ESSAYS ON LABOR ECONOMICS AND INTERNATIONAL TRADE

by

Danyang Zhang

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Economics

West Lafayette, Indiana

May 2022

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Timothy Bond, Co-Chair
Krannert School of Management

Dr. Chong Xiang, Co-Chair
Krannert School of Management

Dr. Farid Farrokhi
Krannert School of Management

Dr. Miguel Sarzosa
Krannert School of Management

Approved by:
Dr. Yong Bao

To my husband

ACKNOWLEDGMENTS

My most sincere gratitude goes to my committee Tim Bond, Chong Xiang, Farid Farrokhi, and Miguel Sarzosa, for their support, advices and guidance. I am lucky to have all of you in my committee. Thanks to Farid for always being so nice and lowering my workload when I am on the job market. Talking with you has always been inspiring for me. Thanks to Miguel for agreeing to be my fourth committee member right before I enter the job market, and for your comments on the potential difficulty I might face for selling my job market paper – it did happen during some interviews and I am lucky to get prepared beforehand. Special thanks to my advisors, Tim Bond and Chong Xiang. Chong is always sharp, farsighted, and great in literature. Thanks for not giving up on me over the long time I was stuck at finding the idea for my job market paper. Tim is witty, humorous, and supportive. Thanks for making time to meet me frequently, even during your sabbatical, and even when I do not have any real progress to discuss. It takes a lot of patience to not kick me out of your office with all my nagging on things that are neither academic nor important. Thanks for all your advices, both academic and non-academic, and I could not have made it without you. Thanks to Yong Bao, Victoria Prowse, Anson Soderbery, Tiefeng Qian, Yihong Tang, and Kuo Feng, for their help on my job market.

Thanks to Yan Zhang, my master advisor, for being such a great role model and friend to me. Despite that I have graduated for so many years, you have never stopped caring for me. You are way more than an advisor to me, but more of a friend and a sister. I am lucky to be your first master student, and without you I would not have chosen this academic career.

Thanks to everyone in my cohort for their friendship and help. Thanks to Chen Wei, who is always optimistic and willing to offer a helping hand. Thanks to Debasmita Das, who is highly talented in collecting information and generous in sharing them. I have always believed in your potential of being a great detective, and shall I become strong enough to be a hitman and want to join the private detective service, you will always be the first one I talk to for partnership. Special thanks to my MFN (Most Favored Nobody) Aaron Fehl, with whom 90% of the talk are mean jokes. Thanks for all the incentive you provided by constantly talking about your millions of new ideas when I am struggling to find one idea for my job market paper; constantly talking about the ten-digit salary job offer you got early in September when I am stressed about applying for jobs; and constantly talking about your slimness when I am trying to lose my never-going-down weight.

I could not imagine how less stressful my life would be without you. Thank you, Aaron, for bringing laughter to my boring Ph.D. life. Thanks to Mouli Modak, for your infectious laughter and for making me the second scariest person in our cohort for Aaron. Thanks to Xuewen Yu, Town Oh, and Xiaotian Liu, for the comfort and information provided when I am on the job market. Thanks to your help, support, and friendship at Purdue for Jixuan Yao, Jiatong Zhong, Xuan Jiang, Jingyi Yang, Shujin Jiang, Danqing Shen, Anton Yang, Jieqiong Jin, Jianing Ding, Xiaoyang He, and Travis Atkinson.

I would like to thank Alma Cortes Selva and Yipu Deng, my coauthors for the second chapter, for being amazing colleagues and friends. Our regular discussions on Saturday 10am have always been refreshing to me. It has also provided good incentive to pull myself out of bed on Saturday mornings, though I hate myself every time I did so. Thank you, Alma, for your constant comfort and encouragement when I was low in spirit. Thanks for dragging me to the gym. Despite the failing effort of all the time spent in gyms for losing my weight, it does help a lot in reducing stress. It was also a great idea to shift our workout to Krannert and Rawls post Covid19. I will never forget our interesting experience of working out in the office, in Ph.D. student lounge, and in the hallways, neither will the unlucky strangers who saw us do so. It is always easier to do weird things with one more person to share the shame. I will pick up our workout routine upon graduation and fight Diana for being your sister when I become strong enough.

Thanks to my parents, Jie Wu and Dezhu Zhang, for their love and support. My parents are college classmates and both biology teachers. They are my role models both in career and life choices: I want to be an instructor since I was a kid, and I ended up marrying my high school classmate too. I have learned a lot especially from my mother, whose motto is "with less power comes less responsibility". This advice is gold in my marriage, and has helped me in getting away from a lot of responsibilities. Thanks to Tianfang Han for being such a supportive and loving husband. I could not have done this without you – without your fantastic cooking skills in non-vegetable, high-calory food, I could not have gained over 60 pounds. Thanks for doing 80% of the housework while I am on the job market. I would love to add a fourth chapter here to analyze why the rest of the 20% of housework should also be done by you. Unfortunately, my acknowledgement and dissertation have been already too long and I would leave that for future research. For the last twelve years, I have always felt so lucky to meet you early in my life and to have you as my partner. I would hope that we will never need to refer to Table 1.4 for relevant policy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

My dissertation is composed of three independent chapters in the field of labor economics and international trade.

The first chapter studies marriage market signaling and women's occupation choice. Despite the general closure of gender disparities in the labor market over the past half century, occupational segregation has been stubbornly persistent. I develop a new model that explains these occupational outcomes through marriage market signaling. Vertically differentiated men have preference over women's unobservable caregiving ability. Heterogenous women choose caregiving occupations to signal their ability to be caregivers. My model generates unique predictions on the influence of marriage market conditions on women's occupational choices. I find empirical support for these predictions using longitudinal data on marriage rates, policy shocks to divorce laws, and shocks to the marriage market sex ratio driven by waves of immigration.

The second chapter investigates Covid19 and consumer animus towards Chinese products. Covid19 has tremendously affected all areas of our lives and our online shopping behaviors have not been immune. China is the first country to report cases of Covid19 and suffers from rising animus in the U.S. In this paper, we study consumer animus towards Chinese products post Covid19 using Amazon data. We tracked all face masks sold on Amazon between Sep. 2019 to Sep. 2020, and collect product information that is available to a real consumer, including reviews. By analyzing both seller-generated (e.g., product name, description, features) and user-generated (e.g., reviews and customer Q&A) content, we collect information on the country-of-origin as well as consumer animus for the products. Under a fully-dynamic event study design, we find that the average rating drops significantly after a product is identified as made in China for the first time, while no such drop is found for products with other countries-of-origin. This negative impact is U-shaped, which quickly expands in the first five weeks, and then gradually fades out within six months. An informative-animus review affects the average rating of a Chinese product both directly (through its own rating) and indirectly (through other future ratings), with both mechanisms supported in data. We also provide strong evidence that the drop in average rating is driven by consumer animus instead of product quality.

The third chapter explores how cultural transmission through international trade affects gender discrimination. In this paper, I propose that international trade helps alleviate gender

discrimination. With imperfect information on workers' ability, there is statistical discrimination towards female workers. Through international trade, culture transmits asymmetrically between firms located in countries with different gender cultures. This cultural transmission benefits women because it transmits only in one direction from more gender-equal cultures to less gender-equal cultures. I prove this by linking the Customs data to the Industrial Firms data of China in 2004, and find that Chinese firms trading with more gender-equal cultures hire a higher fraction of female workers and enjoy higher profits. Similar patterns are not found in Chinese firms trading with less gender-equal cultures. The impact of cultural transmission goes beyond the firms engaged in international trade to have spillover effects onto purely domestic firms. Comparing across skill groups, cultural transmission benefits high-skill female workers more.

CHAPTER 1. MARRIAGE MARKET SIGNALING AND WOMEN'S OCCUPATION CHOICE

1.1 Introduction

While gender disparities have closed markedly over the past half century, progress on occupational segregation has been stubbornly slow (Blau et al.,2013). In 2009, 62.3% of female workers worked in moderately and heavily female occupations, with 33.2% of female workers working in occupations where at least 80% of their coworkers were female (Blau et al., 2013). [1] In 2020, 87.4% of registered nurses and 98.8% of preschool and kindergarten teachers were women. Meanwhile, this share was only 19.4% and 11.6% for software developers and aerospace engineers, respectively. [2]

In this paper, I propose a new explanation for these occupational disparities revolving around incomplete information in the marriage market. I develop a model in which the most attractive men prefer women with strong caregiving traits. Women, in turn, signal their caregiving ability by choosing a caregiving occupation, despite such occupations entailing lower wages. By doing so, women increase their likelihood of marriage and expected quality of their future mates.

I use this model to generate novel predictions on women's occupational choices, which I then test empirically. First, my model predicts that women working in caregiving occupations have a higher marriage rate on average, for which I find strong support in the National Longitudinal Survey of Youth (NLSY). Next, my model predicts that as the marital surplus decreases, women are less likely to sort into caregiving occupations. I confirm this prediction using the wide-spread changes in unilateral divorce laws in the United States beginning in the late 1960s. Finally, my model predicts an increase in the quality pool of marriageable men will increase the return to marriage and encourage women to signal through caregiving occupations. I confirm this prediction

[1] Blau et al. (2013) apply the Blinder-Oaxaca decomposition using PSID data. Apart from occupation, other explanatory variables include education, experience, region, race, unionization, and industry. In total, all these variables explain 62% of the gender gap.

[2] Table A1 in Appendix lists the heavily female occupations (with at least 80% of workers being female) in 2020. Data from CPS table (annual average) was released by the Bureau of Labor Statistics at https://www.bls.gov/cps/cpsaat11.htm. As stated in the CPS table, occupations with less than 50,000 workers are omitted. For interested readers, 35.3% of Economics students entering Ph.D. programs are female. Among economic faculties, 30.6% of assistant professors, 27.4% of associate professors, and 14.7% of full professors were female in 2020. This data is provided by AEA at https://www.aeaweb.org/content/file?id=13749.

using immigration-induced shocks to the male-female sex ratio to ethnically based marriage markets over the period of 1890-1970.

My model is motivated by several empirical regularities. Despite great progress in labor market disparities, women remain the main childcare providers in the family. In 2018, a full-time working woman spent on average 100 more minutes per day on caring for children than a full-time working man — a disparity that remains roughly unchanged since 2005.[3] Caregiving occupations, such as nursing and teaching, are historically female occupations and remain so today. Figure 1.1 shows a positive correlation between a caregiving index and the share of female workers within occupation in 2020.[4] There is a natural overlap between the duty of these occupations and traditional gender norms concerning women's roles in the household (Akerlof and Kranton, 2000). Caregiving occupations, however, are not well compensated given their educational requirements. Figure 1.2 shows the average years of education and the residual hourly wage based on NLSY79 data, collapsed by decile of the caregiving index.[5] While there is a positive baseline correlation between caregiving and hourly wages, the relationship sharply turns negative once accounting for education.

Despite these lower wages, in my model men find women working in such occupations to be more desirable, because they view it as informative of women's ability at care-intense household tasks. This view is empirically supported. In Figure 1.3, I use the NLSY79 Child and Young Adult (NLSY79-CYA) data to look at children's education and mothers' caregiving index.[6] I found children of mothers in more caregiving occupations attain more education as adults. This result is only partly driven by the correlation between mother's education and the caregiving index, and it is robust to controlling for father's education, father's occupation, and net household income. Finally, in my model, women work in caregiving occupations to improve their prospects in the

[3] Appendix Figure A.1 shows the average time (in minutes) spent on childcare per day by gender 2005-2018. The average time is calculated based on data from American Time-Use Survey. This includes primary childcare and secondary childcare but is limited to time spent on caring for children below 13. The sample is limited to respondents in families where a spouse is present, the respondent must work full-time, and there should be at least one kid below 13 in the household.

[4] To measure the childcaring/caregiving level of an occupation, I create an index from O*NET data, where occupations with a larger index engage in more childcaring activities in daily work.

[5] Education is measured by highest grade completed.

[6] Results collapsed by decile. Children's education is measured by the highest grade the children have completed at the last time of interview and children must be at least 18. Mothers' caregiving index and education are averaged where children are under 13.

marriage market. In Figure 1.4, I show this is, again, empirically well founded; women in more caregiving occupations have higher marriage rates.[7]

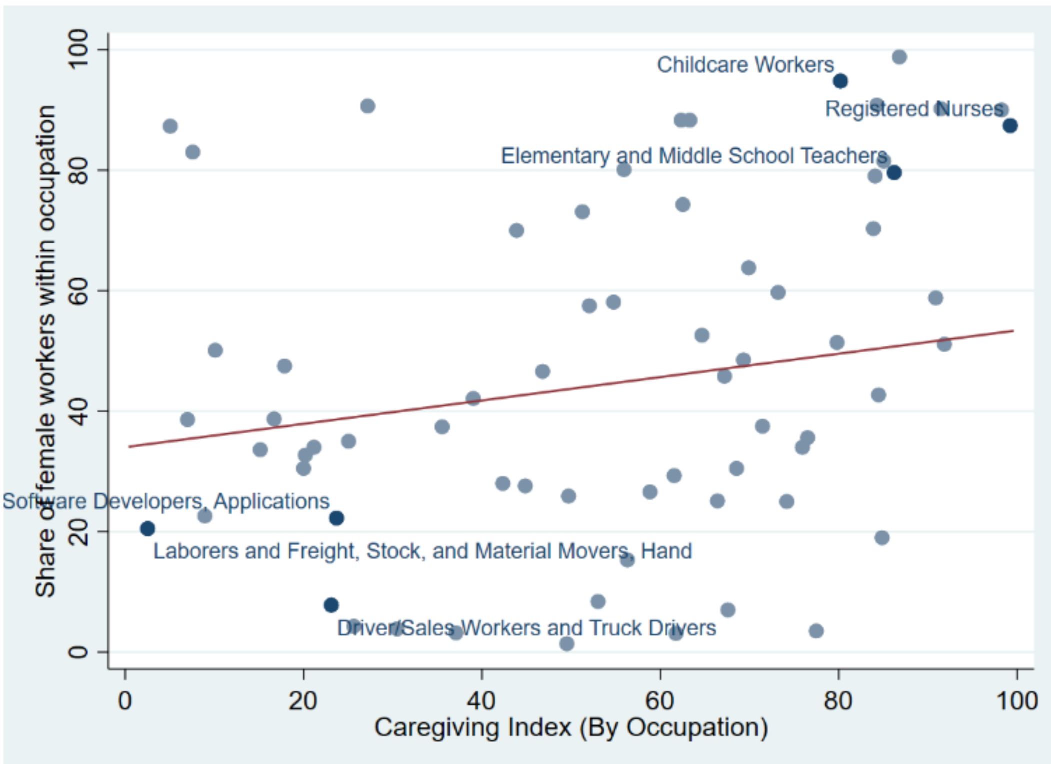

Figure 1.1. Caregiving Index and Share of Female Workers, 2020

*Note: Data from the Bureau of Labor Statistics and O*NET. Only occupations with at least 500,000 workers are kept.*

[7] "Married" is defined here as ever married. This result could be well driven by women switching to caregiving occupation after they get married and would be studied with more caution in empirical parts of this paper.

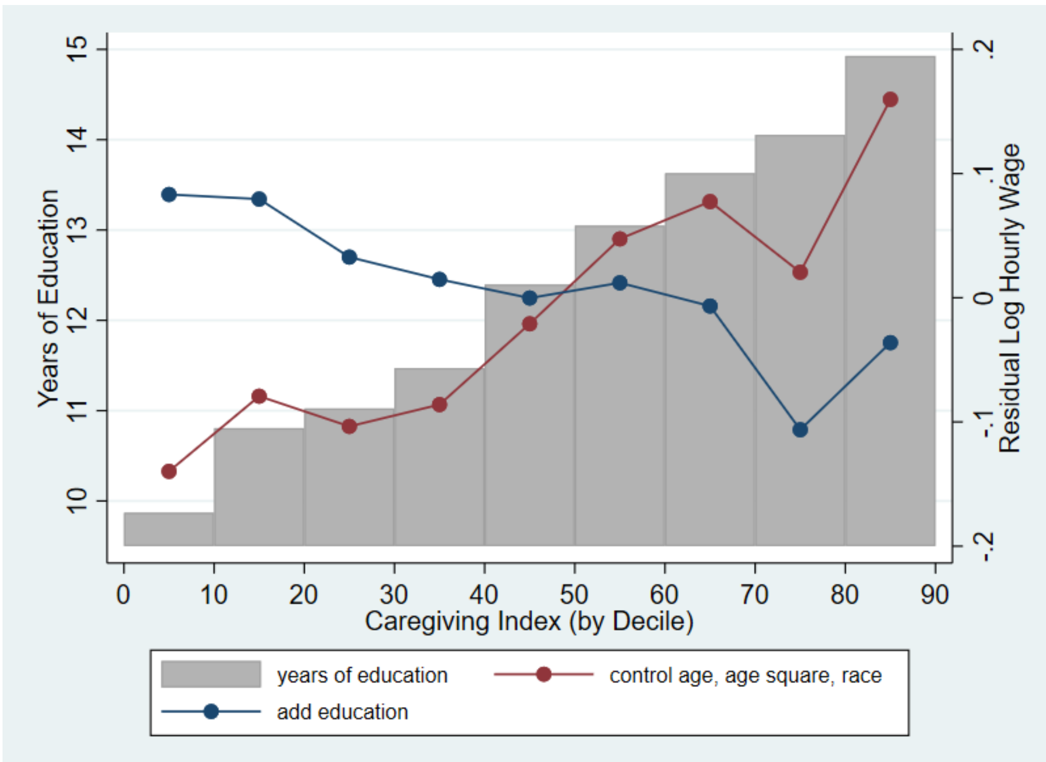

Figure 1.2. Caregiving Index, Education, and Hourly Pay

*Note: Data from NLSY79 and O*NET. On the left y-axis, bars represent years of education averaged within each decile. On the right y-axis, lines represent the residual log of hourly pay averaged within each decile. The red line controls only age, age square, and race. The blue line further controls for years of education and actual working experience.*

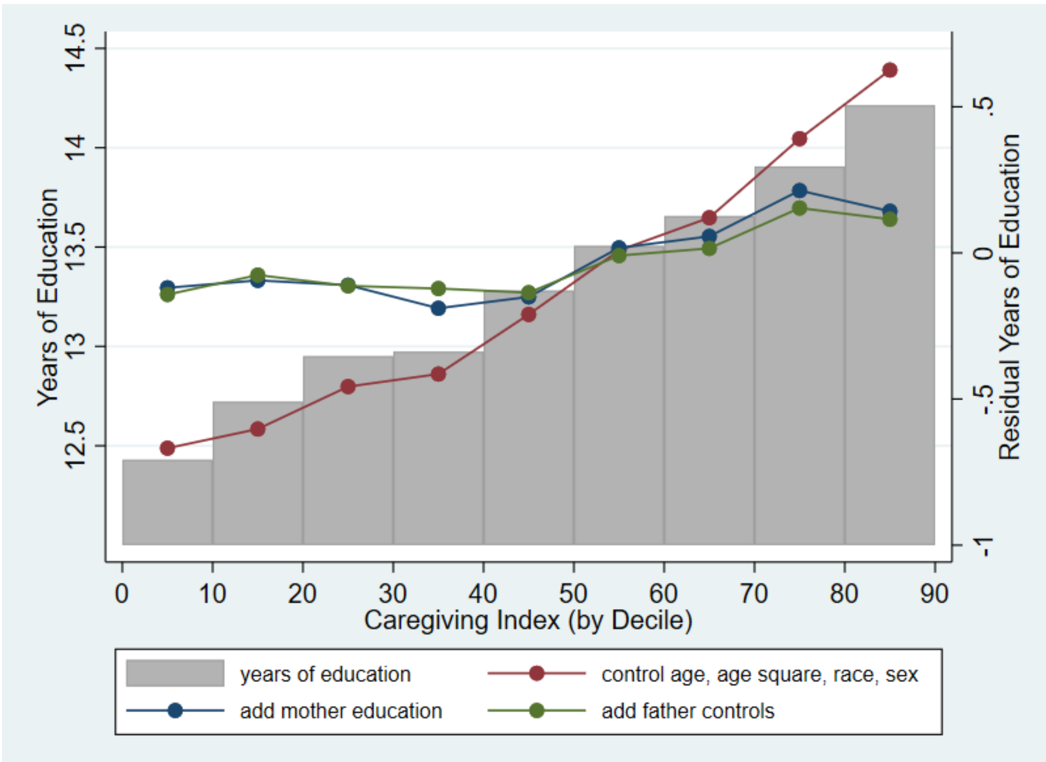

Figure 1.3. Mother Caregiving Index and Children Outcome

*Note: Data from NLSY79 and O*NET. Mothers' caregiving index of occupations are measured while their children are under 13. Children's years of education are measured at the time of last interview for those who are at least 18. On the left y-axis, bars represent children's years of education averaged within each decile. On the right y-axis, lines represent children's residual log of years of education averaged within each decile. The red line controls only age, age square, race and sex. The blue line further controls for years of education for mothers while children are under 13. The green line further adds father's education, father's caregiving index, and the log of household income within the same period.*

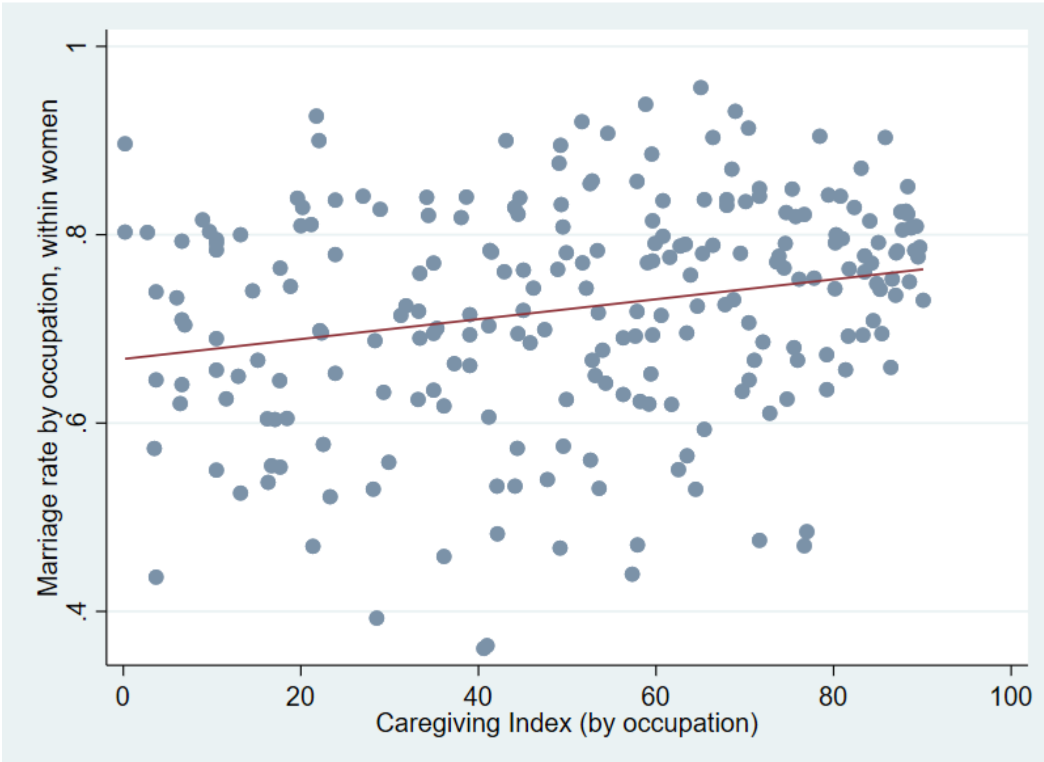

Figure 1.4. Caregiving Index and Marriage Rate for Women

*Note: Data from NLSY79 and O*NET. All occupations with fewer than 20 women are dropped.*

A key feature of my model is that men and women are each vertically differentiated. Specifically, women are vertically differentiated in born preference of caring for children, while men are vertically differentiated in marriageable quality. This follows long literature on the desirability of potential mates on multiple dimensions (e.g. Becker, 1973, 1974; Bergstrom and Bagnoli, 1993; Chiappori et al.,2012; Dupuy and Galichon, 2014; Fisman et al., 2006; Smith, 2006), which is consistent with empirical findings on the narrowness of marriage market (Kirkebøen et al., 2021; Olivetti et al., 2020) and assortative matching on social economic status (Olivetti et al., 2020), income (Schwartz, 2010; Greenwood et al., 2014), education (Schwartz and Mare, 2005; Mare, 1991), credit scores (Dokko, 2015), physical characteristics (Hitsch et al., 2010), and ethnic dimensions (Bisin and Tura, 2019). While I abstract from horizontal differentiation for expository efficiency, my model simply requires that individuals care enough about vertical aspects of partners to take costly action to attract them.

My paper complements existing literature that attempts to explain occupational segregation through preferences, discrimination and social norms, and human capital. Those who point to preferences argue that women desire some job characteristics such as work hour flexibility (Flabbi and Moro, 2012; Wiswall and Zafar, 2018) and job stability (Wiswall and Zafar, 2018), and avoid others such as risk of death (DeLeire and Levy, 2004) and competition (Buser et al., 2014). The literature on discrimination and gender norms stresses the importance of culture, which could either be a demand-side factor, with discriminating employers thinking that men or women are better suited for particular occupations (Kuhn and Shen, 2013); or supply-side factors, with women sorting into particular occupations which might better fit traditional gender roles (Barigozzi et al., 2018). Akerlof and Kranton (2000) discuss the psychological micro-foundation behind such behaviors: when a woman works in a "man's job", women/men incur own/external utility cost in identity. The human capital literature states that occupations might require certain skills which vary by gender such as non-cognitive skills (Cobb-Clark and Tan, 2011) and social skills (Cortes et al, 2018), or that ex ante similar men/woman might make different investments in human capital in anticipation of their future roles within families like caring for children or doing housework (Becker, 1985), or that women with different fertility plans choose different career paths (Adda et al., 2017). Bursztyn et al. (2017) show among MBA students that single women shy away from signaling traits that are undesirable on the marriage market (e.g. ambitious, assertive, pushy); this tendency is not observed among married women, single men, or married men.

The remainder of this paper is organized as follows: Section 1.2 establishes the model and generates my empirical predictions; Section 1.3 discusses my construction of the occupational caregiving index; Section 1.4, 1.5, and 1.6 provide empirical tests of my model; and Section 1.7 concludes and discuss the significance of my findings and future extensions.

1.2 Theoretical Model

Gary Becker argued in "A theory of marriage" that "Children are a major source of the gain from marriage," and thus many current marriage models use children welfare as one important surplus (Chiappori, 2020). In my model, women differ in their preference for caring for children, referred as "caregiving" or "childcaring" preference. The key feature of this model is the imperfect information of women's true preference on the marriage market. Occupation choice is therefore used as signal.

1.2.1 Model Timeline

Within my model, each individual lives for two periods. Before Period One, men and women are born with vertically differentiated types. A woman's true type is private information known only to herself while a man's type is public information on marriage market. In Period One, men and women are young. Women enter the labor market and make occupation choices. Then men and women enter marriage market and form dating matches. Within each dating match, both parties make marriage decisions. In Period Two, men and women are old. Those who did not get married in Period One remain single. For those who did get married in Period One, they face a random divorce shock. The brief timeline of the model is summarized below.

Period Zero: Men and women born with vertically differentiated types.

Period One: Men and women are young.

P1.1 Women enter the labor market and make occupation choices.

P1.2 Men and women enter the marriage market and form stable dating matches.

P1.3 Within dating matches, men and women make marriage decisions.

P1.4 Men and women collect utility from Period One.

Period Two: Men and women are old.

P2.1 Random divorce shock takes place on the marriage market.

P2.2 Men and women collect utility from Period Two, and exit the model.

1.2.2 Model Setup

Period Zero: Men and Women Are Born with Different Types

My model has ρ mass of men (ρ could be larger or smaller than 1) and a unit mass of women. By definition, ρ also denotes sex ratio of men over women.

Women are vertically differentiated in their born preference for caring for children, denoted as their true type A. Share α $(0<\alpha<1)$ of women are born with caregiving type $(A=C)$ and $1\text{-}\alpha$ of women are born with non-caregiving type $(A=N)$. A woman's true type is private information known only to herself and cannot be observed by others. The distribution of all women's true types, however, is public information.

Men are vertically differentiated in their marriage market value, denoted as B. A share β $(0<\beta<1)$ of men have high marriage market value $(B=H)$ and the rest of men have low marriage market value $(B=L)$. There is positive marriage surplus from children's welfare only when a high type man marries a caregiving type woman. Type-*H* men get a payoff U ($U>0$) if marrying a type-*C* woman and get zero payoff if marrying a type-*N* woman. Type-*L* men get zero payoff from marrying either type-*C* or type-*N* women. A man's true type is public information on the marriage market. [8]

Period One: Men and Women Are Young

Before entering into the marriage market, women first enter into the labor market and make occupation choices. My model uses O to denote occupation; women choose between caregiving occupation c and non-caregiving occupation n. The caregiving occupation has high caregiving or childcaring requirements in daily work activities or skills (e.g., nurses or primary school teachers), and it pays workers W_c. The non-caregiving occupation requires little to no caregiving (e.g., equipment operators) and pays W_n. The wage function for women is:

$$W(O) = \begin{cases} W_c & if \;\; O = c \\ W_n & if \;\; O = n \end{cases}$$

There is a systematic wage difference between occupations with $0 < W_c < W_n$, such that the caregiving occupation pays workers lower wages. [9]

Intuitively, a type-*C* woman is more compatible with the caregiving occupation and a type-*N* woman is more compatible with the non-caregiving occupation. Women pay an extra "mismatch" cost of τ *(τ>0)* when they work in an occupation that does not match their born preference. [10] The cost function is:

$$C(A,O) = \begin{cases} 0 & if \;\; (C,c),(N,n) \\ \tau & if \;\; (C,n),(N,c) \end{cases}$$

This mismatch cost could be considered as either a psychological cost of not working in an occupation that matches one's born preference, or a wage punishment of not working in the occupation for which one is best suited. For simplicity, my model assumes the mismatch cost is

[8] This is because a type-*L* man has no incentive to hide his true type since he gets zero payoff from marrying either type of women. A type-*H* man has no incentive to hide his true type either, since he only collects positive marriage surplus when he marries type-*C* women.
[9] This assumption is consistent with data. See Appendix for study of caregiving and wage.
[10] The idea behind this mismatch cost is similar to Akerlof and Kranton (2000), where one earns lower utility when she chooses an activity that does not match her taste.

large enough ($\tau > W_n - W_c$) to negate the marriage market incentives.[11] Under these conditions, all women would match their born preferences with their occupations on the labor market. With marriage market incentives, women choose their occupations to maximize expected lifetime utility, which is composed of three parts: expected gain from marriage, earnings from work, and potential mismatch cost.

Men and women then enter the marriage market and form stable matches of dating. Stable matches are defined such that there will not be a man and woman who both prefer each other to their current match but are not matched. Conditional on the observable variables (e.g., type of men and occupation of women), men and women are each homogenous and there are many potential stable matches. I therefore assume that the actual stable match is randomly assigned. Depending on the relative size of ρ and 1, some men or women would have no dating matches and be single.

Type-*H* men prefer type-*C* women while type-*L* men are indifferent between type-*C* and type-*N* women. However, the true type of a woman is not observed on the marriage market, and a (type-*H*) man forms belief on a woman's true type based on her occupation choice. The occupation choice of a woman therefore serves as a "signal" in the marriage market. Conditional on a woman's occupation choice, the probability that her true type is *C* is $P_C(O,\alpha) = Pr(A = C \mid O,\alpha)$. Considering the mismatch cost, in equilibrium there must be $P_C(c,\alpha) > P_C(n,\alpha) \geq 0$. Men correctly expect these probabilities so that type-*H* men prefer women in caregiving occupations. If a type-*H* man marries a woman working in occupation *O*, his expected surplus from marriage is $P_C(O,\alpha)U$. By Nash bargaining, a type-H man will transfer part of the expected marriage surplus to his wife. In reality, this transfer could be thought of as men benefitting from marrying (highly-likely) caregiving women because the expected welfare for their children is higher and, in turn, the men decide to do more housework, which benefits their wives. For simplicity, assume men transfer half the expected marriage surplus to women. With $P_C(c,\alpha)U/2 > P_C(n,\alpha)U/2 \geq 0$, women in caregiving occupation strictly prefer type-*H* men and women in non-caregiving occupation weakly prefer type-*H* men.

Since type-*H* men strictly prefer women in the caregiving occupation and women in the caregiving occupation strictly prefer type-*H* men, in all potential stable dating matches, when there are more type-*H* men than occupation *C* women, all women in the caregiving occupation are

[11] Relaxing this assumption will not change any predictions from the model but will make the expression more complex.

matched to type-H men; when there are more occupation C women than type-H men, all type-H men are matched to women in the caregiving occupation.

Within all dating matches, both parties make marriage choices ("marry" or "not marry") and a marriage is formed only when both parties choose "marry." For dating matches between type-H men and occupation C women, both parties will choose "marry" due to the positive gain from marriage. For all other dating matches with zero expected gain from marriage, both parties are indifferent between "marry" and "not marry." For simplicity, assume the marriage decision among zero-gain dating matches are randomized, with probability π they would both choose "marry." In reality, this could be understood as both parties taking a random draw (since the stable match is randomly assigned) on how much they like each other (or match quality), with probability π they would both have a positive draw and decide to marry.

Period Two: Men and Women Are Old

In Period Two, single individuals in Period One remain single, including people without dating matches and people with dating matches in which at least one party chooses "not marry." The married individuals from Period One, on the other hand, face a random divorce shock, with probability λ that they will divorce and stop collecting any potential gains from marriage.

1.2.3 Women's Occupation Choice

The key result from the model is the occupation choices of women in P1.1 of the timeline. A woman chooses her occupation to maximize her expected payoff, which includes three parts: wages from work (which varies by occupation), potential mismatch cost (which depends on her born preference and occupation choice), and expected gain from marriage (which depends on her occupation choice, and is also a function of underlying parameters of divorce rate λ, sex ratio ρ, share of women with caregiving preference α, and share of type-H men β). The discount factor of time equals one for simplicity, and women's two-period objective function is:

$$\max_{O\in\{c,n\}} 2W(O) - 2C(A,O) + (2-\lambda)\sum_{B\in\{H,L\}} Pr(B,O,\rho,\alpha,\beta)U_F(B,O,\alpha)$$

As defined above, $W(O)$ is the wage function with $0 < W_c < W_n$, and $C(A,O)$ is the potential mismatch cost, which only takes on positive value τ when occupation choice does not match the born preference. In both periods, women collect the wage and mismatch cost payoffs despite their actual marital status. The third part clarifies the marriage market incentive when women make occupation choices. $Pr(B,O,\rho,\alpha,\beta)$ is the probability that a woman in occupation *O* marries a man of type *B*, and $U_F(B,O,\alpha)$ is her expected gain from that marriage. Based on discussions above:

$$U_F(B,O,\alpha) = \begin{cases} 0 & \text{if } (M,c),(M,n) \\ P_C(O,\alpha)U/2 & \text{if } (T,c),(T,n) \end{cases}$$

Where $P_C(O,\alpha)$, as defined above, is the probability that a woman is type-*C* conditional on her occupation choice *O*. In equilibrium, this belief would be correct because men know the actual distribution of types for all women. The product of $Pr(B,O,\rho,\alpha,\beta)$ and $U_F(B,O)$ is the expected payoff for a woman in occupation *O* marrying a man of type *B*. Summing over *B*, it then gives the one-period expected gain from marriage for women choosing occupation *O*. Further multiplying by $2\text{-}\lambda$ captures the overall expected gain from marriage since with probability λ married couples will divorce in Period Two and collect zero marriage surplus. $Pr(B,O,\rho,\alpha,\beta)$ not only depends on women's occupation choice *O* and men's type *B*, but is also a function of sex ratio ρ, share of women with caregiving preference α, and share of type-*H* men β. The exact value of $Pr(B,O,\rho,\alpha,\beta)$ varies across equilibriums and will be discussed in more detail below. Note that since men's types and women's actual distribution of born types are public information, in equilibria everyone has the correct belief and knows exactly which equilibrium would be realized based on the parameter values.

1.2.4 Equilibrium

The equilibrium for this model is a Perfect Bayesian Equilibrium such that, at each decision point, a man or woman maximizes his or her expected utility given his or her type, (updated) beliefs and other people's strategies; under this equilibrium beliefs are updated following Bayes rule. In equilibrium, men's expectation would be "correct" in the sense that, despite not knowing the true type of a specific woman, a man knows the correct probability that a woman is type-*C* or type-*N* conditional on her occupation choice.

In equilibrium, based on the objective function specified, a type-*C* woman will never work in non-caregiving occupation. With $\tau > W_n - W_c$, the wage difference between the caregiving and non-caregiving occupation is not large enough to compensate for the mismatch cost; with

$P_C(c,\alpha) > P_C(n,\alpha) \geq 0$, there is no marriage market incentive to work in non-caregiving occupation either. Combining these two conditions, a type-*C* woman would never have the incentive to go against her born preference and work in the non-caregiving occupation. For a type-*N* woman, however, she might choose to work in caregiving occupations due to marriage market incentives. [12] Using $\theta \in [0,1]$ to denote the share of type-*N* women who opt into caregiving occupation in equilibrium, I have:

$$P_C(n,\alpha) = \frac{Pr(A=C, O=n)}{Pr(O=n)} = 0$$

$$P_C(c,\alpha) = \frac{Pr(A=C)Pr(O=c \mid A=C)}{\sum_{i=C,N} Pr(A=i)Pr(O=c \mid A=i)} = \frac{\alpha}{\alpha+(1-\alpha)\theta}$$

Therefore, in equilibrium, there is positive gain from marriage for women only when they choose the caregiving occupation and are matched to type-*H* men. $U_F(B,O,\alpha)$ is now:

$$U_F(B,O,\alpha) = \begin{cases} 0 & \text{if } (M,c),(M,n),(T,n) \\ \frac{1}{2}\frac{\alpha U}{\alpha+(1-\alpha)\theta} & \text{if } (T,c) \end{cases}$$

In equilibrium, since type-*N* women are indifferent between choosing caregiving and non-caregiving occupation, this would determine θ. The specific solution of θ would be discussed in separate cases considering the relative size of type-*H* men and women in the caregiving occupation.

Case 1. When There Are "Not Enough" Type-H Men

When $\rho\beta < \alpha + (1-\alpha)\theta$, there are not enough type-*H* men to accommodate all women in the caregiving occupation. As discussed above, both parties will choose "marry" when a woman in the caregiving occupation is matched to a type-*H* man. Therefore, for a woman in the caregiving occupation, the probability of marrying a type-*H* man equals the probability that she is matched with a type-*H* man. Since stable matches are randomly assigned and women know that all type-*C* women and a share θ of type-*N* women work in caregiving occupation in equilibrium, there is

$$Pr(B,O,\rho,\alpha,\beta)\,|_{B=H,O=c} = \frac{\rho\beta}{\alpha+(1-\alpha)\theta}$$

[12] This idea that women send signals on the marriage market to "compete" for good men is consistent with findings in Wilson (1987) and Wilson and Neckerman (1987), where the decreasing availability of marriageable Black men lead to an increased share of children born out-of-wedlock among Black families during the 1970s and 1980s.

The equilibrium condition of equalizing expected payoff for type-*N* women in caregiving and non-caregiving occupations is:

$$2W_c - 2\tau + (2-\lambda)\frac{1}{2}\frac{\alpha U}{\alpha + (1-\alpha)\theta}\frac{\rho\beta}{\alpha + (1-\alpha)\theta} = 2W_n$$

When θ is between 0 and 1, this equation determines θ:

$$\theta = \frac{1}{1-\alpha}\left(\left(\frac{1}{2}\frac{\alpha U\rho\beta(2-\lambda)}{2W_n - 2W_c + 2\tau}\right)^{\frac{1}{2}} - \alpha\right)$$

Using μ to denote the share of women who work in the caregiving occupation in equilibrium such that $\mu = \alpha + (1-\alpha)\theta$, then:

$$\mu = \left(\frac{1}{2}\frac{\alpha U\rho\beta(2-\lambda)}{2W_n - 2W_c + 2\tau}\right)^{\frac{1}{2}}$$

The equilibrium above requires *U* to fall within in specific ranges. When *U* is in a proper range (see Appendix for specific range expression) there is a partial-pooling equilibrium where all type-*C* women and a positive share θ of type-*N* women work in the caregiving occupation. The remaining share $1-\theta$ of type-*N* women work in non-caregiving occupation.

When *U* is too large, the model enters a complete pooling equilibrium. In this case, the marriage market incentive is too strong such that all women work in the caregiving occupation despite their born preference. When *U* is too small, the model enters a separating equilibrium. In this case, the marriage market incentive is too weak to persuade type-*N* women to work in the caregiving occupation to send signals. All type-*C* women work in the caregiving occupation and all type-*N* women work in the non-caregiving occupation. In this paper, I will assume that *U* is within the proper range so that $0<\theta<1$ in equilibrium.

Case 2. When There Are "Enough" Type-H Men

When $\rho\beta \le \alpha + (1-\alpha)\theta$, there are enough type-*H* men to accommodate all women in the caregiving occupation. Since both parties will choose "marry" when a woman in the caregiving occupation is matched to a type-*H* man, the probability that a woman in the caregiving occupation marries a type-*H* man is one.

$$Pr(B, O, \rho, \alpha, \beta)|_{B=H,O=c} = 1$$

The equilibrium condition of equalizing expected payoff for type-N women in caregiving and non-caregiving occupations reduces to:

$$2W_c - 2\tau + (2-\lambda)\frac{1}{2}\frac{\alpha U}{\alpha + (1-\alpha)\theta} = 2W_n$$

And θ simplifies to:

$$\theta = \frac{1}{1-\alpha}\left(\frac{1}{2}\frac{\alpha U(2-\lambda)}{2W_n - 2W_c + 2\tau} - \alpha\right)$$

Use μ to denote the share of women who eventually work in caregiving occupation:

$$\mu = \frac{1}{2}\frac{\alpha U(2-\lambda)}{2W_n - 2W_c + 2\tau}$$

Similar to Case 1, the type of equilibrium depends on range of U (see Appendix). When U is in a proper range, there is a partial pooling equilibrium where all type-C and some type-N women work in the caregiving occupation, while some type-N women work in non-caregiving occupation. When U is too large, all women work in the caregiving occupation in a complete pooling equilibrium. When U is too small, all type-C women work in the caregiving occupation and all type-N women work in the non-caregiving occupation to form a separating equilibrium.

The equilibriums in Case 2 are very similar to those of Case 1, but θ no longer responds to a change in sex ratio ρ and share of type-H men β. This is because now there are "enough" type-H men such that further increase in the population of high type men no longer strengthens the marriage market incentive for women. It could be understood as follows: θ rises with increasing $\rho\beta$ until the point that $\rho\beta = \alpha + (1-\alpha)\theta$ and then θ becomes flat for further increase in $\rho\beta$.

1.2.5 Model Predictions

All predictions below are under the assumption that U is within the "proper range" such that there is a partial-pooling equilibrium.

Prediction 1: Higher Marriage Rate for Women in Caregiving Occupation

Matches between type-H men and women in the caregiving occupation have probability of marriage of one, while in all other dating matches this probability is π. On average, women in the caregiving occupation have higher marriage rate. Intuitively, the preference of type-H men drives this result since type-L men have no preference over women's true types. Because a woman's true type is not observed but occupation conveys valid information on type, for type-H men, preference for type-C women translates into preference for women in the caregiving occupation, which drives the resulting higher marriage rate among women in the caregiving occupation.

Proof: The exact function of the marriage rate varies with underlying parameters. Specifically, to pin down marriage rate, separate discussions should be made in four different cases considering the relative size of $\rho\beta$ and $\alpha+(1-\alpha)\theta$, and relative size of ρ and 1. In all cases, the average marriage rate for women in occupation O, denoted with $P_M(O)$, is:

$$P_M(O) = \sum_{B=H,L} Pr(B, O, \rho, \alpha, \beta)$$

The calculation process is in Appendix. The difference in marriage rate between women in the caregiving and non-caregiving occupations are listed in each case.

Sub-Case 1: $\rho\beta < \alpha+(1-\alpha)\theta \text{ and } \rho < 1$

$$P_M(c) - P_M(n) = \frac{\rho\beta}{\alpha+(1-\alpha)\theta}\left(1-\frac{\rho-\rho\beta}{1-\rho\beta}\pi\right) > 0$$

Sub-Case 2: $\rho\beta < \alpha+(1-\alpha)\theta \text{ and } \rho \geq 1$

$$P_M(c) - P_M(n) = \frac{\rho\beta}{\alpha+(1-\alpha)\theta}(1-\pi) > 0$$

Sub-Case 3: $\rho\beta \geq \alpha+(1-\alpha)\theta \text{ and } \rho < 1$

$$P_M(c) - P_M(n) = 1-\frac{\rho(1-\beta)}{(1-\alpha)(1-\theta)}\pi > 0$$

Sub-Case 4: $\rho\beta \geq \alpha+(1-\alpha)\theta \text{ and } \rho \geq 1$

$$P_M(c) - P_M(n) = 1-\pi > 0$$

In all cases, women in the caregiving occupation have a higher marriage rate than women in the non-caregiving occupation.

Prediction 2: Higher Divorce Rate, Fewer Women Work in Caregiving Occupation

When the exogenous divorce rate is higher, overall fewer women work in the caregiving occupation. Intuitively, when the probability of divorce in Period Two increases, all else being equal, the expected gain from marriage decreases. Therefore, the incentive for type-N women to signal through choosing the caregiving occupation is also lower. In equilibrium, this leads to a lower share of θ within type-N woman. The occupational choices of type-C women, however, are not affected. Other parameters being constant, the overall share of women working in the caregiving occupation, μ, will be lower.

Proof: In Case 1, there is:

$$\frac{\partial \mu}{\partial \lambda} = -\frac{1}{2}\left(\frac{1}{2}\frac{\alpha U \rho \beta}{2W_n - 2W_c + 2\tau}\right)^{\frac{1}{2}} (2-\lambda)^{-\frac{1}{2}} < 0$$

In Case 2, there is:

$$\frac{\partial \mu}{\partial \lambda} = -\frac{1}{2}\frac{\alpha U}{2W_n - 2W_c + 2\tau} < 0$$

In both cases, the share of women working in the caregiving occupation decreases unambiguously.

Prediction 3: Higher Sex Ratio, More Women Work in Caregiving Occupation

Intuitively, when sex ratio is higher so there are more men compared to women, fixing β, the overall number of type-*H* men increases, thus creating an even larger advantage on marriage market for women to work in the caregiving occupation. As discussed above, type-*L* men are indifferent between caregiving and non-caregiving women and the signaling result is driven by the preference for type-*C* women within type-*H* men. A larger number of type-*H* men translates into a higher probability that a woman in the caregiving occupation is matched and married to a type-*H* man. This increases the expected gain from marriage and strengthens the incentive for type-*N* women to opt into the caregiving occupation. Since type-*C* women always work in the caregiving occupation, a higher sex ratio increases the overall share of women working in the caregiving occupation, μ, assuming all other parameters remain constant.

Proof: In Case 1, there is:

$$\frac{\partial \mu}{\partial \rho} = \frac{1}{2}\left(\frac{1}{2}\frac{\alpha U \beta (2-\lambda)}{2W_n - 2W_c + 2\tau}\right)^{\frac{1}{2}} \rho^{-\frac{1}{2}} > 0$$

In Case 2, there is:

$$\frac{\partial \mu}{\partial \rho} = 0$$

Therefore, when the sex ratio is higher, the model predicts a non-decreasing share of women would work in the caregiving occupation. In reality, when the sex ratio is higher, I would expect the data to show a strictly increasing share of women working in the caregiving occupation. This is because marriage markets are narrowly defined in reality. As shown in Kirkebøen et al. (2021), marriage markets are highly localized and can be as small as a college. Since it is hard to precisely pin down a narrowly-defined highly-localized marriage market in data, the overall pattern is averaged between Case 1 and Case 2, and therefore will show an unambiguously increasing pattern in response to a higher sex ratio.

1.3 Data

1.3.1 Data Introduction

This paper combines occupational data with microdata. I use occupational data from O*NET to construct the caregiving/childcaring measurements.[13] These measurements are then linked to microdata based on occupation codes.[14] I use different microdata, which I introduce in later sections, across empirical tests, including National Longitudinal Survey of Youth (NLSY), Current Population Survey, and Decennial Censuses.

Occupational Information Network (O*NET) provides comparable data on occupations and has updated yearly since 2003. It describes features of occupations and workers from six broad dimensions that are either worker-oriented (worker characteristics, worker requirements, and experience requirements) or job-oriented (occupational requirements, workforce characteristics, and occupation-specific information). In this paper, I mainly use the work activity file and the knowledge file of O*NET to construct a caregiving index.[15] The work activity file provides information on typical activities required in each occupation. To make such information comparable across occupations, I use the generalized work activities, which are performed in most occupations.[16] The knowledge file specifies required workers' attributes that are related to work performance, such as knowledge and skill.

[13] The paper is especially concerned with the aspects of caregiving that relate to caring for children. Therefore, when I refer to "caregiving" throughout my model and results, I tie it implicitly to "childcaring."
[14] The systems used to classify occupation vary across the microdata used in this paper, and are linked to the caregiving measurements via the crosswalk of different occupation classification systems.
[15] Ability file, skills file, and work context file are also used to construct other occupation-related skill measures following literature.
[16] In O*NET, the work activities are summarized at three levels: general, intermediate, and detailed.

O*NET data in 2003 is classified using the O*NET-SOC 2000 occupation coding system, which can be linked to 2010 Standard Occupational Classification (SOC) system.[17] It then can be matched with Census Occupation Code (COC) of different years using crosswalks provided by Census Bureau.[18] The O*NET data, in turn, can be linked to microdata using different versions of COC. Details on linking O*NET data and microdata are discussed in each empirical section.

1.3.2 Constructing Caregiving Index

Main Measurement

The caregiving index is constructed from the generalized work activity file, which covers 41 broad categories of activities typically performed in most occupations (see Appendix Table A.2 for the full list of categories). Among these categories, I have chosen the two work activities that are most relevant to caring for children: "Assisting and Caring for Others" and "Training and Teaching Others."[19] For these two work activities, an occupation is measured on both importance scale (on how important a work activity is for an occupation) and level scale (on how intensively a work activity is carried out for an occupation).[20] In this paper, I use the importance scale as my primary measure to construct the caregiving index. The importance scale ranges from 1 (not important) to 5 (extremely important).[21] Following previous literature (e.g., Deming 2017) on constructing indexes from O*NET data, I followed the three steps below to construct the caregiving index:

Step 1: within each dimension, re-scale the rating to 0-10

Step 2: for each occupation, create raw index by averaging over the two chosen dimensions

Step 3: rank occupations by raw index and record each occupation's percentile in the ranking

[17] Specifically, I first match O*NET-SOC 2000 to O*NET-SOC 2010, then match O*NET-SOC 2010 to SOC 2010. The crosswalks are provided by O*NET at https://www.onetcenter.org/taxonomy.html. O*NET-SOC system also has several versions for 2000-2006, 2006 to 2009, and 2009-2010.

[18] The crosswalks between 2010 SOC, 2010 COC, and 2000 COC classification systems are provided by Census Bureau at https://www.census.gov/topics/employment/industry-occupation/guidance/code-lists.html.

[19] As descried by O*NET, "Assisting and Caring for Others" includes "providing personal assistance, medical attention, emotional support, or other personal care to others"; "Training and Teaching Others" includes "identifying the educational needs of others, developing formal educational or training programs or classes, and teaching or instructing others".

[20] Note that I never average over importance and level scales: it is either created based solely on importance scale or solely on level scale.

[21] Specifically, O*NET defines the scale ranges as: 1 "not important", 2 "somewhat important", 3 "important", 4 "very important", 5 "extremely important". In O*NET, the actual data concerning each dimension is continuous instead of just integers between 1 and 5.

This percentile (0-100) is the caregiving index, with 0 being the least caregiving and 100 being the most caregiving among all occupations. Since many samples in this paper cover years as early as 1890, for measurement consistency, I use 2003 as the base year to construct the caregiving index, which is the first year O*NET data is consistently available. Missing values in 2003 are imputed with values from later years.

Table 1.1. Top and Bottom Caregiving Occupations

Top 10 caregiving occupations	Bottom 10 caregiving occupations
Preschool and Kindergarten Teachers	Avionics Technicians
Other Teachers and Instructors	Upholsterers
Occupational Therapists	Tire Builders
Registered Nurses	Furniture Finishers
Licensed Practical and Licensed Vocational Nurses	Rail-Track Laying and Maintenance Equipment Operators
Elementary and Middle School Teachers	Graders and Sorters, Agricultural Products
Personal Care Aides	Electrical Power-Line Installers and Repairers
Special Education Teachers	Food and Tobacco Roasting, Baking, and Drying Machine Operators and Tenders
Dental Assistants	Pile-Driver Operators
Respiratory Therapists	Tool and Die Makers

*Note: Data from O*NET, based on year 2003.*

The ten occupations that scored the highest on the caregiving index and the ten that scored the lowest are listed in Table 1.1. Generally speaking, the most caregiving occupations are teachers, therapists, and nurses; these results are a good match of the general understanding of typical caregiving occupations. The bottom 10 caregiving occupations are typical male-dominated manufacturing occupations.

Figure 1.5 shows the distribution of the caregiving index by gender using National Longitudinal Survey of Youth 1979 data. It is obvious that women on average work in more caregiving occupations than men, but there are workers of both genders over the whole span of caregiving index range and there is not big difference in variance of distribution. Figure 1.6 shows the distribution of the caregiving index within women by marital status. Single and married women have similar distributions, with single women having slightly thicker tails at the high caregiving end. The similarity supports using married women as the comparison group for single women in face of potential shock.

The caregiving index is continuous and varies by occupation. However, to link over time, occupations are combined and divided, which might affect the precision of the continuous measurement. Moreover, in reality, instead of knowing the exact caregiving percentile of all occupations, people on the marriage market might only know the broad range an occupation fall into. Therefore, I also construct a binary measure where occupations in percentiles 50-100 are given a value of one, and the rest of occupations take on a value of zero. This dummy also matches better with the model where occupations are classified into caregiving and non-caregiving categories. In some of the specifications, I also divide occupations into quartiles, with the most caregiving occupations in fourth quartile (75-100 percentile) and the least caregiving occupations in first quartile (0-25 percentile).

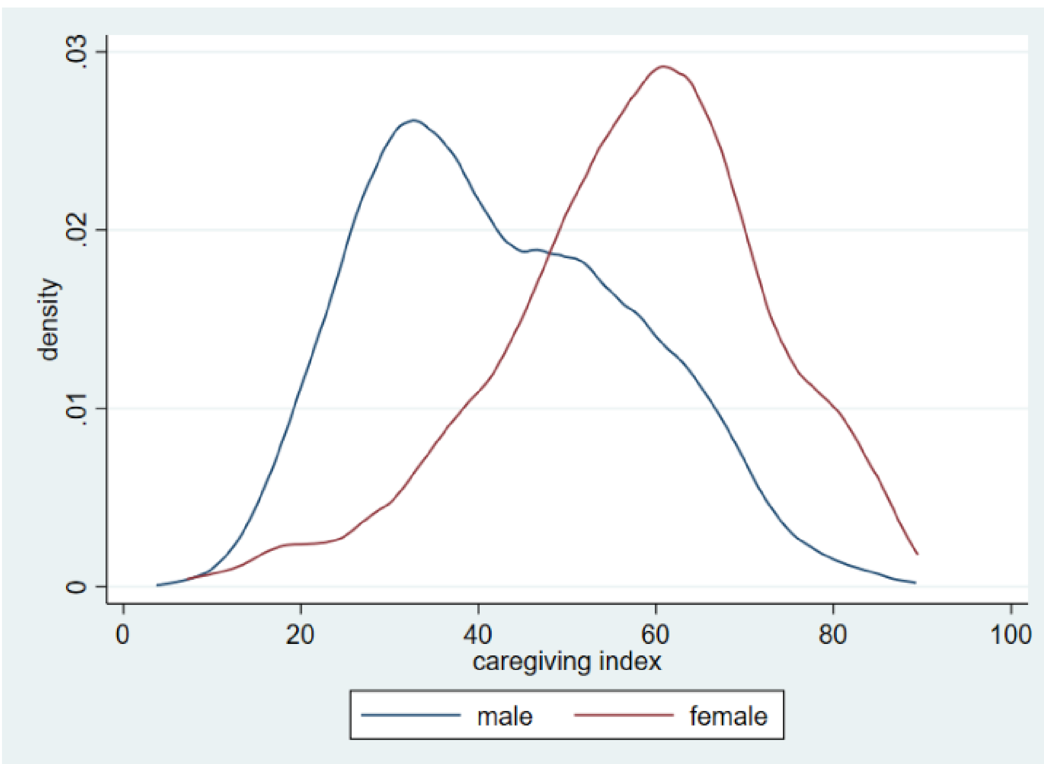

Figure 1.5. Caregiving Index Density by Gender

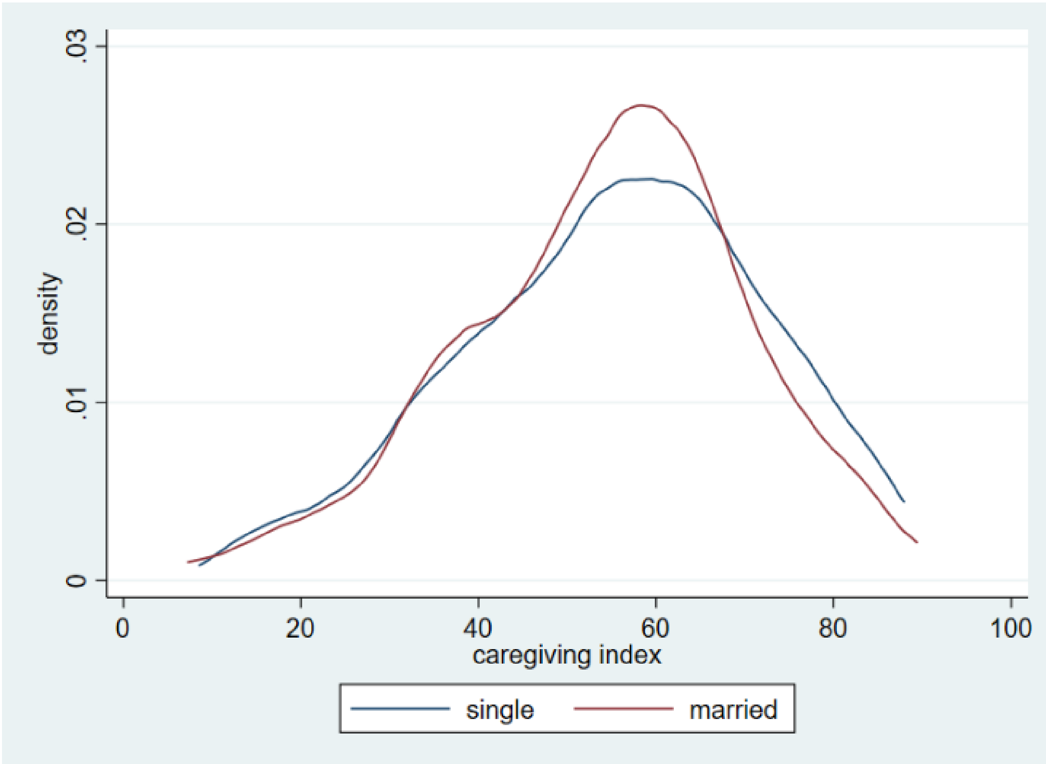

Figure 1.6. Caregiving Index Density within Women

Alternative Measurements

One way to achieve a direct alternative measurement would be to construct the caregiving index based on a level scale. The level scale used in O*NET varies by work activity category, but the level scale in the two categories used in the index (Assisting and Caring for Others, and Training and Teaching Others) both follow a 2-4-6 design.[22] Construction of the alternative caregiving index follows the same three steps as the main index.

I also construct alternative indices with information from the knowledge file. There are 33 categories in the knowledge file (see Appendix Table A.3 for the full list), and I pick the three categories that are most relevant to childcaring: Education and Training, Therapy and Counseling, and Psychology. The construction of this index is same as above to deliver two indices using importance or level scale.

[22] The anchor descriptions for each scale value are listed here. For Assisting and Caring for Others: 2 "Help a coworker complete an assignment," 4 "Assist a stranded traveler in finding lodging," and 6 "Care for seriously injured persons in an emergency room." For Teaching and Training Others: 2 "Give coworkers brief instructions on a simple procedural change," 4 "Teach a social sciences course to high school students," and 6 "Develop and conduct training programs for a medical school."

The three alternative indices are highly correlated with the main caregiving index; the correlation coefficients are 0.96 for the work activity level scale index, 0.84 for the knowledge file importance scale index, and 0.78 for the knowledge file level scale index.[23]

Other Occupational Indices

I follow Deming (2017) to define routine task, non-routine task, and social skill intensity. Deming's (2017) method originates from Autor et. al. (2003) which uses DOT data and Deming (2017) tailors it to the O*NET data. This method uses both level scale and context measure, and data come from several files covering work context, ability, skills, and knowledge.

I also construct five indices measuring other occupational features. To measure flexibility, I have created the work hours index from the "Duration of Typical Work Week," data in the work context file. The time pressure index is constructed based on the "Time Pressure" data from the work context file and is measured by the frequency the job requires the worker to meet strict deadlines. I construct the job stability index from the "Security" data in the work value file, which measures the extent that workers on this job have steady employment. I build the competition index from the "Level of Competition" listed in the work context file, which measures the extent that a job requires the worker to compete or be aware of competitive pressures. I average the six dimensions measured under the "Job Hazard" in the work context file, which describes how often workers are exposed to several hazardous conditions as part of the job, to create the job hazard index. Specifically, these six dimensions include: exposed to radiation; exposed to disease or infection; exposed to high places; exposed to hazardous conditions; exposed to hazardous equipment; and exposed to minor burns, cuts, bites, or stings.

1.4 Caregiving and Marriage Rate

In this section, I test Prediction One, which states that the marriage rate is higher for women working in caregiving occupations. I use the National Longitudinal of Youth's detailed long-term information on the respondents. Empirically, I first study the marriage rate across occupations to directly test Prediction One. Then I study respondents' age at first marriage, which uses a cleaner sample and measures variables at a more relevant time point.

[23] The occupations are measured using the 2002 census occupation classification system here when calculating the correlation coefficients. To save space, results using alternative indices are suppressed from this paper.

1.4.1 Data

National Longitudinal Survey of Youth

The National Longitudinal Survey of Youth 1979 (NLSY79) has surveyed the same respondents on a regular basis since 1979.[24] The respondents were ages 14-21 at the time of the first interview (1979), and there were initially 12,686 respondents covering three subsamples: the main sample, the military sample (dropped after 1984[25]), and the non-Hispanic, non-black, economically disadvantaged sample (dropped since 1990). NLSY covers detailed information about the respondents, including some typical demographic information such as race, gender, age, education, income, marital status, and state of residence.[26] For marital status, NLSY categorizes respondents into married, separated, divorced, widowed, and never married, and it documents respondents' age of first marriage. It also asks respondents some unique questions on their attitudes toward marriage/children, e.g., what is your expected age of marriage, what is the total number of children you expect to have. Detailed work-related information is also provided in NLSY, including hourly wage, actual working experience, occupation, industry, employment status, usual hours worked per week, etc. NLSY also follows children born to NLSY79 mothers (see Appendix for details).

Linking NLSY79 and O*NET Data

NLSY79 provides occupational data on up to five jobs held by respondents since their last interview. In this paper, I focus on the current/most recent job (or CPS job as called in NLSY data). NLSY79 records each respondent's occupation based on census code, which varies by year. To match the occupation system provided in NLSY, I use 3-digit 1970 Census Occupation Code (COC) for data collected 1979-1981, 3-digit 1980 COC for 1982-2000, and 4-digit 2000 COC since 2002. I use the "Dorn Code" (Autor and Dorn, 2013) and IPUMS crosswalks to link data over time.[27] Based on the crosswalks and some small manual revisions, occupations are linked from 1970/1980 to 2000 COC, with combinations and divisions of occupations over time.[28] The

[24] In 1979-1993, respondents were surveyed annually. Since 1994, respondents have been surveyed every two years.
[25] Instead, 201 respondents were randomly drawn to participate in the NLSY military sample after 1984.
[26] Note that state of residence is limited data from NLSY and requires an application.
[27] Thanks to Autor and Dorn (2013) for sharing data at https://www.ddorn.net/data.htm. I use it for crosswalks between 1980-1990 and 1990-2000 COC. The IPUMS crosswalk is provided at https://usa.ipums.org/usa/volii/occ_ind.shtml. I use it for crosswalks between 1970-1980 COC.
[28] I mostly follow Dorn code and make very few changes in my matching. The few changes I do make are to re-separate occupations that have been combined in the Dorn code but have significantly different caregiving levels.

final sample contains 363 unique occupations. For those who are married, NLSY also records spouse's occupation information based on 1970 COC in the years 1979-2000 and 2000 COC since 2002. Industry data in NLSY79 is treated in a similar way, where industrial information is matched across years and consistently measured using 2000 Census Industry Code.

1.4.2 Marriage Rate and Women's Occupation Choices

My model states that men use occupation to collect information on women's childcaring preferences and predicts a higher marriage rate among women working in caregiving occupations. To test this prediction, I run the following regression:

$$Y_{ot} = \alpha + \beta C_o + \beta_X X_{ot} + \delta_t + \varepsilon_{ot} \tag{1.1}$$

Variables are averaged by occupation o and year t. The dependent variable is marriage rate, where respondents are classified as "married" if they have ever been married.[29] The marriage rate thus refers to share of respondents that has ever been married within an occupation-year cell. C is the caregiving measurement, which could be the continuous index, dummy, or quartiles. X denotes the control variables, including AFQT (measuring cognitive skills), age, age squared, education (years completed), share of full-time jobs, share of races (Hispanic or Black), share in each Body Mass Index groups (underweight, normal, overweight, obese), hourly pay, and yearly overall income.[30] Standard errors are clustered at occupation.

Table 1.2 shows the result. As the model predicts, women who work in more caregiving occupations have higher marriage rates as in column 1. Column 2 shows that, on average, women working in caregiving occupations are 2.3% more likely to ever be married, all else being equal. Further breaking down the results into quartiles makes clear that the higher marriage rate is mainly driven by occupations with the highest scores on the caregiving index. Compared to women in the least caregiving occupations in the first quartile, women in the most caregiving occupations in the fourth quartile have a statistically significant 4% higher marriage rate. Women in the second quartile have a non-significant higher marriage rate of 1.3% and women in third quartile have very

[29] Divorced, widowed, and separated respondents are classified as "married" here. This result is robust to measuring marriage rate using the current marital status, where only currently married people are classified as "married".

[30] Ranges of BMI groups: BMI<18.5, underweight; 18.5<=BMI<25, normal; 25<=BMI<30, overweight; BMI>30, obese. Results are very similar with further inclusion of hours usually worked per week, number of children, and presence of any children under 13 in the household, and when only controlling for age, age square, race shares, BMI group, and education.

similar marriage rates compared to least caregiving occupations. These results also provide evidence that the most caregiving occupations convey the strongest signals on women's caregiving preference. In reality, people in the marriage market might not know the exact ranking, but they recognize some occupations – such as registered nurses and K-12 school teachers – as caregiving-heavy, and these are also occupations that women heavily sort into.

Table 1.2. Marriage Rate and Caregiving Level across Occupations

	Women			Men		
	Index	Dummy	Quartile	Index	Dummy	Quartile
Index	0.000578***			-3.67e-05		
	(0.000219)			(0.000154)		
Dummy		0.0228**			0.0137	
		(0.0115)			(0.0145)	
Quartile 4			0.0400**			-0.00882
			(0.0162)			(0.0236)
Quartile 3			0.0132			-0.00837
			(0.0161)			(0.0185)
Quartile 2			0.000155			-0.0382**
			(0.0173)			(0.0160)
Year	Y	Y	Y	Y	Y	Y
Observations	5,383	5,383	5,383	6,814	6,814	6,814
R-squared	0.357	0.357	0.358	0.402	0.403	0.405

*Note: Standard errors (clustered at occupation) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In columns 3 and 6, the least caregiving occupations (the first quartile) is used as the reference group.*

In my model, men do not signal their childcaring preference through occupation like women. Results in column 3 to 6 of Table 1.2 support this asymmetry. The same regression reveals no evidence of a higher marriage rate for men working in caregiving occupations. In fact, men working in the second quartile have a marriage rate that is 3.8% lower than the least caregiving occupations.

The results in Table 1.2 support Prediction One from the model. However, these results could well be driven by women switching to caregiving occupations after marriage. It could also be explained by a positive correlation between caregiving and marriage preferences for women such that women with higher caregiving preferences choose to work in more caregiving occupations. To show that these results are not purely driven by sorting on preference, I compare age of marriage across occupation choices in a cleaner way below.

1.4.3 Age of First Marriage and Women's Occupation Choices

To provide stronger evidence, I analyze women's age of first marriage. Much like the straightforward test of Prediction One in Section 4.2, I now further restrict the sample to single women and use their first job to measure the occupation caregiving indices. This sample avoids the problem of married women switching to more caregiving occupations. NLSY also asked respondents about their marriage age preference, and this information helps control the potential correlation between preference in caregiving and marriage. I use the following specification:

$$Y_i = \alpha + \beta C_i + \beta_E E_i + \beta_X X_i + \delta_t + \varepsilon_i \tag{1.2}$$

The outcome variable is age of first marriage for each respondent i.[31] C is the caregiving measurement based on women's first jobs. Control variables X are averaged across years when women are single[32] and include race, AFQT, years of education, working experience, whether working full-time, yearly personal income, hourly pay, and BMI group.[33] Year fixed effect is included based on the year each woman had her first job. E denotes the expected age of marriage for each respondent. In the first round of the survey (1979), NLSY asked unmarried respondents at what age they expected to marry.[34] This variable helps me to separate women's signaling behavior from their born preference.

The results are listed in Table 1.3.[35] Column 1 shows that single women who work in more caregiving occupations as their first job marry at significantly younger ages in future, even after controlling for their preference on age of first marriage. All else being equal, women working in caregiving occupations on average marry 0.56 years earlier than those in non-caregiving occupations, though the difference is not significant at conventional levels. The quartiles reveal that women working in the most caregiving occupations marry an average of 2 years younger than their counterparts in least caregiving occupations. Women working in the second and third

---

[31] In my final sample, the actual oldest age of first marriage is 58. Therefore, for those who have never been married by the time of last interview, I impute their "age of first marriage" as 60.

[32] Here panel data is not proper because I run the regression while women are single. For women who marry at an older age, they have more years of being single. Therefore, if using panel data, the sample would be biased towards women marrying at an older age and then bias the result.

[33] Results are similar with inclusion of hours usually worked per week, number of children, and presence of any children under 13 in the household. Results are similar if only controlling for race, BMI group, and education.

[34] The question asked respondents "at what age would you like to marry?" There are five answers to choose from for this question, namely: less than 20, age 20-24, age 25-29, age 30 or older, and never.

[35] For expected age of first marriage controls, the omitted group is the group who expect to marry under the age of 20. The results on these controls are oppressed to save space.

quartiles also marry earlier than women working in the least caregiving occupations, but the differences are not significant.

Table 1.3. Caregiving and Age of First Marriage (Expected Age of First Marriage Controlled)

	Women			Men		
	Index	Dummy	Quartile	Index	Dummy	Quartile
Index	-0.0288**			0.0139		
	(0.0124)			(0.0125)		
Dummy		-0.563			0.447	
		(0.435)			(0.488)	
Quartile 4			-2.059**			0.715
			(0.996)			(1.506)
Quartile 3			-0.814			0.273
			(0.816)			(0.639)
Quartile 2			-0.555			-0.181
			(0.840)			(0.549)
Year	Y	Y	Y	Y	Y	Y
Observations	3,040	3,040	3,040	3,663	3,663	3,663
R-squared	0.407	0.406	0.406	0.301	0.301	0.301

*Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In columns 3 and 6, the least caregiving occupations (the first quartile) is used as the reference group.*

Applying the same analysis to men reveals no evidence that men working in more caregiving occupations marry at a younger age. If anything, all else being equal, on average men working in more caregiving occupations marry at an older age, though the results are not significant. Table A.6 in the Appendix lists the results when expected age of marriage is not controlled. Compared to Table 1.3, the results without controlling for expected age of first marriage are very similar but slightly larger in coefficient size.

These results support my theory that marriage market incentives encourage women to choose caregiving occupations. The results match Prediction One: women working in caregiving occupations have higher marriage rates (or marry earlier). The same pattern, however, is not found among men, which supports an asymmetric model in which only women signal through sorting into caregiving occupations.

1.5 Caregiving and Divorce Rate

1.5.1 Data

In this section, I test Prediction Two, which states that fewer women will work in caregiving occupations when the divorce rate is higher. This is because a higher divorce rate lowers the expected gain from marriage, thus women have a weaker incentive to signal through occupation. I use microdata from the Current Population Survey, which has sampled households on a yearly basis since 1962. I use the adoption of unilateral divorce laws as an exogenous policy shock, and I carry out difference-in-difference analysis.

Current Population Survey

IPUMS provides harmonized data from the Current Population Survey (CPS),[36] which covers a representative sample (selected with probabilities) of respondents from occupied households. The survey interviews non-military[37] respondents who are at least 15 years old and collects comprehensive information on demographic variables (e.g., race, gender, age, education, income, marital status, etc.) as well as work-related variables (e.g., occupation, industry, usual hours worked, employment status, etc.). This data has been updated annually since 1962 and provides the publicly available geographic information of respondents (e.g., state of residence). IPUMS has harmonized occupation and industry information across different classification systems over time. Compared to NLSY, CPS has a larger sample size and provides data at a longer time-span that is more relevant for the policy change studied. However, since CPS selects random samples each year, I cannot track respondents over time as in NLSY. In CPS, the same respondents can be tracked for at most 16 months,[38] making within-individual comparisons difficult; while NLSY provides longitudinal data. Compared to the Decennial Censuses used in next section, CPS is annual and can better track the policy changes adopted in different years.

Linking CPS and O*NET

For CPS data, IPUMS provides harmonized occupation data in the 2010, 1990, and 1950 COC. I use the harmonized 2010 COC to best match occupation data from O*NET (which provides

[36] Specifically, this paper uses CPS data from the Annual Social and Economic Supplement.
[37] Respondents in some institutions, such as prisons or nursing houses, are also excluded.
[38] CPS interviews respondents for 4 consecutive months, and then there is an 8-month gap that respondents are not interviewed, followed by interviews for another 4 consecutive months.

more precise matching and more variation across occupations).[39] CPS has occupation data available since 1962, but their occupations categories were too broad in 1962-1967 (only 37 categories) so I drop years before 1968. This drop has minimal impact since only one state (Nevada) changed to a unilateral divorce law policy during this time. In the final sample from 1968-2000, there are 342 unique occupations with at least one worker.[40]

1.5.2 Unilateral Divorce Law

Unilateral divorce laws (UDL for short below) allow for divorce upon the request of one spouse, regardless of consent or behavior of the other. Before the adoption of UDL, in many states, divorce required either mutual consent or proof of some form of marital fault (e.g., adultery or domestic violence). During the 1970s, many states adopted UDL.[41] The timing of reform in divorce laws is considered exogenous. As summarized in Stevenson (2007), the adoption of UDL is "routine policy refinement" that is passed "with little notice or dissent."[42] To demonstrate that divorce law reforms are not correlated with women's economic conditions pre-reform in each state, I include scatter graphs similar to those in Voena (2015) that show states' adoption time of UDL and female labor force participation or wives' income in households before the sample period (1960) (see Appendix Figure A.2 and A.3).[43] I also show the residual occupation index overtime, in which time is measured in the number of years that have passed since the adoption of unilateral divorce laws (see Appendix Figure A.4).[44] Based on the graphs, there is no obvious trend for single women in the 10 years before the adoption of UDL. After adoption of UDL, there is an obvious downward sloping trend in the residual caregiving index. Note that the residual caregiving index is more volatile in earlier years, which might be related to a larger measurement error in earlier years (since the caregiving index is calculated based on O*NET data in 2003). Although I use the

[39] I construct the caregiving index based on O*NET data in 2003, classified by the 2000 O*NET-SOC system. This is linked to 2010 SOC and then linked to 2010 COC. Therefore, I use 2010 COC in CPS for a better match.
[40] This time period is similar to the period used to study the adoption of unilateral divorce law in Voena (2015), which uses 1967-1999. Since I follow Voena (2015) and Stevenson (2007) to define years adopting unilateral divorce laws and different property division laws, the sample period also resembles Voena (2015) to avoid potential change in policy outside of the sample period. Stevenson (2007) uses the 1970 and 1980 decennial censuses data.
[41] Only New Mexico and Alaska adopted unilateral divorce law earlier, in 1933 and 1935, respectively. From 1967-2000, another 34 states have adopted unilateral divorce law, with Ohio, South Dakota, Utah, and West Virginia adopting after 1980.
[42] For a detailed discussion of this, see Jacob (1988).
[43] A slight difference between Voena's graphs and my own is that, instead of directly using female labor force participation rate and share of wives' income in households, I use the rate/share relative to male.
[44] Here age, age square, and race are controlled.

2010 COC, which is already harmonized by IPUMS, the measurement is still more accurate for recent years due to more similarity in the classification system.

Apart from UDL, property division laws for divorce also change over time. I follow Voena (2015) to classify property division law into three different regimes: title-based, where property is divided based on who has the title for it; community property, where jointly-owned community property is equally divided between couples and each spouse keeps their own property; equitable distribution, where the judge has the discretion to achieve "fairness" or "equity" in property division. Within my sample period, 1968-2000, there are three main cases concerning the changes in divorce property laws, as summarized by Voena (2015): always community, always equitable, and change from title-based to equitable.[45] Again, I show scatter graphs similar to Voena (2015) on states' adoption of equitable distribution and female labor force participation or wives' income in households before the sample period (in 1960) (see Appendix Figure A.5 and A.6).[46] Table 1.4 lists a summary of the years that states adopted UDL and equitable distribution property laws, with 33 states changed to UDL in sample period.[47]

Previous literature does not agree on the impact of UDL on the divorce rate, but there is evidence pointing to a decline in marriage duration post UDL (Wolfers, 2006). In my model, λ is the only parameter governing marriage duration, and a shorter marriage duration is equivalent to higher rate of divorce shock in model. Therefore, in this paper, the adoption of UDL serves as a shock of larger λ and will lead to fewer women working in caregiving occupations, as Prediction Two states. Stevenson (2007) uses data from newly married couples and finds that after the adoption of unilateral divorce laws couples are less willing to make marriage-specific investments,[48] which are largely invariant to the prevailing property division laws in each state. I would expect similar reductions in women signaling post UDL when viewing women's occupation choice as one such marriage-specific investment made before marriage.

[45] One special case is Wisconsin, which changed from equitable distribution to community property law in 1986.

[46] A slight difference between Voena's graphs and my own is that, instead of directly using female labor force participation rate and share of wives' income in households, I use the rate/share relative to male.

[47] The classification of states with unilateral divorce laws and property division laws follows Voena (2015) and Stevenson (2007).

[48] Examples include being less likely to support each other (sequentially) through school, being less likely to have children in the first 2 years of marriage, having larger female labor force participation, and being more likely to both working full-time.

Table 1.4. Year Adopting Unilateral Divorce Law and Equitable Distribution Regime

State	Unilateral Divorce	Equitable Distribution		Unilateral Divorce	Equitable Distribution
Alabama	1971	1984	Montana	1973	1976
Alaska	1935	pre-1968	Nebraska	1972	1972
Arizona	1973	community	Nevada	1967	community
Arkansas		1977	New Hampshire	1971	1977
California	1970	community	New Jersey		1974
Colorado	1972	1972	New Mexico	1933	community
Connecticut	1973	1973	New York		1980
Delaware	1968	pre-1968	North Carolina		1981
District of Columbia		1977	North Dakota	1971	pre-1968
Florida	1971	1980	Ohio	1992	1981
Georgia	1973	1984	Oklahoma	1953	1975
Hawaii	1972	pre-1968	Oregon	1971	1971
Idaho	1971	community	Pennsylvania		1980
Illinois		1977	Rhode Island	1975	1981
Indiana	1973	pre-1968	South Carolina		1985
Iowa	1970	pre-1968	South Dakota	1985	pre-1968
Kansas	1969	pre-1968	Tennessee		pre-1968
Kentucky	1972	1976	Texas	1970	community
Louisiana		community	Utah	1987	pre-1968
Maine	1973	1972	Vermont		pre-1968
Maryland		1978	Virginia		1982
Massachusetts	1975	1974	Washington	1973	community
Michigan	1972	pre-1968	West Virginia	1984	1985
Minnesota	1974	pre-1968	Wisconsin	1978	community*
Mississippi		1989	Wyoming	1977	pre-1968
Missouri		1977			

Note: Information on year of introduction of unilateral divorce law and property division regime comes from Voena (2015) and Stevenson (2007). For states with empty years of policy change, it means that those states did not adopt the unilateral divorce policy by the year 2000. For equitable distribution, it listed out the years that property division laws change from title-based to equitable distribution regime. "Community" means these states stick to community property regime throughout the research period. One special case is Wisconsin, which switched from equitable to community property regime in 1986.

1.5.3 Empirical Model and Results

I test Prediction Two using data from the Current Population Survey of 1968-2000, in which Ohio was the last state to adopt UDL in 1992. Note that I restrict the main sample to women who have never married. To accommodate the adoption of a UDL, I use difference-in-difference (DID):

$$C_{ist} = \alpha + \beta Unilateral_{st} + \beta_p Property_{st} + \beta_X X_{ist} + \delta_s + \delta_t + \varepsilon_{ist} \tag{1.3}$$

The dependent variable C denotes the caregiving indices, which vary for individual i by state s and year t. The independent variable of interest is the dummy for adoption of UDL, which varies by state and year. The dummy takes on value of 1 when the state has already adopted UDL in year t and takes on value of 0 in all other cases. Property regime dummies are included to control for varying property division laws. The baseline control variables X include individual age, age square, and race. Stevenson (2007) finds that after the adoption of a UDL, newly married couples are less likely to support each other (sequentially) through school, are less likely to have children, have larger female labor force participation, and are more likely to both be working full-time.[49] To avoid potential correlations between these variables and the dependent variable, when comparing results with married men or women, I further include years of education, whether they have any young children (under age of 13) in household, local (defined by state and year) female labor participation rate, and whether they have a full-time or part-time job.[50] In other specifications, I additionally include yearly income from wages and salary, yearly total personal income from all sources, local unemployment rate (gender-specific), usual hours worked, industry, and industry by year fixed effects.[51] The baseline result is robust to the addition of extra controls. State and time fixed effects are included to account for any potential non-time-varying state differences (e.g., gender norm) and shocks that vary over time but common to all states (e.g., national technology change). Note that the marital status here defines single women as "never married" and married women as "currently married". Women who are separated, divorced, and widowed are tested in the robustness check and are not included as single women here.[52]

The results are shown in Table 1.5. I first run DID for single women using the caregiving index as the dependent variable. To avoid over-controlling column 1 includes only baseline controls (age, age square, and race). After the adoption of UDL, single women work in less caregiving occupations on average, but the decrease is not significant at conventional levels. In column 2 to column 7, the dependent variable is replaced as the caregiving dummy. In column 2, after adoption of UDL, single women are 2.3% less likely to work in caregiving occupations. Based

[49] Voena (2015), however, finds a decrease of female labor force participation in states with equitable property laws.
[50] These four controls are refereed as the Stevenson controls below.
[51] Because women could also signal through industry choice similarly to occupation choice, the inclusion of the industry fixed effect would make our result smaller since in reality the choice of occupation and choice of industry might be correlated.
[52] Because they might distribute differently from never-married women's age ranges, and might not follow the same age cutoff when defining the young and old cohort.

on the model, the signaling theory should work better for young single women than old single women. In this empirical part, the age cutoff between the young and old groups is 38 for women and 40 for men.[53] In column 3, there is a very small and statistically insignificant decrease in the likelihood that old single women work in caregiving occupations, while in column 4, there is a 2.6% decrease (significant at the 5% level) for young single women. In column 5, I further include the Stevenson controls. The inclusion of local labor force participation rates could also control for the selection problem of women entering and leaving the labor market. In column 6, beyond the Stevenson controls, I include extra controls of yearly income from wage and salary, yearly total personal income from all sources, local unemployment rate (gender-specific), and usual hours worked. The inclusion of the local unemployment rate helps mitigate some labor market demand shocks that are common to all female workers. Comparing results in columns 4, 5, and 6, the decrease in the likelihood of working in caregiving occupations for young single women largely does not vary (both in size and significance) with the inclusion of these controls. In column 7, I further include industry and industry by year fixed effects because the choice of industry and occupation could be correlated. This does not change the pattern but makes the decrease smaller in size.

Results in Table 1.5 support Prediction Two: single women are less likely to work in caregiving occupations when the divorce rate is higher. This result is mainly driven by young single women and is robust to the inclusion of many other controls. The result in column 4 is used as a baseline. Appendix Figure A.7 shows the dynamic result where time is measured relative to the year when unilateral divorce law was adopted. There is no obvious pattern in the change of the coefficient size nine years after the law adoption so this paper applies the simple DID design averaging the impact over time instead of a dynamic design.

[53] The age cutoff is chosen to be consistent with next section, in which I follow Angrist (2002) to define young women as under age 33 and young men as under age 35 and use 1890-1970 as research period. According to the Bureau of Labor Statistics, the estimated median age of first marriage for men/women was 23.2/20.8 in 1970, and rose to 26.8/25.1 in 2000. Therefore, for the CPS data in 1968-2000, I use a higher age cutoff (40 for men and 38 for women). Using 1970 decennial data, I find that 97.63% of men have married by 40 and 98.11% of women have married by 38.

Table 1.5. Unilateral Divorce Law and Women's Occupation Choice

	Index	Dummy					
	All single	All single	Old single	Young single	Stevenson controls	Extra controls	Industry controls
Treated	-0.828	-0.0231*	-0.00594	-0.0263**	-0.0267**	-0.0264**	-0.0178**
	(0.660)	(0.0135)	(0.0269)	(0.0123)	(0.0119)	(0.0114)	(0.00767)
Property Regime	Y	Y	Y	Y	Y	Y	Y
Year	Y	Y	Y	Y	Y	Y	Y
State	Y	Y	Y	Y	Y	Y	Y
Observations	132,949	132,949	14,523	118,426	118,426	118,426	117,507
R-squared	0.011	0.015	0.021	0.017	0.045	0.048	0.286

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In column 1, the dependent variable is the caregiving index while in column 2 through 7, the dependent variable is the caregiving dummy. In column 1 and column 2, all single women are included. In column 3, only single women over 38 are included. In column 4 through column 7, only single women under 38 are included. Column 4 is the baseline result, where only age, age square, and race are included. In column 5 through column 7, variables that might change after the adoption of unilateral divorce law for newly married couples (as shown in Stevenson 2007) are further controlled, including years of education, whether there are any young children (under age of 13) in the household, local (defined by state and year) female labor participation rate, and full-time or part-time job. In column 6 and column 7, yearly income from wage and salary, yearly total personal income from all sources, local unemployment rate (gender-specific), usual hours worked are further included. In column 7, industry and industry by year fixed effect are further included.*

I carry out the same analysis for each gender by marital status group in Table 1.6. Column 1 shows the baseline result among young single women. In column 2 to column 4, samples are limited to young married women, young single men, and young married men. In column 2 and 4, the Stevenson controls are included since Stevenson (2007) finds changes in these variables among newly married couples post UDL. Comparing column 1 and 2, I do not observe the same decrease in likelihood of working in caregiving occupations for young married women – the change is very small in size and not significant. This supports the model prediction that signaling behavior on the marriage market applies only to single women, and married women do not have the same incentive. The pattern is also not observed within single or married men, supporting the idea embedded in the model that men do not signal through caregiving occupations like women. The results in column 2 to 4 provide evidence that the pattern observed among young single women cannot be fully explained by general demand-side shocks on the labor market (e.g., technological change, outsourcing of some occupations). Performing the same analysis within separated, divorced, and widowed women yields similar, but smaller, results to young single women (see Appendix).

Table 1.6. Unilateral Divorce Law and Occupation Choice of Young Workers, by Gender and Marital Status

	Dummy			
	Single Female	Married Female	Single Male	Married Male
Treated	-0.0263**	0.000786	-0.00766	0.00776
	(0.0123)	(0.00693)	(0.00744)	(0.00679)
Property Regime	Y	Y	Y	Y
Year	Y	Y	Y	Y
State	Y	Y	Y	Y
Observations	118,426	157,489	151,537	234,785
R-squared	0.017	0.086	0.043	0.167

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In columns 1 and 3, only age, age square, and race are included. In columns 2 and 4, variables that might change after adoption of unilateral divorce law for newly married couples (as shown in Stevenson 2007) are further controlled, including years of education, whether there are any young children (under age of 13) in the household, local (defined by state and year) labor participation rate (gender-specific), and full-time or part-time job.*

A comparison between single and married women helps me solve several potential problems. This comparison shows that fundamental labor market demand shocks cannot be the driving force making single women less likely to work in caregiving occupations because this change is not observed for married women. The pattern within young single women does not vary with the addition of local labor market controls (labor force participation rate and unemployment rate of women), which indicates the result is not driven by women selecting in or out of the labor market. The pattern is also unlikely to be purely driven by sorting based on preference (e.g., work hour flexibility, job stability, risk of death, etc.), since sorting on such preferences would create an equal (if not even larger) decrease among married women. Alternative explanations based purely on employer discrimination or gender norms face the same problem of not explaining the non-result among married women.

1.5.4 Robustness Checks

Triple Difference

To directly treat any potential shock to women that might correlate with the adoption of UDL (e.g., new technology that benefits working women and vary across states and year), I further apply difference-in-difference-in-difference (DDD) using married women as the comparison group:

$$C_{ist} = \alpha + \beta Unilateral_{st} \times M_{ist} + \beta_X X_{ist} + \delta_m + \delta_s + \delta_t + \delta_{sm} + \delta_{tm} + \delta_{st} + \varepsilon_{it} \tag{1.4}$$

Here variables bear the same meaning as in the DID regression model (1.3). Note that adoption of UDL is further interacted with M, which is an individual's marital status. M is 1 for never-married women, and 0 otherwise. In addition to the time and state fixed effects also included in DID, I include the marital status dummy as well as double interaction between time, state, and marital status dummy. Baseline controls include age, age square, and race as in DID, but property division regime drop out with inclusion of state-year fixed effect in DDD. In comparing young single women with young married women, I also include the Stevenson controls in some specifications (except local labor force participation rate of women drops out with inclusion of state-year fixed effects). Note that the inclusion of the state-year fixed effect potentially controls the adoption of UDL by state and year. For a comparison of the trends between single and married young women, see Appendix Figure A.4. There is no obvious pre-trend of caregiving index for single or married young women before UDL. After UDL, however, there are clearly opposite trends across single and married young women.

Table 1.7. Unilateral Divorce Law and Women's Occupation Choice (DDD)

	Index		Dummy	
	Baseline Controls	Stevenson controls	Baseline Controls	Stevenson controls
Treated × Single	-1.313***	-0.972***	-0.0411***	-0.0346***
	(0.305)	(0.303)	(0.00987)	(0.00939)
Property Regime	Y	Y	Y	Y
Marital Status	Y	Y	Y	Y
Year	Y	Y	Y	Y
State	Y	Y	Y	Y
Marital × Year	Y	Y	Y	Y
Marital × State	Y	Y	Y	Y
Year × State	Y	Y	Y	Y
Observations	275,915	275,915	275,915	275,915
R-squared	0.016	0.104	0.016	0.070

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In columns 1 and 3, only age, age square, and race are included in the control variables. In columns 2 and 4, variables that might change after the adoption of unilateral divorce law for newly married couples (as shown in Stevenson 2007) are further controlled, including years of education, whether there are any young children (under age of 13) in the household, and full-time or part-time job. Note that property division law regime and local (by state and year) labor force participation rate of women will drop out with inclusion of state-year fixed effect.*

The results of DDD are shown in Table 1.7. Column 1 and 3 only include baseline controls while column 2 and 4 further include the Stevenson controls. In column 1 and 2, there is a decrease in the average caregiving levels of single over married women after the adoption of UDL when measured using the continuous caregiving index. This decrease is smaller in size with Stevenson controls. In column 3 and 4, the dependent variable is replaced with the caregiving dummy. In column 3, the likelihood of working in caregiving occupations decreases by 4.1% more in single young women compared to married young women after the adoption of UDL, and this difference reduces to 3.5% with Stevenson controls. In all specifications, the decrease is significant at 1%. Results are very similar if replacing state-year fixed effects with the UDL dummy as defined in regression (1.3) (results in Appendix Table A.7).

Overall, the results support my theory of women's signaling behaviors on the labor market, with the incentive unique to single women. Coefficient sizes are larger in DDD than in DID, as shown in Table 1.5. In DID, the average caregiving index decreases by 0.83 and likelihood of working in caregiving occupations decreases by 2.6% for young single women.[54] In DDD, the corresponding decreases are 1.31/0.97 and 4.1%/3.5% without/with Stevenson controls.

Age at Policy Change

Compared to baseline DID, I now change the treated dummy to indicate whether the state has adopted UDL before a woman is 18. This measurement is more likely to change in women's first job choices post UDL than it is to influence them to switch away from their current occupations. Note that CPS collects information on the current status of respondents, and thus in this specification there is implicit assumption that women have remained in the same state since they turned 18 (or that moving across states does not create systematic measurement errors that bias the result). Since the measurement considers whether UDL has already taken place before women are 18, under this design, the current marital status and current young/old age group of women are no longer restricted. The sample size is therefore much larger than the baseline sample in which only young single women are included. Prediction Two is robust to this specification (results in Table 1.8). In column 1, women on average work in less caregiving occupations if the state they reside in adopted UDL before they turned 18. In column 2, women are 2% less likely to

[54] The figure of 0.83 is not reported in Table 1.5. It uses the caregiving index as the dependent variable and is restricted to young single women. This coefficient is not significant at conventional levels.

work in caregiving occupations for the treated group, which is slightly smaller than 2.6% in the baseline specification.

Table 1.8. Unilateral Divorce Law and Women's Occupation Choice – Age at Policy Change

	Caregiving	
	Index	Dummy
Treated	-1.244***	-0.0201***
	(0.279)	(0.00571)
Property Regime	Y	Y
Year	Y	Y
State	Y	Y
Observations	550,234	550,234
R-squared	0.009	0.008

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Controls include age, age square, and race. In column 1, the dependent variable is the caregiving index. In column 2, the dependent variable is the caregiving dummy. Here the treated is measured as whether the state has adopted unilateral divorce policy when the women is 18 (assuming that the current state is the same state as when women is 18). Under this design, the current marital status and the current age of women are no longer restricted, and the sample includes all women.*

Alternative Explanations

In Table 1.5, the very small and non-significant result within young married women conflicts with many alternative explanations to the signaling model. To directly test alternative explanations, I carry out further robustness exercises in Table 1.9. Column 1 lists the baseline result,[55] and column 2 to 6 each includes extra controls on one alternative explanation. As explained in Section 1.3, these controls are constructed from O*NET data similarly to the main caregiving index and vary across occupations.

In column 2, I include routine, math, and social intensity as they are constructed in Deming (2017). Compared to the baseline result, the decrease is a bit smaller in size and significant only at 10%. This result shows that it is unlikely that the baseline pattern is driven purely by demand shock or sorting based on skills. In column 3, I include the flexibility measurement, including work hours (measured by number of hours typically worked in one week) and time pressure (measured by the frequency the job requires the worker to meet strict deadlines). In previous literature,

[55] The result is slightly different than in Table 1.5 because here the sample size has further decreased by about 3500 observations after the inclusion of extra controls.

flexibility is one key feature that drives the occupation decision of women (e.g., Flabbi and Moro, 2012; Wiswall and Zafar, 2018). The result in column 3 is even slightly larger than baseline after controlling for flexibility; this suggests that the pattern is not driven by a change in sorting based on flexibility post UDL either. In column 4, I control job stability, another factor studied in literature which affects women's occupation choice (e.g., Wiswall and Zafar, 2018). The result is robust to the inclusion of job stability, despite slightly smaller size and lower significance. It is not likely that the pattern is driven by women sorting into more stable jobs post UDL. In column 5, I include the competition control, which measures the extent to which a job requires workers to compete or to be aware of competitive pressures. Despite previous findings on occupation choice differences concerning competitiveness across gender (e.g., Buser et al., 2014), it is unlikely that changes in women's preferences for competition post UDL drives the pattern observed because the result in column 5 greatly resembles the baseline result. In column 6, I include the job hazard control, which describes how often workers are exposed to several hazardous conditions as part of the job. [56] The job hazard control proxies for the on-the-job death rate, which is another determinant of occupation choice for women (e.g., DeLeire and Levy, 2004). Inclusion of this control has barely any impact on the pattern observed.

Table 1.9. Unilateral Divorce Law and Women's Occupation Choice – Other Features

	Caregiving Dummy					
	Baseline	Deming measures	Flexibility	Job Stability	Competition	Job Hazard
Treated	-0.0259**	-0.0192*	-0.0297**	-0.0215*	-0.0268**	-0.0260**
	(0.0126)	(0.00989)	(0.0126)	(0.0122)	(0.0126)	(0.0124)
Property Regime	Y	Y	Y	Y	Y	Y
Year	Y	Y	Y	Y	Y	Y
State	Y	Y	Y	Y	Y	Y
Observations	114,849	114,849	114,849	114,849	114,849	114,849
R-squared	0.015	0.444	0.080	0.067	0.022	0.016

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Controls include age, age square, and race. Column 1 is the baseline specification. Column 2 includes routine, math, and social intensity (constructed following Deming 2017). Column 3 includes work hours and time pressure. Column 4 includes job stability. Column 5 includes competition. Column 6 includes job hazard.*

[56] Specifically, it averages over six dimensions: exposed to radiation; exposed to disease or infection; exposed to high places; exposed to hazardous conditions; exposed to hazardous equipment; and exposed to minor burns, cuts, bites, or stings.

Across Traditional/Modern States

The model assumes men prefer childcaring women. In reality, this preference depends on how traditional men are, and therefore I expect stronger signaling in states where men are more traditional. Empirically, I carry out the baseline specification across traditional and modern states. To measure the traditional level, I examine the share of church members (all denominations included) in each state in 1971.[57] States are ranked by share of church members, with above-median states classified as "traditional" and below-median states classified as "modern". Based on the model, I would expect women in traditional states to have a larger decrease in likelihood of working in caregiving occupation post UDL than women in modern states.

Table 1.10. Unilateral Divorce Law and Women's Occupation Choice, by Traditional and Modern States

	Index		Dummy	
	Traditional	Modern	Traditional	Modern
Treated	-1.235***	0.152	-0.0379***	0.00477
	(0.369)	(0.625)	(0.00789)	(0.0111)
Property Regime	Y	Y	Y	Y
Year	Y	Y	Y	Y
State	Y	Y	Y	Y
Observations	54,470	62,979	54,470	62,979
R-squared	0.017	0.009	0.023	0.013

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Controls include age, age square, and race. In columns 1 and 2, the dependent variable is caregiving index. In columns 3 and 4, the dependent variable is caregiving dummy. Columns 1 and column 3 only include the traditional states. Columns 2 and column 4 only include the modern states.*

The results are shown in Table 1.10. In column 1 and 2, young single women on average work in less caregiving occupations post UDL in the traditional states. This pattern is not observed, however, in modern states. In column 3, young single women are 3.8% less likely to work in caregiving occupations post UDL in traditional states, which is larger than the baseline result of 2.6%. In modern states, however, there is no significant change. The results here suggest that women have a stronger incentive to signal in more traditional states.

[57] The church members of each state in 1971 (with census population data in 1970) is provided by Association of Religion Data Archives at https://www.thearda.com/Archive/Files/Downloads/CMS71ST_DL2.asp. The sample period in this section is 1968-2000, and the most relevant "start-of-period" church data should be before 1968. However, the closet church data pre-1968 from the same data source is from 1952, which is a bit too early for my study. I therefore use the year 1971, which is at the very start of my sample period.

Across Birth Cohorts

I also run robustness checks by birth cohorts. Young single women are divided into four subgroups based on their birth decades (before 1950, 1950-1960, 1960-1970, and after 1970) in Table 1.11. The decrease in young single women working in caregiving occupations is larger and more significant in 1950-1960 and 1960-1970. The smaller size and insignificancy of results for before 1950 and after 1980 are driven by the following facts: most states adopted UDL among 1970-1980; the research period starts at 1968; working choices are more relevant after age 18; and the sample only includes young single women under 38. The combination of these four facts (in addition to a smaller sample size) makes it relatively rare to identify variation in women born before 1950 and after 1970. In all time periods, however, there is a negative impact post UDL, though it's not significant for women born before 1950 and after 1970.

Table 1.11. Unilateral Divorce Law and Women's Occupation Choice, by Birth Cohorts

	Caregiving Dummy			
	Before 1950	1950-1960	1960-1970	After 1970
Treated	-0.00867	-0.0559***	-0.0478***	-0.0133
	(0.0184)	(0.0171)	(0.0151)	(0.0133)
Property Regime	Y	Y	Y	Y
Year	Y	Y	Y	Y
State	Y	Y	Y	Y
Observations	9,229	35,447	46,049	27,701
R-squared	0.021	0.010	0.009	0.025

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Controls include age, age square, and race. Here the year fixed effect refers to the birth year of respondents.*

1.6 Caregiving and Sex Ratio

In this empirical part, I test Prediction Three, which states that more women will work in caregiving occupations when the sex ratio (men over women) is higher. This is because a higher sex ratio increases women's expected gain from marriage and single women will have a stronger incentive to signal through occupation. The data in this section comes from Decennial Censuses, and I use first-generation immigration as exogenous shock and construct instrumental variables which vary by ethnic group, immigration period, and state. I then use these instruments to study the occupation choices of young single women among second-generation immigrants.

1.6.1 Data

Decennial Censuses

IPUMS provides Decennial Censuses that are harmonized over time. I use the immigration data available in the 1900-1970 Decennial Censuses. Specifically, I use the 1900 5% sample, 1910 1% sample, 1930 5% sample, 1940 1% sample, 1950 1% sample, 1960 5% sample, and 1970 1% samples. [58] Since the samples represent varying percentages of the whole population across decades, I apply personal weights for each sample. [59] Decennial Censuses provide unique immigration data that is essential to my analysis on shock to sex ratio.

For all survey years in this section, Decennial Censuses collect information on respondents' place of birth. The census record birth states for respondents born within the U.S. and birth countries respondents born outside the U.S. This information is used to identify respondents' immigration statuses as well as immigrants' country-of-birth. In all samples except 1970 form 1 (state, metro, and neighborhood form), the census also asks respondents about their parents' place-of-birth. Based on one's own and one's parents' place-of-birth, respondents are classified into three groups: first-generation immigrant, second-generation immigrant, and natives. Respondents born outside of U.S. are classified as new immigrants or first-generation immigrants. Respondents born in the U.S. with one or both parents born outside of U.S. are classified as second-generation immigrants. Respondents born in U.S. with both parents also born in U.S. are classified as natives. For my empirical tests, I use country-of birth for first-generation immigrants and parents' country-of-birth for second-generation immigrants.

In 1900-1930 and 1970 (form 1 of state, metro, and neighborhood form), the census also asked respondents their immigration year, which is used to identify the immigration flow within each immigration period defined in this paper (every ten years). The immigration year question is missing in the 1940-1960 samples and is thus imputed from 1970 sample. Because 1940 and 1950

[58] There are six 1% sample files for 1970: state form 1, state form 2, metro form 1, metro form 2, neighborhood form 1, and neighborhood form 2. All of these files are used at some point, and I will specify which 1970 1% file is used in the discussion below.

[59] In the main sample, the personal weight variable "perwt" is used. For samples including 1940 and 1950, the sample-line weight "slwt" is used as in Angrist (2002) and Lafortune (2013). This variable is same as "perwt" in all research years but 1940 and 1950.

also have much smaller sample sizes than other years, [60] I drop 1940 and 1950 in main specifications. In robustness checks, I also use an all-year sample and a sample that drops 1960.

Linking Decennial Censuses and O*NET

For Decennial Censuses, since the sample years start as early as 1890, the only proper harmonized occupation classification system IPUMS provides is the 1950 COC. I then crosswalk 2000 COC back to 1950 COC to match with O*NET data. In the final sample, there are 261 distinct occupations with at least one worker.

1.6.2 Construction of Instruments

Method Introduction

Prediction Three states that when the sex ratio (number of men over women) is higher, more women will work in caregiving occupations. Empirically, sex ratio can be endogenous, and thus OLS studies using contemporaneous sex ratio and women's occupation choice could suffer from obvious omitted variable problems. To explain exogenous variation in the sex ratio on the marriage market in literature, researchers have used war (Abramitzky et al., 2011; Bethmann and Kvasnicka, 2013; Brainerd, 2017), son preference and abortion (Abrevaya, 2009), gender-skewed incarceration (Charles and Luoh, 2010; Mechoulan, 2011), local sex ratio of same cohort in previous years (Wei and Zhang, 2011), and immigration (Angrist, 2002; Lafortune, 2013).

In my paper, I explore state-ethnic-cohort specific immigration as a potential source of exogenous variation in sex ratio. To solve potential endogeneity problems, I follow Angrist (2002) and Lafortune (2013) in using instrumental variables constructed from first-generation immigrants and study outcomes among second-generation immigrants. As in Angrist (2002), the arrival of first-generation immigrants is used as an exogenous shock to second-generation immigrants within the same ethnic group. This method is further extended by Lafortune (2013) with the addition of geographical variation in immigration distribution across states. I follow their methods to construct

[60] In the final sample (containing all second-generation immigrants who are in the labor force during 1891-1970), there are only 28,967 observations in 1940 and 33844 observations in 1950. Although 1951-1960 immigration flow data is also imputed from the 1970 census data, there are 603,894 observations in 1960 and the imputation time gap is smaller since 1960 is closer to 1970. In 1910 and 1920, sample sizes are also smaller (with 65,796 and 77,758 observations, respectively) but immigration flow data is directly calculated with no imputation, and would not have large measurement errors as in 1931-1950.

instrumental variables (predicted immigration flow and predicted sex ratio) from the first-generation immigration flows. This method is described in detail below.

Steps on Constructing the Instruments

Using N to denote number of immigrants and c to denote country (country-of-birth), the instrumental variables for ethnic group j in state s during immigration period t are:

$$predicted\ sex\ ratio = \frac{\widehat{N_{jst}^{m}}}{\widehat{N_{jst}^{f}}} = \frac{\sum_{c\in j}\left(\frac{N_{cs,past}}{N_c}\right)\times N_{ct}^{m}}{\sum_{c\in j}\left(\frac{N_{cs,past}}{N_c}\right)\times N_{ct}^{f}}$$

$$predicted\ flow = \widehat{N_{jst}} = \sum_{c\in j}\left(\frac{N_{cs,past}}{N_c}\right)\times\left(N_{ct}^{m}+N_{ct}^{f}\right)$$

Specifically, the construction of instruments contains three steps:

Step 1: Calculate historical shares $N_{cs,past}/N_c$. Calculate what share of first-generation immigrants from each country-of-birth is located in each state historically.

Step 2: Allocate new immigrants to states according to historical shares. For each country-of-birth, calculate the total number of new immigrants arrived during an immigration period in the whole nation by gender (N_{ct}^{m} and N_{ct}^{f}). Within each birth country, allocate the new immigrants to states based on the historical shares.[61] The allocation is done separately by gender to calculate the predicted number of male/female immigrants from country c in state s over period t.

Step 3: Sum the ethnic groups. Each ethnic group j contains several countries.[62] Within each ethnic group, sum up the predicted immigrant number calculated in Step 2. This delivers the predicted number of new male/female immigrants from ethnic group j in state s over period t ($\widehat{N_{jst}^{m}}$ and $\widehat{N_{jst}^{f}}$). This implicitly assumes that people tend to marry immigrants from the same ethnic groups.[63]

[61] This method is valid because the choice of location for immigrants of the same country-of-birth are persistent over time

[62] Note that it is necessary that each ethnic group contains several countries. Since the historical shares for allocating new immigrants do not vary between gender within each country-of-birth, if each ethnic group contains only one country, the predicted sex ratio would always collapse into the actual sex ratio. When each ethnic group contains multiple countries, however, variation in historical shares across different countries can be explored.

[63] Lafortune (2013) shows supporting evidence that the endogamy marriage rate can be higher within ethnic groups.

With $\widehat{N^m_{jst}}$ and $\widehat{N^f_{jst}}$ from Step 3, the two instruments are then constructed: the predicted sex ratio is calculated by dividing the number of male over female immigrants, and the predicted immigration flow is calculated by adding the number of male and female immigrants.

Advantages of the Instruments

There are several advantages of using instruments constructed this way, as discussed in Angrist (2002) and Lafortune (2013). First, variation is explored across different ethnic-state-cohort groups, and many confounding factors that do not vary with one of the dimensions will be absorbed by fixed effects. Second, predicted immigration flow and predicted sex-ratio come from first-generation immigrants but outcomes are studied within second-generation immigrants. This helps to alleviate the confounding factors when immigration shock and outcome variables are contemporaneous. Third, instruments are constructed by allocating new immigrants to states based on historical shares, where time provides exogenous source of variation. The exogeneity of historical shares fails only when previous new immigrants select into states based on expectation of future marriage market or labor market outcomes for their children. Four, changes in immigration flow and sex ratio are largely driven by immigration policy changes, providing extra exogenous power.

1.6.3 Immigration Summary Statistics

Treatment of Immigration Data

As described in Section 3, I use representative samples from the 1900-1970 Decennial Censuses in this section. The actual research period is 1891-1970 (specifically, 1891-1930 and 1951-1970) because I aggregate immigration statistics by every 10 years. Correspondingly, I calculate the historical shares by the start-of-period (1890). To avoid potential measurement errors using sample data, especially when an immigration ethnic-state-cohort group is small, I calculate historical shares based on full-count 1900 census microdata. [64] As in Lafortune (2013), all countries outside of U.S. are grouped into 42 distinct countries-of-birth (see Table A.4 in Data Appendix). [65] These countries-of-birth are then combined into nine different ethnic groups, namely:

[64] Since the 1900 census provides immigration year, I can calculate the historical shares in 1890 based on the 1900 full count data.
[65] Some of the “countries-of-birth” might contain more than one country.

British ancestry, Francophone, Southern Europeans, Hispanics, Scandinavians, Germanic, Russians and others, Other Europeans, and Other countries. See Table A.5 in Data Appendix for which countries are included in each ethnic group. For second-generation immigrants, I define their ethnic groups using their parents' country-of-birth.[66]

Each immigration period is defined by decade, and for simplicity I use "cohort" to refer to a corresponding immigration period.[67] The immigration flows, both endogenous and predicted, are limited to first-generation immigrants who arrive within a specific immigration period. Other variables on second-generation immigrants are measured in the survey year. E.g., for cohort 1970, second-generation immigrants' occupations are measured at year 1970 while the immigration flows are calculated based on first-generation immigrants arriving during immigration period of 1961-1970.

To mimic the actual pool of potential mates on the marriage market, the main endogenous variable used in this paper is the sex ratio for the foreign stock which includes both first-generation and second-generation immigrants. Respondents are further divided into younger cohorts and older cohorts based on their age and gender. As in Angrist (2002), the younger cohort includes men aged 20-35 and women aged 18-33,[68] while the older cohort includes men aged 36-50 and women aged 34-48. The ages of both first and second-generation immigrants are measured at the time of survey. For example, in cohort 1970, the younger cohort of male foreign stock contains all second-generation male immigrants who are 20-35 in 1970 and first-generation male immigrants who arrived at US during 1961-1970 and are age 20-35 in 1970. I mainly focus on the younger cohort since it is more relevant to my study of single women on the marriage market.

Summary Statistics

Table 1.12 summarizes the actual flow and sex ratio of new immigrants in each ethnic group over the main research period (1891-1930 and 1951-1970). In Appendix Table A.8, the same statistics are summarized based on 1900-1930 full-count decennial microdata (with research period of 1891-1930). The two figures are similar, but Appendix Table A.8 does not contain new immigrants arriving during 1951-1970, which were more male dominated on average. Figure 1.7

[66] Following Lafortune(2013), if a second-generation immigrant has both parents or only father born outside of US, then her ethnic group is defined based on father's country-of-birth. If a second-generation immigrant has foreign-born mother and American-born father, then her ethnic group is defined based on mother's country-of-birth.
[67] In this empirical part, descriptions of immigration period and cohort are used interchangeably.
[68] As explained in Angrist (2002), men aged 20-35 and women aged 18-33 have the highest marriage rates

shows the variation more clearly using a scatter graph of sex ratio for each ethnic group within the main research period, where each point represents one state and states are sorted in ascending order of sex ratio.

Table 1.12. Summary Statistics on Younger Cohort

Ethnic group	Flow	Sex Ratio		
	Total (million)	Mean	Max (state)	Min (state)
British ancestry	1.71	0.84	1.84	0.30
Francophone	0.30	0.94	20.28	0.07
Southern Europeans	1.31	1.59	7.58	0.44
Hispanics	0.77	1.06	6.21	0.16
Scandinavians	0.70	1.28	22.11	0.26
Germanic	1.59	1.18	2.77	0.35
Russians and others	1.68	1.24	9.88	0.34
Other Europeans	0.71	1.39	9.06	0.52
Other countries	0.44	1.34	10.81	0.48

Note: Calculated using data from IPUMS-USA decennial samples. The young cohort includes men aged 20-35 and women aged 18-33. The flow numbers are based on 1891-1930 & 1951-1970 first-generation young immigrants. Columns 1 and 2 show the total flow of new immigrants and average sex ratio over the whole research period by ethnic group. Columns 3 and 4 shows the maximum and minimum sex ratio of each ethnic group over the same research period, but by state.

Based on the summary statistics and scatter graph, there sex ratio varies highly across ethnic groups during the research period, ranging from an average of 0.84 for the British ancestry to 1.59 for the Southern Europeans. Most ethnic groups have male-dominated new immigration flow with average sex ratios over 1.2, including Southern Europeans, Scandinavians, Russians and others, Other Europeans, and Other countries (Germanic is 1.18). Only British ancestry has female-dominated new immigration with an average sex ratio of 0.84. Francophone and Hispanics have on average more balanced sex ratio of 0.94 and 1.06, respectively.[69] Within each ethnic group, the sex ratio also varies much across states. Some ethnic groups, such as British ancestry and Germanic, have narrowly distributed sex ratios across states. Other ethnic groups, such as Southern Europeans, Hispanics, and Other Europeans, show great variation in sex ratio across states. Compared to the

[69] In Appendix Table A.8, the average sex ratio is 1.30 for Hispanic immigrants who arrived in 1891-1930. This cohort is also male-dominated.

typically balanced sex ratio among all residents, new immigrants show great variation in sex ratio both across and within ethnic groups.

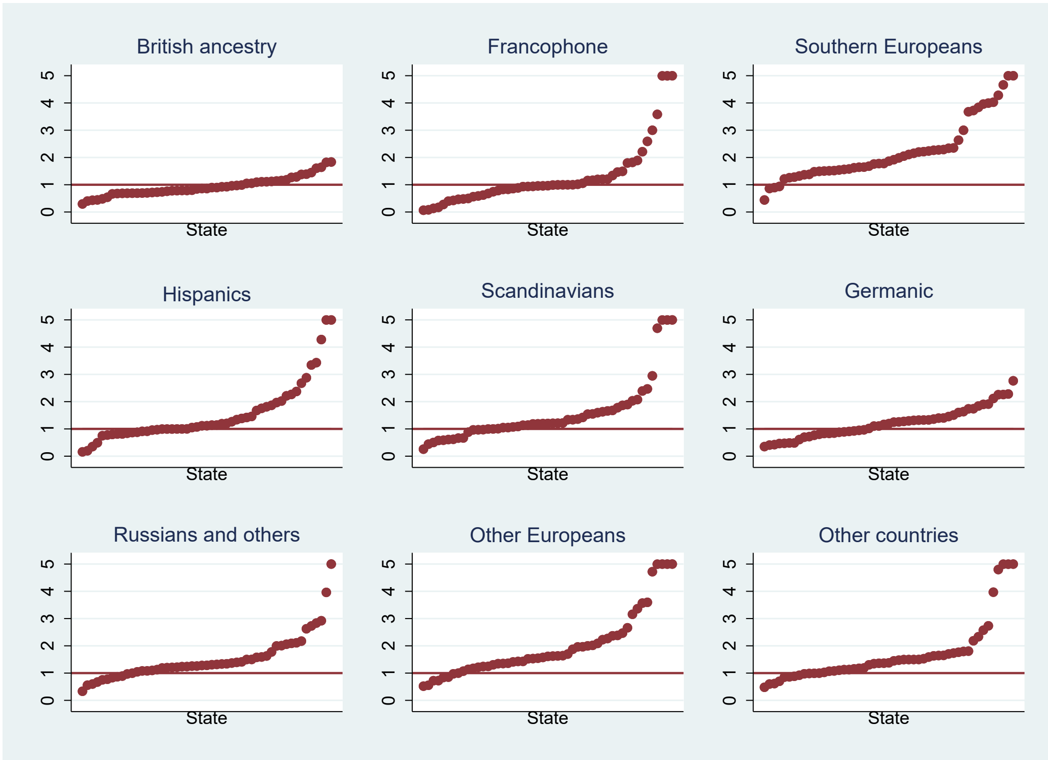

Figure 1.7. Actual Sex Ratio of New Immigrants 1891-1930 and 1951-1970

Note: To narrow the y-axis range, the sex ration in the scatter graph is upper-bounded at 5 (less than 5% observations have sex ratio higher than 5).

1.6.4 Empirical Model and Results

As specified above, the marriage market of second-generation immigrants is defined within each ethnic-state-cohort group. The research period covers 1891-1930 and 1951-1970 for both availability of key variables on immigration and sample size. The empirical strategy is specified below:

$$\frac{N_{jst}^{m}}{N_{jst}^{f}} = \alpha + \gamma_1 \widehat{\frac{\widehat{N_{jst}^{m}}}{\widehat{N_{jst}^{f}}}} + \gamma_2 \left(\widehat{N_{jst}^{m}} + \widehat{N_{jst}^{f}} \right) + \gamma_X X_{ijst} + \delta_j + \delta_s + \delta_t + \delta_{js} + \delta_{jt} + \delta_{st} + v_{ijst} \qquad \text{(1.5.1-Stage 1)}$$

$$C_{ijst} = \alpha + \beta \frac{N^m_{jst}}{N^f_{jst}} + \beta_X X_{ijst} + \delta_j + \delta_s + \delta_t + \delta_{js} + \delta_{jt} + \delta_{st} + \varepsilon_{ijst} \quad \text{(1.5.2-Stage 2)}$$

The outcome variable is the occupation caregiving measurement for each individual i in ethnic group j in state s at immigration period t. In Stage 1, the endogenous sex ratio is instrumented using predicted sex ratio and predicted immigration flow. In Stage 2, the instrumented sex ratio is used to study occupation choice. In both stages, there are individual-level controls of age dummies and nativities of both parents. [70] Stage 2 contains one-dimensional fixed effects of ethnic group, state, and immigration period, as well as two-dimensional fixed effects of interactions between ethnic group, state, and immigration period. Table 1.13 shows the first-stage results.

Table 1.13. First-Stage

	Sex-ratio	
	Flow	Stock
Predicted Sex Ratio	0.622***	0.384***
	(0.167)	(0.107)
Predicted Flow	-0.000780	-0.00116***
	(0.000519)	(0.000207)
Observations	134,066	134,066
Joint F-test	6.660	15.55

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The F-test shows the F-stat for excluded instruments. Here control variables include age dummies, whether mother was born in foreign country and whether father was born in foreign country. Fixed effects include state, immigration period, ethnic groups, and double interaction between these three variables.*

In column 1, endogenous sex ratio is calculated within new immigrants. In column 2, the endogenous sex ratio is calculated based on the whole foreign stock. In both columns, the sample is limited to single young women who are second-generation immigrants and are in the labor force, which is the most relevant group for my study. Column 1 studies how closely the predicted sex

[70] I follow the design of Anriest (2002) and Lafortune (2013) in only using age dummies and origins of both parents. Another reason for only using these controls is that many control variables are missing in the early years of my sample. For example, education, usual hours worked per week, and income from wage and salary are not available before 1940; total personal income is not available before 1950. Employment status is not available in our sample for 1900 and 1920.

ratio correlates with the actual sex ratio in new immigration flows. An increase of one in the predicted sex ratio is correlated with 0.62 increase in the actual sex ratio.[71] Foreign stock is a more relevant marriage pool for second-generation immigrants, and all results going forward use the foreign stock to calculate the endogenous sex ratio. In column 2, a one-unit increase in predicted sex ratio of new immigrants increases the sex ratio of foreign stock by 0.38. The predicted flow has a much smaller impact on sex ratio of foreign stock, where the flow is measured in thousands and the average flow number within a cell is only 15.1 (thousand). The F-stat for the excluded instruments is 15.55, which indicates that the predicted sex ratio and predicted flow are not weak instruments. First-stage results using all young second-generation female immigrants, both single and married, are shown in Appendix Table A.9. The results of both groups are very similar. The 2SLS results are shown in Table 1.14.

Table 1.14. 2SLS – Young Cohort

	Index	Dummy			
	Single Female	Single Female	Married female	Single male	Married male
Stock Sex Ratio	13.38***	0.244***	0.0576	-0.0840	0.00161
	(4.621)	(0.0797)	(0.0825)	(0.131)	(0.00810)
Observations	134,066	134,066	53,481	193,253	253,167

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Here control variables include age dummies, whether mother was born in foreign country and whether father was born in foreign country. Fixed effects include state, immigration period, ethnic groups, and double interaction between these three variables.*

Column 1 shows that when sex ratio is higher post new immigration shock, young single women within the second-generation immigrant group work in more caregiving occupations on average, which supports Prediction Three. When the sex ratio is higher (more men compared to women), the expected gain from marriage is larger and women will have a stronger incentive to signal through working in caregiving occupations. If measured using the caregiving dummy as in column 2, the impact is sizable. Within second-generation immigrants, the likelihood that a single young woman works in caregiving occupation will increase by 24.4% when the sex ratio increases by one unit. The Kleibergen-Paap F-stat[72] equals 15.55, which is larger than the corresponding

[71] A coefficient that is not close to one might be a result of using historical shares that are too early (before 1890). The distribution of new immigrants might change over time, which cannot be fully grasped by the historical shares, especially for later years e.g., 1961-1970.
[72] Here since standard errors are clustered, the Kleibergen-Paap F-stat is better than Cragg-Donald F-stat, according to Baum et al. (2007) in Stata Journal.

Stock-Yogo critical value of 11.59 and thus are not weak at 15% maximum IV size. [73] The Hansen test for over-identification has p-value of 0.64, so there is no evidence of over-identification either. The same analysis is applied to married young women in the labor force, with the coefficient size being much smaller and not significant at 10%. Among single men and married men, there is no significant change in likelihood to work in caregiving occupations either. These results support the model's expectation that signaling only makes sense for single women while married women and men do not have similar incentives on the marriage market. In the following discussion, the result in column 2 of Table 1.14 is used as baseline.

These results support single women's signaling behavior on the labor market. If this pattern is purely driven by alternative explanations such as sorting based on preference, discrimination, and gender norms, there should be a similar (if not stronger) pattern among married women, which is not found in data. The immigration results further provide supporting evidence for the signaling theory against bargaining. If bargaining is the driving force, then woman's bargaining power would increase with a higher sex ratio and they would be less likely to work in caregiving occupations, but this is not supported by data. One potential concern is that if new immigrants sort into similar occupations as the second-generation immigrants, then the arrival of new immigrants directly changes local supply of labor. This is not a big concern here because, even if this cannot be absorbed by the fixed effects, there should be similar pattern among married women or men if results are driven by labor supply shock due to arrival of new immigrants. No such pattern is found in data.

Overall, the results above support Prediction Three of model that when sex ratio is higher (more men compared to women), women will have a stronger incentive to work in caregiving occupations to convey a signal on the marriage market. It also provides evidence for the signaling mechanism and against alternative stories.

1.6.5 Robustness Checks

Placebo of Older Cohort

The signaling story works for all single women but should be more relevant for the younger cohort than the older cohort. I therefore estimate the same regression among the older cohort as a

[73] This means that when the true rejection rate is 5%, the true rejection rate using the IVs in paper is no more than 15%.

placebo test. Based on the signaling story, the decrease in the occupation caregiving level should be smaller and less significant in the older cohort post positive shock to the sex ratio.

The results are listed in Table 1.15. As expected, there is no significant increase in average caregiving occupations for single women in the older cohort, and the size of increase is also smaller than in the younger cohort. The coefficient in column 2 is only half the size of the baseline result and is not significant at 10%. Results within married women, single men and married men are still insignificant. The result matches our expectation that for the older cohort, the marriage market incentive is weaker for occupation choices due to lower marriage rate for these age groups.

Table 1.15. 2SLS – Old Cohort

	Index	Dummy			
	Single Female	Single Female	Married female	Single male	Married male
Stock Sex Ratio	8.933	0.122	-0.0304	0.00652	0.0221
	(7.811)	(0.136)	(0.175)	(0.00669)	(0.0351)
Observations	36,017	36,017	97,227	49,223	379,707

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Here control variables include age dummies, whether mother was born in foreign country and whether father was born in foreign country. Fixed effects include state, immigration period, ethnic groups, and double interaction between these three variables. Here the old cohort includes women between age 34 and 48 and men between age 36 and 50.*

Across Ethnic Groups of Different Endogamy Level

The implicit assumption behind constructing instruments in such a way is that second-generation immigrants tend to marry partners from the same ethnic group. This is supported in Lafortune (2013) by the endogamy rates of cohorts born 1865-1884 in the nine different ethnic groups. This measure has deducted the population share of one's own ethnic group (in 1900) from the actual endogamy rate, so it reflects one's likelihood of marrying a partner from the same ethnic group beyond randomly assigned matches. This rate is gender-specific and is replicated for women in Appendix Table A.10. Based on the endogamy rate, I divide samples into high-endogamy (including South Europeans, Hispanics, Scandinavian, Russians and others, and Other Europe) and low-endogamy ethnic groups (including British ancestry, Francophone, Germanic, and Other countries). The analysis is carried out in each subsample.

Table 1.16. High and Low Endogamy

	High Endogamy	Low Endogamy
Stock Sex Ratio	0.296**	0.141*
	(0.125)	(0.0775)
Observations	51,605	82,441

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Here control variables include age dummies, whether mother was born in foreign country and whether father was born in foreign country. Fixed effects include state, immigration period, ethnic groups, and double interaction between these three variables.*

As shown in Table 1.16, in both groups a higher sex ratio leads to more young single women working in caregiving occupations, as stated in Prediction Three. However, this increase is more significant and twice as large in high-endogamy than in low-endogamy ethnic groups. This is because the actual sex ratio each woman faces cannot be exactly identified due to the narrowly defined marriage market, and instrumented foreign stock sex ratio does a better job in tracking the actual sex ratio faced by women in high-endogamy ethnic groups.

Across Education Groups

As shown in the introduction, occupation caregiving level is positively correlated with workers' average education. I therefore carry out the same analysis across different skill groups by defining high-skill and low-skill caregiving occupations. Occupations are divided into high and low skill based on 1940 full-count census microdata. I rank occupations by share of workers with high-school degree, with above-median occupations classified as high-skill and below-median occupations classified as low-skill. An occupation is high-skill caregiving if it is high-skill and has value of 1 in the caregiving dummy. Low-skill caregiving occupations are similarly defined.

Results are shown in Table 1.17. In column 1, young single women who work in low-skill caregiving occupations are dropped from the sample. In column 2, young single women who work in high-skill caregiving occupations are dropped from the sample. In both columns, young single women who work in non-caregiving occupations are included. The pattern is robust to both high-skill and low-skill caregiving occupations, with the increase in working in high-skill caregiving occupation larger and more significant among young single women. Based on Table 1.17, it is unlikely that selection based on education would be the driving force behind the main results.

Table 1.17. High and Low Skill

	High Skill	Low Skill
Stock Sex Ratio	0.370***	0.218**
	(0.125)	(0.100)
Observations	97,027	78,594

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Here control variables include age dummies, whether mother was born in foreign country and whether father was born in foreign country. Fixed effects include state, immigration period, ethnic groups, and double interaction between these three variables.*

Robustness Tests with Different Samples

For robustness, I test Prediction Three in samples constructed in different ways (results in Table 1.18). All these samples, like the baseline, are limited to young second-generation female immigrations who are single and in the labor force. Column 1 lists the baseline result used as the benchmark. In column 2, I drop the immigration period of 1951-1960 since immigration year in 1960 is imputed from 1970, as explained in data section. The results are very similar to the baseline and a one-unit increase in sex ratio increases the likelihood of single women working in caregiving occupations by 23.3%. In column 3, all years 1891-1970 are included with 1931-1950 added back into sample. The coefficient is similar in size but less significant, due to larger measurement errors. As explained in the data section, 1940 and 1950 data are smaller in sample size [74] and immigration year is imputed, leading to large measurement errors. Instead of using person-level data as in baseline, data in column 4 are aggregated into state-ethnic-cohort cells. There are 1871 non-empty cells with an average of 71.28 observations (non-weighted) per cell. The outcome variable is now the share of single women working in caregiving occupations, and the control variables are correspondingly adapted to the share of these women with foreign-born mothers, the share with foreign-born father, and the share of each age from 18 to 33. For each one-unit increase in sex ratio, the share of single women working in caregiving occupations rises by 21%, which is similar to the baseline result in size but less significant. In column 5, immigration flow data from 1891-1930 are calculated from full-count census microdata in 1900-1930 to potentially avoid measurement error using sample data. However, since such full-count data is not yet publicly

[74] This can be observed through the fact that inclusion of these twenty years only adds about 4,900 observations in the sample (compared to observation numbers in column 1 and column 3).

available for 1970, the immigration flow data from 1951-1970 is still calculated based on the 1970 sample data. The control variables and occupation information still come from sample census data, the same as in the baseline case. Compared to column 1, the sample size has a slight increase of 330 observations and fewer observations are missing immigration flow by ethnic-state-cohort using full-count data. Prediction Three is robust to this sample but the size of increase is slightly smaller than baseline (0.195 compared to 0.244). In column 6, the share varies with immigration period and is chosen as "just before the immigration period" (calculated from 1900-1930 full-count microdata), rather using the 1890 historical shares for construction of instruments for all immigration periods (calculated from 1900 full-count microdata) as before. For example, the historical shares for immigration period 1891-1900 are calculated until 1890, while the historical shares for immigration period 1901-1910 are calculated until 1900. Note that for data availability of full-count censuses, the historical shares for 1951-1970 are based on shares until 1930.[75] The result using rolling historical shares is very similar to the baseline, both in size and significance. Overall, the pattern found in the baseline result is robust to samples constructed in alternative ways, providing further support to Prediction Three.

Table 1.18. Different Samples

	Baseline	Drop 1960	All years	Cells	Full-count	Rolling
Stock Sex Ratio	0.244***	0.233***	0.235*	0.210*	0.195***	0.223***
	(0.0797)	(0.0831)	(0.131)	(0.124)	(0.0724)	(0.0760)
Observations	134,066	116,604	138,960	1,871	134,396	134,066

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Here control variables include age dummies, whether mother was born in foreign country and whether father was born in foreign country. Fixed effects include state, immigration period, ethnic groups, and double interaction between these three variables.*

Robustness Tests with Alternative Specification

I also run robustness tests using alternative specifications in Table 1.19. In the baseline case, I use two instruments (predicted sex ratio and predicted flow) for one endogenous variable (actual foreign stock sex ratio). In column 1, I use the same two instruments but now with two endogenous variables, including foreign stock sex ratio and foreign stock number. The size is similar to the baseline model (which is 0.24) but significant only at 5%. The foreign stock number has very small

[75] The full-count census microdata is publicly available until 1940, but 1940 does not have immigration year data.

and not significant impact on single women's occupation caregiving choices. In column 2, instead of 2SLS, I use Limited Information Maximum Likelihood, which will provide a better estimation with weak instruments. The result is very robust to this specification. In column 3, I show the reduced form, in which whether a single woman works in caregiving occupation is directly regressed on predicted sex ratio and predicted flow. The result is robust to the reduced form, in which a one unit increase in predicted sex ratio leads to a 9.3% increase in the likelihood that single women will work in caregiving occupations (recall that in the first-stage, a one unit increase in predicted sex ratio is correlated with a 0.377 increase in foreign stock sex ratio). Like the foreign stock number in column 1, column 2 similarly shows the small and non-significant impact that the predicted flow number has on single women's occupation. In column 4, I show the OLS result where a one-unit increase in sex ratio is correlated with 6.8% increase in likelihood of working in caregiving occupation for single women, which is much smaller than the IV result.

Table 1.19. Robustness Check

	Two endo	LIML	Reduced	OLS
Stock Sex Ratio	0.267**	0.244***		0.0680***
	(0.103)	(0.0797)		(0.0253)
Stock Number	0.000221			
	(0.000425)			
Predicted Sex Ratio			0.0931***	
			(0.0241)	
Predicted Flow			-0.000233	
			(0.000148)	
Observations	134,066	134,066	134,066	134,066

*Note: Standard errors (clustered at states) in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Here control variables include age dummies, whether mother was born in foreign country and whether father was born in foreign country. Fixed effects include state, immigration period, ethnic groups, and double interaction between these three variables.*

The downward bias of the OLS result might come from bargaining (across genders) in the contemporary generation of immigration. In an ethnic-state-cohort group that has a higher sex ratio for the foreign stock, single women might have higher bargaining power on the marriage market and are less likely to signal through opting into caregiving occupations. This will put a downward bias on the OLS result when first-generation male immigrants have the intention to select into states where men from the same ethnic group have higher bargaining power on the marriage market.

This correlation would be less of a problem in the IV result using first-generation immigration flow to study second-generation outcomes, with allocation across states using historical shares.

1.7 Conclusion

Occupation segregation is an important contributor to the remaining gender wage gap. However, why women sort into certain occupations, typically low-paying ones, remains an open question. Previous literature mainly approaches this issue through three broad explanations: human capital (Becker, 1985; Cobb-Clark and Tan, 2011; Cortes et al, 2018; Adda et al., 2017), gender norms and discrimination (Kuhn and Shen, 2013; Barigozzi et al., 2018), and preferences (Flabbi and Moro, 2012; Wiswall and Zafar, 2018; DeLeire and Levy, 2004; Buser et al., 2014). In this paper, I raise another dimension that might lead to occupation segregation by gender – marriage market signaling. I propose a model that features incomplete information on the marriage market, women born with vertically differentiated preferences on caring for children, and occupations with different caregiving levels. Men value women's caregiving traits, which cannot be observed in the marriage market. In a Perfect Bayesian Equilibrium, some women work in caregiving occupations despite preferring other employment due to expected gains in the marriage market.

The model generates three novel predictions, each of which I tested empirically. Prediction One states that women working in caregiving occupations have higher marriage rates. Using NLSY data, I show that single women whose first job involves more caregiving marry younger, even conditional on their expected age of marriage. Prediction Two states that with lower expected marital surplus, women are less likely to sort into caregiving occupations to signal. I support Prediction Two by studying the adoption of unilateral divorce laws under the DID design using the Current Population Survey 1968-2000. I find young single women are 2.6% less likely to work in caregiving occupations post policy shock. This pattern does not appear among young married women or young men (single or married). Prediction Three states that a higher male-female sex ratio encourages women to signal through caregiving occupations. To test Prediction Three, I construct instrumental variables from first-generation immigrants and apply these instruments to study the occupation choices of second-generation immigrants. Using the Decennial Census 1900-1970, I find that a one unit increase in sex ratio leads to a 24.4% increase in the likelihood of working in caregiving occupations among young single women. This increase is smaller and less significant among older single women and is not found among young married women or young men (single or married).

My model provides new insights into why women's employment in caregiving occupations has remained so persistent. However, several recent papers have found deteriorating labor market conditions for men (Autor et al., 2013; Autor et al., 2019; Cortes et al., 2018). My model predicts that these changes in men's labor market will reduce the incentive for women to signal through occupational choice. Analyzing the impact on the next cohort of women will provide an important avenue for future research. Likewise, there have been various policies, implemented or proposed, across the Western world on providing for free or subsidized childcare. To the extent that these policies change men's preference for partners on the marriage market, we may expect to see less occupational segregation among future cohorts.

1.8 References

Abramitzky, Ran, Adeline Delavande, and Luis Vasconcelos (2011). Marrying Up: The Role of Sex Ratio in Assortative Matching. *American Economic Journal: Applied Economics*, *3*(3), 124-57.

Abrevaya, Jason (2009). Are There Missing Girls in the United States? Evidence from Birth Data. *American Economic Journal: Applied Economics*, *1*(2), 1-34.

Adda, Jérôme, Christian Dustmann, and Katrien Stevens (2017). The Career Costs of Children. *Journal of Political Economy*, *125*(2), 293-337.

Akerlof, George A., and Rachel E. Kranton (2000). Economics and Identity. *Quarterly Journal of Economics*, *115*(3), 715-753.

Angrist, Josh (2002). How Do Sex Ratios Affect Marriage and Labor Markets? Evidence from America's Second Generation. *Quarterly Journal of Economics*, *117*(3), 997-1038.

Autor, David H., Frank Levy, and Richard J. Murnane (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *Quarterly Journal of Economics*, *118*(4), 1279-1333.

Autor, David H., David Dorn, and Gordon H. Hanson (2013). The China Syndrome: Local labor Market Effects of Import Competition in the United States. *American Economic Review*, *103*(6), 2121-68.

Autor, David, David Dorn and Gordon Hanson (2019). When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men. *American Economic Review: Insights*, *1*(2), 161-78.

Autor, David and David Dorn (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, *103*(5), 1553-97.

Barigozzi, Francesca, Helmuth Cremer, and Kerstin Roeder (2018). Women's Career Choices, Social Norms and Child Care Policies. *Journal of Public Economics*, *168*, 162-173.

Baum, Christopher F., Mark E. Schaffer, and Steven Stillman (2007). Enhanced Routines for Instrumental Variables/Generalized Method of Moments Estimation and Testing. *The Stata Journal*, *7*(4), 465-506.

Becker, Gary S (1973). A Theory of Marriage: Part I. *Journal of Political Economy*, *81*(4), 813-846.

Becker, Gary S (1974). A Theory of Marriage: Part II. *Journal of Political Economy*, *82*(2, Part 2), S11-S26.

Becker, Gary S (1985). Human Capital, Effort, and the Sexual Division of Labor. *Journal of Labor Economics*, *3*(1, Part 2), S33-S58.

Bergstrom, Theodore C., and Mark Bagnoli (1993). Courtship as a Waiting Game. *Journal of Political Economy*, *101*(1), 185-202.

Bethmann, Dirk, and Michael Kvasnicka. (2013). World War II, Missing Men and out of Wedlock Childbearing. *The Economic Journal*, *123*(567), 162-194.

Bisin, Alberto, and Giulia Tura (2019). Marriage, Fertility, and Cultural Integration in Italy. NBER Working Paper No.26303.

Blau, Francine D., Peter Brummund, and Albert Yung-Hsu Liu (2013). Trends in Occupational Segregation by Gender 1970–2009: Adjusting for the Impact of Changes in the Occupational Coding System. *Demography*, *50*(2), 471-492.

Blau, Francine D., and Lawrence M. Kahn (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, *55*(3), 789-865.

Brainerd, Elizabeth (2017). The Lasting Effect of Sex Ratio Imbalance on Marriage and Family: Evidence from World War II in Russia. *Review of Economics and Statistics*, *99*(2), 229-242.

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais (2017). 'Acting Wife': Marriage Market Incentives and Labor Market Investments. *American Economic Review*, *107*(11), 3288-3319.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek (2014). Gender, Competitiveness, and Career Choices. *Quarterly Journal of Economics*, *129*(3), 1409-1447.

Charles, Kerwin Kofi, and Ming Ching Luoh (2010). Male Incarceration, the Marriage Market, and Female Outcomes. *The Review of Economics and Statistics*, *92*(3), 614-627.

Chiappori, P. A. (2020). The Theory and Empirics of the Marriage Market. *Annual Review of Economics*, *12*, 547-578.

Chiappori, Pierre-André, Sonia Oreffice, and Climent Quintana-Domeque (2012). Fatter Attraction: Anthropometric and Socioeconomic Matching on the Marriage Market. *Journal of Political Economy*, *120*(4), 659-695.

Cobb-Clark, Deborah A., and Michelle Tan (2011). Noncognitive Skills, Occupational Attainment, and Relative Wages. *Labour Economics*, *18*(1), 1-13.

Cortes, Guido Matias, Nir Jaimovich, and Henry E. Siu (2018). The "End of Men" and Rise of Women in the High-Skilled Labor Market. NBER Working Paper No. 24274

DeLeire, Thomas, and Helen Levy (2004). Worker Sorting and the Risk of Death on the Job. *Journal of Labor Economics*, *22*(4), 925-953.

Deming, David J (2017). The Growing Importance of Social Skills in the Labor Market. *Quarterly Journal of Economics*, *132*(4), 1593-1640.

Dokko, Jane, Geng Li, and Jessica Hayes (2015). Credit Scores and Committed Relationships.

Dupuy, Arnaud, and Alfred Galichon (2014). Personality Traits and the Marriage Market. *Journal of Political Economy*, *122*(6), 1271-1319.

England, Paula, Michelle Budig, and Nancy Folbre (2002). Wages of Virtue: The Relative Pay of Care Work. *Social problems*, *49*(4), 455-473.

Fisman, Raymond, Sheena Iyengar, Emir Kamenica, and Itamar Simonson (2006). Gender Differences in Mate Selection: Evidence from a Speed Dating Experiment. *The Quarterly Journal of Economics*, *121*(2), 673-697.

Flabbi, Luca, and Andrea Moro (2012). The Effect of Job Flexibility on Female Labor Market Outcomes: Estimates from a Search and Bargaining Model. *Journal of Econometrics*, *168*(1), 81-95.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles and J. Robert Warren. *Integrated Public Use Microdata Series, Current Population Survey: Version 8.0* [dataset]. Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/D030.V8.0

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos (2014). Marry Your Like: Assortative Mating and Income Inequality. *American Economic Review*, *104*(5), 348-53.

Harris, Matthew C., and Christopher J. Cronin (2017). The Effects of Prospective Mate Quality on Investments in Healthy Body Weight among Single Women. *Economics & Human Biology*, *24*, 164-183

Hawley, Peggy (1971). What Women Think Men Think: Does It Affect Their Career Choice?. *Journal of Counseling Psychology.* 18.3 193.

Hitsch, Gunter J., Ali Hortaçsu, and Dan Ariely (2010). Matching and Sorting in Online Dating. *American Economic Review*, *100*(1), 130-63.

Jacob, Herbert (1988). *Silent revolution: The Transformation of Divorce Law in the United States*. University of Chicago Press.

Kirkebøen, Lars, Edwin Leuven, and Magne Mogstad (2021). College as a Marriage Market. NBER Working Paper No. 28688

Kuhn, Peter, and Kailing Shen (2013). Gender Discrimination in Job Ads: Evidence from China. *Quarterly Journal of Economics*, *128*(1), 287-336.

Lafortune, Jeanne (2013). Making yourself attractive: Pre-Marital Investments and the Returns to Education in the Marriage Market. *American Economic Journal: Applied Economics*, *5*(2), 151-78.

Mare, Robert D. (1991). Five Decades of Educational Assortative Mating. *American Sociological Review*, 15-32.

Mechoulan, Stéphane (2011). The External Effects of Black Male Incarceration on Black Females. *Journal of Labor Economics*, *29*(1), 1-35.

Olivetti, Claudia., M. Daniele Paserman, Laura Salisbury, and E. Anna Weber (2020). Who Married, (to) Whom, and Where? Trends in Marriage in the United States 1850-1940. NBER Working Paper No. 28033

Schwartz, Christine R. (2010). Earnings Inequality and the Changing Association between Spouses' Earnings. *American Journal of Sociology*, *115*(5), 1524-1557.

Schwartz, Christine R., and Robert D. Mare. (2005). Trends in Educational Assortative Marriage from 1940 to 2003. *Demography*, *42*(4), 621-646.

Smith, Lones (2006). The Marriage Model with Search Frictions. *Journal of Political Economy*, *114*(6), 1124-1144.

Ruggles, Steven, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler and Matthew Sobek. *IPUMS USA: Version 11.0* [dataset]. Minneapolis, MN: IPUMS, 2021. https://doi.org/10.18128/D010.V11.0)

Stevenson, Betsey (2007). The Impact of Divorce Laws on Marriage-Specific Capital. *Journal of Labor Economics*, *25*(1), 75-94.

Voena, Alessandra (2015). Yours, Mine, and Ours: Do Divorce Laws Affect the Intertemporal Behavior of Married Couples?. *American Economic Review*, *105*(8), 2295-2332.

Wei, Shang-Jin, and Xiaobo Zhang (2011). The Competitive Saving Motive: Evidence from Rising Sex Ratios and Savings Rates in China. *Journal of Political Economy*, *119*(3), 511-564.

Wilson, William J. (1987). *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. University of Chicago Press.

Wilson, William J. and Kathryn M. Neckerman (1987). Poverty and Family Structure: The Widening Gap between Evidence and Public Policy Issues. In S. H. Danziger & D. H. Weinberg (Eds.), *Fighting Poverty: What Works and What Doesn't* (pp. 232–259). Harvard University Press.

Wiswall, Matthew, and Basit Zafar (2018). Preference for the Workplace, Investment in Human Capital, and Gender. *Quarterly Journal of Economics*, *133*(1), 457-507.

Wolfers, Justin (2006). Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results. *American Economic Review*, *96*(5), 1802-1820.

CHAPTER 2. COVID19 AND CONSUMER ANIMUS TOWARDS CHINESE PRODUCTS: EVIDENCE FROM AMAZON DATA

2.1 Introduction

"Masks from China, no, no, no !!!!!!"

– A consumer review on Amazon.com

Covid19 has tremendously affected all areas of our lives. As of April 2022, there have been more than 487 million reported cases and more than 6.14 million deaths from Covid19 worldwide[76]. In the United States, the total unemployment rate surged from 3.5% in January 2021 to 14.7% in April 2021[77]. The impact of Covid19, however, is not equally born by all. Researchers have studied the inequality in Covid19 impacts across genders (Adams-Prassl et al., 2020; Alon et al., 2020), races (Amuedo-Dorantes et al., 2021; Couch et al, 2020), immigration-status (Borjas and Cassidy, 2020), education (Adams-Prassl et al., 2020), and work arrangements (Adams-Prassl et al., 2020). This unequal burden is especially true for Asian Americans. Between March 28th to April 25th, 2020, Asian Americans witnessed a 6900% increase in initial unemployment claims in New York, compared to a rise of 1840%, 1260%, and 2100% for the white, black, and Hispanic/Latino workers.[78]

In this paper, we study the animus towards China from a new dimension – online shopping. Unfortunately, online platform has not been a pure land from discrimination. Previous literature has shown the discrimination towards certain racial or ethnic groups via online platforms such as eBay (Ayres et al., 2015), Airbnb (Kakar et al., 2018; Edelman and Luca, 2014), Blocket[79] (Ahmed and Hammarstedt, 2008), local online retailing website (Doleac and Stein, 2013), and online carpooling markets (Tjaden et al., 2018).

Since China was the first country to report cases of Covid19, it has led many to relate Covid19 with China. To make matters worse, there have been prominent political leaders that have used language to stigmatize China with blame and fear around Covid-19, including the ex-

[76] Data reported by John Hopkins at https://coronavirus.jhu.edu/map.html.
[77] Data from Bureau of Labor Statistics https://www.bls.gov/charts/employment-situation/civilian-unemployment-rate.htm. The unemployment rate has declined after April 2021, back to 3.8% as of February 2022.
[78] Change is measured compared with similar time in 2019 (March 30th to April 27th). Reported by CNN Business at https://www.cnn.com/2020/05/01/economy/unemployment-benefits-new-york-asian-americans/index.html.
[79] Blocket.se is one of the largest buy and sell sites for the housing market in Sweden.

president of the United States. On March 16th, 2020, the then U.S. President Donald Trump first called Covid19 the "Chinese virus" in his posts on Twitter and has used that language ever since in media interviews, at rally speeches, and on social media. Fox news, Newsmax and other fringe networks outlets have continued with this rhetoric. This rising animus towards China is further generalized to Chinese and Asians. Hahm et al. (2021) records the rise of anti-Asian (including Asian and Asian Americans) discrimination found within the Adult Resilience Experiences Study (CARES). Lu and Sheng (2020) finds an immediate rise in anti-Asian animus following the exogenous arrival of first Covid19 case locally, reflected by derogatory racial epithets in Google searches and Twitter posts. On employment, Amuedo-Dorantes et al. (2021) records declining entrepreneurship among Asian immigrants compared to non-Hispanic whites after January 2020, and find a substantial increase in exits from self-employment for Asian immigrants. More generally, Bartoš et al. (2021) provides evidence for rising hostility against foreigners from the EU, U.S., and Asia due to the Covid19 Crisis in the Czech Republic. This rising hostility towards foreign people can be partly explained from an evolutionary psychological perspective, where the chronic and contextually aroused feelings of vulnerability to disease motivate negative reactions to foreign (but unfamiliar) immigrants (Faulkner et al., 2004).

The social media has further promoted the propagation of this animus. Croucher et al. (2020) studies social media use and finds that the more social media users believe that their most-used daily social media are fair/accurate/giving facts, the more likely that they believe Chinese pose a threat to U.S. He et al. (2021) analyzes anti-Asian hate speech on Twitter, and finds that in 2020, users who are exposed to hateful content are highly-likely to become hateful.

This phenomenon of rising animus, however, is not unique to Covid19. Historically, researchers have observed rising animus or even hate crime towards certain ethnic groups post profound negative events, such as Arabs/Muslims post 9/11 attack (Kaushal et al., 2007); Germans post WWI (Ferrara and Fishback, 2020), Asians/Arabs post 7/7 attack (Hanes and Machin, 2014); and Muslims post Jihadi attacks (Ivandic et al., 2019). [80]

After Covid19, consumers might hold animus towards Chinese products, either out of prejudice (e.g., blaming China for the breakout of Covid19) or health concerns (e.g., worrying that Chinese products might carry the virus, which risk is low according to Centers for Disease Control

[80] On July 7th, 2005, there was a serious of suicide bombs attacks on London's public transport system.

and Prevention[81]). In this paper, we investigate how the animus towards China affects the Chinese products. To that end, we compile data covering all face masks sold on Amazon between September 1st, 2019 to September 7th, 2020, including the consumer reviews. We collect information on country-of-origin of a product from both seller-generated (e.g., product name, description, feature) and buyer-generated information (e.g., reviews and customer Q&A).

To avoid the problems raised in Goodman-Bacon (2021) concerning the inclusion of two-way fixed effects under the DID design when treatment time varies, we apply the fully-dynamic event study design as suggested by Borusyak and Xavier (2017). Under this design, we find that, despite no change in quality, the average rating of a product drops after being identified as made in China for the first time. This negative impact is U-shaped, which quickly expands in the first five weeks, and gradually fades out within six months. By further splitting Chinese product into high and low reputation using its average rating before the identification, we find that the U-shape decline in product average rating is driven by the high reputation ones. The same pattern is not found among products made in the U.S. or other countries-of-origin.

The negative impact of the informative reviews can be explained by the direct (via its own rating) and indirect (via ratings given by other future reviewers) mechanism. The direct impact persists overtime through the high correlation of product average rating from day to day, and decreases in size unambiguously over time. The indirect impact is U-shaped, which first expands in size with more future consumers seeing the review, and then shrinks in size with fewer new consumers see and get affected by the review due to a rising time cost. The explanations via direct and indirect mechanisms are then supported by studying the negative impacts (lagged) informative reviews have on product average rating. Our results are not driven by quality difference between Chinese and non-Chinese products, which is potentially controlled by the product fixed effects. We provide further evidence against the quality story by analyzing the informative reviews and collect information on whether it contains any complaint about product quality. Results on total impacts and indirect impacts are very similar using the informative-animus reviews without quality complaints.

The remainder of this paper is organized as follows: Section 2.2 describes our data and treatment; Section 2.3 specifies our event study design, shows the results, and explains the

[81] Refer at CDC: https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/surface-transmission.html.

mechanisms behind the results; Section 2.4 provides further evidence and discussions for the mechanisms raised in Section 2.3; Section 2.5 concludes our finding and discuss its significance.

2.2 Data

2.2.1 Data Introduction

Amazon is the largest online retailer. In 2021, it is estimated to account for 41% of the total retail ecommerce sales in the U.S., followed by Walmart Inc. (6.6%), eBay (4.2%), and Apple (4.0%)[82]. Unlike eBay, Amazon does not have a character limit on its reviews, which promotes the richness of information expressed in the consumer reviews.[83] We therefore choose Amazon as the online platform to study, and collect data both from Keepa.com and by ourselves.

Product Information

Keepa.com is an online platform tracking products listed on Amazon. For each product, it provides its current and historical information, including product name, product description, price, sales rank, total number of customer ratings, average rating, ASIN, seller information, first day the product is listed, etc.[84] ASIN is Amazon's Standard Identification Number, which uniquely identifies a product that is very narrowly defined (e.g., the same face mask sold by the same seller but of different color or size could be given two separate ASINs). This level of precision gives us the confidence that the product is identical over time under the same ASIN.[85] In this paper, an ASIN and a product is used interchangeably.

One highlight of Keepa.com is that it tracks products (identified by the same ASIN) overtime and provides high-frequency data[86]. For convenience of use, we aggregate data on a daily or weekly basis to construct the panel data that vary by ASIN and date/week. Despite the great richness of data, Keepa.com does not retain information for delisted products. Once a product gets

---

[82] Data estimated by eMarketer at https://www.emarketer.com/.
[83] There is an 80-character limit on reviews posted on eBay. Compared to Amazon, reviews that express animus towards China on eBay and Yelp.com are very rare.
[84] Amazon does not reveal the actual sales of a product to the consumers. Instead, it provides the sales rank of a product under a specified category.
[85] There are some cases that multiple sellers are selling the same product under the same ASIN (e.g., on September 7th, 2020, only 24.02% of face masks sold on Amazon have multiple sellers). This, however, would not affect our analysis since we study animus towards Chinese products, and the key is to track products instead of sellers.
[86] Most products' information is updated several times daily (see https://keepa.com/#!faq).

delisted from Amazon, Keepa.com will also remove its data within two weeks, leading to data selection concerns. In our data, however, data selection is not a big problem. We tracked 1630 face masks that are listed on Amazon on September 18th, 2020, and find that only 129 (or 7.91%) of them get delisted until August 15th, 2021.

To better study customer animus towards Chinese products, we focus on the face mask, a product that might be sensitive to many customers post-Covid19. We restricted our sample to products listed on Amazon that meets the following three criteria: (1) have the key word "face mask" in product name; (2) have at least 3 ratings; and (3) fall under the "Health and Household" category. We then manually examine these products to exclude irrelevant ones that pass the three filters and further drop the ones with missing key variables.[87] The number of products retained in the final sample comes down to 1400.

Review Information

Keepa.com does not provide data on customer review details. We therefore manually collected all review data that are publicly available to Amazon customers for the products we are tracking. This data includes, for each review, reviewer name, date of review, review headline, review body, review rating, country of the reviewer, and helpful votes.[88] For each product, we also collect the information on top (positive or critical) reviews.[89] For details on Amazon's regulation of consumer reviews, see Data Appendix.

Matching Products with Reviews

We link the products and reviews by ASIN and date. In our sample, the first face mask sold on Amazon which is now still in operation dates back to July 18th, 2011. We restrict our research period from September 1st, 2019 to September 7th, 2020 for ease of comparison before and after the shock of Covid19 and also for consistency of Amazon policy.[90] In the final sample, we have

[87] Typical examples of such products include face mask straps, face mask filters, and ear savers for face mask.
[88] For an existing review, customers could click on "helpful" button to support the review. Reviews with the highest number of helpful votes would be displayed at the top of the review pages under "top positive review" or "top critical review", depending on the rating given by the reviewer.
[89] For most of the products in our data, Amazon listed the top positive and top critical reviews.
[90] Our choice of study period covers the time pre and post the breakout of Covid19 – January 23rd, 2019. This is the time that city of Wuhan (China) is sealed and the time that people worldwide are aware of this virus. Also see https://www.marketplacepulse.com/articles/amazon-replaces-reviews-with-ratings for the change in policy on Amazon on the rating policy since September 2019. Before the change, customers need to leave a comment if they want to leave a rating for a product. After the change, customers can rate a product without leaving a review text. We therefore choose the time period after the change of this policy.

1400 products and over 70,000 reviews. Note that when we collapse the data into panel of date/week by product, the panel is unbalanced due to arrival of new products during the research period. Figure 2.1 shows the arrival of new face masks sold on Amazon by day. There is a rising number of new face masks sold on Amazon after the breakout of Covid19 (Jan. 23rd, 2019), which peaks on June 8th, 2020, and decreases after that.

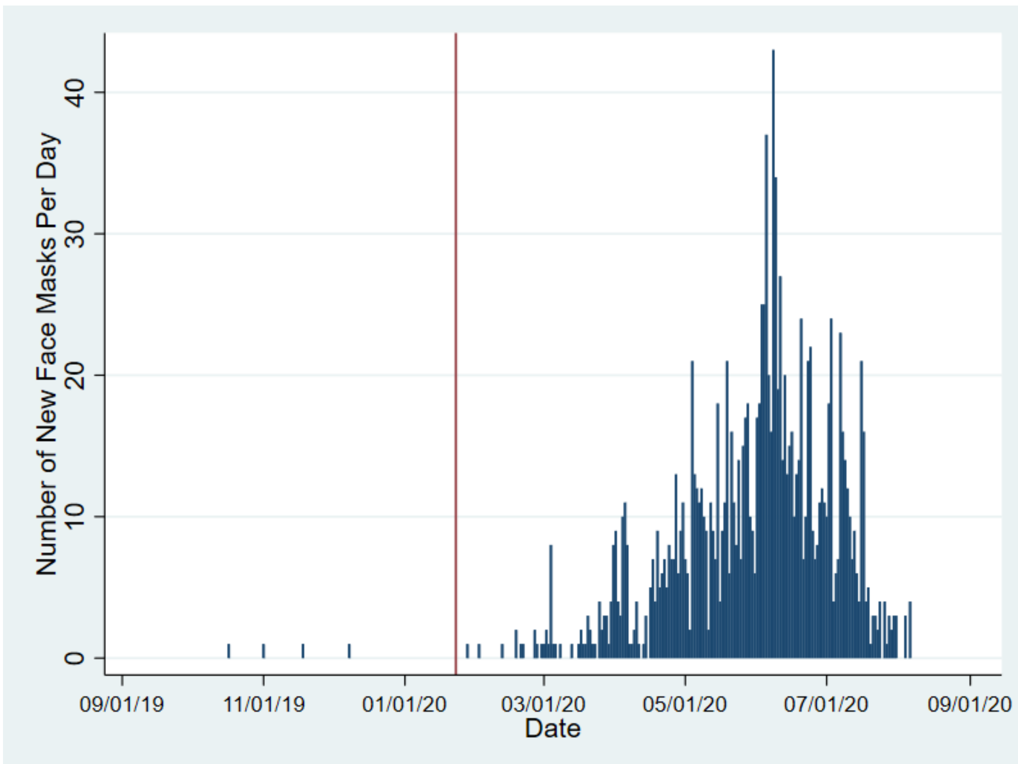

Figure 2.1. Number of New Face Masks Sold on Amazon by Day

2.2.2 Data Treatment

Identify Chinese Products

Amazon does not require sellers to reveal the countries-of-origin for products. Therefore, we collect information on a product's revealed origin country from several sources – product name, product description/details/features, consumer reviews, and customer question and answer (Q&A). All this information is also available to real customers. [91]

[91] Through our analysis, we try to mimic a real customer and avoid using any information which is beyond the reach of real customers.

We first analyze the product name and descriptions/details/feature and collect information on whether a product is made in China. Since this information is provided by the seller herself and does not vary over time, we classify the product as self-identified as Chinese and there is no change in the revealed origin country of the product over time in this case. We then analyze the texts of reviews (including the review headline and review body) and pick out all reviews that mention China.[92] Within these reviews, we manually go through them and collect information on whether a review actually identifies the product as made in China.[93] We also use information from customer Q&A, where typically a customer poses a question and sellers or other customers would leave an answer. We manually collect information on country-of-origin and date that a product is first credibly revealed as Chinese.[94] Note that not all products have the customer Q&A section, but "Is it made in China" is a very common question for most face masks with customer Q&A.[95]

Combining all the information sources above, we are able to identify 386 products made in China out of 1400 face masks as of September 7th, 2021. Since customers might hold animus towards Chinese products, only 24 products are self-identified as Chinese, with 93.8% of the Chinese products either identified from reviews or customer Q&A.

There could be cases where Chinese products are not yet identified as of the last day of our research period, but this case shall not be problematic for our study. We carry out the research trying to look exactly from the view of real consumers and we have access to exactly the same information as real consumers do. If we do not observe a product to be made in China, neither do real consumers. And if consumers do not know that a product is Chinese, they would not have hold animus towards the product either. What matters eventually is not whether a product is actually

[92] Specifically, we picked out all reviews that mention China or Chinese, in upper or lower case. We also tried racial slurs against Chinese (e.g., Ching Chong), but did not spot such cases in our data since Amazon removes comments with offensive language. We find a few reviews which use the expression of "foreign spelling" but we did not count such products as Chinese since it could be from any other non-English-speaking countries. There is also one rare case where customer refers to China as "the communist country". In most of the cases, the use of China/Chinese is adequate to identify Chinese products.
[93] There are a few cases where a review mentions China but does not identify the product reviewed as a Chinese product. E.g., "Overpriced with a bit of price gouging and excessive shipment cost but what can one do these days? At least they are not from China."
[94] If the seller or manufacturer replies and mentions that the product is made in China, we think this information is credible. If it is the consumers that mentioned a product is made in China, we check the answers of other consumers to make sure the information is credible. For credible answers, we then track the first date that it identifies a product as a Chinese product. Typical non-credible answers could include answers like "I guess this is made in China" or competition out of malicious purpose e.g., "Do not buy this product, it is made in China. If you want to buy safe products satisfying FDA standard, you can find them at (links for other sellers)".
[95] In our sample, about 65% of the products has customer Q&A.

made in China, but whether a product is "revealed" or "identified" and known to consumers as Chinese.

Types of Reviews

We analyze the review content (including both the headline and body of a review) and divide reviews into three categories based on the information provided and attitudes expressed towards China or Chinese products: non-informative, informative-neutral, and informative-animus.[96] The non-informative reviews do not provide any information on the Chinese identity of a product. The informative reviews, on the other hand, reveal the Chinese identity of a product. The informative reviews are further divided into "neutral" and "animus" depending on whether expressing animus towards China or Chinese products. For the informative reviews, we also collect information on whether a review expresses any complaint about the quality of the product.

In data, among a total of 70,136 reviews, there are 1201 reviews that mention China/Chinese, within which only 132 reviews are non-informative.[97] Among the 1069 informative reviews, 826 reviews (or 77.3%) express animus towards China or Chinese products, and only 243 reviews (or 22.7%) are neutral. As for product quality, 25.8% (or 276) of the informative reviews contain complaints about quality, while 74.2% (or 793) of informative reviews are not about product quality. See Table B.1 in Appendix for examples of each type of informative reviews. Despite the small number of informative reviews, their impacts could be larger since they catch more attention from other customers reflected by more helpful votes[98]. Within our sample, a review on average gets 2.58 helpful votes. For an informative review, however, this number is 5.84, with informative-neutral reviews get 4.28 helpful votes and informative-animus reviews get 6.30 helpful votes.

Figure 2.2 shows the distribution of ratings within all Chinese products identified by the end of our research period, by non-informative and informative reviews. For the non-informative reviews, 50.04% give 5-star rating while only 20.36% give 1-star rating. This pattern is reversed within informative reviews, where 56.22% give 1-star rating and only 12.07% give 5-star rating.

---

[96] Specifically, we first use keyword of China/Chinese to pick out the reviews which mentions China. Within these reviews, we manually go through them to collect information on three questions: Does this review express animus towards China/Chinese products? Can we identify this product as Chinese? Does this review contain any complaint about quality of the product?
[97] An example of non-informative review that mentions China/Chinese is (review body): "Made in USA and much better than ones we've bought made in China."
[98] By default, customers are more likely to see reviews with more helpful votes since reviews are shown in descending order in terms of number of helpful votes they get. This could be changed by the consumer if she wishes to review the comments in order of date instead of by number of helpful votes.

Within informative reviews, on average, the neutral ones give higher ratings than the animus ones, and the share of animus reviews decreases by ratings. From 1-star to 5-star ratings, the share of animus reviews among the informative reviews are, respectively: 96%; 81%; 64%; 45%; and 23%. Table B.2 in Appendix shows some examples of informative-animus reviews from consumers at each rating.

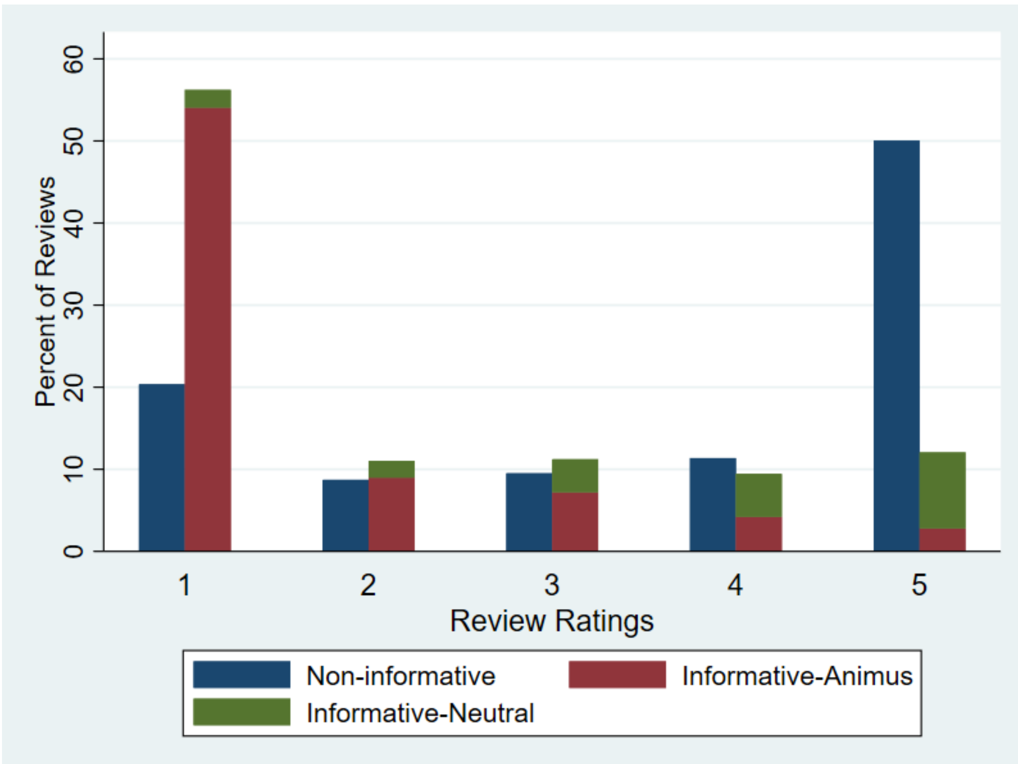

Figure 2.2. Rating Distribution of Chinese Products by Review Category

Note: Sample is limited to Chinese products identified by the end of research period. Reviews are divided into non-informative, informative-animus, and informative-neutral, with the last two categories stacked.

Other Countries of Origin

We apply similar analysis for other countries to collect information on products' countries-of-origin by first searching the country names and then manually analyzing the reviews and customer Q&A to correct mis-information.[99] See Figure 2.3 for summary of country-of-origin.

[99] Specifically, to correct for cases where a country name is mentioned in the review but the review does not identify the country-or-origin of the product; or cases where the customer Q&A mentions the country-or-origin but is not credible. We use the name list of countries from the World Bank to search within the relevant reviews. Considering

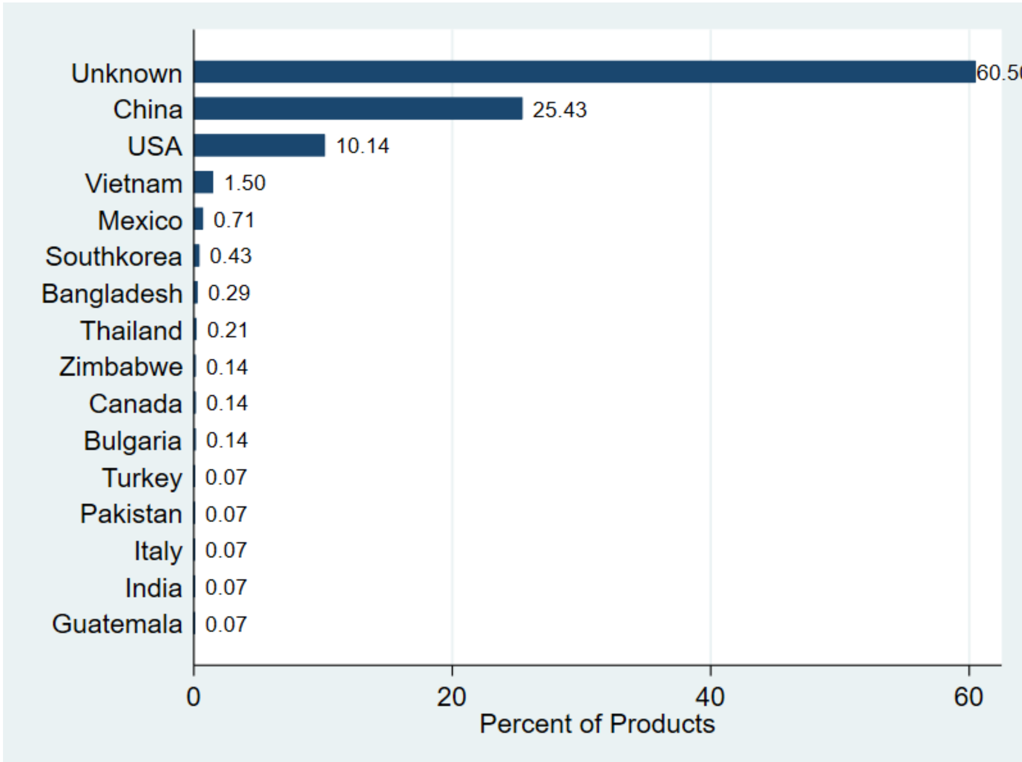

Figure 2.3. Share of Products by Country-of-Origin

By the end of our research period, 60.50% of the products' countries-of-origin remain to be unknown. For the known ones, the face masks sold on Amazon in the U.S. are revealed to come from 21 countries, including U.S. itself. Out of the 1400 face masks in our sample, China is the biggest country-of-origin identified (25.43%), following by USA (10.14%), with the rest of the countries summing up to less than 4% of the face masks.

2.2.3 Summary Statistics on Chinese Products

Table 2.1 below compares the products from China and other countries (including the unknown) at both the first and last day in the research period. Comparing the price, there is no evidence that Chinese products charge lower prices after Covid19. If anything, the average price of face masks sold on Amazon is slightly higher for Chinese ones at the last day of the sample

cases where a country is mentioned in the reviews but under an alternative name, we slightly change the name list to improve the chances a country name is identified. If a product is said to come from Korea without further specification, we will take it as South Korea. For United States, we use the key words of, in upper and lower cases, United States, USA, U.S., and America. A product is only identified as U.S. if it is confirmed in reviews or Q&A as made and shipped within the U.S., and not identified with other countries-of-origin.

period. For average rating, Chinese face masks have similar average rating to those from other countries, both at the start and end of the research period.[100] On average, the Chinese products have more sales than non-Chinese ones, but this gap becomes smaller at the end of research period. Chinese face masks receive more ratings than non-Chinese products, and the number of ratings greatly expanded during the research period for all products.[101] At September 1st, 2019, Chinese sellers are on average in business for shorter time, which is reversed by September 7th, 2020, due to the arrival of new products during the research period.

Table 2.1. Summary Statistics (Mean Only)

	Sep. 1st, 2019		Sep. 07th, 2020	
	Chinese	Others	Chinese	Others
Price	16.27	16.71	18.17	17.84
Average Rating	4.06	4.11	4.06	4.08
Rescaled Sales Rank	55.47	37.65	53.63	48.56
Number of Rating	21.96	3.57	458.43	161.84
Days in Business	697	728	175	113
ASIN count	27	16	396	1014

Note: Summary statistics are based on the first day (September 1st, 2019) and last day (September 7th, 2020) of our sample. Amazon does not provide actual number of sales, and only provides the sales rank under a specific product category. In this table the sales rank is re-scaled from 0 to 100, where a larger index means more sales.

2.3 Event Study

2.3.1 Empirical Model

In this section, we perform an event study on the impact on average rating for a product to be identified as made in China for the first time. Goodman-Bacon (2021) has pointed out problems associated with the inclusion of two-way fixed effects under the DID design when treatment time varies. We therefore apply a fully-dynamic design of event study in this part, as suggested by Borusyak and Jaravel (2017). Regression below specifies the empirical model:

$$Y_{it} = \alpha + \sum_{k=-\infty}^{\infty} \beta_k Treat_{ik} + \beta_X X_{it} + \theta_i + \theta_t + \varepsilon_{it} \tag{2.1}$$

[100] Note that this is not inconsistent with our main results as the effect of identification on average rating is short-lived.
[101] This could be related to the rising of sales of face masks or that the share of consumers who leaves a rating changes overtime. However, since we do not observe the actual number of sales of a product from Amazon, we cannot distinguish between the two cases.

The data are aggregated on week-product level as panel. The dependent variable is the average rating, which varies by product (ASIN) i and time (week) t. $Treat_{ik}$ is a dummy variable, which equals one if at time t, product i is k weeks from first time ever being identified as made in China from reviews. Specifically, $k = t - K + 1$ where K denotes the time when a product is first time ever being identified as Chinese. The time difference k can be positive or negative/zero, depending whether time t is after or before/equals identification time K, and k equals one at the first week a product is ever identified as Chinese. Control variables X_{it} include price and sales rank[102]. θ_i is product fixed effect, which absorbs potential quality difference across products[103]. θ_t is time fixed effect, which absorbs any potential common supply-side (e.g., scarcity of product) or demand-side variations (e.g., change of policy on requirement of wearing face masks). This event study design provides good exogenous variation under the assumption that conditional on time and product fixed effect as well as the price and sales of a product, the arrival of a review that first time reveals the Chinese identity of a product is random.

2.3.2 Empirical Results

Main Result of Event Study

Results of event study are shown in Figure 2.4. There is no evidence of declining trend of average rating before a product is identified as Chinese. However, once a product is identified as made in China, the average rating declines. This drop persists over time and fades out in six months. At the first several weeks (before week -39), there is a large rise in average rating, and in the last several weeks (after week 30), there is a sudden drop of average rating. Both these patterns are a bit inconsistent with the data in between. This may be due to sample size for the first and last several weeks, and these coefficients are driven by just a few products (e.g., for time difference before -41 and after 27 weeks, there are fewer than five products). Appendix Figure B.1 shows the count of products for each time difference k (in weeks). For the first week being even identified as Chinese, the confidence interval is much wider than other weeks close in time. This may be due to

[102] Both controls take logs. The sales ranks are limited under the same category of Health and Household, so they are comparable. The ideal variable would be actual sales but unfortunately Amazon does not provide data on the actual sales of products. We therefore use sales rank to proxy for actual sales on Amazon, with larger rank meaning smaller actual sales amount.
[103] Since ASIN can precisely identify a product even up to color and size, we are confident that the quality of the product under the same ASIN does not vary over time.

collapsing day into weeks, and the first week therefore contains a mixture of "will-be" and "just" treated Chinese products.[104]

The result provides evidence for consumer animus towards Chinese products. The average rating first drops and slowly recovers since the informative reviews affects average rating both directly and indirectly. The direct mechanism refers to the impact of a review on the average rating through its own rating. As shown in Figure 2.2, most of the informative reviews are informative-animus and give low ratings, which directly lower the average rating of a product. The direct impact decreases over time unambiguously.[105] The indirect mechanism refers to the impact of a review on the average rating through ratings given by other (future) customers. Since over time, a growing number of (future) consumers see an existing informative review, the indirect impact of the review first increases. However, with the arrival of new reviews, consumers are also less likely to see an existing informative review due to higher time cost going back to more outdated reviews. Therefore, the indirect impact first grows over time and then decreases. Combining the direct and indirect impact therefore gives us a U-shaped total impact. It is unlikely that this pattern is driven by difference between the product quality of Chinese products compared to others. Since ASIN is very narrowly defined, which ensures the precise tracking of the same face mask, the product fixed effect will absorb any quality difference between products. The actual quality of a product does not change upon the time point that it is identified as Chinese, and as shown in the data section, 74.2% of the informative reviews are not about quality.

In the next section, we will provide more evidence and explanation of the direct and indirect impacts of the informative reviews, as well as having more discussion to distinguish our results from explanation via product quality.

---

[104] To be specific, weeks are collapsed relative to the September 1st, 2019, the start time of the sample. The time in week is therefore "absolute time" that can reflect actual time, and therefore the inclusion of time fixed effect can account for time-varying information such as product scarcity. To be consistent, the "weeks since being identified as a Chinese product" is also measured using this "absolute time" and then subtract the week being first identified and plus one. Therefore, the "first week" will be collapsed within a mixture of will-be-treated and just-treated days. E.g., a product is identified as Chinese in the sixth day of week N, then week N is the first week for identification. However, the average rating (as well as prices and sales rank) in first week contains information of (averaged across) both the will-be-treated (day one to five of the week) and just-treated (day six and seven of the week) days. Therefore, the variance is larger and confidence interval is wider in the first week of identification.

[105] However, we cannot precisely pin down the impact of one extra rating on the average rating of a product due to ambiguity of Amazon's algorithm in calculation. Before 2015, Amazon uses simple/unweighted average to calculate the average rating. In 2015, Amazon switched to a more complicated machine-learning algorithm. See Appendix for more details.

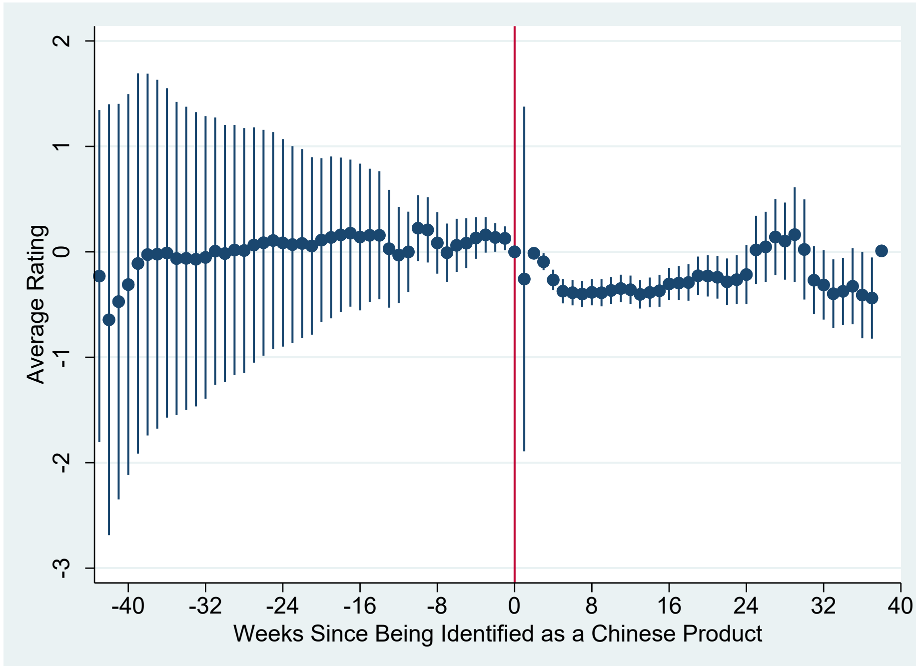

Figure 2.4. Event Study for Chinese Products

Heterogenous Impacts across Subgroups

We then study the heterogenous impacts of consumer animus across high and low reputation Chinese products. Based on the average rating for the week right before a product is identified as Chinese (at $k = 0$), products are divided into high-reputation (above-median) and low-reputation (below-median) groups. Results are shown in Figure 2.5 and 2.6.

The Event Study results among all Chinese products are driven by Chinese products of higher reputation. As shown in Figure 2.5, before being identified, Chinese products with high reputation have a clear increasing pattern. Once identified, its average rating decreases, with the negative impacts fading out within 6 months. For the low-reputation products, however, being identified as Chinese has little impact, since they already had low average rating even before the identification. Similar as above, results at both ends of the time have fewer products and should be interpreted with caution. [106]

[106] Specifically, for the high-reputation group, weeks before -29 and after 24 have fewer than 5 products. For the low-reputation group, weeks before -37 and after 23 have fewer than 5 products.

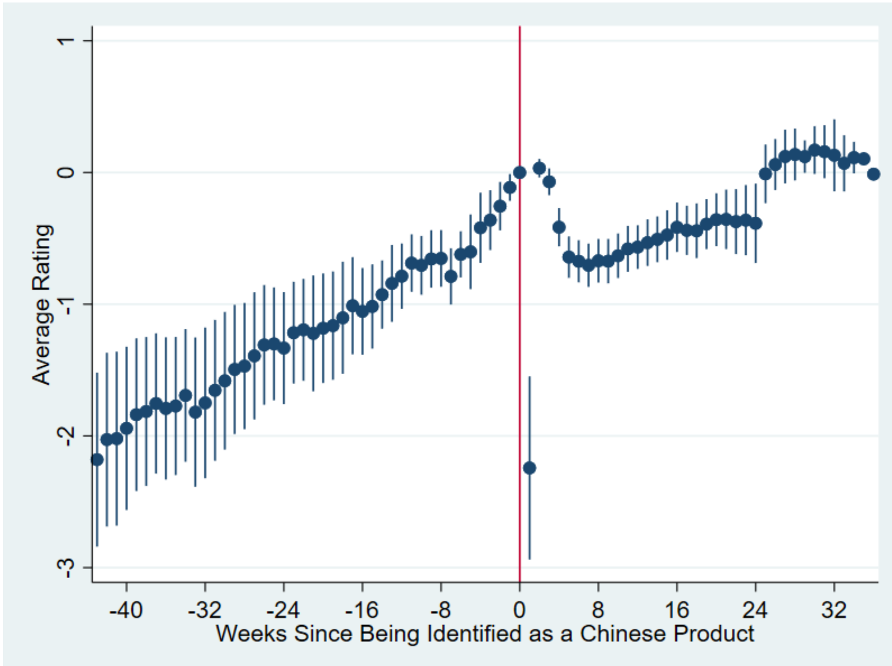

Figure 2.5. Event Study for Chinese Products – High Reputation

Note: Chinese products are divided into high-reputation (above-median) and low-reputation (below-median) ones depending on their average rating at the week right before a product is identified as Chinese.

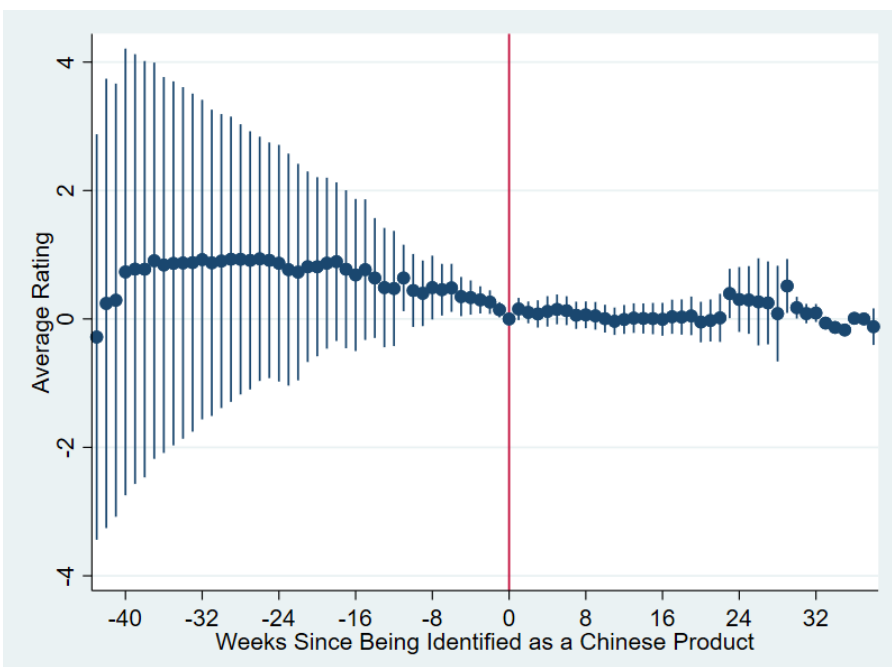

Figure 2.6. Event Study for Chinese Products – Low Reputation

Note: Chinese products are divided into high-reputation (above-median) and low-reputation (below-median) ones depending on their average rating at the week right before a product is identified as Chinese.

Placebo Using Other Countries

We perform the same event study using, respectively, products from U.S. and all other countries (excluding U.S. and China). Placebo results are shown in Figure 2.7 and 2.8.

There is no similar pattern of a drop in average rating for U.S. and other countries-of-origin products as the Chinese ones. If anything, there is a declining trend in the average rating for these products, which stops when their countries-of-origin are identified. The placebo result further supports our interpretation that consumers hold animus towards China and Chinese products. Note that as shown in Figure 2.7 and 2.8, since only 10.14% products are U.S. and less than 3.93% of the products have other countries-of-origin (besides China and U.S.), the sample sizes of the placebo tests are also much smaller than the main result. [107]

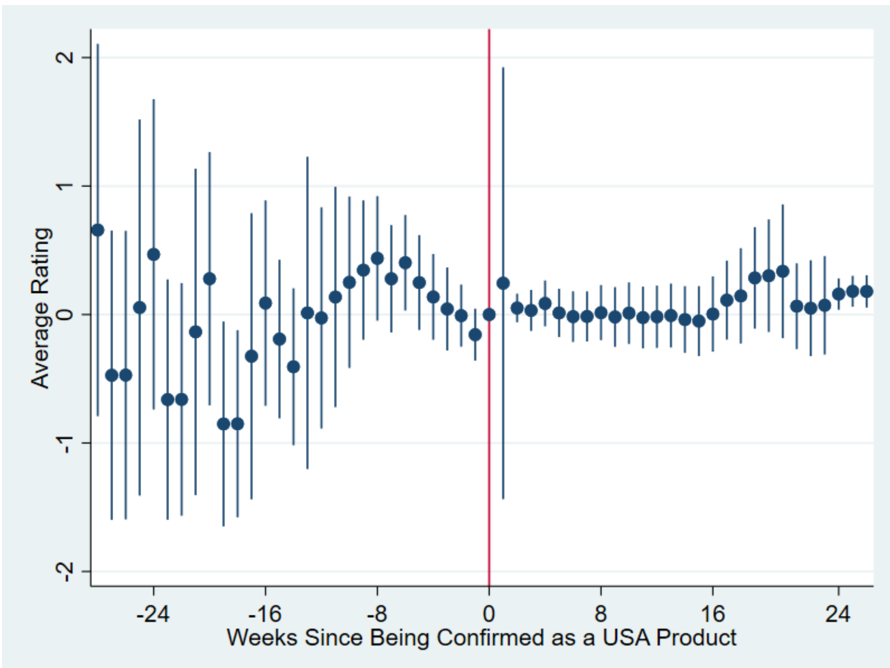

Figure 2.7. Placebo Event Study of Non-Chinese Products – U.S.

Note: Non-Chinese products are further divided into U.S. products and products from all other countries besides China and U.S. A product is classified as U.S. only when it is both identified/confirmed as U.S. and there is no information suggesting the product has other countries-of-origin in the reviews.

[107] For U.S. products, weeks before -15 and after 22 have fewer than 5 products. For other countries' products, weeks before -3 and after 10 have fewer than 5 products.

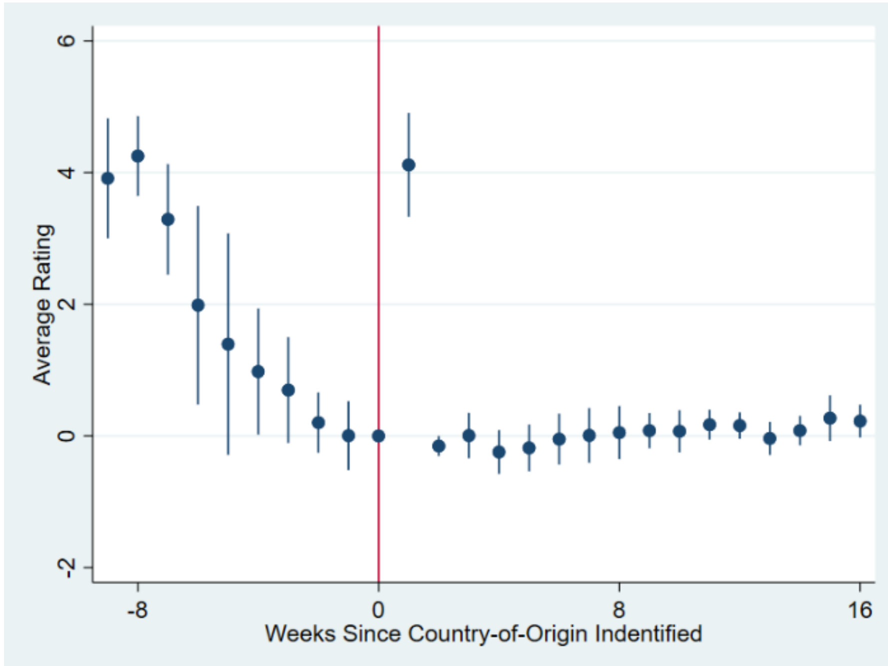

Figure 2.8. Placebo Event Study of Non-Chinese Products – Other Countries

Note: Non-Chinese products are further divided into U.S. products and products from all other countries besides China and U.S.

Robustness Checks

We carry out two robustness exercises. First, we further combine information from customer Q&A in identifying the time that a product is ever identified as Chinese. Results are in Figure 2.9. The pattern is similar to our main result, with drop in average rating after a product being identified as Chinese for the first time, and this negative impact fades out within 6 months. Compared to the main result, however, there is decreasing trend of average rating before the identification.

For the second robustness check, we use the cumulative share of 1-star rating reviews as the dependent variable, instead of product average rating. Results are in Figure 2.10. There is a rise in the cumulative shares of 1-star rating reviews after a product is identified as made in China for the first time, which fades out in 6 months. However, there is also pre-trend of increasing share of 1-star rating reviews before the identification.

Overall, results are similar to the main analysis, but with some pre-trends.

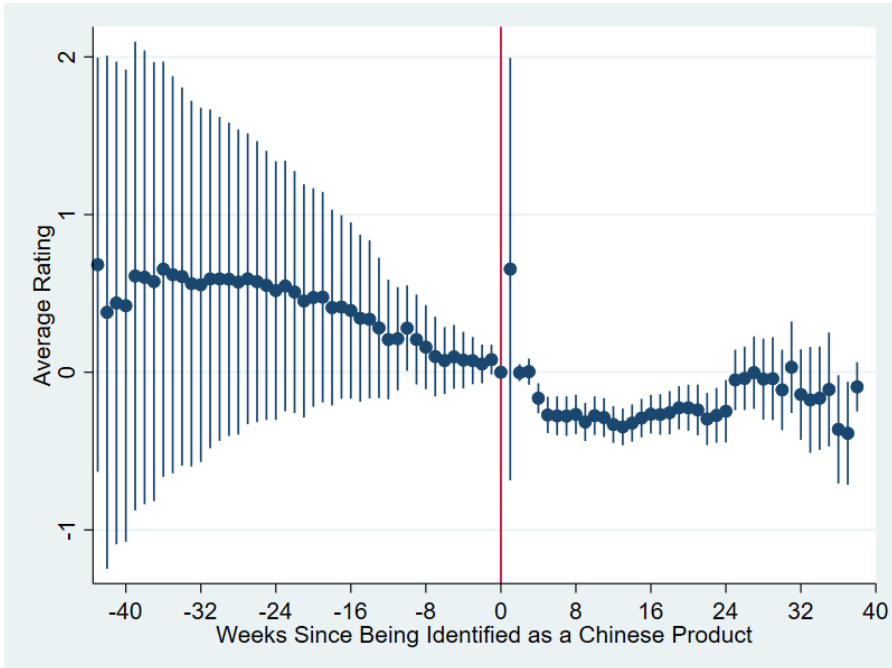

Figure 2.9. Event Study for Chinese Products – Combining Reviews and Customer Q&A

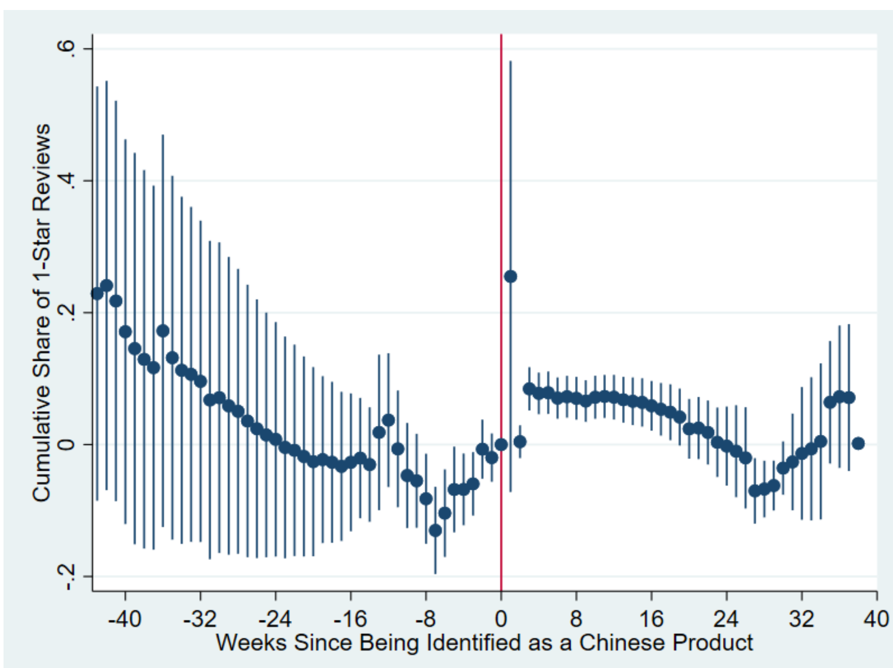

Figure 2.10. Dependent Variable Using Share of 1-Star Reviews

Event Study on Price

We carry out the same analysis on price. The dependent variable is now the log of price, and controls are now average rating and sales. Product and time fixed effects are included as in the main analysis.

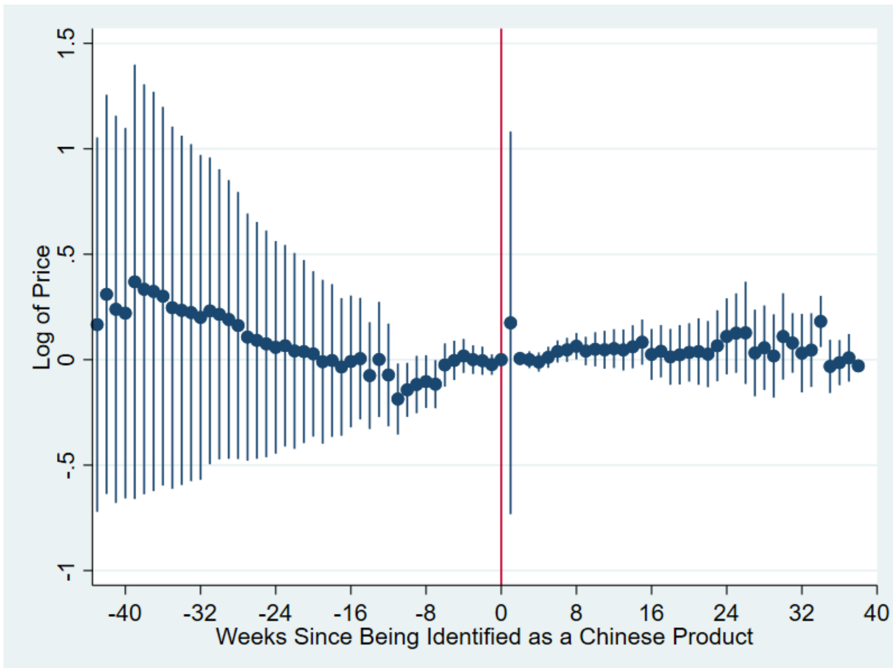

Figure 2.11. Event Study for Chinese Products – Dependent Variable Using Log of Price

As shown by results in Figure 2.11, there is no drop in price as we found in average rating. There is no evidence that sellers lower the price in response to products being identified as made in China. This is different from the finding in Lucking-Reiley et al. (2007) about the measurable effect of seller's rating on the prices on eBay. The difference in findings might relate to the product scarcity of face masks, especially during the first several months after the breakout of Covid19. Due to scarcity, consumers have limited ability to switch away from a Chinese face mask even if they hold animus towards it, and sellers therefore do not have the incentive to lower the price. Cabral and Xu (2021) studied the price gouging behavior of 3M masks sellers on Amazon due to product scarcity, and found that between mid-January to mid-March 2020, these masks charge 2.72 times higher price than Amazon sold them in 2019. We will provide more evidence in the empirical part below on product scarcity and consumer's response to it.

2.4 Direct and Indirect Impacts

As explained in Section 2.3, an informative review can have direct and indirect impacts on the product average rating. In this section, we will provide more discussion and evidence on these direct and indirect impacts. We will also provide more evidence on the animus consumers hold towards Chinese products, further distinguishing our results from the quality story, and discuss the logic behind consumers' purchasing behavior with rising animus.

2.4.1 Impacts of Informative Reviews – Direct and Indirect

There are two mechanisms by which an informative review can affect the average rating of a product. The direct mechanism specifies that an informative review affects the average rating of a product by directly going into the calculation of the average rating [108]. The indirect mechanism specifies the impact an informative review has on the average rating by affecting the (future) ratings left by other consumers.

For the direct impact, as shown in the data section, since a much higher share of informative reviews give 1-star ratings, they have an unambiguous negative direct impact on the average rating. In the Appendix, we show that conditional on the price, sales, average rating, as well as the product and time fixed effect, the ratings are 1.35-star lower for the informative reviews and 1.82-star lower for the informative-animus reviews, which is a large difference for a rating system that ranges between one and five. The direct impact decreases over time with arrival of new reviews and the informative review being more outdated. [109] In the Appendix, we also show that the product average rating is highly correlated over time in our sample (correlation coefficient equals 0.97). [110] Through the high correlation of average rating, the negative direct impact persists, but decreases in size over time.

For the indirect impact, an informative review can affect the future ratings via other (future) reviewers who read the existing informative review. Previous researchers have found similar "indirect impacts" where one negative opinion invites another. Cabral and Hortacsu (2010) finds

[108] To be specific, the rating of the review is used in the calculation of the average rating.
[109] Essentially, it is the weight assigned to the informative review in calculation of average rating decreases over time. However, since Amazon uses a machine-learning algorithm of calculating product average rating after 2015, and that Amazon might have more information than what is publicly available to a customer, we cannot specify the change of this weight overtime.
[110] This controls for price, sales rank, product fixed effect, and time fixed effect. The correlation is 0.98 without these controls.

that after a seller receives the first negative feedback, subsequent negative feedback arrives 25% more rapidly than the first one, leading to an increase in the negative feedback rate. Moe and Trusov (2011) also find that the previously posted ratings significantly affect future rating behavior. Specific to animus, He et al. (2021) finds users exposed to hateful contents on Twitter and highly likely to become hateful. There could be several cases for the indirect impact to work. First, an informative review provides information which might not be previously known or noted by a consumer.[111] A discriminative consumer could choose to leave a low rating for the product after knowing a product is Chinese from an informative review. The informative review does not need to be the first review ever to identify the Chinese product, since consumers will pay higher time cost to find more outdated reviews.[112] Second, since 77.3% of the informative reviews express animus towards China or Chinese products, this might increases the likelihood that a future consumer leaves a low rating out of animus either due to increase in animus towards China or decrease in cost of expressing animus. There is no evidence that consumers are more likely to leave a review (informative or not) after the product is identified as made in China. If anything, the average ratio of number of reviews over number of ratings see a small decreases for a Chinese product after the identification (see Appendix).[113] In both the "extra information" case and "higher likelihood of expressing animus" case, the future consumer could also choose to click on "Helpful" to increase the helpful votes for the informative review. The increase of number of helpful votes would affect the weights assigned to the informative review in calculation of average rating, as well as increase the probability that another future consumer will see this existing informative review.[114]

Comparing the direct and indirect mechanism, the direct mechanism is a "one-time shock" which persists over time via correlation of average rating overtime; the indirect mechanism,

[111] There could be cases when a product is already self-identified as Chinese, or that previous reviews or customer Q&A has already mentioned the product is made in China, but a consumer does not notice the information.
[112] A consumer can choose to look at the reviews sorted by date so that more outdated reviews might have a higher cost to be seen by a future consumer. As regulated by Amazon, each review page only contains 10 reviews. We consulted the customer service and up to our knowledge, there is no method that a customer can adjust the number of reviews per page shown to her.
[113] However, there is no significant decrease observed for products of other countries-of-origin.
[114] See https://www.feedbackwhiz.com/blog/how-does-amazon-calculate-product-ratings/ for what might affect the weight assigned to a review rating to the calculation of average rating of a product. These factors might include, verified purchase, number of helpful votes, age of the review, and the richness and length of the review text.

however, has "several shocks" with more and more future consumers seeing the informative reviews.

2.4.2 Empirical Model and Results

To provide empirical evidence on the change of direct and indirect impact overtime, we use the empirical models below:

$$Y_{it} = \alpha + \beta S_{i,t-m} + \beta_X X_{it} + \theta_i + \theta_t + \varepsilon_{it} \tag{2.2}$$

$$Y_{it} = \alpha + \beta_1 S_{i,t-m} + \beta_2 Y_{i,t-m} + \beta_X X_{it} + \theta_i + \theta_t + \varepsilon_{it} \tag{2.3}$$

To better observe the variation of impacts over time, data here is collapsed into product by day panel (instead of week as in the event study).[115] The dependent variable here is the product average rating, which varies by product i and day t. $S_{i,t-m}$ denotes the share of informative (or informative-animus) reviews of product i on day t-m ($m \in [0, t-1]$), which is the m-day lag of share of informative reviews. $Y_{i,t-m}$ denotes the m-day lag of the product average rating. Similar as in the event study analysis, controls contain log of price and sales rank, and we include the product and time (day) fixed effect.

Comparing these two specifications, β captures the total impact of informative reviews, both directly and indirectly; β_1 captures only the indirect impact, since $Y_{i,t-m}$ has controlled for the direct impact which persists via the high correlation of average rating. Results using informative reviews are shown in Figure 2.12 and 2.13, with Figure 2.12 showing total impacts and Figure 2.13 showing indirectly impacts.[116]

As explained above, the total impact is U-shaped over time, which size first expands over the first week and then gradually shrinks (at a faster speed during week 2 to 7, and then at a slower speed). The indirect impacts show a similar pattern as the total impact, but of a smaller size. In first week, the indirect impact expands with many more future consumers seeing the informative review; then during the second to fourth week, the indirect impact remains at similar size with

---

[115] Note that in the event study, we only care for the first time ever a product is identified as Chinese and collapsing into weeks does not really lose much information. Here, however, we take into account all informative reviews, and daily data will give us more variation.
[116] Since data is now collapsed into product-day (instead of week), I only show the first 12 weeks here for clarity of data.

more future consumers continuing to see the informative review; after the fourth week, the size of the indirect impact shrinks overtime due to fewer and fewer future consumers see the review due to rising time cost of going through the product review pages. This provides supporting evidence of our story of how an informative review affects future consumers over time. In Appendix Figure B.2 and B.3, we carry out the same analysis using the share of informative-animus reviews. The patterns are similar but that the impacts of informative-animus reviews are more negative. This is because, on one hand, the informative-animus reviews give on average lower ratings, and thus have larger direct impact; on the other hand, the indirect impact is also larger, since (as discussed above), the informative-animus reviews can work through the "higher likelihood of expressing animus" besides the "extra information" indirect mechanism. This explanation is supported by larger (in size) indirect impacts of informative-animus reviews (Figure B.3) compared to the informative ones (Figure B.2).

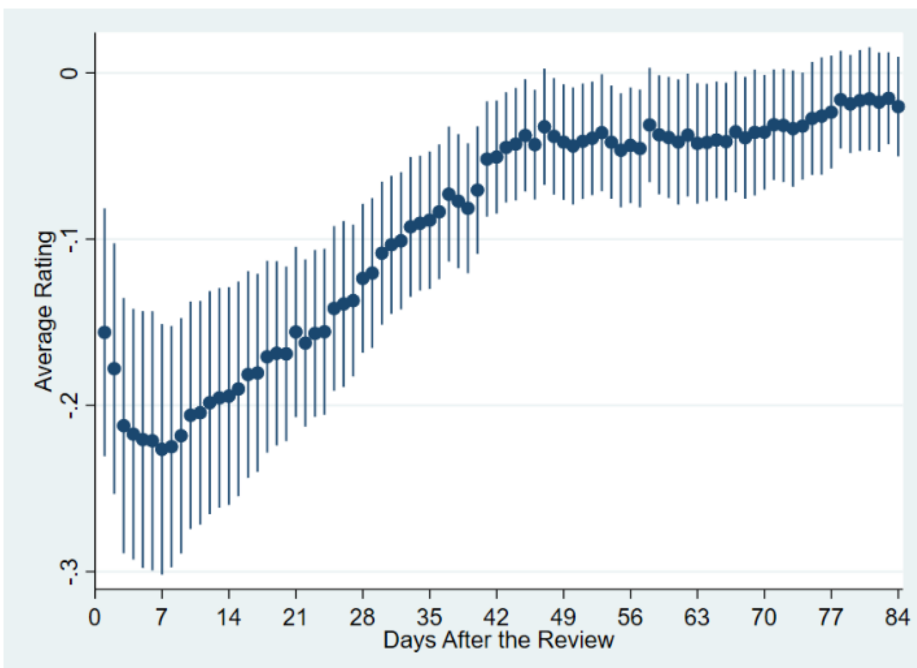

Figure 2.12. Average Rating and Share of Informative Reviews (Total Impact)

Note: Standard errors clustered by ASIN. Controls include price, sales rank, and lag of share of informative-animus reviews, with the x-axis showing the days for this lag.

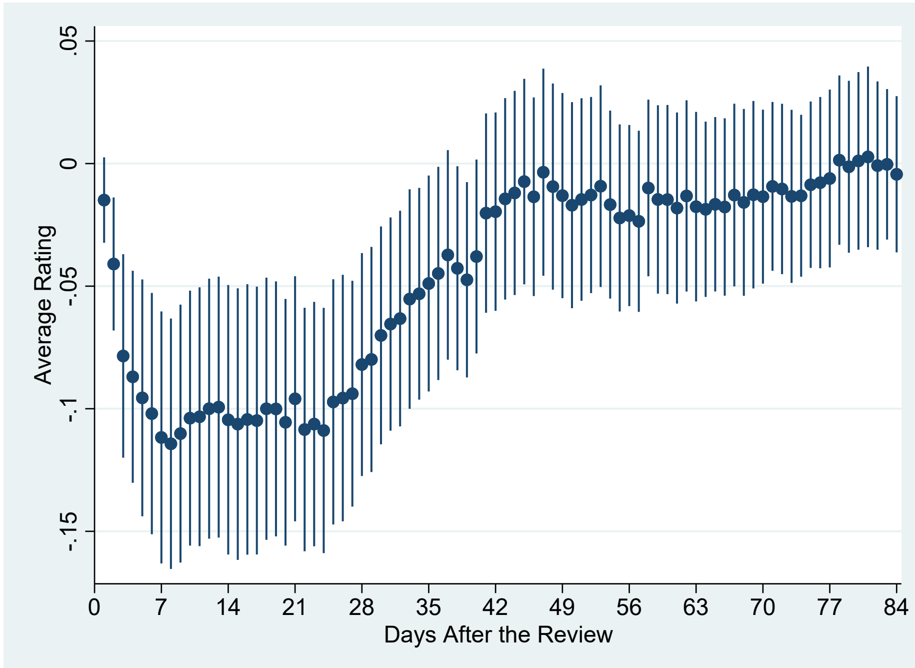

Figure 2.13. Average Rating and Share of Informative Reviews (Indirect Impact)

Note: Standard errors clustered by ASIN. Controls include price, sales rank, lag of share of informative-animus reviews, and corresponding lag of average rating, with the x-axis showing the days for this lag.

2.4.3 Animus versus Quality

One might worry that instead of the animus towards China, the negative impact of informative reviews is driven by the lower quality of products made in China. As we have shown in all previous results, this is unlikely to be the case since the quality difference has already been controlled by the inclusion of product fixed effect. In this part, we provide further evidence against the quality explanation.

For the informative reviews, we analyze the texts of reviews and collect information on whether a review contains any complaint about the quality of the product. As stated in the data section, 74.2% of informative reviews are not about product quality. To show that animus instead of product quality drives the story, we now carry out the analysis again using the share of informative-animus reviews that do not contain any complaint about the quality of the product. Results are shown in Figure 2.14 and 2.15, which is very similar to the pattern using all informative-animus reviews (in Appendix). If anything, using informative-animus reviews that do not contain any quality complains give slightly larger-size coefficients at the peak than using all informative-animus reviews. These results give strong evidence that our findings in this paper are not driven by the quality story, but rather show consumer animus towards China and Chinese products.

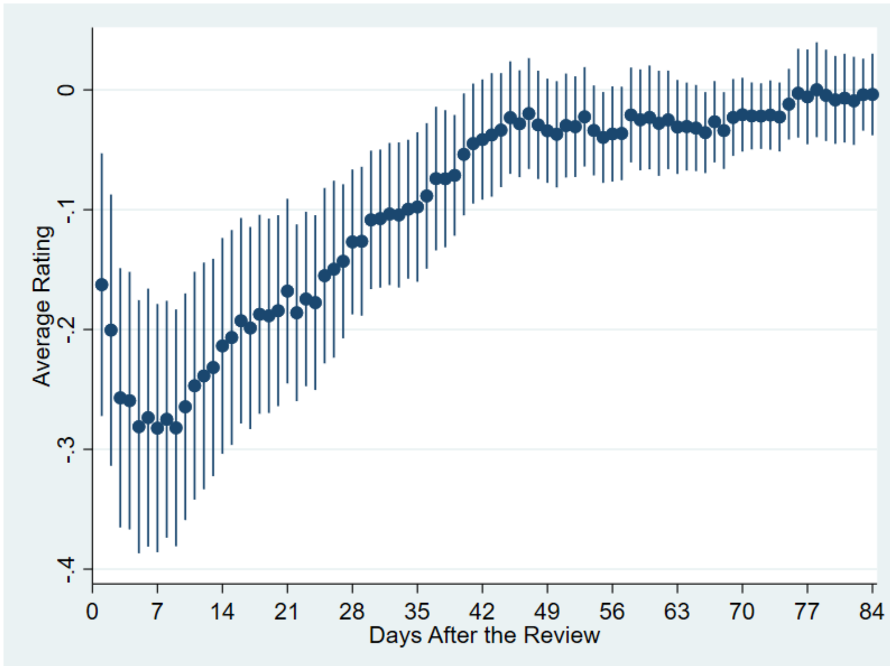

Figure 2.14. Average Rating and Share of Informative-Animus Reviews with No Quality Complaint (Total Impact)

Note: Standard errors clustered by ASIN. Controls include price, sales rank, and lag of share of informative-animus reviews that do not contain any complaint about product quality, with the x-axis showing the days for this lag.

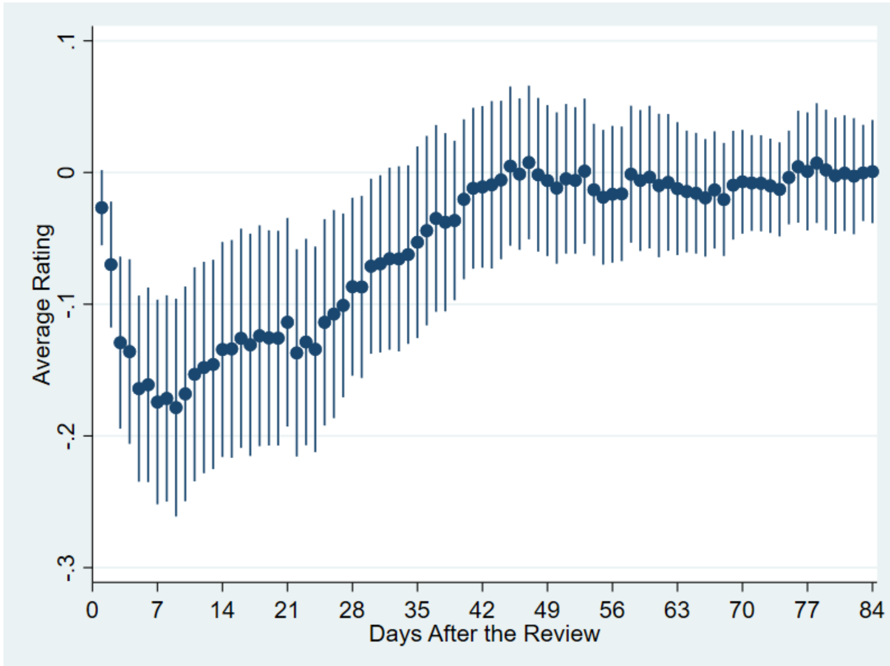

Figure 2.15. Average Rating and Share of Informative-Animus Reviews with No Quality Complaint (Indirect Impact)

Note: Standard errors clustered by ASIN. Controls include price, sales rank, lag of share of informative-animus reviews that do not contain any complaint about product quality, and corresponding lag of average rating, with the x-axis showing the days for this lag.

2.4.4 Animus and Purchase – Information Friction and Product Scarcity

It is an interesting question why consumers buy the Chinese products at first place if they hold animus towards them. Why do they choose to buy a Chinese product and then leave a low rating or a review expressing the animus, instead of simply buying a non-Chinese product? There are mainly two reasons behind it: information friction and product scarcity.

The information friction explanation states that the consumers simply do not know that a product is made in China before they come across the informative review. This could either because there is no information revealing the Chinese identity of the product; or that there is such information, but the consumer missed it due to the time cost of collecting country-of-origin information for a product. As a piece of evidence for information friction, among the 386 Chinese products identified as of September 7th, 2021, only 24 products are self-identified. The product scarcity explanation states that due to the scarcity of face masks, especially during the first several months after the breakout of Covid19, there is limited choice on Amazon, and consumers are restricted from switching away from Chinese products. Cabral and Xu (2021) recorded and studied such scarcity in 3M face masks and price gouging behavior of sellers on Amazon mid of January to mid of March, 2020. In our sample, there are only 43 products at the start of the research period, with 62.8% of them made in China. Appendix Figure B.1 shows the arrival of new products by day within the research period. There is also supporting evidence for these two explanations from the reviews (see Table B.3 in Appendix for some typical examples).

2.5 Conclusion

Covid19 has tremendously affected all areas of our lives and our online shopping behaviors have not been immune. China is the first country to report cases of Covid19, and suffers from rising consumer animus towards its products, either out of prejudice or health concerns. Amazon, the largest online shopping platform, has witnessed this rising consumer animus.

In this paper, we provide evidence of this rising consumer animus towards Chinese products post Covid19 and study its impact on the product average rating on Amazon. We collect information on all face masks sold on Amazon between September 1st, 2019 and September 7th, 2020, including the consumer reviews.[117] Using the same information that is available to a real consumer, including seller-generated and user-generated information, we identify the countries-of-origin of the products.[118] By analyzing the text of reviews, we further divide reviews into non-informative and informative ones, depending on whether it identifies a product to be made in China. The informative reviews are then divided into animus and neutral ones, depending on whether it expresses animus towards China or Chinese products.

Under a fully-dynamic event study design, we find that, despite no change in quality, the average rating of a product drops after being identified as Chinese for the first time. This negative impact is U-shaped, which quickly expands in the first five weeks, and gradually fades out within six months. By further splitting Chinese product into high and low reputation using its average rating before the identification, we find that the U-shape decline in product average rating is driven

[117] Specifically, all face masks that meets the three filtering criteria: (1) product name contains “face mask” (2) at least three ratings (3) under “Health and Household” category.
[118] Specifically, the seller-generated information includes product name, product features, and product description or details. The user-generated information here refers to the consumer reviews. We also use information from customer Q&A, which can either be seller-generated or user-generated information since both sellers and users can respond to a raised question.

by the high reputation ones. Similar patterns are not found among products made in the U.S. or other countries-of-origin.

The negative impact of the informative reviews can be explained by the direct (via its own rating) and indirect (via ratings given by other future consumers) mechanism. The direct impact persists overtime through the high correlation of product average rating from day to day, and decreases in size unambiguously over time. The indirect impact is U-shaped, which first expands in size with more future consumers seeing the review, and then shrinks in size with rising time cost for the review to be seen and thus to affect fewer new consumers. The explanations via direct and indirect mechanisms are then supported by studying the negative impacts (lagged) informative reviews have on product average rating. Our results are not driven by quality difference between Chinese and non-Chinese products, which is potentially controlled by the product fixed effects. We provide further against the quality story by analyzing the informative reviews and collect information on whether it contains any complaint about product quality. Results on total impacts and indirect impacts are very similar using the informative-animus reviews without quality complaints.

The findings in this paper about consumer animus towards China and product average rating on Amazon provides another dimension to look at the impact of the rising animus towards China in the U.S., besides what is recorded in the literature (e.g., Hahm et al., 2021; Lu and Sheng, 2020; Amuedo-Dorantes et al., 2021). The impact might go beyond product average rating to affect the profits (e.g., via price and sales) of Chinese sellers on Amazon with product scarcity being less of a concern over time, and that consumers can more easily switch away from Chinese products. This is supported by literature, where researchers study the impact of reviews on product demand, price, and revenue (e.g., Chevalier and Mayzlin, 2006; Luca, 2016; Cabral and Hortacsu, 2010; Lucking-Reiley et al., 2007). The negative impacts of reviews with animus provides support for platforms of online retailers on screening of reviews, e.g., Amazon removes reviews with offensive language. Our study also bears realistic meanings and can be extended to other political or health events which might increase consumer animus towards products associated with a certain country or region, such as the Russo-Ukrainian War and the following boycott of Russian products.

2.6 References

Adams-Prassl, A., Boneva, T., Golin, M., & Rauh, C. (2020). Inequality in the Impact of the Coronavirus Shock: Evidence from Real Time Surveys. *Journal of Public economics*, *189*, 104245.

Ahmed, A. M., & Hammarstedt, M. (2008). Discrimination in the Rental Housing Market: A Field Experiment on the Internet. *Journal of Urban Economics*, *64*(2), 362-372.

Alon, T., Doepke, M., Olmstead-Rumsey, J., & Tertilt, M. (2020). *The Impact of COVID-19 on Gender Equality* (No. w26947). National Bureau of Economic Research.

Amuedo-Dorantes, C., Borra, C., & Wang, C. (2021). Asian Discrimination in the Coronavirus Era: Implications for Business Formation and Survival.

Ayres, I., Banaji, M., & Jolls, C. (2015). Race Effects on eBay. *The RAND Journal of Economics*, *46*(4), 891-917.

Bartoš, V., Bauer, M., Cahlíková, J., & Chytilová, J. (2021). Covid-19 Crisis and Hostility Against Foreigners. *European Economic Review*, *137*, 103818.

Becker, G. S. (2010). *The Economics of Discrimination*. University of Chicago press.

Borjas, G. J., & Cassidy, H. (2020). *The Adverse Effect of the COVID-19 Labor Market Shock on Immigrant Employment* (No. w27243). National Bureau of Economic Research.

Borusyak, K., & Jaravel, X. (2017). Revisiting Event Study Designs. *Available at SSRN 2826228*.

Cabral, L., & Hortacsu, A. (2010). The Dynamics of Seller Reputation: Evidence from eBay. *The Journal of Industrial Economics*, *58*(1), 54-78.

Cabral, L., & Xu, L. (2021). Seller Reputation and Price Gouging: Evidence from the COVID-19 Pandemic. *Economic Inquiry*, *59*(3), 867-879.

Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, *43*(3), 345-354.

Couch, K. A., Fairlie, R. W., & Xu, H. (2020). Early Evidence of the Impacts of COVID-19 on Minority Unemployment. *Journal of Public Economics*, *192*, 104287.

Croucher, S. M., Nguyen, T., & Rahmani, D. (2020). Prejudice toward Asian Americans in the COVID-19 Pandemic: The Effects of Social Media Use in the United States. *Frontiers in Communication*, 39.

Doleac, J. L., & Stein, L. C. (2013). The Visible Hand: Race and Online Market Outcomes. *The Economic Journal*, *123*(572), F469-F492.

Edelman, B. G., & Luca, M. (2014). Digital Discrimination: The Case of Airbnb.com. *Harvard Business School NOM Unit Working Paper*, (14-54).

Faulkner, J., Schaller, M., Park, J. H., & Duncan, L. A. (2004). Evolved Disease-Avoidance Mechanisms and Contemporary Xenophobic Attitudes. *Group Processes & Intergroup Relations*, *7*(4), 333-353.

Ferrara, A., & Fishback, P. V. (2020). *Discrimination, Migration, and Economic Outcomes: Evidence from World War I* (No. w26936). National Bureau of Economic Research.

Goodman-Bacon, A. (2021). Difference-in-Differences with Variation in Treatment Timing. *Journal of Econometrics*, *225*(2), 254-277.

Hahm, H. C., Xavier Hall, C. D., Garcia, K. T., Cavallino, A., Ha, Y., Cozier, Y. C., & Liu, C. (2021). Experiences of COVID-19-Related Anti-Asian Discrimination and Affective Reactions in a Multiple Race Sample of US Young Adults. *BMC Public Health*, *21*(1), 1-11.

Hanes, E., & Machin, S. (2014). Hate Crime in the Wake of Terror Attacks: Evidence from 7/7 and 9/11. *Journal of Contemporary Criminal Justice*, *30*(3), 247-267.

He, B., Ziems, C., Soni, S., Ramakrishnan, N., Yang, D., & Kumar, S. (2021, November). Racism is a Virus: Anti-Asian Hate and Counterspeech in Social Media during the COVID-19 Crisis. In *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 90-94).

Ivandic, R., Kirchmaier, T., & Machin, S. J. (2019). Jihadi Attacks, Media and Local Hate Crime.

Kakar, V., Voelz, J., Wu, J., & Franco, J. (2018). The Visible Host: Does Race Guide Airbnb Rental Rates in San Francisco?. *Journal of Housing Economics*, *40*, 25-40.

Kaushal, N., Kaestner, R., & Reimers, C. (2007). Labor Market Effects of September 11th on Arab and Muslim Residents of the United States. *Journal of Human Resources*, *42*(2), 275-308.

Moe, W. W., & Trusov, M. (2011). The Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing Research*, *48*(3), 444-456.

Lu, R., & Sheng, S. Y. (2020). How Racial Animus Forms and Spreads: Evidence from the Coronavirus Pandemic.

Luca, M. (2016). Reviews, Reputation, and Revenue: The Case of Yelp.com. *(March 15, 2016). Harvard Business School NOM Unit Working Paper*, (12-016).

Lucking-Reiley, D., Bryan, D., Prasad, N., & Reeves, D. (2007). Pennies from eBay: The Determinants of Price in Online Auctions. *The Journal of Industrial Economics*, *55*(2), 223-233.

Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *The American Economic Review*, *62*(4), 659-661.

Tjaden, J. D., Schwemmer, C., & Khadjavi, M. (2018). Ride with Me — Ethnic Discrimination, Social Markets, and the Sharing Economy. *European Sociological Review*, *34*(4), 418-432.

CHAPTER 3. LEARNING FROM INTERNATIONAL TRADE: ASYMMETRIC CULTURAL TRANSMISSION AND GENDER DISCRIMINATION

3.1 Literature Review

To the best of my knowledge, the closely related papers are Tang and Zhang (2017), Bøler et al. (2018), Zhang and Dong (2008), Dong and Zhang (2009), Black and Brainerd (2004), and Maystre et al. (2014).

Tang and Zhang (2017) is the closest paper to mine, studying how FDI affects the employment of female workers through cultural transmission based on Chinese experience [119]. One important difference is that here I propose and prove the asymmetric transmission of culture, which is the key of why trade can alleviate discrimination. In their paper, however, this asymmetry is not discussed. Comparably speaking, the countries having FDI in China tend to be concentrated in countries with better gender-equal culture than China while for export and import, the destination and origin countries are more dispersed in their gender-equal culture [120]. Another difference is that I model cultural transmission as update of information compared to spread of taste in their paper [121].

Bøler et al. (2018) studies how flexibility of working hours affects the relative wage for the exporters in Norway, but in many of the specifications, they control for the Gender Gap Index, which is also used in my paper. In their paper, they control GGI to consider the difference in in women's social status between Norway and the destination country in case this makes female employees in Norway harder to do business with customers in those destination countries. Our papers are different in the following aspects. Firstly, their paper stresses the flexibility of working hours as the reason for wage difference while mine stresses the gender discrimination as the underlying reason. Secondly, their paper only studies relative wage while mine studies both relative wage and employment, thus having a more complete picture of female workers' labor market outcomes. Thirdly, I propose and proves the asymmetric transmission of culture. In their paper, GGI is never significant while in my paper I find strong evidence that culture transmission

[119] For a very similar paper studying transmission of culture via FDI and impact on gender outcomes, but based on Japanese experience, see Kodama et al. (2016).
[120] For comparison of the FDI source country and trade partner countries, see Table C.1 in the appendix.
[121] Though they mentioned that they also build a statistical discrimination model that has the same implications as the taste-based one, which would be available upon request.

matters. The insignificance of cultural transmission in their results is exactly explained by the asymmetry of cultural transmission, which is established in my paper [122]. Lastly, their study is restricted within exporters but my paper also studies the spillover effects to the non-exporters, and I further distinguish the impacts on the direct exporters from the non-direct exporters.

Zhang and Dong (2008) and Dong and Zhang (2009), they compared female workers' productivity-adjusted relative wage across different firm types in China. Compared to my paper, they only find that exporters pay female workers higher relative wage without explaining why this is the result. They do not specify cultural transmission as the potential mechanism as in my paper, and they simply compared exporters to other type of firms without further disguising the destination countries, which is done in my paper. Another drawback of their paper is that the sample size is too small so the result might suffer from limited representativeness and estimation power [123], while the sample used in my paper is much larger.

Black and Brainerd (2004) tests one of the Becker (1957) predictions. It shows that faced with trade shocks, the residual gender wage gap in concentrated industries decreases more than that in the competitive industries. The mechanism they test in their paper is that competition can limit employers' ability to discriminate female workers, while in my paper trade affects the labor market through transmission of culture. Both of our papers study how trade affects the labor market, but the underlying mechanisms in two papers are completely different [124]. Moreover, their study is on industry level, comparing concentrated industries with competitive industries, while my study is on firm level, making within-industry comparisons.

Maystre et al. (2014) studies trade's role in the transmission of culture [125]. But compared to my paper, their paper focuses on how trade liberalization biases people's preference while mine stresses trade's impact on the labor market. They study how culture changes and they stop there, but in my paper, change of culture is only the underlying mechanism, and I care more about how the change of culture further affects the labor market. The important result of their paper is that trade liberalization will shift local culture to global ones, thus leading to cultural convergence. My

[122] This is because Norway is one of the best countries in respecting women, ranking 2nd among the 115 countries in GGI in 2006. As a result, most countries have a worse gender-equal culture than Norway and thus cultural transmission does not take place. Comparably speaking, China is a better country to study asymmetric of cultural transmission since it ranks 63 among the 115 countries surveyed, having many countries on both sides.
[123] Their sample only includes about 2000 firms in total and located in only 5 cities in China.
[124] In my paper, I include industry HHI to control for competition.
[125] For discussion on cultural transmission and its impact, see Bisin and Topa (2003), Bisin and Verdier (2001, 2005, 2011).

major conclusion is, however, that trade can improve female worker's labor market outcome through alleviation of discrimination. The underlying stories are also different: they assume that culture is attached to certain products so it is through consumption decision that people's cultural preference changes while in my paper cultural transmission happens through interaction of firms.

The contribution of this paper is that it finds another potential mechanism of how trade affects exporting country's labor market — through asymmetric cultural transmission. It proposes and proves, through both a simple model and empirical results, the fact that trade can alleviate the gender discrimination by improving women's employment and wage in the exporting country. It further stresses that the asymmetric transmission of cultural is the key to this result — by interacting with importers in countries with more gender-equal culture, the exporter would be less discriminative towards female workers; however, interacting with importers from less gender-equal cultures will not make exporters more discriminative to female workers. In this paper, under a statistical discrimination set-up, this asymmetry is a natural result of the information updating for exporters.

The remainder of this paper is organized as follows: Section 3.2 specifies the theoretical model and generates predictions; Section 3.3 describes the data and treatment; Section 3.4 specifies the empirical model and discusses the results; and Section 3.5 concludes.

3.2 Theoretical Model

Phelps (1972) is the first paper which establishes a simple model about statistical discrimination. Overtime, this model has been developed, modified, and applied in many discrimination papers. The model I propose here in my paper is also based on statistical discrimination.

3.2.1 Model Setup

It is assumed that due to the existence of searching cost, each period one worker is only randomly matched to one firm [126] (Note that in this paper firms and not distinguished from

[126] Specifically, each period one worker can only be matched to one firm, but one firm can be matched with multiple workers. The number of workers matched to a firm, however, is exogenously given and out of firms' control.

employers, and they would be used interchangeably in the following discussion). All workers live for only one period.

Suppose all firms have the same task that a worker is either qualified or not qualified for, which is denoted by indicator variable I.

$$I_i = \begin{cases} 1 & \text{if worker i is qualifed for the task} \\ 0 & \text{if worker i is not qualifed for the task} \end{cases}$$

If worker is qualified, a positive payoff μ_q is generated; otherwise, a negative payoff of μ_n is generated.

The proportion of workers who are qualified for the job is denoted by q, which falls between 0 and 1. For simplicity, male and female workers are assumed to have the same q such that they are equally productive in general. (However, this assumption can be easily relaxed without changing any predictions from the model.) Firms observe q but they do not observe each worker's qualification. Instead, they receive a signal S_i revealing worker i's type, and then make hiring decision solely based on the value of signal. The signal is either good or bad, and firms only hire workers with a good signal.

The quality of the signal is denoted as P, which equals the probability that the signal is correct. A higher P then means the signal is of better quality.

$$P = prob(S_i = good \mid I_i = 1) = prob(S_i = bad \mid I_i = 0)$$

In Aigner and Cain (1977), Lundberg and Startz (1983), and Oettinger (1996), they all assume that the test score equals a worker's true ability plus an error term, and the variance of the error term is larger for the black workers. As a result, the test score is less reliable for the black workers compared to their white counterparts. Following their idea about the difference in the reliability of signal across races, in my paper I apply this assumption on genders such that men have better information quality. In my model, instead of a test score, I simplify the information into a signal taking on only two values, and P is the only variable that controls the reliability of the signal. Therefore, I assume that the quality of signal is better for men than for women such that men have a higher P. For standardization, I assume P=1 for men, hinting that a firm can perfectly observe whether a man is qualified or not [127]. For women, however, I assume 0<P<1 to reflect the lack of experience in assessing women's capability (compared to men). As a result, all the male workers hired are actually qualified, while there are some hired female workers who are not qualified. For notation simplicity in the following analysis, the proportion of hired female workers who are actually qualified is denoted as T.

$$T = prob(I_i = 1 \mid S_i = good) = \frac{Pq}{Pq + (1-P)(1-q)}$$

Note that T is increasing in P, so that firms with better signal end up having higher proportion of qualified female workers among the hired ones.

$$\frac{\partial T}{\partial P} = \frac{q(1-q)}{\left[Pq + (1-P)(1-q)\right]^2} > 0$$

To ensure that the signal is actually informative, P needs to be large than a half. This is because firms always have the outside option of not using the signal at all and randomizing the hiring decision for women. Therefore, to make the signal useful, we need T to be larger than the randomized outcome. Note that in the randomized case, the proportion of actually qualified women among the hired ones simply equals the true proportion of qualified women, which is q.

$$T = \frac{Pq}{Pq + (1-P)(1-q)} > q \Rightarrow P > \frac{1}{2}$$

With the basic set-up of the model above, now I turn to the analysis of the implications from this model.

3.2.2 Firm Measurements

Employment Fraction

We first study the employment fraction of women within a firm. Here I use L_m and L_f to denote the number of male and female workers hired by firms, and M and F to denote the number of male and female initially matched to firms, respectively [128]. Note again that for men, due to the

[127] A more natural and less restrictive assumption could be that P for men also lies between 0 and 1, but is larger than P for women. However, this will not change the nature of our analysis or predictions of this model but only adds in extra coefficients here and there. Therefore, to keep the model in a simpler form, P=1 is assumed for men.
[128] For a related paper using firm-level data to study gender discrimination, see Kawaguchi (2007).

existence of perfect signal, all qualified male workers would be hired; for women, however, only those who send the good signal are hired despite of their true qualification.

$$L_m = M * prob(S_i = good) = Mq$$
$$L_f = F * prob(S_i = good) = FPq + F(1-P)(1-q)$$

Within a firm, the female employment fraction S would then be:

$$S = \frac{L_f}{L_f + L_m} = \frac{Pq + (1-P)(1-q)}{Pq + (1-P)(1-q) + \frac{M}{F}q}$$

Here note that M/F is exogenous to firms and varies between industries and regions[129]. Within an industry and region, M/F is fixed. It is then obvious from the above equation that the female employment fraction is decreasing in the male-female ratio in an industry. Moreover, female employment fraction is an increasing function of information quality P as long as over half of the female applicants are qualified, which I would assume to be true.

$$\frac{\partial S}{\partial P} = \frac{(2q-1)q}{\left(Pq + (1-P)(1-q) + \frac{M}{F}q\right)^2}\frac{M}{F} > 0 \text{ as long as } q > \frac{1}{2}$$

Relative Wage

Now we analyze the implication for relative wage of female over male workers derived from this model. The payment schedule is designed as follows: employers and workers bargain over the wage, and any payoff of firms are shared between workers and employers. Therefore, employer would hire a worker as long as he/she generates a positive expected payoff (in our model, the hiring decision is made solely based on the signal, so whoever gives a good signal generates positive expected payoff, while those who give a bad signal generate negative expected payoff. Conditions are discussed in appendix A. 1. which guarantees the validity of this statement). In this payment schedule, neither workers nor employers have the incentive to deviate[130]. Sigma is

[129] The matched male-female worker ratio varies across industries to reflect the fact that some industries might more intensively use female labor than other industries. It also varies across regions to reflect the variation of gender composition in local labor supply.

[130] For an employer, the only decision he can make is whether to offer a worker the job. If he refuses to offer the job to a worker with good signal, then he gets 0 payoff while he could have got a positive expected payoff. If he offers the job to a worker with bad signal, then he gets negative expected payoff instead of a 0 payoff. For a worker, the only decision he can make is whether to accept the job offer. If he declines the offer, he then gets a 0 payoff while he could have got a positive payoff, which is his wage. Therefore, neither the employer nor the worker has the incentive to deviate.

defined as the share of payoff, which falls between 0 and 1, that is enjoyed by the worker. The wage of a worker is simply the product of δ and his/her expected payoff, as is defined below. (Note that since the employers know q, it can also derive the true T, which is the probability that a hired female worker is actually qualified). Note that the model is essentially partial equilibrium and workers take their wages as given. Workers do not respond to the difference in wages across firms of different types. [131]

$$w_m = \delta\mu_q$$
$$w_f = \delta\left(T\mu_q + (1-T)\mu_n\right)$$

It is obvious that

$$w_m = \delta\mu_q > \delta\left(T\mu_q + (1-T)\mu_n\right) = w_f \text{ since } \mu_q > \mu_n$$

Therefore, female workers receive lower wages than male workers with imperfect information, and relative wage of female workers over male workers, R, can be derived.

$$R = \frac{w_f}{w_m} = \frac{\delta\left(T\mu_q + (1-T)\mu_n\right)}{\delta\mu_q} = T + (1-T)\frac{\mu_n}{\mu_q}$$

This relative wage is then increasing in P.

$$\frac{\partial R}{\partial P} = \frac{\partial R}{\partial T}\frac{\partial T}{\partial P} = \left(1 - \frac{\mu_n}{\mu_q}\right)\frac{\partial T}{\partial P} > 0$$

Therefore, firms with higher P give women higher wages relative to men [132].

Note that since workers only live for one period, after one period all current workers exit the model and firm rehire new workers. Therefore, there would be no information updating across periods in this model. However, due to data limitation, the wage prediction of the model cannot be empirically tested. [133]

[131] To be specific, equally productive female workers in domestic firms do not move to international trade firms in response of higher wages (as shown in section 3.2.3). Moreover, due to search friction, workers not hired by one firm will remain unemployed instead of searching for jobs at another firm.

[132] For a discussion on typical explanations on gender wage gap, see Blau and Kahn (2000, 2017). For studies specific to the gender wage gap in China, see Chi and Li (2008), Liu et al. (2000), Zhang et al. (2008), and Maurer-Fazio and Hughes (2002). For economic background in China about reforms and wage determination, see Meng (2012) and Meng and Kidd (1997).

[133] As explained in the next section, the firm-level variables come from Chinese Industrial Firm Database. It only contains the total amount of wage paid to all workers but not gender-specific average wage. Therefore, I cannot test model predictions on gender wage gap.

Profit

For an employer, its profit is the sum of shared payoffs from all workers:

$$\pi = Mq(1-\delta)\mu_q + F(1-\delta)\left(Pq\mu_q + (1-P)(1-q)\mu_n\right)$$

Profit per worker, U, is then written out as:

$$U = \frac{\pi}{L_m + L_f} = (1-\delta)\mu_q + (1-\delta)(\mu_n - \mu_q)\frac{(1-P)(1-q)}{Pq + (1-P)(1-q) + \frac{M}{F}q}$$

Given M and F, a firm's profit per worker is increasing in P:

$$\frac{\partial U}{\partial P} = \frac{(1-\delta)(\mu_q - \mu_n)(1-q)q(1+\frac{M}{F})}{\left(Pq + (1-P)(1-q) + \frac{M}{F}q\right)^2} > 0$$

Therefore, firms with better information quality would be more profitable.

As comparison, a taste-based discrimination model is briefly described in the appendix A.2, which would generate different predictions compared to the statistical discrimination model I use in my paper.

3.2.3 Trade and Learning

As is mentioned above, P reflects the ability of the firm to correctly estimate a female worker's capability, specifically, whether the female worker is qualified for the task or not. A higher P means a firm makes better estimate of a female worker's ability. I therefore assume that in countries with more gender-equal culture, firms have better information about a female worker's true ability since the gender gap is smaller and thus firm are more experienced in dealing with female workers. By exporting to firms in other countries, the exporters are assumed to be able to observe and learn how the importing firms estimate their female workers [134], but with a discount factor F(.), which lies between 0 and 1. F(n) is an increasing function of interaction between exporters and importers, and n reflects the intensity of interaction, which could be number of interactions in reality. Since the profits of a firm is increasing in P, an exporter then has the incentive to update its own information after it observes how an importer gets its workers' signals.

[134] For culture change in response to information, see Fernández (2013).

Exporters will compare its own signal and the importer's signal according to the information quality, and choose the signal with higher P.

$$P_{\text{exporter}}' = \max\left\{P_{\text{exporter}}, F(n) * P_{\text{importer}}\right\}$$

Therefore, by exporting to firms in destination countries with more gender-equal culture, exporters might be able to update its signal quality to have a better estimate of a female worker's true capability. However, firms which export to countries with worse gender-equal culture would have no updates in P since their own signal has better quality than the importer's, so they would continue adopting their original signal. Moreover, since the cultural transmission takes place through firms' interactions, this cultural transmission process could also happen between exporters and non-exporters with interactions. There would then be spillover effects on the non-exporters which are in the same industries or regions with the exporters.

$$P_{\text{non-exporter}}' = \max\left\{P_{\text{non-exporter}}, P_{\text{exporter}}'\right\}$$

3.2.4 Comparison between Different Skill Groups

I then compare different skill groups. Suppose that both men and women are divided into two subgroups based on some characteristics (e.g., education), with one group having higher proportion of qualified workers. Note that the assumption that men and women are equally productive is kept so that the true proportions of qualified workers of each subgroup are still identical across genders. Here we use H to denote the group with higher fraction of qualified workers q_H (the high-skill group), and L to denote the subgroup with lower fraction of qualified workers q_L (the low-skill group).

For the female employment fraction in a firm, it can be derived that:

$$\frac{\partial S}{\partial q} = \frac{-\frac{M}{F}(1-P)}{\left(Pq + (1-P)(1-q) + \frac{M}{F}q\right)^2} < 0$$

Since $q_H > q_L$, it is then predicted that the female employment fraction within the high-skill groups should be lower than that of the low-skill group.

I then study how female employment fraction react to the change in P differently across skill groups. Taking the cross derivative with respect to P and q:

$$\frac{\partial^2 S}{\partial P \partial q} = \frac{(4q-1)(1-P) + q(2P-1) + \frac{M}{F}q}{\left(Pq + (1-P)(1-q) + \frac{M}{F}q\right)^3} > 0 \text{ when } q > \frac{1}{2} \text{ and } P > \frac{1}{2}$$

Note that we have already derived above that the female employment fraction is increasing in P. Combining these two results, the female employment fraction increases more when q is higher. Therefore, with $q_H > q_L$, it is expected that when exporting to firms with better gender-equal culture, the female employment fraction among the high skill group would have larger increase than the low skill group.

Similar analysis is carried out for the relative wage:

$$\frac{\partial R}{\partial q} = \frac{\partial R}{\partial T}\frac{\partial T}{\partial q} = (1 - \frac{\mu_n}{\mu_q})\frac{P(1-P)}{\left(Pq + (1-P)(1-q)\right)^2} > 0$$

$$\frac{\partial^2 R}{\partial P \partial q} = \frac{(1-2q)\left(Pq + (1-P)(1-q)\right) - 2q(1-q)(2P-1)}{\left(Pq + (1-P)(1-q)\right)^3} < 0 \text{ when } \frac{1}{2} < q < 1 \text{ and } P > \frac{1}{2}$$

Since $q_H > q_L$, based on the above results, comparing with the low-skill group, high-skill group have larger relative wage of female workers over male workers but smaller increase in relative wage when exporting to firms with better gender-equal culture.

3.2.5 Model Implications

The testable implications from this model are summarized below:

Implication 1

Female employment fraction within a firm is higher for the low-skill workers compared to the high-skill group. The relative wage of female workers over male workers, however, would be larger for the high-skill group.

Implication 2

Trading with firms in countries with larger P (more gender-equal culture) would make the exporters/importers hire a higher fraction of female workers, pay female workers higher relative wage, and have higher profits. However, trading with firms located in countries with smaller P (less gender-equal culture) would have no impact. The impact of cultural transmission through trade on the labor market is thus asymmetric.

Implication 3

When trading with firms with larger P, the female employment fraction increases more in the high-skill group. Relative wage, on the other hand, increases more in the low-skill group.

Implication 4

Cultural transmission through trade would have spillover effects on the firms not engaged in international trade (purely domestic firms).

3.3 Data

3.3.1 Data Introduction

In this paper we use multiple databases, but the primary data comes from Chinese Customs Database, the Chinese Industrial Firm Database and the Global Gender Gap Report 2006

The Chinese Customs Database provides transaction-level data on imports and exports. It contains information including date of transaction, quantity, price, 8-digit HS code, destination country/region and name, ownership and ID of the firm [135].

The Chinese Industrial Firm Database is a firm-level database constructed by the National Bureau of Statistics in China [136]. It records firm's basic information (e.g., name, phone number, location, industry, etc.), financial status (capital, inventory, profit, total wage, etc.) and other information related to production and sales (output, sales, employee number, etc.). It surveys all industrial firms in the mainland of China with sales higher than 5 million RMB [137], covering firms in mining, manufacturing and production and supply of electricity, gas, and water industry.

Gender Gap Index (GGI) comes from the Global Gender Gap Report 2006 provided by World Economic Forum [138]. GGI covers 115 countries with over 90% of the world's population. It is composed based on 14 ratios on gender gap classified into four sub-indexes: economic participation and opportunity, education attainment, health and survival, and political empowerment (see Table C.2 in appendix for detailed information on how GGI is constructed).

[135] For challenges using this data, see Brandt et al. (2014) and Nie et al. (2012).
[136] Each record corresponds to a legal unit. For the detailed information on legal unit, see Brandt et al. (2012)
[137] About 604098 dollars using the 2004 exchange rate of 8.2768 between RMB and dollar. (Exchange rate data comes from World Bank at https://data.worldbank.org/indicator.) However, 5 million RMB is not a hard rule. See Brandt et al. (2012) for detailed information on this.
[138] See https://www.weforum.org/reports/the-global-gender-gap-report-2017. Boler et al. (2018) also use GGI to measure gender culture.

GGI lies between 0 and 1, where a higher index means more gender-equal culture. The top and bottom 10 countries with respect to GGI value is listed in Table 3.1 (for the full rank, see Table C.3). Among the 115 countries, China ranks 63 with GGI value of 0.656. The fact that China ranks almost in the middle of the GGI distribution gives me enough space to test the asymmetric transmission of culture. For distribution of all countries with GGI data, see Figure 3.1.

Table 3.1. Top and Bottom 10 Countries/Regions in GGI Value

Top 10 countries			Bottom 10 countries		
Rank	Country	GGI	Rank	Country	GGI
1	Sweden	0.8133	106	Mauritania	0.5833
2	Norway	0.7994	107	Morocco	0.5826
3	Finland	0.7958	108	Iran	0.5802
4	Iceland	0.7813	109	Egypt	0.5785
5	Germany	0.7524	110	Benin	0.5778
6	Philippines	0.7516	111	Nepal	0.5477
7	New Zealand	0.7509	112	Pakistan	0.5433
8	Denmark	0.7462	113	Chad	0.5246
9	United Kingdom	0.7365	114	Saudi Arabia	0.5241
10	Ireland	0.7335	115	Yemen	0.4594

Note: Data from Global Gender Gap Report 2006.

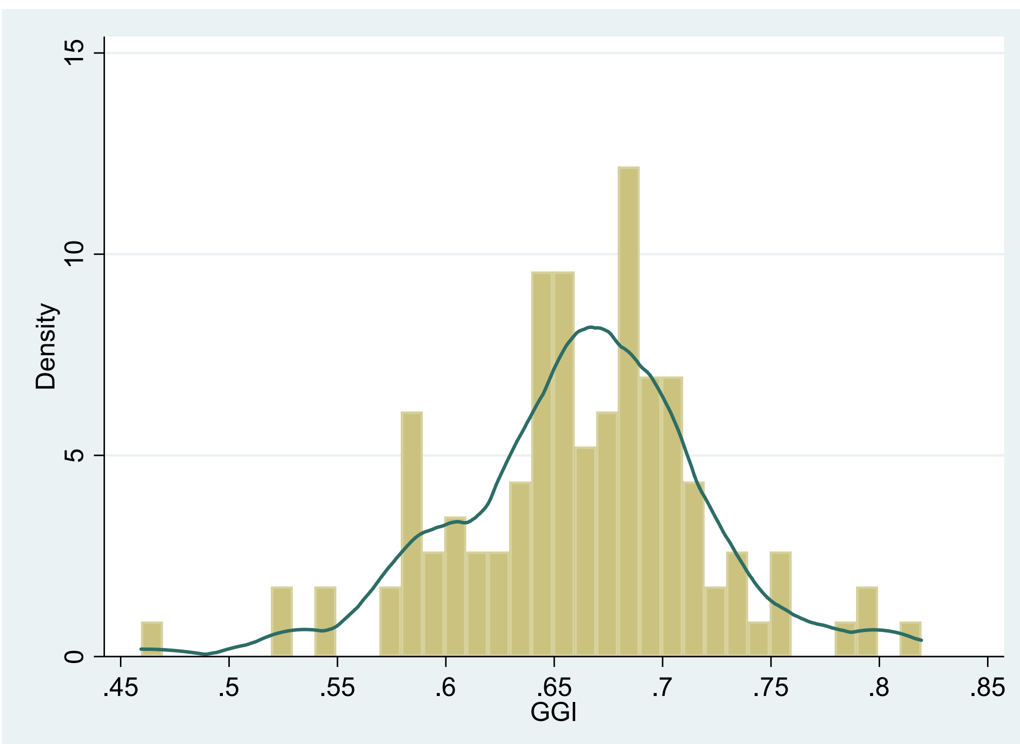

Figure 3.1. Distribution of GGI

3.3.2 Data Treatment

I first merge customs data with GGI data according to country name. Based on the records in the customs data, China exports to 229 destinations and imports from 210 origins in 2004, both covering all the 114 countries with GGI data [139]. Note that in Chinese customs data, Hong Kong, Macao and Taiwan (for simplicity, HMT), three regions in China, are listed as exporting destinations or importing origins. Since all three regions are part of China, they do not have their own GGI value, but 19.66% of the export value and 13.70% of the import value (measured in dollars) are between HMT and other regions in China. To keep those transactions in my dataset, I impute the GGI value of HMT [140]. In the baseline analysis, I assume that HMT has the same GGI value as China mainland [141]. For all other countries that China trade with but do not have GGI value, altogether they only account for 2.02% of the export value and 2.67% of the import value in 2004. Therefore, exclusion of those transactions will not be a big problem when using GGI data. Based on this merged customs-GGI database, I construct the firm-level information quality index (the P in my model) or discrimination index, which is the maximum GGI of the origin/destination which a Chinese firm trades with.

I then merge the customs-GGI database with the industrial firms' database. Unfortunately, the identification number of firms in Customs Database and Industrial Firm Database are constructed in different ways and cannot be matched. Following Yu (2015) I adopt the two-step matching. In the first step, I match the two databases based on firm name with 52,309 successful matches (46,243 for exporters and 37,299 for importers, separately [142]). In the second step, for all unmatched firms in the first step, I then match them based on the last seven-digit phone number and postcode [143], which successfully matches another 2,064 firms (1,919 for exporters and 1,456 for importers, separately). Applying this matching method, about 53.16% of the trade value in 2004 is successfully matched, with a higher matched fraction for export (56.65%) than import (49.88%). The reason for the low matching fraction is twofold: firstly, as I have mentioned above,

[139] Since China is among the 115 countries and regions with GGI data, for Chinese firms, there are only 114 destination/origin countries with GGI data to trade with.
[140] In one of the robustness checks, however, I exclude HMT trade to show that my results are not driven by the inclusion of HMT trade.
[141] In one of the robustness checks, I impute the GGI value for HMT using their historical background. See the discussion in robustness check for how the imputation is done.
[142] Note that the sum of numbers of successful match using exporters only and importers only is larger than the successful match using both exporters and importers. This is because many firms are both exporters and importers.
[143] Based on the final sample in my paper, there are 21533 distinct postcodes.

Industrial Firm Database only covers firms with sale over 5 million RMB, so that the exporters/importers with less sales cannot be matched; secondly, there are purely trading firms in China that serve only as intermediaries without doing production, and thus are not included in the Industrial Firm Database either. The GGI-customs-industrial matched database is then my final sample to use. As is discussed in Brandt et al. (2012), I drop the firms with fewer than 8 workers since these firms are under a different legal regime. See Table 3.2 for the summary statistics of the final sample.

Table 3.2. Summary Statistics

	Exporters and Importers				Local Firms	
	More		Less			
	Observation	Mean	Observation	Mean	Observation	Mean
female fraction	45,988	0.497	8,035	0.479	215,197	0.377
female fraction (high)	44,065	0.382	7,468	0.376	175,676	0.285
female fraction (low)	45,787	0.505	8,004	0.484	214,504	0.383
skill ratio	45,988	0.151	8,035	0.142	215,197	0.116
firm age	45,988	8.864	8,035	7.816	215,197	9.765
firm size	45,988	462	8,035	222	215,197	199
output	45,988	171	8,035	55	215,197	53
computer count	45,988	47	8,035	15	215,197	12
foreign capital	45,988	0.276	8,035	0.349	215,197	0.026
HMT capital	45,988	0.263	8,035	0.181	215,197	0.038
GGI	45,988	0.750	8,035	0.639		
processing trade	45,988	0.332	8,035	0.205		
destination number	45,988	9.941	8,035	1.594		

Note: Skill ratio is the fraction of high-skill workers in all employee, where high-skill workers are defined as workers with a college degree, bachelor's degree, master's degree or above. The low/high skill female ratio is measured by the number of female low/high skill workers over the number of all low/high skill workers in a firm. Foreign fraction and HMT fraction are the fractions of capital that come from foreign countries and Hong Kong, Macao and Taiwan, separately. Firm age is the number of years a firm has operated. Firm size is the number of employees in a firm. Output and export are measured by monetary value in million RMB. GGI is the maximum of the destination GGI, as is used in the main regressions. Processing trade is the fraction of trade value that is processing trade.

According to Table 3.2, as is shown in previous literature about exporters' characteristics, the exporters in my sample are on average larger, produce higher output, hire more skilled workers, and depend on FDI more (both from foreign and from HMT) as their capital resources. One thing to note is that exporters also have a higher fraction of female workers, both for high and low skill. Comparing the second and third row, the statistics shows that the female employment fraction among the high skill group is lower than that of the low skill group for both exporters and all firms, which supports *Implication 1* from my model.

3.4 Empirical Approach and Results

3.4.1 Empirical Model

Due to the limitation of data, I cannot directly observe the wage bill and other personal characteristics of each individual worker, so I cannot decompose the wage into the explained and unexplained part as some of the typical discrimination paper do in their empirical analysis section. In my paper, instead, female employment fraction and relative wage are used as a reflection of the level of statistical discrimination.

To see how discrimination affects the female employment ratio in a firm, we run the following regression:

$$female\_fraction_i = \alpha + \rho GGI_i + V\lambda + \beta_{industry} + \delta_{region} + \varepsilon \tag{3.1}$$

Here the dependent variable is the female employment fraction in each firm i. GGI is the destination information quality faced by each exporter, and thus regression (3.1) uses sample covering only the exporters. Industry and province fixed effects are included to take into consideration the selection across industries and regions. V is firm-level control variables including skill ratio, output, firm age, firm size, output, processing trade, ownership. See appendix A. for how these control variables are constructed. Note that since the data is cross-sectional, firm fixed effect is not controlled and empirical results might suffer from potential confounding factors. Moreover, without clearly identified exogenous variation, the empirical results are not causal, but instead suggestive and only show correlation. There is implicit identification assumption that across firms, the choices of export destination/import origin countries are not systematically correlated with the female employment fraction of the firms.

3.4.2 Empirical Results and Explanation

Basic Results

The basic regression results are listed in Table 3.3. [144] The industry fixed effects are included at 2-digit level (39 industries) and regional fixed effects are at province level (31 provinces in the mainland of China). Note that it is of great importance to control for industry and province fixed effects, since there would be selection across industries and provinces. For calculation of GGI, I take into consideration of HMT's historical background. Historically, all these three regions have been seized and colonized by other countries: Taiwan is colonized by Japan (with GGI 0.645, ranks 79) during 1895-1945, Hong Kong is colonized by United Kingdom (with GGI 0.737, ranks 9) during 1842-1997, and Macao is colonized by Portugal (with GGI 0.692, ranks 33) during 1553-1999. I therefore impute these countries GGI value for HMT. In robustness checks, I also impute China's GGI value for HMT or simply drop all transactions with HMT. The results don't change much.

The fixed effects are not included in column (1) but included in column (2), but the coefficient doesn't change much with inclusion of the fixed effects. Based on the results, trading with more gender-equal cultures would increase the female fraction within a firm overall, as is predicted by the model. However, when I further distinguish trade to countries with more gender-equal cultures from the less gender-equal cultures compared to China, it shows that the positive effects trade has on female fraction is driven merely by trading with more gender-equal cultures. Trading with less gender-equal culture, however, would have no impact. This supports our main prediction from the model that the transmission of culture is asymmetric since update of information only happens when the new information is better than the existing one. To save space, in other regression I would only report the variables of interest.

Since for the gender gap index, the most relevant sub-index is the "Economic Participation and Opportunity", I also run the basic regression using only this sub-index. See appendix for the results. Another potential problem is that people worry that the female employment fraction of the product might be correlated with the product quality, I hence further control for the product price. See appendix for results.

[144] Note that firms with all workers of the same gender are always excluded. For one thing, the matched male-female worker ratio cannot be 0 or infinity. For another, these firms are not responsive to the change in information quality at all.

Table 3.3. Trade and Female Fraction

	Female Fraction			
	All	All	More	Less
GGI	0.147***	0.146***	0.162***	-0.0225
	(0.0292)	(0.0217)	(0.0328)	(0.129)
GGI*Firm age	-0.00345***	-0.00207***	-0.00209***	-0.00205***
	(0.000266)	(0.000206)	(0.000211)	(0.000642)
ln(Output)	-0.0568***	-0.0438***	-0.0440***	-0.0396***
	(0.00204)	(0.00146)	(0.00156)	(0.00288)
ln(Firm Size)	0.0878***	0.0633***	0.0635***	0.0607***
	(0.00250)	(0.00188)	(0.00192)	(0.00409)
ln(Firm Age)	0.00460*	0.00959***	0.0107***	0.00316
	(0.00272)	(0.00241)	(0.00254)	(0.00538)
Skill Ratio	-0.185***	-0.134***	-0.132***	-0.148***
	(0.0122)	(0.00783)	(0.00845)	(0.0173)
Processing Trade	0.0647***	0.0477***	0.0439***	0.0678***
	(0.0139)	(0.00571)	(0.00605)	(0.00823)
FDI (Foreign)	0.0373***	0.0347***	0.0359***	0.0307***
	(0.00541)	(0.00392)	(0.00408)	(0.00732)
FDI (HMT)	-0.0149***	-4.31e-05	0.00378	-0.0160**
	(0.00570)	(0.00341)	(0.00357)	(0.00800)
ln(computer)	-0.0199***	-0.0114***	-0.0118***	-0.00732***
	(0.00231)	(0.00121)	(0.00125)	(0.00276)
Constant	0.198***	-0.00103	-0.0172	0.0885
	(0.0194)	(0.0717)	(0.0848)	(0.108)
Ownership	Yes	Yes	Yes	Yes
Industry Fixed	No	Yes	Yes	Yes
Province Fixed	No	Yes	Yes	Yes
Observations	53,671	53,671	45,701	7,970
R-squared	0.213	0.448	0.438	0.514

*Note: Standard errors in the parenthesis, clustered at county level. *** p<0.01, ** p<0.05, * p<0.1.*

Another prediction the model makes is that firms with better information quality would enjoy higher profits. Here the dependent variable is basically the log of profits for a firm. Since there are some firms with zero or negative profits. To deal with this problem, the dependent variable is calculated as below[145]:

$$\text{outcome variable} = \begin{cases} \ln(\textit{profit}) & \text{if} \quad \text{profit} > 1 \\ 0 & \text{if} \quad -1 \leq \text{profit} \leq 1 \\ -\ln(-\textit{profit}) & \text{if} \quad \text{profit} < -1 \end{cases}$$

Similar to Table 3.3, here all columns apart from column (1) have controlled for fixed effects. Based on the results, overall firms trading with more gender-equal culture enjoy higher profits.

[145] Note that in the final sample, there are not many firms lies within the profit range of (-1,1), so it is not a big problem treating all these firms having a 0 profit.

When we split the sample into trading with more and less gender-equal cultures, only firms trading with more gender-equal cultures enjoy higher profits.

For comparison, here I also use other two ways dealing the outcome variable. One is that for all firms with zero or negative profits, I impute 0.001 for their profits, and then take log based on the treated profits. The other way is that I add a constant to profits of all firms so that all firms now have positive profits, and then I take log of the treated profits [146]. The results are similar using the other two methods. For results of other two ways of dealing with log of profits, see Appendix.

Table 3.4. Profit

	ln(Profit)			
	All	All	Only Better	Only Worse
GGI	4.372***	3.475***	6.678***	2.393
	(0.660)	(0.614)	(1.030)	(4.036)
GGI*Firm age	-0.199***	-0.193***	-0.198***	-0.232***
	(0.0124)	(0.0123)	(0.0133)	(0.0294)
Ownership	Yes	Yes	Yes	Yes
Industry Fixed	No	Yes	Yes	Yes
Province Fixed	No	Yes	Yes	Yes
Observations	52,951	52,951	45,063	7,888
R-squared	0.076	0.095	0.104	0.067

*Note: Standard errors in the parenthesis, clustered at county level. *** p<0.01, ** p<0.05, * p<0.1.*

Subgroups

As is shown in the model, groups with higher proportion of qualified workers should experience larger change in female fraction faced with change in information quality, which is Implication *3*. I then test this implication by dividing workers into high-skill and low-skill groups based on their education level. Those with a college degree, bachelor's degree, master's degree and above are classified as high-skill workers while other workers are classified as low-skill workers. Within each firm, I then calculate the female fraction for high-skill and low-skill workers and run regressions separately for each skill group.

Based on the results in Table 3.5, trading with more gender-equal culture has significant effects on the employment fraction of female workers only for the high-skill group (for low skill

[146] In my paper, here I looked at the smallest profit a firm has, which is negative, and I add the absolute value of this number to profits of all firms.

group, the result is marginally insignificant with P value of 0.115), while trading with less gender-equal culture have on impact in both skill groups.

Comparing the coefficients in column (1) and column (3), it is obvious that trading with more gender-equal culture has larger impact on the female fraction for the high-skill workers, which supports *Implication 3* from the model.

Table 3.5. Trade and Female Fraction (by Skill Group)

	Female Fraction (High Skill)		Female Fraction (Low Skill)	
	Only Better	Only Worse	More	Less
GGI	0.226***	0.165	0.106***	-0.237
	(0.0307)	(0.132)	(0.0377)	(0.166)
GGI*Firm age	-3.26e-05	0.000466	-0.00247***	-0.00210***
	(0.000199)	(0.000680)	(0.000236)	(0.000736)
Ownership	Yes	Yes	Yes	Yes
Industry Fixed	Yes	Yes	Yes	Yes
Province Fixed	Yes	Yes	Yes	Yes
Observations	37,266	5,687	37,266	5,687
R-squared	0.136	0.156	0.415	0.482

*Note: Standard errors in the parenthesis, clustered at county level. *** p<0.01, ** p<0.05, * p<0.1.*

Spillover Effects

Although the non-exporters and non-importers do not directly interact with firms in other countries, they do interact with the local exporters and importers in the same region or same industry, and thus there would be spillover effects for the firms not engaged in international trade. Similar with the case of exporters/importers, here the regional/industry GGI for the domestic firms is the maximum of exporter/importer GGI within a region/industry. The 6-digit area code is used as region identifier and there are 2846 distinct regions in the final sample, with only 68 regions (2101 firms) having regional GGI that is lower than China's GGI. As for industry, the 4-digit industry code is used, which contains 524 unique values, with only 3 industries (74 firms) having industry GGI lower than China's GGI. Since the numbers of regions and industries with less gender-equal culture than China are too small, here I do not split sample into "only more gender-equal culture" and "only less gender-equal culture" again. Instead, according to the model, for those regions and industries with GGI lower than China, I use China's GGI for these regions and industries since these firms will continue use their original information quality, which equals to the GGI of China in empirical analysis. The results of regional and industrial spillover effects are listed in Table 3.6.

The basic pattern still holds for the spillover effects. That is, increase of female fraction occurs only in regions with regional GGI higher than China. One thing to note is that the industrial spillover effect is not significant for trading with more gender-equal cultures while regional spillover effect is. This might be that firms located in distant regions within the same industry do not really interact with each other, and thus there is not spillover effects within industry. Another problem of the industrial spillover effect is there are only 524 different industries at 4-digit level, with only 3 industries having average GGI lower than China, which makes the sample greatly unbalanced. The insignificant result might be driven by this mis-aggregation so results in the last two columns might be reliable to study for industrial spillover effects.

Table 3.6. Spillover Effects

	Female Fraction					
	Region			Industry		
	max	average		max	average	
		Better	Worse		Better	Worse
GGI	0.128***	0.143***	0.0409	0.505***	0.389	1.135
	(0.0314)	(0.0552)	(0.136)	(0.154)	(0.284)	(0.805)
GGI*Firm age	-0.00129***	-0.00142***	-0.000542*	-0.00123***	-0.00135***	0.000992
	(0.000115)	(0.000131)	(0.000300)	(0.000217)	(0.000238)	(0.00119
Ownership	Yes	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes	Yes
Province Fixed	Yes	Yes	Yes	Yes	Yes	Yes
Observations	211,038	193,513	17,525	211,038	210,395	643
R-squared	0.443	0.436	0.468	0.443	0.443	0.324

*Note: Standard errors in the parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Column (1) and (2) are clustered at county level, and column (3) and (4) and clustered at 4-digit industrial code level.*

Based on the results, it is obvious that the cultural transmission does have spillover effects to firms in the same region.

Robustness Check

In the robustness check, I use five other methods to calculate the GGI faced by each firm. The summary statistics of the main regression GGI and three robustness GGIs are listed in Table 3.7.

Table 3.7. Summary Statistics for Different Calculation of GGIs

	Observation			Mean	Min	Max
	All	Better	Worse			
GGI	54,023	45,988	8,035	0.73	0.52	0.81
GGI (China)	54,023	42,181	5,794	0.73	0.52	0.81
GGI (frequency)	54,023	42,181	5,794	0.72	0.52	0.81
GGI (no HMT)	51,783	42,181	8,787	0.73	0.46	0.81
GGI (no worse)	54,023	42,181	0	0.73	0.67	0.81
GGI (dummy)	54,023	45,988	8,035	0.85	0	1

Note: In the first row, GGI is used in the main regressions. It is the maximum of all origin/destination GGIs considering the historical background and impute the value of the countries that seized Hong Kong, Macao, and Taiwan before China formally took them back. The five other GGIs are alternative methods used in the robustness checks. GGI (China) imputes China's GGI for Hong Kong, Macao, and Taiwan. GGI (frequency) further uses the transaction frequency Chinese firms have with the origin/destination countries. GGI (no HMT) drop the transaction to Hong Kong, Macao and Taiwan. GGI (no worse) imputes the GGI value of China to all the firms trading with GGI index lower than China. GGI (dummy) is a dummy that takes 1 if the destination GGI is larger than China and 0 otherwise.

Empirical esults are listed in Table 3.8. In the first two columns, GGI (China) imputes China's GGI for Hong Kong, Macao, and Taiwan. For column (3) and (4), GGI (frequency) is calculated considering transaction frequency. It is nature to think that Chinese firms cannot learn all their customers'/suppliers' information through one single transaction. Therefore, here I calculate the frequency of Chinese firms interact with some origin/destination country, and how much a Chinese firm can learn is an increasing function of transaction frequency. The GGI is then calculated as the maximum of the origin/destination information that is successfully learned by the Chinese firms (see Appendix for details). For column (5) and (6), trade with Hong Kong, Macao and Taiwan are dropped compared to imputing other countries' GGI value as in the main regressions. For column (7), I impute China's GGI for all firms that end up trading with only less gender-equal cultures. In last column, instead of using the GGI value, I simply make trading with more or less gender-equal cultures a dummy variable, where trading with more gender-equal culture is 1.

Table 3.8. Trade and Female Fraction (Alternative GGI Calculations)

Panel A	Female Fraction			
	China for HMT		Frequency	
	More	Less	More	Less
GGI	0.119***	0.0700	0.136***	0.0536
	(0.0306)	(0.115)	(0.0322)	(0.120)
GGI*Firm age	-0.00207***	-0.00264***	-0.00209***	-0.00264***
	(0.000213)	(0.000545)	(0.000214)	(0.000544)
Ownership	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes
Province	Yes	Yes	Yes	Yes
Observations	41,924	11,747	41,924	11,747
R-squared	0.433	0.502	0.433	0.502
Panel B	**Female Fraction**			
	Drop HMT		No Less	
	More	Less		More
GGI	0.119***	0.131	0.108***	0.119***
	(0.0306)	(0.112)	(0.0216)	(0.0306)
GGI*Firm age	-0.00207***	-0.00250***	-0.00207***	-0.00207***
	(0.000213)	(0.000592)	(0.000206)	(0.000213)
Ownership	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes
Province	Yes	Yes	Yes	Yes
Observations	41,924	9,527	53,671	41,924
R-squared	0.433	0.517	0.448	0.433

*Note: Standard errors in the parenthesis, clustered at county level. *** p<0.01, ** p<0.05, * p<0.1.*

From the comparison, we can see that the alternative GGIs resemble the GGI that is used in the main regressions. Once taking the transaction frequency into consideration, the GGI (frequency) is a bit lower than GGI, since now firms cannot learn all of trade partners' information through one single transaction.

Based on the results, the asymmetric cultural transmission pattern is quite robust to alternative calculations of GGI. Therefore, when Chinese firms trade with countries of more gender-equal culture, they would hire a higher fraction of female workers due to the update of information about individual female workers' ability. Trading with countries of less gender-equal culture, however, would have no impact.

Placebo Tests

As is predicted by the model, a firm is going to learn from the best information source it can get, and thus I use the maximum of all destination GGI in my empirical study. As comparison, here I use the weighted average of all destination GGI, where the weight is the trade value. Based on the result, GGI is not significant no matter trading with more or less gender-equal cultures. This also gives extra credit to the previous results, and showed that the fact that the asymmetry of the results is not driven the sample size since here the sample size is more balance but GGI is not significant in either column.

Table 3.9. Placebo Results

	Female Fraction	
	More	Less
GGI	0.0515	0.0915
	(0.0499)	(0.0793)
GGI*Firm age	-0.00201***	-0.00315***
	(0.000266)	(0.000369)
Ownership	Yes	Yes
Industry	Yes	Yes
Province	Yes	Yes
Observations	30,160	23,511
R-squared	0.421	0.482

Note: Standard errors in the parenthesis, clustered at county level.
**** p<0.01, ** p<0.05, * p<0.1.*

3.5 Conclusion

In this paper, using Chinese customs data and firm level data in 2004, and by measuring gender equality using Gender Gap Index, I proved that firms trading with destinations with more gender equal culture would hire a higher fraction of female workers and enjoy higher profits. This is because firm learn about estimating female workers' productivity through international trade, and there would be information update only when Chinese firms trade with countries with better information about female workers. This transmission of culture is thus naturally asymmetric where there is no impact on firms which trading with less gender-equal cultures. Moreover, this alleviation of gender discrimination is more enjoyed by high skill female workers.

3.6 References

Aigner, D. J., & Cain, G. G. (1977). Statistical Theories of Discrimination in Labor Markets. *ILR Review*, *30*(2), 175-187.

Arrow, K. (1973). The Theory of Discrimination. *Discrimination in labor markets*, *3*(10), 3-33.

Becker, G. S. (1957). The Economics of Discrimination: An Economic View of Racial Discrimination. University of Chicago.

Bisin, A., & Topa, G. (2003). Empirical Models of Cultural Transmission. *Journal of the European Economic Association*, *1*(2-3), 363-375.

Bisin, A., & Verdier, T. (2001). The Economics of Cultural Transmission and the Dynamics of Preferences. *Journal of Economic Theory*, *97*(2), 298-319.

Bisin, A., & Verdier, T. (2005). Cultural Transmission. *The New Palgrave Dictionary of Economics*.

Bisin, A., & Verdier, T. (2011). The Economics of Cultural Transmission and Socialization. In *Handbook of Social Economics* (Vol. 1, pp. 339-416). North-Holland.

Black, S. E., & Brainerd, E. (2004). Importing Equality? The Impact of Globalization on Gender Discrimination. ILR Review, 57(4), 540-559.

Blau, F. D., & Kahn, L. M. (2000). Gender Differences in Pay. *Journal of Economic Perspectives*, *14*(4), 75-99.

Blau, F. D., & Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, *55*(3), 789-865.

Bøler, E. A., Javorcik, B., & Ulltveit-Moe, K. H. (2018). Working Across Time Zones: Exporters and the Gender Wage Gap. *Journal of International Economics*, 111, 122-133.

Brandt, L., Van Biesebroeck, J., & Zhang, Y. (2012). Creative Accounting or Creative Destruction? Firm-Level Productivity Growth in Chinese Manufacturing. *Journal of Development Economics*, *97*(2), 339-351.

Brandt, L., Van Biesebroeck, J., & Zhang, Y. (2014). Challenges of Working with the Chinese NBS Firm-Level data. *China Economic Review*, *30*, 339-352.

Charles, K. K., Guryan, J., & Pan, J. (2009). Sexism and Women's Labor Market Outcomes. Unpublished Manuscript, Booth School of Business, University of Chicago.

Chi, W., & Li, B. (2008). Glass Ceiling or Sticky Floor? Examining the Gender Earnings Differential Across the Earnings Distribution in Urban China, 1987–2004. *Journal of Comparative Economics*, *36*(2), 243-263.

Dong, X. Y., & Zhang, L. (2009). Economic Transition and Gender Differentials in Wages and Productivity: Evidence from Chinese Manufacturing Enterprises. *Journal of Development Economics*, 88(1), 144-156.

Fernández, R. (2013). Cultural Change as Learning: The Evolution of Female Labor Force Participation over a Century. *American Economic Review*, *103*(1), 472-500.

Kawaguchi, D. (2007). A Market Test for Sex Discrimination: Evidence from Japanese Firm-Level Panel Data. *International Journal of Industrial Organization*, *25*(3), 441-460.

Kodama, N., Javorcik, B. S., & Abe, Y. (2016). Transplanting Corporate Culture across International Borders: FDI and Female Employment in Japan. *Research Institute of Economy, Trade and Industry (RIETI) Working Paper*.

Liu, P. W., Meng, X., & Zhang, J. (2000). Sectoral Gender Wage Differentials and Discrimination in the Transitional Chinese Economy. *Journal of Population Economics*, *13*(2), 331-352.

Lundberg, S. J., & Startz, R. (1983). Private Discrimination and Social Intervention in Competitive Labor Market. *American Economic Review*, *73*(3), 340-347.

Maurer-Fazio, M., & Hughes, J. (2002). The Effects of Market Liberalization on the Relative Earnings of Chinese Women. *Journal of Comparative Economics*, *30*(4), 709-731.

Maystre, N., Olivier, J., Thoenig, M., & Verdier, T. (2014). Product-Based Cultural Change: Is the Village Global?. *Journal of International Economics*, *92*(2), 212-230.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6), 1695-1725.

Meng, X. (2012). Labor Market Outcomes and Reforms in China. *Journal of Economic Perspectives*, *26*(4), 75-102.

Meng, X., & Kidd, M. P. (1997). Labor Market Reform and the Changing Structure of Wage Determination in China's State Sector During the 1980s. *Journal of Comparative Economics*, *25*(3), 403-421.

Nie, H., Jiang, T., & Yang, R. (2012). A Review and Reflection on the Use and Abuse of Chinese Industrial Enterprises Database. *The Journal of World Economy*, 35(5), 142-158.

Oettinger, G. S. (1996). Statistical Discrimination and the Early Career Evolution of the Black-White Wage Gap. *Journal of Labor Economics*, *14*(1), 52-78.

Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *American Economic Review*, *62*(4), 659-661.

Tang, H., & Zhang, Y. (2017). Do Multinationals Transfer Culture? Evidence on Female Employment in China.

Yu, M. (2015). Processing Trade, Tariff Reductions and Firm Productivity: Evidence from Chinese Firms. *The Economic Journal*, *125*(585), 943-988.

Zhang, J., Han, J., Liu, P. W., & Zhao, Y. (2008). Trends in the Gender Earnings Differential in Urban China, 1988–2004. *ILR Review*, *61*(2), 224-243.

Zhang, L., & Dong, X. Y. (2008). Male–Female Wage Discrimination in Chinese Industry Investigation Using Firm-Level Data. Economics of Transition, 16(1), 85-112.

APPENDIX A. MARRIAGE MARKET SIGNALING AND WOMEN'S OCCUPATION CHOICE

A.1 Data Appendix

Table A.1: Occupations with at Least 80% Female Workers in 2020

Share	Occupation Title	Share	Occupation Title
99	Preschool and kindergarten teachers	86	Paralegals and legal assistants
96	Medical records specialists	85	Social workers, all other
95	Childcare workers	85	Healthcare support occupations
94	Speech-language pathologists	85	Court, municipal, and license clerks
94	Dental hygienists	85	Phlebotomists
94	Skincare specialists	84	Therapists, all other
94	Dental assistants	84	Tailors, dressmakers, and sewers
93	Medical secretaries and administrative assistants	84	Interior designers
93	Secretaries and administrative assistants, except legal, medical, and executive	84	Diagnostic medical sonographers
93	Executive secretaries and executive administrative assistants	84	Legal secretaries and administrative assistants
91	Dietitians and nutritionists	83	Librarians and media collections specialists
91	Hairdressers, hairstylists, and cosmetologists	83	Office clerks, general
90	Home health aides	83	Hotel, motel, and resort desk clerks
90	Medical assistants	82	Psychiatric technicians
90	Licensed practical and licensed vocational nurses	82	Hosts and hostesses, restaurant, lounge, and coffee shop
90	Child, family, and school social workers	82	Eligibility interviewers, government programs
89	Nursing assistants	82	Personal care aides
89	Billing and posting clerks	82	Floral designers
88	Receptionists and information clerks	81	Special education teachers
88	Maids and housekeeping cleaners	81	Substance abuse and behavioral disorder counselors
88	Veterinary technologists and technicians	81	Massage therapists
88	Nurse practitioners	80	Tellers
87	Registered nurses	80	Other psychologists
87	Bookkeeping, accounting, and auditing clerks	80	Teaching assistants
87	Healthcare social workers	80	Flight attendants
87	Payroll and timekeeping clerks	80	Social and human service assistants
86	Occupational therapists	80	Elementary and middle school teachers
86	Library assistants, clerical	80	Travel agents

Note: Data from CPS table (annual average) released by the Bureau of Labor Statistics at https://www.bls.gov/cps/cpsaat11.htm. As given in the CPS table, occupations with less than 50,000 workers are omitted.

Table A.2: Full List of Categories Specified in the Generalized Work Activities File

Analyzing Data or Information	Judging the Qualities of Things, Services, or People
Assisting and Caring for Others	Making Decisions and Solving Problems
Coaching and Developing Others	Monitor Processes, Materials, or Surroundings
Communicating with Persons Outside Organization	Monitoring and Controlling Resources
Communicating with Supervisors, Peers, or Subordinates	Operating Vehicles, Mechanized Devices, or Equipment
Controlling Machines and Processes	Organizing, Planning, and Prioritizing Work
Coordinating the Work and Activities of Others	Performing Administrative Activities
Developing Objectives and Strategies	Performing General Physical Activities
Developing and Building Teams	Performing for or Working Directly with the Public
Documenting/Recording Information	Processing Information
Drafting, Laying Out, and Specifying Technical Devices, Parts, and Equipment	Provide Consultation and Advice to Others
Establishing and Maintaining Interpersonal Relationships	Repairing and Maintaining Electronic Equipment
Estimating the Quantifiable Characteristics of Products, Events, or Information	Repairing and Maintaining Mechanical Equipment
Evaluating Information to Determine Compliance with Standards	Resolving Conflicts and Negotiating with Others
Getting Information	Scheduling Work and Activities
Guiding, Directing, and Motivating Subordinates	Selling or Influencing Others
Handling and Moving Objects	Staffing Organizational Units
Identifying Objects, Actions, and Events	Thinking Creatively
Inspecting Equipment, Structures, or Material	Training and Teaching Others
Interacting with Computers	Updating and Using Relevant Knowledge
Interpreting the Meaning of Information for Others	

*Note: Data from O*NET. Refer to https://www.onetonline.org/find/descriptor/browse/Work_Activities/ for a detailed explanation.*

Table A.3: Full List of Categories Specified in the Knowledge File

Administration and Management	History and Archeology
Biology	Law and Government
Building and Construction	Mathematics
Chemistry	Mechanical
Clerical	Medicine and Dentistry
Communications and Media	Personnel and Human Resources
Computers and Electronics	Philosophy and Theology
Customer and Personal Service	Physics
Design	Production and Processing
Economics and Accounting	Psychology
Education and Training	Public Safety and Security
Engineering and Technology	Sales and Marketing
English Language	Sociology and Anthropology
Fine Arts	Telecommunications
Food Production	Therapy and Counseling
Foreign Language	Transportation
Geography	

*Note: Data from O*NET. For a detailed explanation, refer to https://www.onetonline.org/find/descriptor/browse/Knowledge/.*

Table A.4: Country-of-Birth List

Country-of-birth	Countries or places included
1	Africa
2	Atlantic Islands
3	Australia & New Zealand
4	Austria
5	Belgium
6	Czechoslovakia
7	English Canada
8	European Canada
9	Central America
10	China
11	Northern Europe n.s., Liechtenstein, Monaco, Albania, Andorra, Gibraltar, Malta, San Marino, Southern Europe n.s., Bulgaria, Yugoslavia, Central Europe n.s., Eastern Europe n.s., Western Europe n.s., Europe n.s.
12	Cuba
13	Denmark & Iceland
14	England & United Kingdom n.s.
15	Finland
16	France
17	Germany (include part of French)
18	Greece
19	Hungary
20	India
21	Ireland
22	Italy
23	Japan
24	Luxembourg
25	Mexico
26	Netherlands
27	Norway
28	Korea (North and South), Iran, Maldives, Nepal, Middle East/Asia Minor (apart from Syria and Turkey)
29	All the rest (immigrants not otherwise specified)
30	Pacific Islands
31	Poland
32	Portugal
33	Romania
34	Russian Empire
35	Scotland
36	South America
37	Spain
38	Sweden
39	Switzerland
40	Israel/Palestine, Syria, Turkey
41	Wales
42	West Indies

Note: The grouping of countries-of-birth follows Lafortune (2013).

Table A.5: Ethnic Group Composition

	Ethnic Group	Countries-of-birth
1	British ancestry	Australia, New Zealand, English Canada, England, Scotland, Wales, Northern Ireland, United Kingdom (n.e.c.)
2	Francophone	Belgium, Austria, European Canada, France
3	Southern Europeans	Italy, Portugal, Spain
4	Hispanics	Central America, Cuba, South America, West Indies, Mexico
5	Scandinavians	Denmark, Iceland, Finland, Norway, Sweden
6	Germanic	Austria, Germany, Switzerland, Luxembourg, Netherlands
7	Russians and others	Poland, Romania, Russian Empire
8	Other Europeans	Hungary, Czechoslovakia, Greece, Europe (n.e.c.)
9	Other countries	China, India, Japan, Korea (North and South), Iran, Maldives, Nepal, Middle East/Asia Minor, Israel/Palestine, Syria, Turkey, Africa, Atlantic Islands, Pacific Islands All other immigrants (n.e.c.)

Note: The grouping of ethnic groups follows Lafortune (2013).

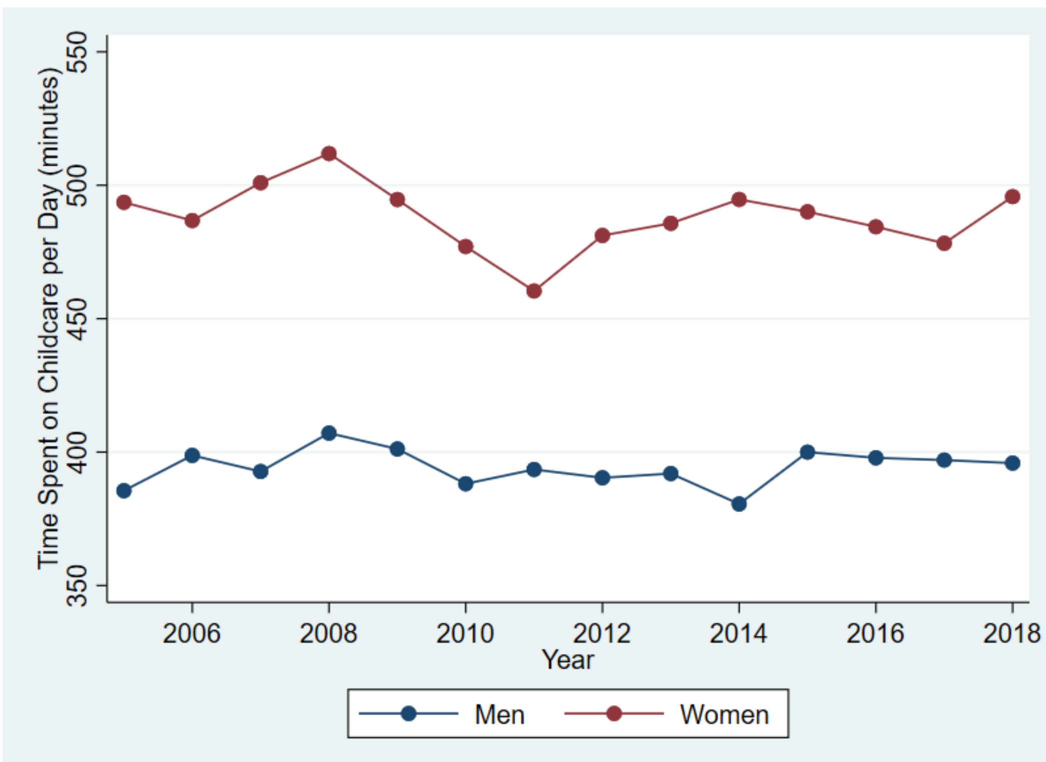

Figure A.1: Time Spent on Childcare Per Day by Gender, 2005-2018

Note: Data from American Time-Use Survey. The sample is limited to respondents in families in which a spouse is present, respondents work full-time, and at least one child below age 13 is in the household. Childcare time here includes both primary and secondary childcare.

A.2 Model Appendix

A.2.1 Parameter Ranges for Different Equilibria

Case 1: $\rho\beta < \alpha + (1-\alpha)\theta$

The proper range of U is defined such that $0<\theta<1$ in a partial-pooling equilibrium.

$$0 < \frac{1}{(1-\alpha)}\left(\left(\frac{1}{2}\frac{\alpha U \rho\beta(2-\lambda)}{2W_n - 2W_c + 2\tau}\right)^{\frac{1}{2}} - \alpha\right) < 1$$

$$\Rightarrow \alpha < \left(\frac{1}{2}\frac{\alpha U \rho\beta(2-\lambda)}{2W_n - 2W_c + 2\tau}\right)^{\frac{1}{2}} < 1$$

$$\Rightarrow 2\alpha^2 < \frac{\alpha U \rho\beta(2-\lambda)}{2W_n - 2W_c + 2\tau} < 2$$

$$\Rightarrow \frac{4\alpha\left(W_n - W_c + \tau\right)}{\rho\beta(2-\lambda)} < U < \frac{4\left(W_n - W_c + \tau\right)}{\alpha\rho\beta(2-\lambda)}$$

If U is too large, θ would equal 1 in a complete pooling equilibrium. In this case:

$$U \geq \frac{4\left(W_n - W_c + \tau\right)}{\alpha\rho\beta(2-\lambda)}$$

If U is too small, θ would equal 0 in a separating equilibrium. In this case:

$$U \leq \frac{4\alpha\left(W_n - W_c + \tau\right)}{\rho\beta(2-\lambda)}$$

Case 2: $\rho\beta \geq \alpha + (1-\alpha)\theta$

The proper range of U is defined such that $0<\theta<1$ in a partial-pooling equilibrium.

$$0 < \frac{1}{1-\alpha}\left(\frac{1}{2}\frac{\alpha U(2-\lambda)}{2W_n - 2W_c + 2\tau} - \alpha\right) < 1$$

$$\Rightarrow \alpha < \frac{1}{2}\frac{\alpha U(2-\lambda)}{2W_n - 2W_c + 2\tau} < 1$$

$$\Rightarrow \frac{4\left(W_n - W_c + \tau\right)}{2-\lambda} < U < \frac{4\left(W_n - W_c + \tau\right)}{\alpha(2-\lambda)}$$

If U is too large, θ would equal 1 in a complete pooling equilibrium. In this case:

$$U \geq \frac{4\left(W_n - W_c + \tau\right)}{\alpha(2-\lambda)}$$

If U is too small, θ would equal 0 in a separating equilibrium. In this case:

$$U \le \frac{4\left(W_n - W_c + \tau\right)}{2-\lambda}$$

A.2.2 Calculation of Marriage Rate in Different Cases

Sub-Case 1: $\rho\beta < \alpha + (1-\alpha)\theta \ and \ \rho < 1$

$$\begin{aligned} P_M(c) &= \frac{\rho\beta}{\alpha+(1-\alpha)\theta}\times 1 + \left(1-\frac{\rho\beta}{\alpha+(1-\alpha)\theta}\right)\frac{\rho(1-\beta)}{1-\rho\beta}\times\pi + \left(1-\frac{\rho\beta}{\alpha+(1-\alpha)\theta}\right)\left(1-\frac{\rho(1-\beta)}{1-\rho\beta}\right)\times 0 \\ &= \frac{\rho\beta}{\alpha+(1-\alpha)\theta} + \left(1-\frac{\rho\beta}{\alpha+(1-\alpha)\theta}\right)\frac{\rho(1-\beta)}{1-\rho\beta}\pi \\ P_M(n) &= 0\times 1 + \frac{\rho(1-\beta)}{1-\rho\beta}\times\pi + \left(1-\frac{\rho(1-\beta)}{1-\rho\beta}\right)\times 0 \\ &= \frac{\rho(1-\beta)}{1-\rho\beta}\pi \end{aligned}$$

Sub-Case 2: $\rho\beta < \alpha + (1-\alpha)\theta \ and \ \rho \ge 1$

$$\begin{aligned} P_M(c) &= \frac{\rho\beta}{\alpha+(1-\alpha)\theta}\times 1 + \left(1-\frac{\rho\beta}{\alpha+(1-\alpha)\theta}\right)\times\pi + 0\times 0 \\ &= \frac{\rho\beta}{\alpha+(1-\alpha)\theta} + \left(1-\frac{\rho\beta}{\alpha+(1-\alpha)\theta}\right)\pi \\ P_M(n) &= 0\times 1 + 1\times\pi + 0\times 0 \\ &= \pi \end{aligned}$$

Sub-Case 3: $\rho\beta \ge \alpha + (1-\alpha)\theta \ and \ \rho < 1$

$$\begin{aligned} P_M(c) &= 1\times 1 + 0\times\pi + 0\times 0 \\ &= 1 \\ P_M(n) &= 0\times 1 + \frac{\rho(1-\beta)}{(1-\alpha)(1-\theta)}\times\pi + \left(1-\frac{\rho(1-\beta)}{(1-\alpha)(1-\theta)}\right)\times 0 \\ &= \frac{\rho(1-\beta)}{(1-\alpha)(1-\theta)}\pi \end{aligned}$$

Sub-Case 4: $\rho\beta \ge \alpha + (1-\alpha)\theta \ and \ \rho \ge 1$

$$\begin{aligned} P_M(c) &= 1\times 1 + 0\times\pi + 0\times 0 \\ &= 1 \\ P_M(n) &= 0\times 1 + 1\times\pi + 0\times 0 \\ &= \pi \end{aligned}$$

A.3 Empirical Appendix

Table A.6: Caregiving and Age of First Marriage

	Women			Men		
	Index	Dummy	Quartile	Index	Dummy	Quartile
Index	-0.0320**			0.0154		
	(0.0125)			(0.0125)		
Dummy		-0.619			0.458	
		(0.436)			(0.489)	
Quartile 4			-2.246**			0.721
			(1.003)			(1.479)
Quartile 3			-0.939			0.404
			(0.821)			(0.641)
Quartile 2			-0.653			-0.107
			(0.848)			(0.548)
Year	Y	Y	Y	Y	Y	Y
Observations	3,055	3,055	3,055	3,699	3,699	3,699
R-squared	0.397	0.397	0.397	0.296	0.296	0.296

*Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In columns 3 and 6, the least caregiving occupations (the first quartile) is used as the reference group.*

Table A.7: Unilateral Divorce Law and Women's Occupation Choice (DDD)

	Index		Dummy	
	Part controls	Stevenson controls	Part controls	Stevenson controls
Treated × Single	-1.369***	-1.037***	-0.0411***	-0.0349***
	(0.306)	(0.292)	(0.0102)	(0.00950)
Treated	0.464	0.0256	0.0130*	0.00404
	(0.457)	(0.461)	(0.00758)	(0.00662)
Property Regime	Y	Y	Y	Y
Marital Status	Y	Y	Y	Y
Year	Y	Y	Y	Y
State	Y	Y	Y	Y
Marital × Year	Y	Y	Y	Y
Marital × State	Y	Y	Y	Y
Year × State	Y	Y	Y	Y
Observations	275,915	275,915	275,915	275,915
R-squared	0.011	0.100	0.012	0.066

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In columns 1 and 3, only age, age square, and race are included in the control variables. In columns 2 and 4, variables that might change after adoption of unilateral divorce law for newly married couples (as shown in Stevenson 2007) are further controlled, including years of education, whether there are any young children (under age of 13) in the household, local (by state and year) labor force participation rate of women, and full-time or part-time job.*

Table A.8: Summary Statistics on Younger Cohort – Full Count Data

Ethnic group	Flow	Sex Ratio		
	Total	Mean	Max (state)	Min (state)
British ancestry	1.37	0.89	5.77	0.70
Francophone	0.14	0.98	4.11	0.43
Southern Europeans	1.05	1.68	25.33	1.20
Hispanics	0.37	1.30	7.00	0.56
Scandinavians	0.58	1.33	6.56	0.62
Germanic	1.27	1.31	11.30	0.93
Russians and others	1.39	1.27	9.06	1.06
Other Europeans	0.54	1.47	30.87	0.94
Other countries	0.23	2.39	6.90	1.57

Note: Calculated using data from decennial censuses full-count microdata from IPUMS. The young cohort includes men aged 20-35 and women aged 18-33. The flow numbers are based on 1891-1930 first-generation young immigrants, and are measured in millions. Columns 1 and 2 show the total flow of new immigrants and average sex ratio over the whole research period by ethnic group. Columns 3 and 4 shows the maximum and minimum sex ratio of each ethnic group over the same research period, but by state.

Table A.9: First-Stage – All Young Women (Single and Married)

	Sex-ratio	
	Flow	Stock
Predicted Sex Ratio	0.568***	0.377***
	(0.148)	(0.108)
Predicted Flow	-0.000666	-0.00121***
	(0.000552)	(0.000219)
Observations	185,451	185,451
Joint F-test	6.860	14.82

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The F-test shows the F-stat for excluded instruments. Here control variables include age dummies, whether mother was born in foreign country and whether father was born in foreign country. Fixed effects include state, immigration period, ethnic groups, and double interaction between these three variables. In column one, the sex ratio is calculated within new immigrants. In column two, the sex ratio is calculated based on the foreign stock.*

Table A.10: Endogamy Rate for Women

Ethnic group	Endogamy Rate
British ancestry	0.305
Francophone	0.304
Southern Europeans	0.615
Hispanics	0.592
Scandinavians	0.507
Germanic	0.430
Russians and others	0.602
Other Europeans	0.498
Other countries	0.207

Note: Data are replicated from Table 1 of Lafortune (2013).

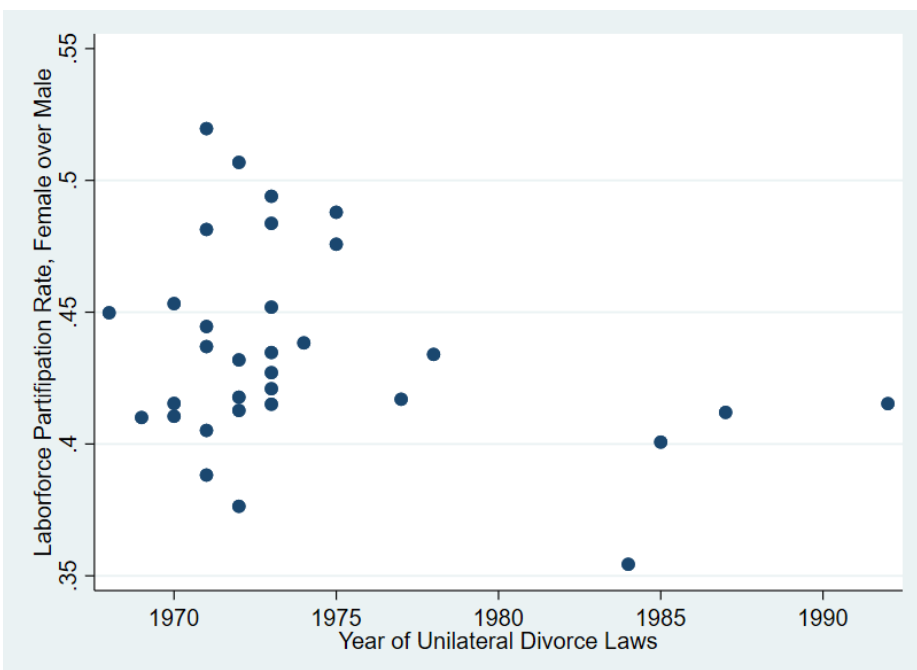

Figure A.2: Timing of Unilateral Divorce Law Adoption and State Characteristics – Female Labor Force Participation Rate over Men, 1960

Note: Data from Decennial Censuses 1960 5% sample (provided by IPUMS).

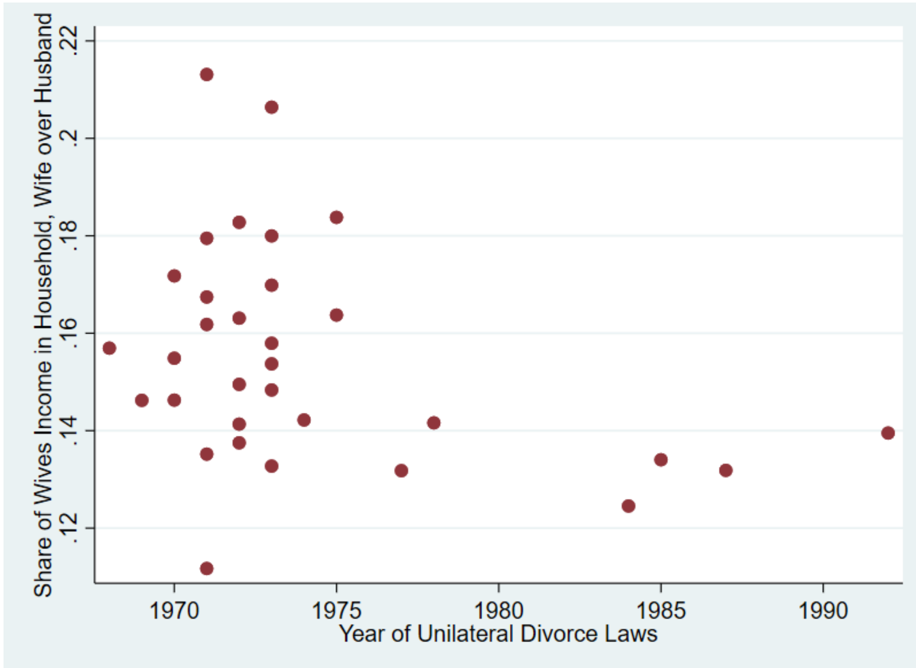

Figure A.3: Timing of Unilateral Divorce Law Adoption and State Characteristics – Share of Wives' Income in Household over Husband, 1960

Note: Data from Decennial Censuses 1960 5% sample (provided by IPUMS).

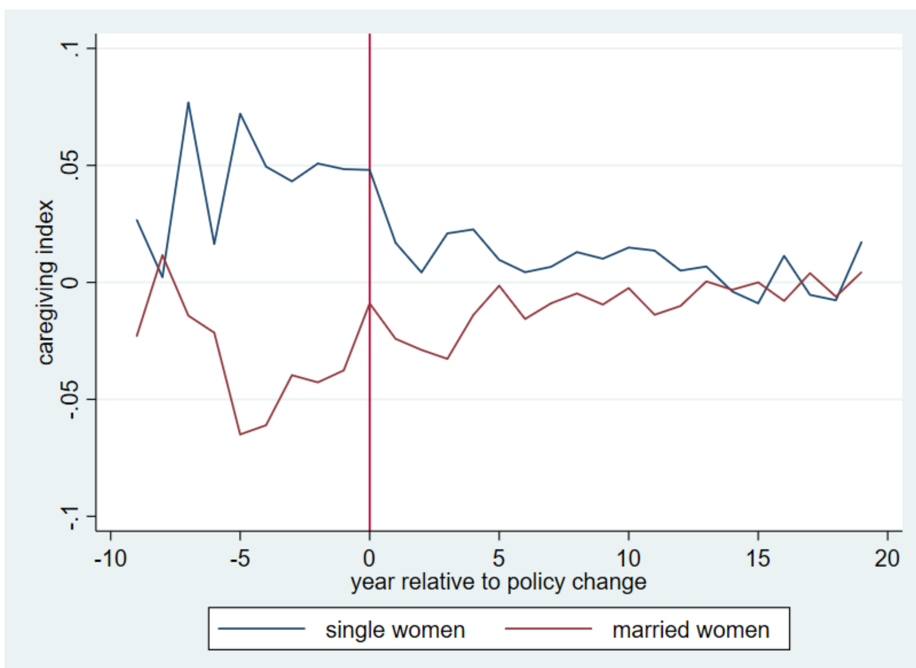

Figure A.4: Residual Caregiving Index Relative to Unilateral Divorce Law Adoption

Note: Controls include age, age square, and race. Time is measured relative to the adoption of unilateral divorce law and one year before the adoption of unilateral divorce law is set as year zero.

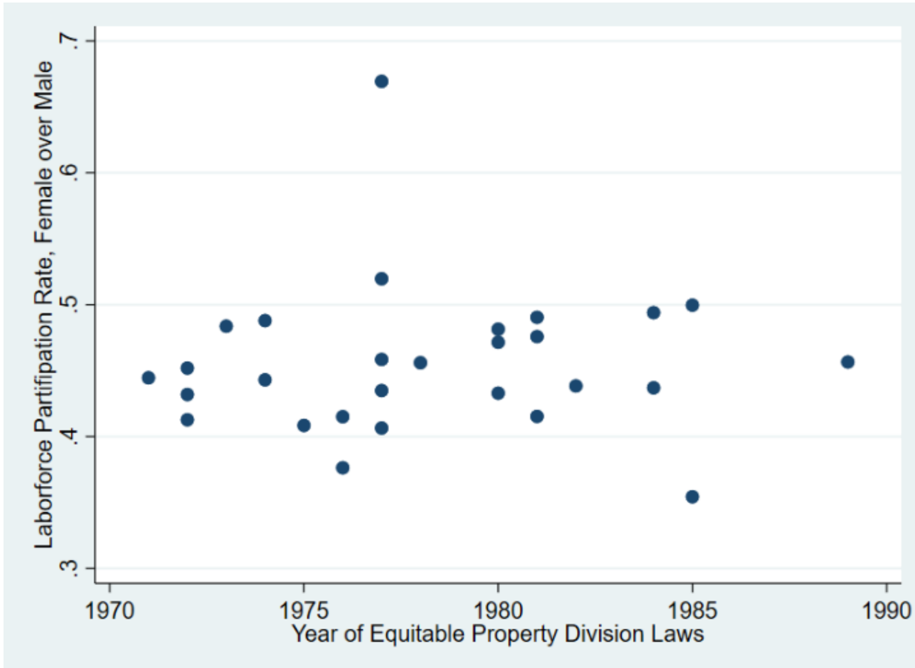

Figure A.5: Timing of Equitable Division Law Adoption and State Characteristics – Female Labor Force Participation Rate over Men, 1960

Note: Data comes from IPUMS-USA, 1960 5% sample.

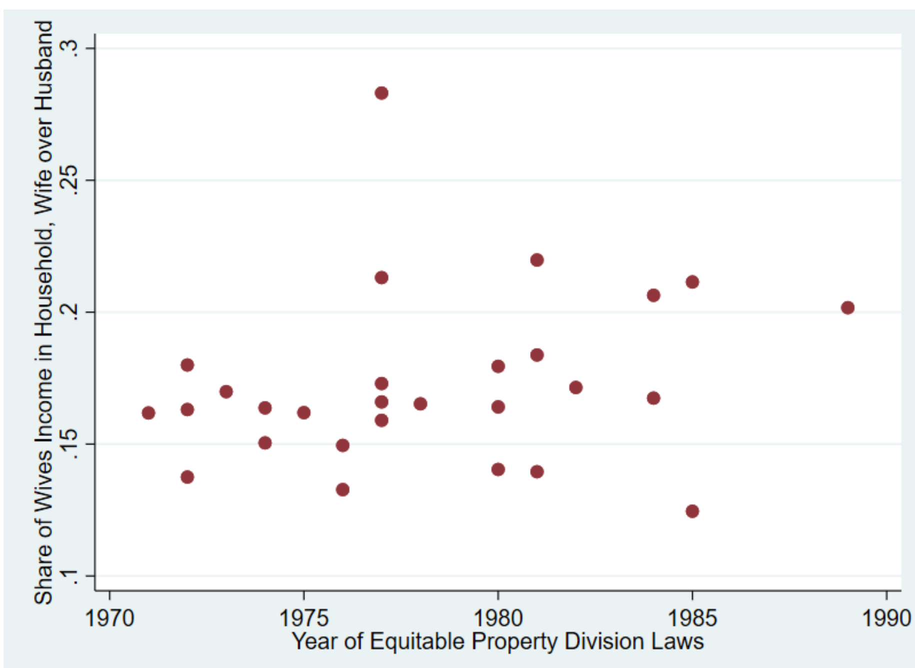

Figure A.6: Timing of Equitable Division Law Adoption and State Characteristics – Share of Wives' Income in Household over Husband, 1960

Note: Data comes from IPUMS-USA, 1960 5% sample.

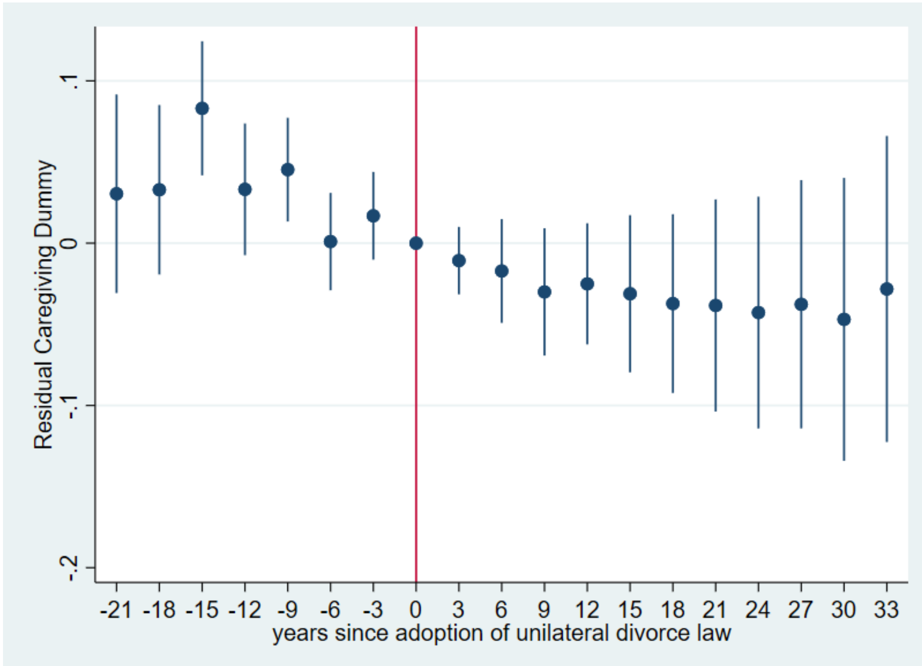

Figure A.7: Years Since Adoption of Unilateral Divorce Law

Note: Sample is restricted to single women. The residual caregiving dummy is calculated when age, age square, race, and property division law are controlled. Here one year before the adoption of unilateral divorce law is set as year zero.

A.4 Section Appendix

A.4.1 Caregiving Occupations Pay Lower Wages

The model assumes $w_n > w_c$, which means that the caregiving occupations pay a lower wage than the non-caregiving ones. England et. al. (2002) shows that all else being equal, care work pays less than other occupations. This result is replicated using the caregiving index I constructed. Compared to England et. al. (2002) which directly identifies specific occupations as care work occupations, my use of O*NET measurements provides a more standard and comparable way to rank the occupations in their caregiving features. [147] I also include in regressions other occupation features commonly used as important factors impacting wages. England et. al. (2002) have similar considerations about occupation features but they used their own measures based on DOT. [148] I

[147] See England et. al. (2002) appendix Table A1 for what occupations are identified as care work in their paper.
[148] They include three occupation features: cognitive skill (created in 134-135 of England (1992), *Comparable worth: Theories and evidence*), physical strength, and physical hazards.

show that $W_n > W_c$ in data not only to support the model assumption, but also to argue that women's signaling behavior on the marriage market has real-world implications: women sort into more caregiving occupations to signal their family orientation on the marriage market, but caregiving occupations pay less. This means that women's marriage market signaling behavior contributes to gender wage gap through the choice of occupation.

To replicate the results using the England et. al. (2002) care work definition and then use our own caregiving index, the following empirical regression is applied:

$$Y_{it} = \alpha + \beta Caregiving_{it} + \beta_X X_{it} + \delta_i + \delta_t + \delta_{ind} + \varepsilon_{it} \tag{A.1}$$

The outcome variable is the natural log of hourly pay here. The variable of interest is the caregiving index of the occupation individual i is working in at time t. Individual, time, and industry fixed effect would be included. Individual fixed effect would take care of the unobserved difference across individuals (e.g., preference, ability), time fixed effect would remove the "common shocks" to all individuals that vary by year (e.g., financial crisis), and industry fixed effect would absorb the wage differentials across industries. [149] Control variables here include four controls that will be omitted once year and individual fixed effect are both included: AFQT, age, age square, sex, and race of respondent; demographic variables that vary over time including education, actual working experience, marital status, number of kids, whether full-time job, and hours usually work per week; occupation features of routine task, non-routine (math) task, and social skill intensity (following Deming 2017 which origins from Autor et. al.'s 2003 method) or cognitive intensity, communication intensity, and manual intensity (following Ottaviano et. al. 2013). [150]

From Table A.11, we see results similar to those in England et. al. (2002) in that more caregiving occupations pay lower hourly wages even when other occupations features are already considered. This pay gap is robust to the inclusion of individual and industry fixed effects. Note that women might sort into certain industries similarly to how they sort into more caregiving occupations to signal on the marriage market, but since the focus of this paper is sorting across occupations, I still include the industry fixed effect in regressions below, which reduces the size of the coefficients.

---

[149] Occupation is often studied together with industry in research on the gender wage gap.
[150] In the main results, I only show when routine task, non-routine task (math), and social skill intensity are controlled. Results using cognitive intensity, communication intensity, and manual intensity are in the appendix. The patterns are very similar using both methods, but the latter method gives us smaller size coefficients of caregiving indexes.

Table A.11: Hourly Pay and Occupation Caregiving Index

	ln(hourly pay)			
	England (2002)	Caregiving Index		
Caregiving Index	-0.0381***	-0.0038***	-0.0028***	-0.0023***
	(0.0115)	(0.0001)	(0.0001)	(0.0001)
Year Fixed	Y	Y	Y	Y
Individual Fixed	Y	N	Y	Y
Industry Fixed	Y	N	N	Y
Observations	69,952	92,407	95,454	95,454
R-squared	0.5826	0.3130	0.5394	0.5559

*Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.*

In Table A.12 the occupations are divided into quartiles based on their caregiving rank. On average, a female worker suffers a 4.6% hourly pay decrease if she switches from a non-caregiving to a caregiving occupation within the same industry. This pattern is clearer looking at results by quartiles. A female worker gets lower hourly pay by 15.7%, 8.3%, and 5.5% respectively in occupations of quartile 4/3/2 compared to herself working in the least caregiving occupations. Married women suffer more from the penalty of working in caregiving occupations than single women. A married woman earns 20% lower by switching from the least to most caregiving occupations, compared to only 11% for a single woman.

Table A.12: Hourly Pay and Occupation Caregiving Level

	ln(hourly pay)					
	All Women		Single Women		Married Women	
Dummy	-0.0457***		-0.0441***		-0.0739***	
	(0.0051)		(0.0099)		(0.0133)	
Quartile 4		-0.1568***		-0.1095***		-0.2002***
		(0.0095)		(0.0200)		(0.0244)
Quartile 3		-0.0829***		-0.0822***		-0.1400***
		(0.0070)		(0.0159)		(0.0199)
Quartile 2		-0.0546***		-0.0460***		-0.0895***
		(0.0053)		(0.0138)		(0.0161)
Year	Y	Y	Y	Y	Y	Y
Individual	Y	Y	Y	Y	Y	Y
Industry	Y	Y	Y	Y	Y	Y
Observations	95,454	95,454	17,764	17,764	20,233	20,233
R-squared	0.5547	0.5558	0.6194	0.6195	0.5743	0.5753

*Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In columns 2, 4, 6, the least caregiving occupations (the first quartile) is used as the reference group.*

When first collapsing into individuals and then running the same regression without year or fixed effects, the results would be larger (Table A.13), which might be unobserved ability that correlates choice of caregiving occupations that was previously deducted by the individual fixed effects.

Table A.13: Hourly Pay and Occupation Caregiving Level

	ln(hourly pay)		
	All	Single women	Married women
Caregiving Dummy	-0.3476***	-0.1687***	-0.2828***
	(0.0216)	(0.0269)	(0.0310)
Observations	9,502	3,395	3,610
R-squared	0.4322	0.3789	0.4335

*Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.*

A.4.2 Caregiving Occupation as Valid Signal

One unique feature of NLSY is that it follows the children born to NLSY79 mothers in the NLSY Child and Young Adult data. This data contains information on all children born to NLSY female respondents, amounting to a total of 11,530 children as of 2016. NLSY therefore provides unique detailing information both for the children and their mothers by linking the NLSY79 and NLSY Child and Young Adult data. These children are tracked since their birth, and a wide range of child-specific information is recorded, covering demographic variables, health conditions, skills (e.g., math, reading, verbal), habits (e.g., drinking, smoking, use of marijuana), and psychological results (e.g., depression, self-esteem). Since 1986, these children (aged 0 to 23 in 1986) and mothers have been interviewed every two years. In 1994, the Young Adult Survey was first carried out for children who are at least 15,[151] where the questions more resemble those asked to their mothers in NLSY79. The relationship between mothers' occupation and children's outcomes are studied in the Appendix.

As is stressed in the model, men care about the childcaring preferences of women and women then use their occupation choice partly as a signal. For this signal to be valid, single women who work in caregiving occupations should have some "better childcaring outcomes" in the future

[151] There are some special years to note. In 1998, only young adult in 15-20 are surveyed. Since 2010, all children above age of 30 are surveyed every 4 years. Since 2016, those who are 12 and above are surveyed in the Young Adult sample too.

after they get married. Otherwise, if a single woman working as a registered nurse has no difference in childcaring-related outcomes than a single woman working as an avionics technician, then the signal is not trustworthy and would not be valid in equilibrium. I therefore pick three childcaring-related outcome variables – number of children in future after getting married, spending more time caring for children, and having children of higher performance in future. Again, I run the same regression as in wage study and average all variables within occupation-year cells. For the caregiving index, I use the occupation information for single women, and see their future outcomes after they get married to partly avoid the direct reverse causality problem.

Caregiving Occupations and Number of Children in Future

Results on having children in future are shown below in Table A.14. For selection concerns, I further controlled their expressed preference on children. In NLSY, respondents are asked in the first interview in 1979 about their intention on having kids: “How many children do you want to have?” In column 1 to 3, I control the share of single women who say they want at least one child; and in column 4 to 6, I control the share of single women who want to have one/two/three/four/five or more children. Even after controlling for number of children respondents desire when they are single, women working in more caregiving occupations end up more likely to have at least one child, and also on average have more children, despite small difference.

Table A.14: Child Decision and Occupation Choice

	Child Decision					
	Have at least one child			Number of children		
Index	0.000601***			0.00133***		
	(0.000222)			(0.000483)		
Dummy		0.0321***			0.0567**	
		(0.0108)			(0.0228)	
Quartile 4			0.109***			0.0449***
			(0.0367)			(0.0172)
Quartile 3			0.0820**			0.0484***
			(0.0356)			(0.0165)
Quartile 2			0.0577			0.0231
			(0.0373)			(0.0173)
No. of children desired	Y	Y	Y	Y	Y	Y
Year	Y	Y	Y	Y	Y	Y
Observations	2,923	2,923	2,923	2,926	2,926	2,926
R-squared	0.459	0.460	0.460	0.461	0.460	0.461

*Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In column 3 and 6, the least caregiving occupations (the first quartile) is used as the reference group.*

Caregiving Occupations and Time Spent on Childcare

Time spent on caring for children is not provided in NLSY. Here I use data from American Time-Use Survey (ATUS) from the Bureau of Labor Statistics. I didn't find evidence that women working in childcaring occupations actually spend more time caring for children.

Starting from 2003, ATUS provides the diary of respondents on their time expenditure. Respondents are randomly selected from CPS sample [152] for those who are over 15, and will be asked how they have spent their last 24 hours [153] for eight weeks in a row. ATUS therefore provides data on time (in minutes) that respondents spend on caring for children on an average day. Using ATUS data combined with O*NET data, I study whether, among working mothers with at least one child under 13, women working in more childcaring occupations spend more time caring for children. Here only time spent on children under 13 counts, and results on time spent on primary and secondary childcare are both included. [154]

[152] Specifically, only households that completed their final (eighth) CPS interview are considered.
[153] Here the 24-hour window is from 4am the previous day to 4am on the interview day.
[154] According to explanation of ATUS, "Primary childcare activities include time spent providing physical care; playing with children; reading with children; assisting with homework; attending children's events; taking care of children's health needs; and dropping off, picking up, and waiting for children. Passive childcare done as a primary activity (such as "keeping an eye on my son while he swam in the pool") also is included. Secondary childcare is care for children under age 13 that is done while doing an

Table A.15: Time Spent on Childcare and Occupation Choice

	ln(Time spent on childcare)		
Index	0.000146		
	(0.000908)		
Dummy		0.000721	
		(0.0434)	
Quartile 4			0.0469
			(0.0719)
Quartile 3			0.0266
			(0.0703)
Quartile 2			0.0416
			(0.0699)
Year	Y	Y	Y
Observations	3,337	3,337	3,337
R-squared	0.224	0.224	0.224

*Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In column 3, the least caregiving occupations (the first quartile) is used as the reference group.*

Note that since ATUS does not follow respondents over a long time-span like NLSY data, here we cannot use the occupation information while the respondent is single and study the outcome variable after she gets married. Instead, here the results are simply contemporary. Table A.15 shows this result using ATUS data from 2003 to 2019. Based on the results there is no evidence that, among married women with kids under 13, the women who work in more caregiving occupations would spend more time caring for children. However, this result needs to be explained with caution. As is mentioned, the data is contemporary and is within working moms. If single women working in more caregiving occupations are more likely to drop out of the market after getting married and spend more time caring for children, this would not be reflected in the sample we use.

Caregiving Occupations and Child Outcomes

If it is not about the "quantity" of childcare that these occupations send signals about, then what about "quality"? I then turn to the differences in child outcome between occupations.

activity other than primary childcare, such as cooking dinner. Secondary childcare estimates are derived by summing the durations of activities during which respondents had at least one child under age 13 in their care while doing other things."

Table A.16: Measurement for Child Outcomes

measurement	explanation	Age
Health		
Overall health rating	Mother rating of child health	0-15
Whether have health problem	Mom felt or told child needed help for mental/behavior/emotional problem	0-13
Whether need treatment	Child has condition, requires treatment by medical professional	0-13
Whether need medicine	Child has condition that requires medicine	0-13
Whether need equipment	child has condition that requires special equipment	0-13
Ability		
Motor & social percentile	motor & social development: percentile score	0-3
Location memory percentile	memory for location: percentile score	0-3
Behavioral problem percentile	behavioral problems index: total percentile score (lower is better)	4-15
Picture vocabulary percentile	Peabody picture vocabulary test (ppvt): total percentile score	4-15
Verbal memory words percentile	verbal memory for words percentile score	3-6
Verbal memory story percentile	verbal memory for story percentile score	3-6
Math percentile	Piat math percentile	6-15
Cognitive stimulation percentile	Cognitive stimulation percentile score	0-13
emotional support percentile	emotional support percentile score	0-13
reading recognition percentile	reading recognition percentile	0-13
reading comprehension percentile	reading comprehension percentile	0-13
Habit		
Whether smoke	Has child ever smoked a cigarette	0-13
Smoke frequency	How often in past 30 days smoked cigarettes	0-13
Whether drink	Has child ever drunk alcohol	0-13
Drink frequency	How often in last year gotten drunk	0-13
Whether marijuana	Has child ever used marijuana	0-13
Marijuana frequency	How often used marijuana in past 30 days on average	0-13

Note: Data from NLSY.

As mentioned in the Data section, one highlight of NLSY data is that it follows the children born to NLSY79 mothers. NLSY provides many measurements considering ability, habits, health, and achievements of children. For explanation of these measurements, see Table A.16 and Table A.17.

Table A.17: Measurement for Adult Outcomes

measurement	explanation	Age
Health		
Whether receive help	whether receive help for behavioral, emotional, or family problems last year	0-13
Depression percentile	depression CES-D percentile (lower is better)	0-13
Self-esteem percentile	Rosenberg esteem percentile	0-13
Control percentile	control over life, Pearlin percentile	0-13
Outcome		
ln(yearly income)	Cover income both from wage & salary, and from farm & business	0-13
Education (in years)	Years of education reported as of last time interviewed	0-13
Habit		
Whether smoke	Have you ever smoked a cigarette	0-13
Smoke frequency	How often in past 30 days smoked cigarettes	0-13
Whether drink	How many times in last year respondent has gotten drunk	0-13
Drink frequency	On average how often respondent drank in the past 12 months	0-13
Whether marijuana	Have you ever used marijuana	0-13
Marijuana frequency	How often used marijuana in past 30 days on average	0-13

Note: Data from NLSY.

To make the measurement of mother's occupation relevant to the measurements, the time-span of occupation information is tailored to each measurement. For example, the measurement of children's motor development is only relevant for children under the age of 4. Therefore, the occupation information only considers mother's occupation while the children is between 0 and 4. For measurements that are not specific to an age range, we use the mother's information from when the children were under 13.

I study child outcomes in three broad categories: health, ability, and habits. Here I controlled for spouse's occupation caregiving index (if there is a spouse), AFQT of mother, education of mother and spouse (if there is a spouse), net family income, whether mother is employed, whether mother is employed in a full-time job, years that mother reports non-missing caregiving index of occupation, and sex and gender of child. For all the variables related to mothers (and potential spouse), I use the related information in the corresponding time span when children are eligible for the outcome measurement. Table A.18 shows that for child outcomes that are under age of 13, it seems that children of mothers working in more childcaring occupations have slightly better health results, achieve better ability in most measurements of different age ranges, and have no significant differences in habits of smoking, drinking, and marijuana use.

Table A.18: Mother Occupation Caregiving Index and Child Outcome

Panel A: Child Health (all ages)						
	Health Rating	Need Treatment	Need Medicine	Need Equipment		
Caregiving	0.0211**	-0.000583	0.00348	-0.000184		
(dummy)	(0.00970)	(0.00258)	(0.00267)	(0.00147)		
Observations	16,822	40,491	39,979	39,232		
R-squared	0.037	0.009	0.018	0.005		
Panel B: Child Ability (Percentile)						
	Motor and Social	Location Memory	Behavioral Problem	Picture Vocabulary	Verbal Memory Words	Verbal Memory Story
Caregiving	3.460***	1.435	-1.769***	0.984**	2.529***	2.754**
(dummy)	(0.727)	(1.484)	(0.338)	(0.425)	(0.946)	(1.266)
Observations	7,079	1,490	31,937	16,250	3,682	2,180
R-squared	0.050	0.036	0.081	0.338	0.094	0.047
Panel B – continued: Child Ability (Percentile)						
	Math	Cognitive Stimulation	Emotional Support	Reading Recognition	Reading Comprehension	
Caregiving	1.293***	3.582***	2.138***	1.180***	0.897***	
(dummy)	(0.332)	(0.299)	(0.319)	(0.338)	(0.338)	
Observations	27,656	39,477	37,070	28,865	24,749	
R-squared	0.235	0.220	0.160	0.199	0.272	
Panel C: Child Habit						
	Whether Smoke	Smoke Frequency	Whether Drink	Drink Frequency	Whether Marijuana	Marijuana Frequency
Caregiving	-0.00308	0.0594	0.0171**	0.0197*	0.000242	0.0860*
(dummy)	(0.00669)	(0.0717)	(0.00678)	(0.0111)	(0.00350)	(0.0510)
Observations	11,912	1,153	12,903	10,421	12,084	893
R-squared	0.102	0.141	0.089	0.046	0.042	0.251

*Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. For Panel C, mother's habit is also controlled (e.g., whether smoke, smoking frequency, and age of first smoke).*

For adult outcomes, I still use the occupation information for mothers when the children are between 0 and 13, but see their outcome variables after the children are 13 and older. Controls are similar here except that year fixed effects and children's ages are also included in regression of children income or education. Based on results in Table A.19, children with mothers who worked in more caregiving occupations when they were young (age 0 to13 have lower levels of depression and higher feelings of control over life when they become adults (after the age of 15). These children do not seem to earn higher incomes than children whose mothers worked in low caregiving occupations, but they do receive slightly more years of education. As for life habits, there seems to be no difference in smoking, drinking, and use of marijuana for those children. If anything, they seem to be less likely to smoke when they become adults.

Based on the results, considering health, ability, and life habits of children under 13, as well as their health, income and education, and life habits after they becomes adults (after age 15), it seems that mothers working in more caregiving occupations have better outcomes in children. This then provides the incentive for men to make different marriage decisions based on the occupation choices of women. That is to say, women's occupation choices serve as valid signals on the marriage market.

Table A.19: Mother Occupation Caregiving Index and Adult Outcomes

Panel A: Adult Health						
	Whether Receive help	Depression Percentile	Self-Esteem Percentile	Control Percentile		
Caregiving	0.00298	-0.899*	0.856*	1.226**		
(dummy)	(0.00273)	(0.489)	(0.471)	(0.564)		
Observations	47,650	15,337	15,582	11,337		
R-squared	0.017	0.022	0.050	0.019		
Panel B: Adult outcome						
	ln(Yearly Income)	Education (in Years)				
Caregiving	0.0111	0.161***				
(dummy)	(0.0136)	(0.0115)				
Observations	31,678	153,783				
R-squared	0.552	0.244				
Panel C: Adult Habit						
	Whether Smoke	Smoke Frequency	Drink Frequency1	Drink Frequency2	Whether Marijuana	Marijuana Frequency
Caregiving	-0.00421	-0.0878***	0.0106	0.0485*	0.00176	0.00188
(dummy)	(0.00551)	(0.0300)	(0.0271)	(0.0253)	(0.00516)	(0.0468)
Observations	23,901	10,838	6,779	39,481	27,943	5,665
R-squared	0.084	0.232	0.080	0.137	0.071	0.255

*Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. For Panel C, mother's habit is also controlled (e.g., whether smoke, smoking frequency, and age of first smoke).*

A.4.3 Policy Shock for Separated, Divorced, and Widowed Women

In all previously showed specifications, women's martial statuses are measured as either single, which includes only women who have never been married, or married, which includes women who are currently married. Here I run the baseline specification within women who are currently separated, divorced, and widowed, categories which are not included in any sample in this empirical part. Note that here I include all women and no longer restrict it to the same group of young women, defined as under the age of 38. The reason is that the average age of never-married women is younger than women who are separated, divorced, and widowed, and thus using 38 as the age cutoff is no longer appropriate, as shown in Appendix Figure F1.

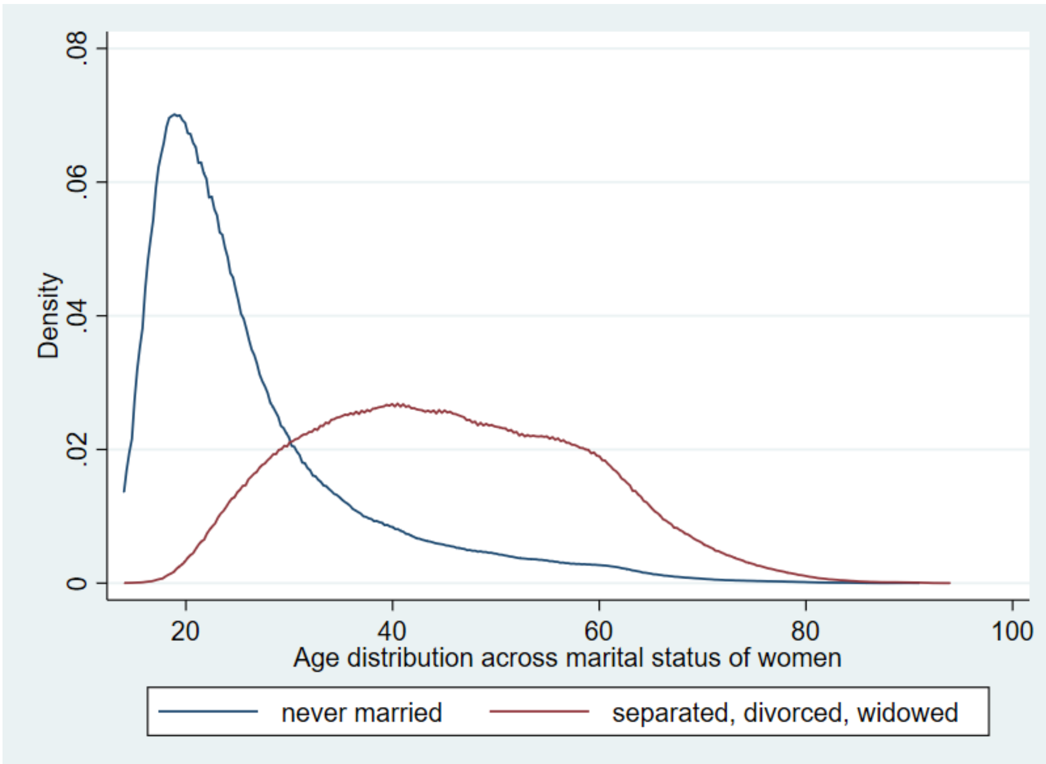

Figure A.8: Age Distribution Among Working Women, by Marital Status

Note: Data from CPS, 1968-2000.

Results are shown in Table A.20, where column 1 and column 3 are within single women while column 2 and column 4 are within separated, divorced, and widowed women. There is a decrease in working in caregiving occupations post policy shock among separated, divorced, and widowed too, but smaller in size than with single women. Note that the results in column 1 and 3 are less significant with inclusion of old single women, and results would be slightly larger in size and more significant if restricting to young single women. For the separated, divorced, and widowed group, the results are slightly larger and less significant if restricting to young women defining as under 38.

Table A.20: Unilateral Divorce Policy and Women's Occupation Choice for Separated, Divorced, and Widowed

	Caregiving Index		Caregiving Dummy	
	Single	Separated, divorced, widow	Single	Separated, divorced, widow
Treated	-0.828	-0.745**	-0.0231*	-0.0144**
	(0.660)	(0.345)	(0.0135)	(0.00702)
Property Regime	Y	Y	Y	Y
Year	Y	Y	Y	Y
State	Y	Y	Y	Y
Observations	132,949	101,550	132,949	101,550
R-squared	0.011	0.012	0.015	0.012

*Note: Standard errors (clustered at states) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Controls include age, age square, and race. In columns 1 and 2, the dependent variable is the caregiving index. In columns 3 and 4, the dependent variable is the caregiving dummy. Columns 1 and 3 use all single women. Columns 2 and 4 use all women who are separated, divorced, and widowed.*

APPENDIX B. COVID19 AND CONSUMER ANIMUS TOWARDS CHINESE PRODUCTS: EVIDENCE FROM AMAZON DATA

B.1 Data Appendix

Table B.1: Examples of Informative Reviews by Category

Category	Count	Example
Informative-animus	826	They ship from China you know where the virus first broke out
Informative-neutral	243	Reasonable price for standard basic face mask made in China.
Informative with quality complaints	276	Don't buy it, this is poison mask made in China, after using it for couple of minutes it gave me headache, confusion
Informative without quality complaints	793	The masks look good, but are made in China, and the box says Non-Medical!

Table B.2: Examples of Informative-Animus Reviews at Each Rating

Rating	Review Title	Review Body
5	comfortable mask, finally	Liked the mask. Disliked the fact that it is made in China!
4	Made in China!?	These masks seem to be working ok. The biggest disappointment was that they were made in China. After the Covid Pandemic, we are very suspicious of ANY items made in China.
3	Straight Outta China	Came straight from a factory in China, not exactly what I was looking for during a pandemic that started there. Seem to be VERY cheaply made.
2	china	they ship from china you know where the virus first broke out
1	Crap	Made in china!!! Nuff said

Table B.3: Evidence from Reviews on Why Buying a Chinese Product Despite Animus

Review Title	Review Body
Good product but made in China	… Should have researched the manufacturer more fully. Had I known they were made in China, most likely would not have purchased this.
Made in CHINA	Warning: these come directly from China. Package is written in Chinese. I have thrown them away. Make your own mask; it will be safer.
FYI These are made in China	Sold by a US company, but made in China.
Fast shipping--nice mask	I questioned getting a mask from China, but couldn't find one in the US. …
It’s Just “okay”, here’s why	When I purchased this 2 weeks ago, it had enough good reviews and the fact that there was only few to select from on Amazon hat WASNT made in China. I purchased few. …
Works well for me	I find it troubling that it's so difficult to find masks on Amazon that is not from China. I spent nearly an hour digging into every surgical mask trying to find ones not from China. …

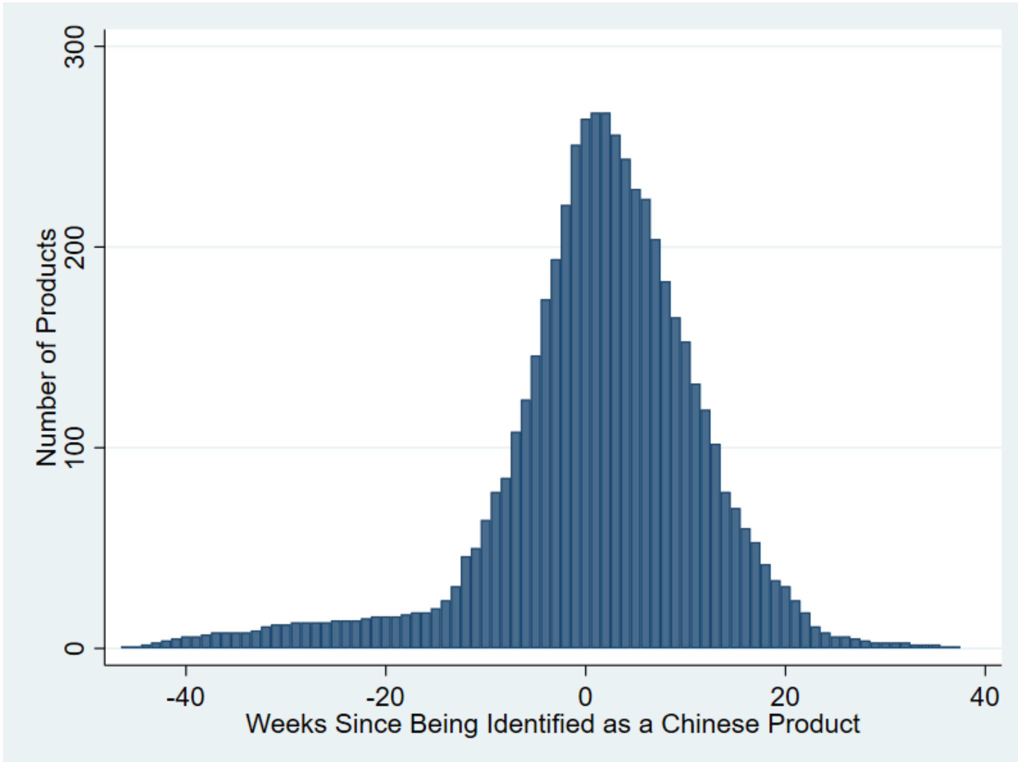

Figure B.1: Number of Products by Time Difference k (in Weeks)

B.2 Empirical Appendix

B.2.1 Direct Impact – Ratings of Informative Reviews

The sample is limited within Chinese products that have been identified by the end of research period. Below is the empirical model.

$$Y_{itc} = \alpha + \beta D_{itc} + \beta_X X_{itc} + \theta_i + \theta_t + \varepsilon_{itc} \tag{B.1}$$

Data is on review level, varying by product i, time t, and review c. The dependent variable is the rating of a review. D is dummy here, denoting either being informative or informative-animus. Other variables bear the same meaning as in the main regression.

Table B.4: Ratings of Informative Reviews

	Rating of a Review	
	(1)	(2)
Informative	-1.346***	
	(0.0721)	
Informative-Animus		-1.817***
		(0.0706)
Controls	Yes	Yes
Date	Yes	Yes
ASIN	Yes	Yes
Observations	37,037	37,037
R-squared	0.161	0.169

*Note: Standard errors in parentheses, clustered by ASIN, *** p<0.01, ** p<0.05, * p<0.1 The sample is limited to Chinese products that are identified by the end of research period. Controls include price, sales rank, and average rating of a product. In column (1), the dummy denotes whether a review is informative. In column (2), the dummy denotes whether a review is informative-animus.*

B.2.2 Direct Impact – Correlation of Product Average Rating Over Time

For this regression, we use the full sample (including products of all countries-of-origin, including the unknowns). Below is the empirical model.

$$Y_{it} = \alpha + \beta Y_{it-1} + \beta_X X_{it} + \theta_i + \theta_t + \varepsilon_{it} \qquad \text{(B.2)}$$

Data is panel, and varies by product i and time (day) t. The dependent variable is the average rating of product i on day t, and Y_{it-1} is the lag of average rating of product i. Other variables bear the same meaning as in the main regression.

Table B.5: Correlation of Product Average Rating by Day

	Average Rating	
	(1)	(2)
Lag Average Rating	0.983***	0.968***
	(0.000645)	(0.00128)
Controls	No	Yes
Date	No	Yes
ASIN	No	Yes
Observations	150,541	150,170
R-squared	0.981	0.981

*Note: Standard errors in parentheses, clustered by ASIN, *** p<0.01, ** p<0.05, * p<0.1 The sample includes all products in our data. In column (1), no controls or fixed effects are included. In column (2), price, sales rank, product fixed effect, and date fixed effect is controlled.*

B.2.3 Identification of Country-of-Origin and Review-Rating Ratio

For this regression, the sample is limited to products with known countries-of-origin. Below is the empirical model.

$$Y_{it} = \alpha + \beta D_{it} + \beta_X X_{it} + \theta_i + \theta_t + \varepsilon_{it} \tag{B.3}$$

Data is panel, and varies by product i and time (day) t. The dependent variable is the ratio of number of reviews over number of ratings for product in on day t, and D_{it} is the dummy of whether the product's country-of-origin has been ever identified for product i on day t. Other variables bear the same meaning as in the main regression.

Table B.6: Review-Rating Ratio and Country-of-Origin Identification

	Review-Rating Ratio	
	(1)	(2)
Whether Identified	-0.0123**	-0.00469
	(0.00484)	(0.0117)
Controls	Yes	Yes
Date	Yes	Yes
ASIN	Yes	Yes
Observations	49,673	20,852
R-squared	0.107	0.132

*Note: Standard errors in parentheses, clustered by ASIN, *** p<0.01, ** p<0.05, * p<0.1 Controls include price, sales rank, and average rating. In column (1), only Chinese products are included (identified by the end of the research period). In column (2), products of known countries-of-origin besides China are included.*

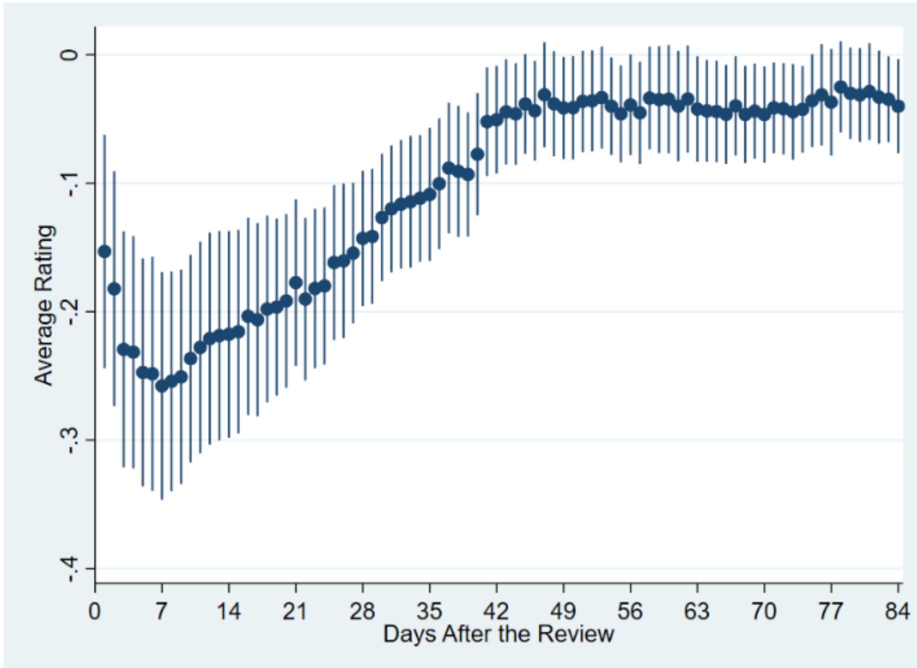

Figure B.2: Average Rating and Share of Informative-Animus Reviews (Total Impact)

Note: Standard errors clustered by ASIN. Controls include price, sales rank, and lag of share of informative-animus reviews, with the x-axis showing the days for this lag.

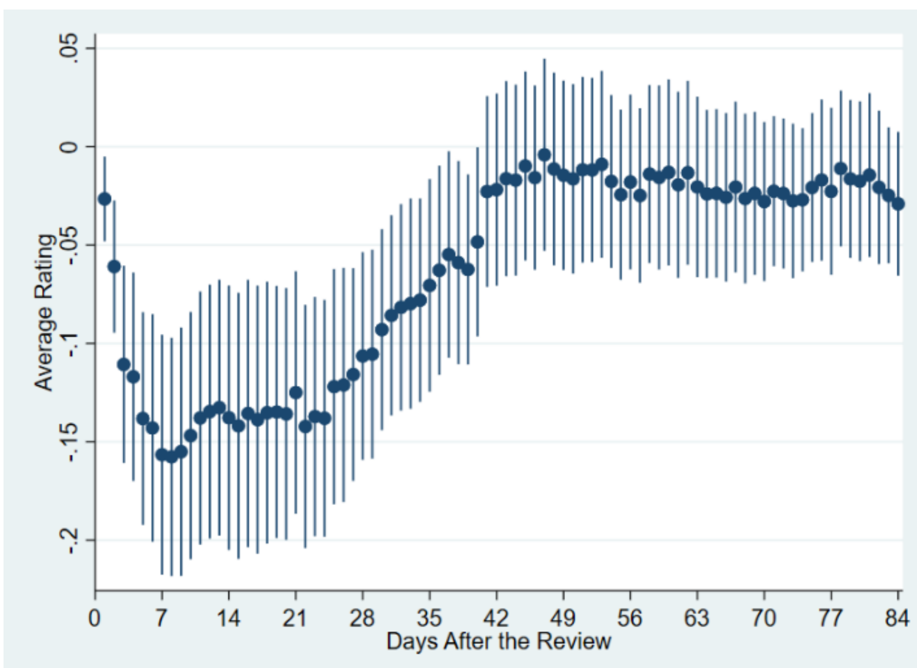

Figure B.3: Average Rating and Share of Informative-Animus Reviews (Indirect Impact)

Note: Standard errors clustered by ASIN. Controls include price, sales rank, lag of share of informative-animus reviews, and corresponding lag of average rating, with the x-axis showing the days for this lag.

B.3 Amazon Appendix

B.3.1 Review Regulation on Amazon

A consumer can leave a comment without buying the product, but such comment will not have the "verified purchase" label, and highly likely will be removed by Amazon (and it needs approval to be displayed in this case). Therefore, most of the reviews/comments are from consumers who actually purchase the product. [155] Note that a product review is different from seller feedback. A consumer can rate a seller or provide seller feedback, but this is available only when a consumer clicks into the seller details. In this paper, we focus on consumer reviews.

Offensive language would be removed by Amazon, and therefore we did not find any reviews with discriminative names for Chinese. [156]

After a consumer orders from a third-party, she can leave a review or ratings within 90 days from the date of order. If a consumer leaves a review before the arrival of the product, it will not have the "verified purchase" label.

For the calculation of average rating of a product, Amazon now uses a machine-learning model instead of simple/unweighted average. [157] This algorithm is not revealed to the public and applies multiple criteria on review authenticity. Amazon does not take into account ratings without "verified purchase" into calculation of product average rating until more details are added (e.g., texts, images, or videos).

---

[155] Look at an example at Amazon.com: Customer reviews: Black Disposable Face Masks, 100 Pack Black Face Masks 3 Ply Filter Protection. On March 16th, 2022, out of a total of 4844 reviews, 4703 reviews (or 97%) are verified purchases.

[156] For more details on regulations of reviews for Amazon, refer to https://www.amazon.com/gp/help/customer/display.html?nodeId=G5T39MTBJSEVYQWW.

[157] Refer to https://www.amazon.com/gp/help/customer/display.html?nodeId=GQUXAMY73JFRVJHE for Amazon's own explanation.

APPENDIX C. LEARNING FROM INTERNATIONAL TRADE: ASYMMETRIC CULTURAL TRANSMISSION AND GENDER DISCRIMINATION

C.1 Data Appendix

Table C.1: Top 20 Countries or Regions of FDI Source, Export, and Import

	FDI		Export		Import	
Rank	Country/region	percent	Country/region	percent	Country/region	percent
1	Hong Kong (China)	31.33	United States	21.06	Japan	16.80
2	Virgin Islands	11.10	Hong Kong (China)	17.01	Taiwan (China)	11.55
3	South Korea	10.30	Japan	12.38	South Korea	11.09
4	Japan	8.99	South Korea	4.68	United States	7.97
5	United States	6.50	Germany	4.01	Germany	6.92
6	Taiwan (China)	5.14	Netherlands	3.12	China	5.39
7	Cayman Islands	3.37	United Kingdom	2.53	Malaysia	3.25
8	Singapore	3.31	Taiwan (China)	2.27	Singapore	2.50
9	Samoa	1.86	Singapore	2.14	Russian Federation	2.16
10	Germany	1.75	France	1.67	Hong Kong (China)	2.10
11	Netherlands	1.34	Italy	1.55	Thailand	2.05
12	United Kingdom	1.31	Russian Federation	1.53	Australia	2.05
13	Australia	1.09	Australia	1.49	Philippines	1.62
14	France	1.08	Canada	1.37	Brazil	1.54
15	Canada	1.01	Malaysia	1.36	India	1.37
16	Mauritius	0.99	United Arab Emirates	1.15	France	1.37
17	Macao (China)	0.90	Indonesia	1.05	Saudi Arabia	1.34
18	Bermuda	0.70	India	1.00	Canada	1.31
19	Malaysia	0.64	Belgium	0.99	Indonesia	1.29
20	Italy	0.46	Thailand	0.98	Italy	1.15
Top 3	52.74		50.45		39.44	
Top 5	68.23		59.14		54.33	
Top 10	83.66		70.87		69.72	
Top 20	93.19		83.35		84.80	

Note: FDI data comes from the China Statistical Yearbook published by National Bureau of Statistics in China at http://www.stats.gov.cn/. Export and import data are calculated based on customs data of China in 2004. For FDI, Virgin Islands (2), Cayman Islands (7), Samoa (9), Mauritius (16), Bermuda (18) are typical free ports in investments. Considering the real source of the investment from these countries/regions, then the FDI source countries might be even more concentrated than the data shows in the table. Another to note is that China is the fifth largest origin of its own import. This is because many firms try to benefit from the tax refund policy for export so there is a fairly high amount of re-import in China.

Table C.2: Calculation of GGI and Four Sub-Indexes

Sub-indexes	Ratios	Weights
Economic Participation and Opportunity	Female labor force participation over male value	0.199
	Wage equality between women and men for similar work (converted to Female-over-male ratio)	0.310
	Estimated female earned income over male value	0.221
	Female legislators, senior officials, and managers over male value	0.149
	Female professional and technical workers over male value	0.121
	Total	1
Educational Attainment	Female literacy rate over male value	0.191
	Female net primary level enrolment over male value	0.459
	Female net secondary level enrolment over male value	0.230
	Female gross tertiary level enrolment over male value	0.121
	Total	1
Health and Survival	Female healthy life expectancy over male value	0.307
	Sex ratio at birth (converted to female over male ratio)	0.693
	Total	1
Political Empowerment	Women with seats in parliament over male value	0.310
	Women at ministerial level over male value	0.247
	Number of years of a female head of state (last 50 years) over male value	0.443
	Total	1

Table C.3: Rank of All Countries by GGI

Rank	Country	Rank	Country	Rank	Country
1	Sweden	40	Thailand	79	Japan
2	Norway	41	Argentina	80	Gambia
3	Finland	42	Mongolia	81	Malawi
4	Iceland	43	Lesotho	82	Ecuador
5	Germany	44	Poland	83	Cyprus
6	Philippines	45	Trinidad and Tobago	84	Madagascar
7	New Zealand	46	Romania	85	Zambia
8	Denmark	47	Uganda	86	Kuwait
9	United Kingdom	48	Ukraine	87	Bolivia
10	Ireland	49	Russian Federation	88	Mauritius
11	Spain	50	Slovak Republic	89	Cambodia
12	Netherlands	51	Slovenia	90	Tunisia
13	Sri Lanka	52	Kyrgyz Republic	91	Bangladesh
14	Canada	53	Czech Republic	92	Korea, Rep.
15	Australia	54	Georgia	93	Jordan
16	Croatia	55	Hungary	94	Nigeria
17	Moldova	56	Luxembourg	95	Guatemala
18	South Africa	57	Venezuela	96	Angola
19	Latvia	58	Ghana	97	Algeria
20	Belgium	59	Dominican Republic	98	India
21	Lithuania	60	Peru.	99	Mali
22	Colombia	61	Albania	100	Ethiopia
23	United States	62	Nicaragua	101	United Arab Emirates
24	Tanzania	63	China	102	Bahrain
25	Jamaica	64	Paraguay	103	Cameroon
26	Switzerland	65	Singapore	104	Burkina Faso
27	Austria	66	Uruguay	105	Turkey
28	Macedonia	67	Brazil	106	Mauritania
29	Estonia	68	Indonesia	107	Morocco
30	Costa Rica	69	Greece	108	Iran
31	Panama	70	France	109	Egypt
32	Kazakhstan	71	Malta	110	Benin
33	Portugal	72	Malaysia	111	Nepal
34	Botswana	73	Kenya	112	Pakistan
35	Israel	74	Honduras	113	Chad
36	Uzbekistan	75	Mexico	114	Saudi Arabia
37	Bulgaria	76	Zimbabwe	115	Yemen
38	Namibia	77	Italy		
39	El Salvador	78	Chile		

C.2 Model Appendix

C.2.1 Conditions to Guarantee the Validity of Hiring Rules

Now we consider the restrictions the payoffs need to satisfy for firms to follow the rule that a worker is hired only when the signal is good. It must be that for a worker with good signal, the expected payoff is positive while for a worker with a bad signal, the expected payoff is negative.

$$\begin{aligned}&\text{expected payoff for hiring a female worker with good signal}\\&\quad = prob(I_i = 1 \mid S_i = good) * \mu_q + prob(I_i = 0 \mid S_i = good) * \mu_n\\&\quad = T\mu_q + (1-T)\mu_n > 0\\&\text{expected payoff for hiring a female worker with bad signal}\\&\quad = prob(I_i = 1 \mid S_i = bad) * \mu_q + prob(I_i = 0 \mid S_i = bad) * \mu_n\\&\quad = \frac{(1-P)q\mu_q + P(1-q)\mu_n}{(1-P)q + P(1-q)} < 0\end{aligned}$$

We therefore derive the condition for payoffs to let the firm fully trust the signal:

$$\begin{cases}\mu_q > \dfrac{1-T}{T}(-\mu_n) = \dfrac{(1-P)(1-q)}{Pq}(-\mu_n)\\[2ex] \mu_q < \dfrac{P(1-q)}{(1-P)q}(-\mu_n)\end{cases} \Rightarrow \frac{(1-P)(1-q)}{Pq}(-\mu_n) < \mu_q < \frac{P(1-q)}{(1-P)q}(-\mu_n)$$

For the condition to be meaningful, we need the upper limit to be larger than the lower limit:

$$\frac{(1-P)(1-q)}{Pq}(-\mu_n) < \frac{P(1-q)}{(1-P)q}(-\mu_n) \Rightarrow P > \frac{1}{2}$$

Therefore, the condition we derived above is plausible only when P is larger than a half, which is always satisfied as long as the signal is informative.

C.2.2 Taste-Based Discrimination Model

Another commonly used model in discrimination is taste-based discrimination models. Under the same setup as in this paper, I briefly show here that a taste-based discrimination model would generate opposite predictions on profits compared to the statistical model I use in this paper.

If taste-based, then there is a distaste parameter α for female workers. Similarly, a firm only hires a female worker with good signal, but this requires that the distaste for female workers cannot be too large: The expected utility for hiring a female work with good signal is:

$$
\begin{aligned}
&= prob(I_i = 1 \mid S_i = good) * \mu_q + prob(I_i = 0 \mid S_i = good) * \mu_n - \alpha \\
&= T\mu_q + (1-T)\mu_n - \alpha > 0
\end{aligned}
$$

As long as the original α is not too large, female fraction within a firm takes the same function form as in statistical discrimination set-up with:

$$\frac{\partial L_m}{\partial \alpha} = 0,\ \frac{\partial L_f}{\partial \alpha} = 0,\ \frac{\partial T}{\partial \alpha} = 0$$

The wage paid to female workers, however, is lower than the statistical discrimination case:

$$w_f = \delta\left(T\mu_q + (1-T)\mu_n - \alpha\right)$$

Note that the distaste affects the employer's utility, but does not directly affect a firm's profit in monetary value. The expected profit a firm gets from a female worker is then:

$$
\begin{aligned}
&= T\mu_q + (1-T)\mu_n - \delta\left(T\mu_q + (1-T)\mu_n - \alpha\right) \\
&= (1-\delta)\left(T\mu_q + (1-T)\mu_n\right) + \delta\alpha
\end{aligned}
$$

The profit of the firm is written as:

$$\pi = L_m(1-\delta)\mu_q + L_f\left((1-\delta)\left(T\mu_q + (1-T)\mu_n\right) + \delta\alpha\right)$$

Profit per worker is then:

$$U = \frac{\pi}{L_m + L_f} = \frac{L_m(1-\delta)\mu_q + L_f\left\{(1-\delta)\left(T\mu_q + (1-T)\mu_n\right) + \delta\alpha\right\}}{L_m + L_f}$$

As mentioned above, when α is not large, hiring decision does not change with distaste parameter: $\partial L_m/\partial\alpha = 0,\ \partial L_f/\partial\alpha = 0,\ \partial T/\partial\alpha = 0$. It is obvious that profit per worker is increasing in α. Therefore, firms earn smaller profits per worker when exporting to more gender-equal cultures (smaller α) and larger profits per worker when exporting to less gender-equal cultures (larger α). Under the same model setup, the change in profit per worker in the taste-based discrimination case is different from what would be predicted in the statistical discrimination case.

C.2.3 Calculation of GGI (Transaction Frequency)

If considering the transaction frequency, GGI can be calculated as:

$$GGI_i = \max_c \{F(n_{ic}) * (GGI_c - GGI_{China}) + GGI_{China}\}$$

Where GGI_c is the GGI of origin/destination country c and GGI_{China} is the GGI value of China. Here n_{ic} is the total number of transactions firm i has with country c, and F(.) is an increasing function of transaction number with the value of F(.) lying between 0 and 1. In my paper, for simplicity, I use the cumulative distribution function of normal distribution as F(.).

C.3 Empirical Appendix

C.3.1 Conditions to Guarantee the Validity of Hiring Rules

Skill Ratio

Skill ratio is measured by number of high-skill workers over the total number of employees in a firm. A worker is identified as high-skill worker as long as he/she has a college degree, bachelor's degree, master's degree or above.

Output

Here I use log of output. Output is the monetary value of the total production of the firm, measured in 1000 RMBs.

Firm Age

Here I use the log of firm age, where firm age is the number of years a firm has operated. In Industrial Firm Database, the opening year of the firm is recorded. Since the main data I use in this paper is 2004, firm age thus equals to 2004 minus the opening year plus one.

Firm Size

Here I use log of firm size. Firm size is measured by the number of employees a firm hires at the end of a year.

Processing Trade

In customs data, it reports the trade mode for each transaction. Here processing trade is the fraction of trade value that belongs to processing trade for an exporter/importer in a year.

Ownership

Following Brandt et al. (2012), I classify firm into different ownership groups based on their registration type: state-owned, hybrid/collective, private, foreign, HMT, and others. The limited liability corporations and shareholding corporations are classified into one of the above groups based on their registered capital.

C.3.2 Empirical Results

Table C.4: Sub-Index

	Female Fraction			
	All	All	Only Better	Only Worse
GGI	0.0281*	0.0785***	0.0899***	0.0251
	(0.0170)	(0.0120)	(0.0260)	(0.0476)
GGI*Firm age	-0.00356***	-0.00208***	-0.00215***	-0.00247***
	(0.000271)	(0.000211)	(0.000214)	(0.000736)
Ownership	Yes	Yes	Yes	Yes
Industry Fixed	No	Yes	Yes	Yes
Province Fixed	No	Yes	Yes	Yes
Observations	53,671	53,671	45,631	8,040
R-squared	0.212	0.448	0.437	0.516

*Note: Standard errors in the parenthesis, clustered at county level. *** p<0.01, ** p<0.05, * p<0.1.*

Table C.5: Controlling Price

	Female Fraction			
	ln(price)		Relative Price	
	More	Less	More	Less
GGI	0.168***	0.0195	0.163***	-0.0379
	(0.0326)	(0.131)	(0.0328)	(0.129)
GGI*Firm age	-0.00192***	-0.00200***	-0.00209***	-0.00203***
	(0.000213)	(0.000637)	(0.000211)	(0.000643)
Ownership	Yes	Yes	Yes	Yes
Industry Fixed	Yes	Yes	Yes	Yes
Province Fixed	Yes	Yes	Yes	Yes
Observations	45,701	7,970	45,701	7,970
R-squared	0.441	0.514	0.438	0.514

*Note: Standard errors in the parenthesis, clustered at county level. *** p<0.01, ** p<0.05, * p<0.1.*

Table C.6: Profit (Impute 0.001)

	ln(Profit)			
	All	All	Only Better	Only Worse
GGI	4.546***	3.657***	6.389***	3.145
	(0.679)	(0.631)	(1.052)	(4.364)
GGI*Firm age	-0.199***	-0.193***	-0.198***	-0.235***
	(0.0116)	(0.0115)	(0.0124)	(0.0285)
Ownership	Yes	Yes	Yes	Yes
Industry Fixed	No	Yes	Yes	Yes
Province Fixed	No	Yes	Yes	Yes
Observations	52,951	52,951	45,063	7,888
R-squared	0.091	0.111	0.120	0.071

*Note: Standard errors in the parenthesis, clustered at county level. *** p<0.01, ** p<0.05, * p<0.1.*

Table C.7: Profit (Add a constant)

	ln(Profit)			
	All	All	Only Better	Only Worse
GGI	0.00253	0.00385*	0.0124**	0.00107
	(0.00351)	(0.00225)	(0.00527)	(0.00188)
GGI*Firm age	-2.11e-05	-3.12e-05	-4.21e-05	-5.96e-05
	(5.55e-05)	(4.38e-05)	(4.78e-05)	(4.22e-05)
Ownership	Yes	Yes	Yes	Yes
Industry	No	Yes	Yes	Yes
Province	No	Yes	Yes	Yes
Observations	52,950	52,950	45,062	7,888
R-squared	0.043	0.064	0.076	0.658

*Note: Standard errors in the parenthesis, clustered at county level. *** p<0.01, ** p<0.05, * p<0.1.*

VITA

Danyang Zhang

Education

Ph.D. in Economics, Purdue University, May 2022 (Expected)

Dissertation Committee: Timothy Bond (co-chair), Chong Xiang (co-chair), Farid Farrokhi, Miguel Sarzosa

M.S. in Economics, Purdue University, August 2019

M.S. in Economics, Central University of Finance and Economics, June 2016

B.A. in Economics, Central University of Finance and Economics, July 2014

Research Interests

Labor Economics, International Trade, Applied Microeconomics

Working Papers

"Marriage Market Signaling and Women's Occupation Choice" (Job Market Paper)

"COVID-19 and Consumer Animus towards Chinese Products: Evidence from Amazon Data", with Alma Cortes and Yipu Deng

"Learning from International Trade: Asymmetric Cultural Transmission and Gender Discrimination"

"Export, Wage Distortion, and Income Inequality – A Study based on Chinese Experience", with Yan Zhang, Zhuang Miao, and Kuo Feng

Work in Progress

"Gender Inequality within the Household: Evidence from Childcare Prices", with Timothy Bond

"Reputation, Optimal Quality Investment and Firm's Export Decisions", with Jiatong Zhong

Conferences and Seminars

"Marriage Market Signaling and Women's Occupation Choice"

2022: American Economic Association Annual Meeting (Poster), Online

2021: Southern Economic Association Annual Meeting, Houston, TX

"COVID-19 and Consumer Animus towards Chinese Products: Evidence from Amazon Data", with Alma Cortes and Yipu Deng

2022: APPAM Fall Research Conference, Austin TX

2021: Southern Economic Association Annual Meeting, Houston, TX; Annual Conference on Global Economic Analysis, Online

2020: Southern Economic Association Annual Meeting, Online

"Learning from International Trade: Asymmetric Cultural Transmission and Gender Discrimination"

2019: Southern Economic Association Annual Meeting, Fort Lauderdale, FL; Krannert Doctoral Research Symposium, Purdue University

2018: Krannert Doctoral Research Symposium, Purdue University

"Export, Wage Distortion, and Income Inequality – A Study based on Chinese Experience"

2015: Economics Forum for Young Scholars in China, Xiamen, China; China Trade Research Group Annual Meeting, Dalian, China

2014: China Society of World Economics Annual Meeting, Hangzhou, China; Forum on Discipline Development of International Trade in China, Beijing, China

Honors and Awards

IPUMS Excellence in Research Award (Best Graduate Student Work), 2022

James A. Papke Graduate Fellowship Award, 2021

Krannert Distinguished Teaching Award, 2021 & 2020

Krannert Doctoral Research Fund, 2021

Krannert Ph.D. Research Symposium Award, 2019

Outstanding Master Graduate of Beijing (China), 2016

Outstanding Thesis Award of Central University of Finance and Economics, 2016

National Scholarship for Masters (China), 2015

Outstanding Graduate of Central University of Finance and Economics, 2014

Teaching Experience

Course Instructor: International Trade, Summer 2021 (Evaluation: 4.4/5), Summer 2020 (5.0/5); Microeconomics, Summer 2018 (4.5/5)

Recitation Instructor: Principles of Economics, Spring 2017

Teaching Assistant (Undergraduate): International Trade, Spring 2022, Spring 2020, Spring 2019, Spring 2018; Economics of Race and Gender Discrimination, Fall 2020; International Economics, Spring 2019; Intermediate Microeconomic Theory, Fall 2016; Macroeconomics, Fall 2016

Teaching Assistant (Master): Wage Discrimination, Fall 2020; International Economics, Summer 2019; Computing for Analytics, Summer 2017; Intermediate Economics II: Macroeconomics, Spring 2017

Teaching Assistant (Ph.D.): Mathematical Analysis for Economists, Fall 2017; Microeconomic Theory, Fall 2017

Research Experience

Research Assistant to Professor Timothy Bond, Fall 2020 – present

Research Assistant to Professor Chong Xiang, Summer 2017 – present

Research Assistant to Professor Farid Farrokhi, Fall 2018 - Spring 2020

Research Assistant to Professor Yaroslav Rosokha, Summer 2017

Research Assistant to Professor Yan Zhang, Fall 2012 - Summer 2016