Pansexual	25%	0%	0%	3%
Strength of Religious Beliefs:				
Not at all religious: 1	25%	31%	25%	23%
2	0%	15%	50%	25%
Somewhat religious: 3	25%	38%	0%	15%
4	0%	8%	13%	7%
Very religious: 5	25%	0%	0%	2%
Average strength:	3.00	2.25	2.00	2.14
Political Party:				
Democrat	75%	85%	75%	77%
Republican	25%	0%	13%	2%
Independent	0%	8%	0%	13%
Other	0%	0%	13%	3%

Non-binary and female participants were more likely than male participants to have tattoos. 43% of non-binary participants, 33% of female participants, and only 18% of male participants had tattoos. This trend was consistent across visible (14% and 6% vs. 0%) and semi-visible (29% and 15% vs. 11%) tattoos, and female participants were more likely to have hidden tattoos than male participants (13% vs. 7%). The sociodemographic distributions were relatively consistent across tattoo status. According to Table 4, significant differences occurred in class year, GPA, work study eligibility, and race. Tattooed participants were more likely to be seniors and none of the first-year participants had tattoos, possibly due to age restrictions. Participants with visible and hidden tattoos were more likely to have 3.00 to 3.69 GPAs whereas participants with semi-visible tattoos and no tattoos were most likely to have GPAs between 3.70 and 3.99, suggesting that some tattooed participants are less intelligent than non-tattooed participants.

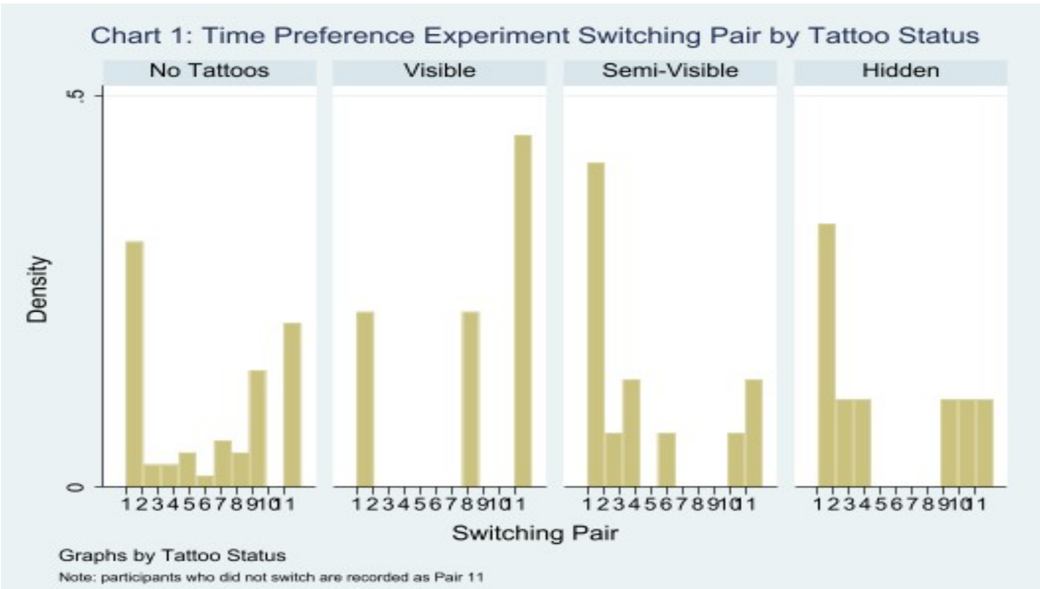
Tattooed participants were more likely to be eligible for work study through financial aid, suggesting that they have lower income than non-tattooed participants. Participants with visible and hidden tattoos were less likely to identify as white than participants with semi-visible tattoos and no tattoos. Additionally, strength of religious beliefs increased with tattoo visibility, and both visibly and semi-visibly tattooed participants had greater religious scores than non-tattooed participants.

Table 5 below reports summary statistics for some of the tattoo specific questions. According to the table, participants with visible and semi-visible tattoos were more likely to have family members and friends with tattoos than participants with hidden or no tattoos. Also, all tattooed participants were more likely to get a tattoo within the next year than non-tattooed participants.

Table 5: Tattoo Summary Statistics, by tattoo status	Tattooed Visible	 Semi	 Hidden	Non-tattooed
Family that has tattoos:				
None	25%	23%	50%	38%
Few	0%	46%	25%	47%
Some	50%	23%	25%	13%
Many	0%	0%	0%	2%
Most/All	25%	8%	0%	0%
Friends that have tattoos:				
None	0%	8%	0%	8%
Few	0%	8%	25%	33%
Some	25%	54%	50%	43%
Many	50%	15%	25%	13%
Most/All	25%	15%	0%	2%
Likelihood of getting a tattoo within the next year:				
Not at all likely	50%	0%	13%	23%
Slightly unlikely	0%	0%	0%	40%
Neither likely nor unlikely	25%	23%	38%	12%
Somewhat likely	0%	8%	0%	5%
Very likely	25%	69%	50%	3%

ii. Time Preference Questions

Chart 1 below reports the results from the first experiment that measured time discounting. Participants that did not switch from the present-oriented option (Option A) to the future-oriented option (Option B) were recorded as switching in Pair 11. As seen in Chart 1, there are no significant differences between tattooed and non-tattooed participants or across tattoo status, suggesting that there is no measurable difference in time discounting between tattooed and non-tattooed individuals, and tattoos cannot accurately predict time preferences. Visibly tattooed participants appear to switch later than all other participants, yet this result may be due to the small visible sample size (*N* = 4). All participants responded rationally, switching from Option A to Option B only once, but interestingly, 3 participants reported preferring Option B for all 10 pairs, even when both payments equaled $1.00, indicating extreme future-orientation (1 semi-visible, 2 non-tattooed).



I found little difference in participants' mean switching pair by tattoo status. Contrary to Ruffle and Wilson (2019), participants with semi-visible and hidden tattoos switched earlier than non-tattooed participants, at pair 4.62 (SD = 3.54) and 5.38 (SD = 3.67) on average compared to pair 6.18 (SD = 3.73). On average, visibly tattooed participants switched the latest, switching at pair 8.00 (SD = 3.67), yet this result may be due to the small sample size ($N = 4$).

Table 6 below summarizes the results from the self-reported risky behaviors questions. Larger mean scores indicate greater likelihood of engaging in the stated behaviors.

Table 6: Risky Behaviors Mean Score (SD), by tattoo status	Tattooed			Non-tattooed
	Visible	Semi	Hidden	
Financial:				
Ordering delivery	3.00 (1.22)	3.69 (0.99)	2.50 (1.00)	2.70 (1.32)
Spending without thinking	3.25 (1.79)	2.08 (0.83)	2.00 (1.00)	2.40 (1.37)
Health:				
Smoking or vaping	2.00 (1.73)	3.23 (1.85)	2.38 (1.80)	1.57 (1.24)
Not exercising	2.75 (1.09)	2.77 (1.42)	2.63 (1.49)	2.62 (1.43)
Social:				
Private or personal posts	3.00 (1.58)	2.69 (1.20)	2.13 (1.17)	2.00 (1.21)
Controversial posts	2.75 (1.79)	2.92 (1.21)	1.88 (1.17)	2.12 (1.27)
Ethical:				
Cheating on exam	2.25 (0.83)	1.77 (1.12)	1.13 (0.33)	2.02 (1.10)
Stealing small item	2.00 (1.73)	2.69 (1.59)	2.13 (1.45)	1.52 (0.94)

Overall, every mean score was below 4 ("Somewhat Likely"), indicating participants were not likely to participate in the risky behaviors and suggests that this specific population was not risk-taking. All tattooed participants were more likely to smoke or vape more than three days a week, not exercise regularly, post private or personal information on social media, and steal small items from Target than non-tattooed participants. Participants with visible tattoos were more likely to engage in all 8 behaviors than non-tattooed participants, supporting Ruffle and Wilson's (2019) findings. In addition, participants with semi-visible tattoos had higher mean scores than non-tattooed participants for every behavior except spending without thinking about recent expenditures and cheating on an exam.

Below, Table 7 reports the CRT results by tattoo status. For the first CRT question, non-tattooed and visibly tattooed participants were equally likely to answer the question correctly (50%) and more likely to answer correctly than participants with semi-visible and hidden tattoos (46% and 38%), but those with visible tattoos were the only participants that selected the unintuitive and incorrect (wrong) answer, suggesting greater impulsivity. For the second question, all participants with hidden tattoos answered correctly, followed by non-tattooed participants (82% correct), semi-visibly tattooed participants (77% correct), and visibly tattooed participants (75% correct). Again, those with visible tattoos were the only participants who selected the wrong answer, further suggesting greater impulsivity. For the third question, participants with visible tattoos answered correctly most often (75%), followed by non-tattooed participants (47%), participants with hidden tattoos (38%), and those with semi-visible tattoos (15%). In addition, those with semi-visible tattoos were the only participants that selected the wrong answer, suggesting greater impulsivity. Overall, visibly tattooed participants were most likely to answer all three questions correctly (50%), followed by non-tattooed participants (35%), and those with semi-visible and hidden tattoos (15% and 13%). Visibly and semi-visibly tattooed participants were most likely to answer all three questions incorrectly (25% and 23%), either intuitive or wrong, followed by non-tattooed participants (13%). No participants with hidden tattoos answered all three questions incorrectly. It is important to state that the visibly tattooed results may be due to the small sample size ($N = 4$). Ignoring the visibly tattooed results, participants without tattoos were more likely to answer correctly for questions 1 and 3, were more likely to answer all three questions correctly, and less likely to answer all three questions incorrectly than tattooed participants.

Table 7: CRT Summary Statistics, by tattoo status	Tattooed			Non-tattooed
	Visible	Semi	Hidden	
CRT 1: Burger and fries				
Correct	50%	46%	38%	50%
Intuitive	25%	54%	63%	50%
Wrong	25%	0%	0%	0%
CRT 2: Passing 3rd place				
Correct	75%	77%	100%	82%
Intuitive	0%	23%	0%	18%
Wrong	25%	0%	0%	0%
CRT 3: 15 printers				
Correct	75%	15%	38%	47%
Intuitive	25%	69%	63%	53%
Wrong	0%	15%	0%	0%
All Correct	50%	15%	13%	35%
All Incorrect	25%	23%	0%	13%

Because the CRT questions have become popular within psychology and behavioral economics fields, I expected some students to be familiar with the CRT questions. Although I changed the details of the questions, the fundamental theory behind each of the questions is the same and would be easier to answer correctly for students who had already heard them before. To control for familiarity with the CRT questions, I asked the participants to indicate which questions they had heard before and were familiar with. Then, I tested the relationship between answering correctly and familiarity for each question, described below in Table 8. Contrary to my expectations, familiarity did not significantly predict response accuracy for any of the questions. Therefore, familiarity will not be included in my final regression analysis.

Table 8: CRT Accuracy Explained by Familiarity with the CRT Questions			
	(1)	(2)	(3)
VARIABLES	CRT 1	CRT 2	CRT 3
Familiar with CRT question	0.0540	0.00278	-0.214
	(0.123)	(0.0838)	(0.137)
Constant	0.468***	0.822***	0.464***
	(0.0641)	(0.0575)	(0.0593)
Observations	85	85	85
R^2	0.002	0.000	0.029
$R^2_{adjusted}$	-0.00972	-0.0120	0.0169
F	0.192	0.00110	2.444
RSS	21.17	12.35	20.16
Standard errors in parentheses			
*** p<0.01, ** p<0.05, * p<0.10			

b. Regression Results

Table 9 below summarizes the regression results, using tattoo status to explain time preferences. Table 9 is an abbreviated version of the actual regression analyses I conducted, and a full report is detailed at the end of this paper in Table 11. Regressions (5), (6), and (7) have been omitted from Table 9 for brevity, as well as the following sociodemographic characteristics: class year, academic major, GPA, hometown regional area, eligibility for work study, gender, white, bisexual, strength of religious beliefs, political party, and likelihood of getting a tattoo within the next year. Regression (5) used the same variables as Regression (4) but controlled for participants' total number of tattoos. Regression (6) added a non-linear control for total number of tattoos (i.e., total number of tattoos^2). Regression (7) added a control for risk preference, measured by the likelihood of engaging in risky, present-oriented behaviors. All the results omitted from Table 9 showed no statistical significance. Tattoo status was divided into hidden, semi-visible, and visible categories, each represented by corresponding variables, where the constant represents non-tattooed participants. The dependent variable, time preference, is measured by the pair that participants switched from a present-oriented payment (Option A) to a future-oriented payment (Option B) in the first experiment. Participants who did not switch to Option B and chose the present-oriented Option A for each scenario were recorded as switching in Pair 11.

Table 9: Time Preference Explained by Tattoo Status, Controlling for Risk Taking, Impulsivity (CRT), and Sociodemographic Characteristics (Abbreviated)
Dependent Variable is the Switching Pair from the Time Discounting Experiment

VARIABLES	(4)	(8)	(9)
Hidden	0.266	-2.979	-5.102
	(0.403)	(3.903)	(4.237)
Semi-Visible	-0.817*	-4.588*	-5.398*
	(0.486)	(2.437)	(2.663)
Visible	0.632	2.018	0.537
	(0.698)	(2.132)	(2.323)
Asian		-4.052	-5.140*
		(2.675)	(2.657)
Queer		3.197	8.181*
		(3.205)	(4.101)
Family with Tattoos		-2.535**	-2.300*
		(1.113)	(1.131)
Friends with Tattoos		-2.846*	-4.666**
		(1.419)	(1.721)
CRT 1 Correct			-11.16**
			(3.958)
CRT 2 Correct			1.421
			(3.088)
CRT 3 Correct			0.210
			(2.426)
All Correct			6.932
			(4.249)
Constant	6.046***	35.19**	48.12**
	(0.440)	(15.49)	(20.62)
Observations	85	53	53
R^2	0.059	0.665	0.797
$R^2_{adjusted}$	0.0237	-0.0259	0.187
F	1.681	0.963	1.306
RSS	1137	256.8	155.7

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

Positive coefficients indicate an increase in the switching pair (i.e., greater present orientation) and negative coefficients indicate a decrease in the switching pair (i.e., greater future orientation). If my findings are consistent with the existing literature, the tattoo status variables should all have significant positive coefficients, increasing in magnitude as visibility increases, arguing that tattooed participants are more shortsighted than non-tattooed participants and the strength of their present orientation increases with tattoo visibility.

Regression (4) reports the results from a regression using tattoo status to explain time preferences without controlling for any other variables. The Hidden and Visible coefficients are statistically insignificant, and the Semi-Visible coefficient is slightly significantly negative ($p < 0.10$), indicating that participants with semi-visible tattoos switch 0.817 pairs earlier than non-tattooed participants. This finding is inconsistent with previous research and suggests that not only are tattooed individuals not more likely to be present-oriented but may even be more future-oriented than non-tattooed individuals. To be as consistent with Ruffle and Wilson's (2019) study, I conducted a series of regressions combining my Hidden and Semi-Visible variables and the results are reported at the end of this paper in Table 12. I failed to replicate their results, finding that the semi-visible results were not strong enough to show statistical significance when combined with the hidden results, and both the Hidden and Visible coefficients were statistically insignificant. Regressions (11), (12), and (13) from the combined series also displayed no statistically significant coefficients.

Regression (8) reports the abbreviated results from a regression using tattoo status to explain time preferences, controlling for total number of tattoos, non-linear total number of tattoos, risk taking behavior, and sociodemographic characteristics. All significant findings are reported in Table 9 and the coefficients are highlighted in yellow. The slightly significant negative correlation between Semi-Visible and time preference remained, decreasing to -4.588 and indicating that semi-visibly tattooed participants switched 4.588 pairs earlier than non-tattooed participants when controlling for total number of tattoos, risk preference, and sociodemographic characteristics. Family and friends with tattoos also displayed significantly negative coefficients. On a scale from 1 to 5, 1 indicating "Not at all likely" and 5 indicating "Very likely", a 1-point increase in the likelihood that participants' family members had tattoos led to a 2.535 decrease ($p < 0.05$) in switching pair (i.e., switched 2.535 pairs earlier). Similarly, a 1-point increase in the likelihood that participants' friends had tattoos led to a 2.846 decrease ($p < 0.10$) in switching pair (switching 2.846 pairs earlier). Contrary to the existing research, these two findings suggest that holding tattoo status constant, having family or friends with tattoos increases future-orientation. Regression (14) from the combined series found similar results, finding significant negative coefficients for the combined variable (semi-visible and hidden) and family and friends with tattoos variables. It also found a slightly significant positive relationship between the non-linear total number of tattoos variable and time preferences, indicating that an increase in the number of tattoos led to switching 0.928 pairs later. Although Ruffle and Wilson (2019) did not find a significant coefficient on their $tattoos^2$ variable, this finding does suggest that having more tattoos increases shortsightedness at a non-linear rate.

Regression (9) replicates the previous regression but adds a control for impulsivity, measured by participants' accuracy on each CRT question. The negative correlation between semi-visible tattoos and time preference remains slightly significant ($p < 0.10$) and has decreased again to -5.398, further suggesting that participants with semi-visible tattoos are significantly more future-oriented than non-tattooed participants. The significance of family and friends with tattoos also is seen in this regression, with the family members coefficient decreasing in significance ($p < 0.10$) and magnitude to -2.300 and the friends coefficient increasing in significance ($p < 0.05$) and magnitude to -4.666. I also found significant coefficients on the Asian ($p < 0.10$), Queer ($p < 0.10$), and CRT 1 Correct ($p < 0.05$) variables. According to my results, Asian or Asian American participants switched 5.140 pairs earlier than non-Asian and non-white participants, indicating greater future orientation. Participants that identified as Queer switched 8.181 pairs later than non-Queer and non-Bisexual participants, indicating greater present orientation. This was the only statistically significant positive coefficient I found in my regressions. Lastly, participants who answered the first CRT question correctly switched 11.16 pairs earlier than participants who answered incorrectly. This finding is consistent with the literature, indicating that less impulsive participants (i.e., answering correctly) are more future-oriented. Yet, I did not see this relationship with the other two CRT questions or with participants who answered all three questions correctly. In the combined series, regression (15) found similar results, finding significant negative coefficients for the combined variable ($p < 0.05$), family ($p < 0.10$) and friends ($p < 0.05$) with tattoos variables, Asian variable ($p < 0.10$), and CRT 1 Correct variable ($p < 0.05$). It also found a significant positive coefficient for the Queer variable ($p < 0.10$), and similar to regression (14), it found a significant positive coefficient for the tattoos^2 variable ($p < 0.05$). In addition, regression (15) found a slightly significant positive relationship between nonbinary participants and time preference, finding nonbinary participants switched 7.897 pairs later than male participants ($p < 0.10$). This suggests that nonbinary participants are more present-oriented than male participants. In the following Gender Analysis section, I will further explore gender differences in my data. Regressions (8), (9), (14), and (15) have fewer observations due to “Prefer not to answer” responses to the sociodemographic questions.

i. Gender Analysis

Consistent with Ruffle and Wilson's (2019) study, I found that female participants were more likely to have tattoos than male participants. I also found that nonbinary participants were more likely to have tattoos than female and male participants. For this reason, I chose to conduct a separate regression series, separating results by gender. Unfortunately, separating results by gender reduced the sample sizes of each tattoo category for each gender. For example, none of the male participants reported having any visible tattoos. For these reasons, in order to have complete tattoo status coefficients, I did not include all control variables in the Male and Nonbinary regressions.

Table 10: Time Preference Explained by Tattoo Status by Gender
Dependent Variable is the Switching Pair from the Time Discounting Experiment

VARIABLES	(16) Female	(17) Male	(18) Nonbinary
Hidden	-27.91	-11.05	0.344
	(9.213)	(6.277)	(1.757)
Semi-Visible	-24.44	-8.981	-1.328
	(7.342)	(5.863)	(1.889)
Visible	-23.28		0.281
	(7.987)		(1.858)
Tattoos	Yes	Yes	No
Risk Preference	Yes	Yes	No
Sociodemographic	Yes	No	No
Impulsivity	Yes	Yes	No
Constant	18.25	-1.481	7.250**
	(17.75)	(4.592)	(1.931)
Observations	30	28	7
R^2	0.990	0.772	0.569
$R^2_{adjusted}$	0.715	0.488	0.137
F	3.597	2.715	1.318
RSS	4.178	92.55	44.75

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

Summarized in Table 10 above, I found no significant relationships between tattoo status and time preference for female, male, or nonbinary participants. This finding indicates that tattoos cannot accurately predict time preferences and this trend is consistent across gender. Tattoo status does not predict time preferences any more for women, men, or nonbinary individuals.

I also conducted a combined regression series consistent with Ruffle and Wilson's (2019) tattoo status, summarized in Table 13 at the end of the paper. These results showed slightly significant negative relationships between hidden tattoos and time preferences for female and male participants ($p < 0.10$), suggesting that there is a correlation for women and men, but tattoos cannot predict time preferences of nonbinary individuals. Female participants with hidden tattoos switched 20.98 pairs earlier than non-tattooed female participants. This finding does not make sense in the context of the experiment that only included a total of 10 pairs but can be interpreted as extreme future orientation for these specific individuals. Male participants with hidden tattoos switched 9.75 pairs earlier than non-tattooed male participants. The magnitude of the hidden coefficient was greater for female participants than male participants, suggesting that the relationship between hidden tattoos and time preferences is stronger for women than men.

V. DISCUSSION

Macalester College students were less likely to have tattoos than the average American, with only about 30% of participants with tattoos. This may be due to age restrictions as many first-year students may not be or recently turned 18, the legal age requirement to get a tattoo in most states. Of the participants with tattoos, only 16% had visible tattoos, suggesting that the placement of tattoos is still stigmatized in this population. Initially, I found that tattooed participants were more likely to be women and nonbinary, have lower GPAs, need greater financial assistance, and were less likely to identify as white and heterosexual. This is consistent with previous research that argues that people with tattoos are less intelligent, poorer, and more likely to be minorities than non-tattooed participants. Additionally, I found that tattooed participants were more likely to engage in some risky, present-oriented behaviors and were more impulsive. Yet, I did not see any significant differences in time discounting in the first experiment across tattoo status, contrary to Ruffle and Wilson's (2019) study. After further analysis, I found little ($p < 0.10$) to no significant relationships between tattoo status and time preferences. The only significant tattoo status was my semi-visible category, and it showed a negative correlation with time preferences, arguing that semi-visible tattoos lead to greater future orientation, not shortsightedness. This finding persisted after controlling for the total number of tattoos, risk preference, sociodemographic characteristics, and impulsivity as well as after combining the semi-visible and hidden variables, consistent with the Ruffle and Wilson (2019) design. When looking at differences across gender, I found no significant relationships between tattoo status and time preferences for female, male, and nonbinary participants, arguing that tattoos equally cannot predict time preferences for women, men, or nonbinary individuals.

I did find significant coefficients for some control variables, including Asian, Queer, Family with Tattoos, Friends with Tattoos, and CRT 1 Correct. All coefficients expect for Queer were significantly negative, arguing that Asian or Asian American participants, participants with family and friends with tattoos, and participants who answered the first CRT question correctly were more future-oriented than the other participants. There may be differences in time discounting across race and ethnicity due to different cultural views on the present and the future that could explain why Asian or Asian American participants were more future-oriented. Considering the significant negative relationship between semi-visible tattoos and time preferences, the negative relationship between family members and friends with tattoos and time preferences holding tattoo status constant argues that even if participants do not have tattoos themselves, having exposure to people with tattoos can impact their time preferences, influencing them to be more future-oriented. Yet, the significance of the correlation between tattoos and time preferences was weak, therefore it is hard to confirm this theory. The negative CRT 1 Correct coefficient is consistent with the existing theory, showing that participants with less impulsivity were more future-oriented, but the other CRT question coefficients were not statistically significant. This suggests that there were errors in my survey design, or the theory is incorrect for assuming that impulsivity is positively correlated with shortsightedness. The Queer coefficient was significantly positive, arguing that participants that identified as Queer were more present-oriented than the other participants.

VI. CONCLUSION

a. Summary of study

Tattoos have a long history of stigmatization and to this day are still associated with criminals and low social status. Despite the present bias against tattoos, more and more people in the US are choosing to get tattoos, suggesting they are more present-oriented because they are favoring the present reward of having a tattoo and disregarding the future discriminatory disadvantages of having a tattoo. Further, researchers Ruffle and Wilson (2019) found a significant relationship between tattoos and time preferences, supporting the claim that people who choose to get tattoos are more present-oriented and shortsighted than people who choose not to get tattoos. Despite following their research design closely, I was unable to replicate their findings in a Macalester College student population. I did see trends similar to the established literature initially, such as tattooed participants were more likely to need financial assistance, have lower GPAs, and belong to minority groups (e.g., gender, race and ethnicity, sexual orientation). Yet, my regression analysis did not support the Ruffle and Wilson (2019) research. I did not find that tattooed participants were more present-oriented than non-tattooed participants, even after controlling for total number of tattoos, risk preference, impulsivity, and sociodemographic characteristics. On the contrary, the only significant relationship between tattoo status and time preferences was for participants with semi-visible tattoos and they were significantly more future-oriented than non-tattooed participants. This finding persisted in the combined analysis that I conducted with tattoo status consistent with the Ruffle and Wilson (2019) study. Not only did I fail to replicate their findings, but I also found the opposite relationship for participants with hidden (i.e., semi-visible and hidden) tattoos. Considering my data only shows a significant relationship for participants with semi-visible tattoos, not hidden or visible tattoos, I argue that there are significant differences between individuals who choose to get hidden tattoos (e.g., chest, back, upper legs) and those who choose to get semi-visible tattoos (e.g., arms, wrists, lower legs) and making a distinction between hidden and semi-visible tattoos is important to isolate their separate effects.

b. Limitations

My results suggests that there is little to no difference in time preferences between people with and without tattoos and provides no significant evidence that people with tattoos are more shortsighted than people without tattoos. Still, my findings may differ from previous research due to the specificity of my sampling population. I chose to focus only on Macalester College students because I was interested in how the existing theory presented itself in a younger, more liberal population and I wanted to control for Macalester College specific characteristics such as eligibility for work study through financial aid. This limited my sample age to college undergraduate students, very different from the Ruffle and Wilson (2019) study, and limited my sample to a predominantly liberal population that may not experience the negative bias against tattoos and therefore having a tattoo would not necessarily be related to time preferences.

Additionally, my sample size was much smaller than Ruffle and Wilson's (2019) sample. Only having 25 tattooed participants, only 4 of which had visible tattoos, reduced the accuracy of my regressions and my results may not be representative of the entire tattooed and non-tattooed population at Macalester College. Instead, my results may depend to heavily on a small portion of the larger group that behaves significantly differently than the rest of their peers.

c. Areas for future research

As my study shows, more research is required to examine the relationship between tattoos and time preferences across different age groups and cultures. Although Ruffle and Wilson (2019) controlled for age in their study, my findings argue that tattoos are not a strong predictor of time preferences among liberal, undergraduate college students. Tattoos are becoming more commonplace and my study suggests that the differences between people who get tattoos and those who do not are disappearing. This phenomenon may be linked to age and younger people show fewer differences across tattoo status than older generations because their behaviors are changing with the times. There may also be differences across cultures, liberal vs. conservative, different racial groups, or Western vs. Eastern philosophies. Overall, the research on this topic would benefit greatly from more studies attempting to replicate the Ruffle and Wilson (2019) study, separating results by different age groups and conducting research over time to conclusively determine if there is a link between tattoos and time preferences and if the relationship is changing across time.

Additional Figures:

Table 11: Time Preference Explained by Tattoo Status, Controlling for Risk Taking, Impulsivity (CRT), and Sociodemographic Characteristics (Full Regression)
Dependent Variable is the Switching Pair from the Time Discounting Experiment

VARIABLES	(4)	(5)	(6)	(7)	(8)	(9)
Hidden	0.266	0.194	-0.342	-0.373	-2.979	-5.102
	(0.403)	(0.788)	(1.110)	(1.148)	(3.903)	(4.237)
Semi-Visible	-0.817*	-0.896	-1.379	-1.408	-4.588*	-5.398*
	(0.486)	(0.886)	(1.133)	(1.171)	(2.437)	(2.663)
Visible	0.632	0.590	0.299	0.0701	2.018	0.537
	(0.698)	(0.803)	(0.910)	(0.951)	(2.132)	(2.323)
Tattoos		0.110	-0.0263	0.106	-0.265	0.125
		(1.028)	(1.051)	(1.101)	(2.779)	(2.519)
$Tattoos^2$			0.162	0.112	0.661	1.121
			(0.235)	(0.253)	(0.637)	(0.679)
Order Delivery				0.423	0.753	1.017
				(0.357)	(0.824)	(0.813)
Buy without Thinking				-0.126	-0.425	-0.593
				(0.335)	(0.635)	(0.586)
Smoking or Vaping				0.156	0.495	0.795
				(0.363)	(0.911)	(0.878)
Not Exercising				0.300	0.0656	-0.799
				(0.303)	(0.791)	(0.800)
Post Private Info				0.182	-0.197	0.360
				(0.411)	(0.840)	(0.963)
Post Controversial Info				0.360	-0.274	-0.588
				(0.371)	(0.808)	(0.848)
Copy Test Answers				0.489	0.441	-1.454
				(0.451)	(1.077)	(1.445)
Steal Small Item				-0.378	0.0395	1.270
				(0.397)	(0.811)	(0.904)
Class Year					0.968	0.787
					(0.985)	(0.937)
Control for Econ Majors					-4.807	1.631
					(3.618)	(4.113)
GPA					-3.410	-3.962
					(2.264)	(2.634)
International Students					-6.042	4.356
					(5.629)	(7.264)
Midwest Students					0.852	0.577
					(3.426)	(3.638)
West Students					-5.646	-4.022
					(4.213)	(3.832)
Southwest Students					-0.173	0.297
					(4.983)	(4.871)
Southeast Students					-3.515	-5.320
					(4.113)	(4.229)

Northeast Students					-1.301	-1.837
					(3.979)	(3.921)
Eligible for Work Study					-1.627	0.949
					(2.080)	(2.230)
Female					-0.197	-1.946
					(1.766)	(1.874)
Nonbinary					3.633	7.741
					(4.196)	(4.844)
White					-4.202	-1.516
					(3.246)	(3.079)
Asian					-4.052	-5.140*
					(2.675)	(2.657)
Bisexual					1.896	1.969
					(2.051)	(1.966)
Queer					3.197	8.181*
					(3.205)	(4.101)
Strength of Religious Beliefs					-0.0455	0.578
					(0.782)	(0.857)
Democrat					0.285	-1.571
					(3.317)	(3.749)
Independent					-0.469	-5.858
					(4.153)	(5.047)
Likelihood of Tattoo in Year					0.839	-1.191
					(0.916)	(1.164)
Family with Tattoos					-2.535**	-2.300*
					(1.113)	(1.131)
Friends with Tattoos					-2.846*	-4.666**
					(1.419)	(1.721)
CRT 1 Correct						-11.16**
						(3.958)
CRT 2 Correct						1.421
						(3.088)
CRT 3 Correct						0.210
						(2.426)
All Correct						6.932
						(4.249)
Constant	6.046***	6.032***	6.123***	2.749	35.19**	48.12**
	(0.440)	(0.463)	(0.483)	(1.782)	(15.49)	(20.62)
Observations	85	85	85	85	53	53
R^2	0.059	0.059	0.064	0.165	0.665	0.797
$R^2_{adjusted}$	0.0237	0.0117	0.00512	0.0118	-0.0259	0.187
F	1.681	1.248	1.086	1.077	0.963	1.306
RSS	1137	1137	1130	1009	256.8	155.7

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

Table 12: Time Preference Explained by Tattoo Status Consistent with Ruffle and Wilson (2019), Controlling for Risk Taking, Impulsivity (CRT), and Sociodemographic Characteristics
Dependent Variable is the Switching Pair from the Time Discounting Experiment

VARIABLES	(10)	(11)	(12)	(13)	(14)	(15)
Hidden	-0.211	-0.210	-0.819	-0.850	-4.813*	-5.479**
	(0.206)	(0.735)	(1.051)	(1.089)	(2.365)	(2.469)
Visible	1.048	1.049	0.679	0.455	1.758	0.468
	(0.633)	(0.734)	(0.865)	(0.905)	(2.049)	(2.159)
Tattoos		-0.00238	-0.157	-0.0513	-0.125	0.162
		(1.031)	(1.050)	(1.099)	(2.719)	(2.408)
$Tattoos^2$			0.191	0.145	0.928*	1.176**
			(0.235)	(0.253)	(0.445)	(0.449)
Order Delivery				0.389	0.569	0.988
				(0.357)	(0.750)	(0.740)
Buy without Thinking				-0.146	-0.403	-0.584
				(0.336)	(0.623)	(0.561)
Smoking or Vaping				0.161	0.746	0.846
				(0.364)	(0.794)	(0.720)
Not Exercising				0.317	0.154	-0.787
				(0.304)	(0.763)	(0.764)
Post Private Info				0.269	-0.396	0.313
				(0.407)	(0.757)	(0.836)
Post Controversial Info				0.302	-0.225	-0.567
				(0.370)	(0.789)	(0.797)
Copy Test Answers				0.462	0.293	-1.519
				(0.452)	(1.030)	(1.278)
Steal Small Item				-0.360	-0.144	1.247
				(0.398)	(0.737)	(0.850)
Class Year					1.045	0.801
					(0.958)	(0.896)
Control for Econ Majors					-4.902	1.691
					(3.549)	(3.931)
GPA					-3.520	-4.043
					(2.216)	(2.440)
International Students					-6.279	4.547
					(5.514)	(6.805)
Midwest Students					0.193	0.515
					(3.185)	(3.466)
West Students					-6.176	-4.071
					(4.044)	(3.670)
Southwest Students					-1.111	0.102
					(4.643)	(4.388)
Southeast Students					-4.446	-5.533
					(3.738)	(3.638)
Northeast Students					-1.678	-1.929
					(3.858)	(3.696)
Eligible for Work Study					-1.452	1.019
					(2.022)	(2.064)
Female					0.0879	-1.880
					(1.669)	(1.714)

Nonbinary					3.894	7.897*
					(4.098)	(4.471)
White					-4.833	-1.609
					(3.013)	(2.859)
Asian					-4.103	-5.200*
					(2.626)	(2.510)
Bisexual					1.952	1.969
					(2.012)	(1.896)
Queer					2.663	8.123*
					(3.022)	(3.923)
Strength of Religious Beliefs					-0.223	0.538
					(0.711)	(0.752)
Democrat					-0.724	-1.791
					(2.804)	(3.075)
Independent					-1.696	-6.135
					(3.543)	(4.237)
Likelihood of Tattoo in Year					0.917	-1.208
					(0.890)	(1.113)
Family with Tattoos					-2.597**	-2.289*
					(1.088)	(1.087)
Friends with Tattoos					-2.893*	-4.703**
					(1.391)	(1.629)
CRT 1 Correct						-11.20**
						(3.799)
CRT 2 Correct						1.340
						(2.892)
CRT 3 Correct						0.211
						(2.339)
All Correct						6.879
						(4.070)
Constant	6.019***	6.019***	6.127***	2.817	38.52**	49.23**
	(0.442)	(0.465)	(0.485)	(1.789)	(14.19)	(17.41)
Observations	85	85	85	85	53	53
R^2	0.037	0.037	0.045	0.146	0.658	0.796
$R^2_{adjusted}$	0.0131	0.000955	-0.00327	0.00364	0.0108	0.244
F	1.559	1.027	0.932	1.026	1.017	1.442
RSS	1164	1164	1154	1031	262.2	155.8

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

Table 13: Time Preference Explained by Tattoo Status Consistent with Ruffle and Wilson (2019) by Gender
Dependent Variable is the Switching Pair from the Time Discounting Experiment

VARIABLES	(19) Female	(20) Male	(21) Nonbinary
Hidden	-20.98*	-9.750*	-0.461
	(6.870)	(5.454)	(0.285)
Visible	-17.73		1.081
	(6.360)		(0.635)
Tattoos	Yes	Yes	No
Risk Preference	Yes	Yes	No
Sociodemographic	Yes	No	No
Impulsivity	Yes	Yes	No
Constant	8.065	-1.307	7.280**
	(15.64)	(4.437)	(1.731)
Observations	30	28	7
R^2	0.979	0.768	0.537
$R^2_{adjusted}$	0.692	0.519	0.306
F	3.408	3.079	2.323
RSS	9.041	94.22	47.98

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

<u>References</u>

Adams, Josh. "Marked difference: Tattooing and its association with deviance in the United States." *Deviant Behavior* 30, no. 3 (2009): 266-292.

Brallier, Sara A., Karen A. Maguire, Daniel A. Smith, and Linda J. Palm. "Visible tattoos and employment in the restaurant service industry." *International Journal of Business and Social Science* 2, no. 6 (2011): 72-76.

Broussard, Kristin A., and Helen C. Harton. "Tattoo or taboo? Tattoo stigma and negative attitudes toward tattooed individuals." *The Journal of Social Psychology* 158, no. 5 (2018): 521-540.

Coller, Maribeth, and Melonie B. Williams. "Eliciting individual discount rates." *Experimental Economics* 2, no. 2 (1999): 107-127.

Dillingh, Rik, Peter Kooreman, and Johannes Jan JM Potters. "Tattoos, life style and the labor market." (2016).

Frederick, Shane. "Cognitive reflection and decision making." *Journal of Economic Perspectives* 19, no. 4 (2005): 25-42.

Ruffle, Bradley J., and Anne E. Wilson. "The truth about tattoos." *Economics Letters* 172 (2018): 143-147.

Ruffle, Bradley J., and Anne E. Wilson. "Tat will tell: Tattoos and time preferences." *Journal of Economic Behavior & Organization* 166 (2019): 566-585.

Thomson, Keela S., and Daniel M. Oppenheimer. "Investigating an alternate form of the cognitive reflection test." *Judgment and Decision Making* 11, no. 1 (2016): 99.

Weber, Elke U., Ann-Renee Blais, and Nancy E. Betz. "A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors." *Journal of Behavioral Decision Making* 15, no. 4 (2002): 263-290.