The London School of Economics and Political Science

Essays in Gender Economics

Jay Euijung Lee

A thesis submitted to the Department of Economics
for the degree of Doctor of Philosophy

September 30, 2020

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of approximately 35,000 words.

Statement of conjoint work

I confirm that Chapter 2 was jointly co-authored with Martina Zanella, and I contributed 50% of this work.

Acknowledgments

The PhD journey brought me many pieces of joy, but it was not always easy. It is the people that I had by me that gave me the strength to reach its end.

I am deeply grateful to my supervisors, Maitreesh Ghatak and Guy Michaels, for their guidance and support. I am where I am as a researcher thanks to their invaluable advice, mentorship, and continued patience. I would also like to thank the other members of my supervisory team, Oriana Bandiera and Robin Burgess. My research benefitted immensely from their commitment in the lead up to and during the job market. I am forever indebted to Tim Besley and Gharad Bryan, for offering so much of their time and care. I am also grateful to many others for insightful discussions on my research, including Francesco Caselli, Alessandro Gavazza, Ethan Ilzetzki, Chad Jones, Camille Landais, Alan Manning, Rachel Meager, Rachel Ngai, Torsten Persson, Steve Pischke, Johannes Spinnewijn, Silvana Tenreyro, John Van Reenen, and Alwyn Young. I would also like to thank seminar participants and colleagues at various conferences and workshops, STICERD, and the Economics Department, and the Ronald Coase Institute.

I thank my friends at LSE for their thoughtful feedback and the times of fun together. I especially thank Shan Aman-Rana, Adrien Bussy, Svetlana Chekmasova, Dita Eckardt, Krittanai Laohakunakorn, Martina Zanella, and Celine Zipfel. Room 4.06 was an awesome office thanks to Karun Adusumilli, Patrick Coen, Weihan Ding, Ruth Fortmann, Will Matcham, Wolfgang Ridinger, and Claudio Schilter.

I also thank Lubala Chibwe, Mike Rose, Nic Warner, Anna Watmuff, and Mark Wilbor for excellent administrative and IT help.

I am extremely grateful to my friends, Jeong Young Hong, Jisoo Hong, Haelin Kim, Danbee Lee, Yoon Lee, and Ksenia Zvonova, for being my emotional anchor. You enriched my world outside of research.

Words cannot describe how grateful I am to my mom, dad, and my brother, Hojae, for their endless love and encouragement. I am blessed to have you. Finally, I want to thank Horng Chern Wong, for teaching me what true love means. I cannot wait for our future together.

Abstract

This thesis investigates how gender norms limit women's opportunities or hinder their performance, using a combination of macro and micro approaches. The first chapter quantifies the cost for aggregate economic output of conservative norms, which glue married women to the home. I develop a model of education, marriage, and labor force participation. I use the model and U.S. census data from 1940 to 2010 to estimate the homemaker norms wedge, which affects how married women discount their market wage when choosing between market work and home production. I then use the model to conduct counterfactual exercises, by taking current American families and assigning them different values of norms wedges. With norms wedges of 1940, aggregate output in 2010 falls by 3.5%; half of this decline is driven by lower education levels and the other half by an inefficient allocation of labor. The second chapter studies how highly male-dominated political parties react to the introduction of gender quotas in the context of South Korean municipal council elections. The results show that parties that are more intensely affected by the quota initially counter its effect by putting forth fewer female candidates where they can. However, those parties gradually increase the number of female candidates and three election cycles later, they field more female candidates – over and above the quota rules. The evidence suggests that this evolving response stems from parties becoming more favorably disposed towards women candidates, as they learn about the competence of women councilors. The third chapter estimates the afore-mentioned homemaker norms wedges for various countries around the world. It demonstrates that there is substantial cross-country variation and shows that differences in cultures regarding gender roles is an important explanatory factor.

Contents

List of Figures

List of Tables

Chapter 1

Marriage and Misallocation: Evidence from 70 Years of U.S. History

1.1 Introduction

Many women shift their time from the labor market to home production upon marriage (Lundberg and Pollak, 2007).[1] If there are efficiency reasons to specialize between income-generating activity and home production, this shift might enhance productivity (Becker, 1981, 1991; Pollak, 2013). On the contrary, the shift might represent misallocation. The need to fulfill the traditional gender role of the homemaker, who stays home to look after the household, might prevent some married women from following their comparative advantage and working in the market.

By how much do traditional gender norms in marriage constrain aggregate output? This paper aims to quantify the effect on aggregate output of the change in the "homemaker" gender role, in the U.S. between 1940 and 2010. While there is ample micro evidence on how gender roles curtail the market work of women (e.g. Field et al., 2019; Bedi et al., 2018; Couprie et al., 2017), little is known about the aggregate implications of gender norms. Existing papers on the aggregate implications of gender differences such as labor market and educational market discrimination (Hsieh et al., 2019) and nonmarket time (Erosa et al., 2017) do not distinguish between *married and single* women, despite markedly disparate labor market outcomes (Goldin, 2006; Blau and Kahn, 2007). I contribute by focusing on gender roles associated with *marriage*.

I develop a model featuring education, marriage, and labor supply choices to quantify the consequences of gender roles in marriage. Gender roles are measured, through the lens of the model, as a composite force that makes the labor force participation of married women diverge from that of single women, besides wage differentials. The model structure is validated by a reduced form analysis, which uses county-level variation in World War 2 casualties that increased female labor force participation and consequently weakened tra-

[1] I show this explicitly later in section 1.2.

ditional gender norms. The model is matched to the education, marriage, and labor force participation patterns in the U.S. decennial census, decade by decade, to track how the magnitude of gender roles change over time. With the model, I then compute counterfactuals to determine by how much aggregate output is affected by changing gender roles and dissect the underlying channels.

My central finding is that gender roles have changed significantly in the U.S. and that gender norms have important output effects. If gender norms had stayed at the level of 1940, aggregate market output in 2010 would be lower by 4.8% and aggregate total (market and nonmarket) output would be lower by 3.5%. Gender norms matter more for the subpopulation of married women, whose labor force participation is directly affected. Married women in this counterfactual cumulatively have a 13.0% lower market output and 6.5% lower total output.

To establish these findings, I start by presenting empirical facts that motivate my focus on traditional gender roles as the distinguishing factor between married and single women. I use the U.S. decennial census to establish the first motivating fact: married women's labor force participation that is *not* accounted for by standard observables, such as age, education, race, and the number of children, rises over time to catch up with the stable counterpart for single women. This disparate trend highlights the importance of unobservable variables as drivers of married women's labor force participation, including traditional gender roles that affect married women but not single women. The second motivating fact, derived from the Panel Survey of Income Dynamics (1968-2015), is that individuals undergo stark changes in their time use right in the first year of marriage. The share of housework hours[2] relative to paid work hours falls sharply for men and rises sharply for women getting married in the 1970s. For later marriages, however, there are no such sharp changes in time use upon marriage. This finding highlights that being married shifts the responsibility of house chores to women, but in a way that weakens over time.

With these motivating facts in mind, I develop a structural model for two purposes. I first use it as a measurement tool to quantify by how much gender roles affect married women's labor supply choice, featured as a parameter in the model. Then I use the model to conduct counterfactuals to gauge the importance of gender roles in marriage for various aggregate measures related to labor supply, earnings, marriage, education, and most importantly, output.

Individuals in my model make three sets of choices over the course of their life cycle. First, individuals choose their level of education as a forward-looking investment decision. Second, they enter the marriage market, a frictionless transferable utility set up in the style of Becker (1973), where individual types are defined by their education levels. They decide on which spousal type to get married to or to stay single, and then draws a family composition category (e.g. number of children) according to match-specific empirical probabilities. Third, individuals draw market and home abilities, and households make the dichotomous labor supply choice of whether to work in the market or on home production

[2]The answer to the question, "About how much time does (he/she) (do you) spend on this housework in an average week–I mean time spent cooking, cleaning, and other work around the house?"

for each individual. Gender roles are modeled as a disutility that a married couple gets when the wife works in the market. This disutility factors into the wife's labor supply decision as a "norms wedge" that lowers the value of her market wage. Therefore in my set up, gender roles directly affect labor supply choice, and also indirectly affect marriage and educational choice in anticipation.

My theoretical contribution is fourfold. Firstly, I augment a tractable form of selection into labor activity by individuals with heterogeneous abilities, derived from the Trade literature (Eaton and Kortum, 2004; Hsieh et al., 2019), with concerns over fulfilling gender roles (Akerlof and Kranton, 2000; Bertrand, Kamenica, and Pan, 2015). Secondly, I embed this form of selection, previously used to study *individual* choices, into a model of *household* decision-making. Thirdly, I ensure that the household economic utilities resulting from the labor activity choice are fully consistent with models of educational choice and marriage market matching (Choo and Siow, 2006; Chiappori, Salanié, and Weiss, 2017; Chiappori, Costa-Dias, and Meghir, 2018). Lastly, the recursive structure of my model simplifies the parameter identification procedure and allows me to manage a very large number of household types.

I calibrate the model to match the education, marriage, labor force participation patterns in the U.S. decennial census decade by decade, assuming that the data is a reflection of the model equilibrium. The practical advantage of my model is that it is not demanding on the data, as the only variables needed are market wage, labor force participation, marital status, education and children, all of which become available from 1940.[3] As the model is fitted decade by decade, model parameters other than norms wedges flexibly account for secular changes in the gender wage gap, gender differences in home productivity, propensity of marriage, assortativeness of marriage matching by education, and educational attainment.

I find that married women faced a 44% norms wedge on the market wage in 1940, which declined to 25% by 2010. To cross-check whether these wedges correlate with more conventional measures of gender norms, I repeat the calibration at the *state level* and regress state-level averages of the norms wedges on the state-level average answers to attitudinal surveys related to gender roles in marriage.[4] I find that the states that answer more conservatively in attitudinal surveys are also the ones with higher norms wedges.

I use the model to conduct a counterfactual, where I consider what would have happened in 2010, had gender norms not changed since 1940. I first find that the number of completed school years of women drops by 1.4% and that of men by 0.8%. This is the result of marriage becoming less attractive and thus the marriage-market returns to education falling. The effect on women is compounded by falling labor-market returns to education. The marriage rate indeed falls by 32.2%. As higher-earning women are affected by more with higher norms wedges, the selection into marriage by education becomes more

[3]This feature of the model implies that the model can be easily applied to many other settings.

[4]I put together various surveys in the Roper Polls Database. Questions include whether one approves of a married woman working if she has a husband capable of supporting her and whether it is more important for a wife to help her husband's career than to have one herself, and many others.

negative for women. Due to assortative matching on education, the selection into marriage by education becomes more negative for men as well. Moreover, married women's labor force participation and cumulative market earnings drop by 17.5% and 13.0%, respectively. On aggregate, the total market output of the economy falls by 4.8% and the total market *and* nonmarket output drops by 3.5%. The effect on total output amounts to a half of the drop in total output that would be seen if married women of 2010 made labor force participation choices based on 1940 wages and home productivities. The finding of smaller effects on output than on labor force participation echoes Hsieh et al. (2019)'s findings on the aggregate effects of occupation-specific preferences that vary by gender.

The counterfactual also implies that the reduction in the gender norms wedges between 1940 and 2010 accounts for many well-documented empirical trends in the United States. Specifically, these are a) rising married female labor force participation rate, b) rise in wife's share of household income, c) faster growth of educational attainment of women relative to men, and d) increasingly positive selection into marriage by education of both men and women (Bar et al., 2018; Juhn and McCue, 2016; Case and Deaton, 2017).

Since the counterfactual results depend on the model structure, I next perform a reduced form exercise to validate the model. For lack of a direct test of model predictions when norms wedges fall, I explore the effects of a shock that *indirectly* affects norms and check that other variables change in the expected direction. Inspired by Fernández et al. (2004), I consider WW2 draftee casualties as a temporary positive shock to female labor force participation that propagates over the long term through weaker gender norms. Underlying this story is the idea of cultural transmission through exposure (Bisin and Verdier, 2000). I employ a difference-in-difference estimation strategy on the U.S. decennial census comparing high-casualty counties with low-casualty counties in each decade relative to 1940, the last decade before the WW2 shock. The results indicate higher female labor force participation in the high-casualty counties every decade from 1950, but in a way that indicates a spike in 1950, a slight drop in 1960, and a gradual increase over the next decades. At the same time, attitudes become gradually less conservative in the high-casualty counties. Other variables, namely labor force participation by gender and marital status, marriage rate, education, and wages, gradually evolve in a way that is consistent with model predictions when norms wedges fall.

The WW2 reduced form result also allows me to extend the structural model to have norms evolving dynamically in response to past female labor force participation. I augment the model to have *economywide* norms evolve in response to *economywide* female labor force participation in the past decade, a relationship that I estimate based on the reduced form coefficients. This extension of the model is compatible with how I identified norms wedges previously, as long as individuals take norms as given and do not internalize the effect of their labor supply choice on the norms of future generations. The model extension further enables me to conduct *dynamic* counterfactuals, on how a shock would affect an economy *over time*. A simple thought experiment of females temporarily getting paid male wages in 2010, shows that the economy of 2010 stabilizes within three decades at a different equilibrium with higher female labor force participation and lower norms wedges. The exercise illustrates how temporary policies encouraging female labor force participation can have permanent effects.

The analysis in this paper is based on historical data from the United States. Yet, numerous countries in the world are experiencing similar trends as the U.S.: gender attitudes are becoming less conservative,[5] and married women's labor force participation is catching up with single women's. These countries include not only the richer OECD countries but also low- and middle-income countries in Eastern Europe and Latin America. At the same time, one in ten countries of the world still has a lower female labor force participation rate than 1940 U.S. (International Labor Organization, 2019). Thus, this paper is informative about the potential growth consequences and the underlying channels of cultural change in other countries that currently operate under traditional gender roles or are moving away from it.

Contributions to related literature A large literature pioneered by Restuccia and Rogerson (2008) and Hsieh and Klenow (2009) study the aggregate implications of various forms of misallocation. A growing number of papers focus on gender differences as a source of misallocation of talent. The most relevant papers are Hsieh et al. (2019), which looks at gender discrimination in the educational and labor markets distorting occupational choice, and Erosa et al. (2017), which studies the gender differences in nonmarket time using married couples only. I add to this literature by focusing primarily on the difference between *married* and *single* women. As I integrate the marriage market matching into the model, I can explore a new set of channels behind the aggregate output implications, such as selection into marriage and marriage market returns to education.

I also contribute to a large body of work that seeks to explain the dramatic rise in married women's labor force participation in the U.S. The explanations proposed thus far can be broadly categorized into two branches: technological progress and cultural change. The first branch includes the invention of birth control pills (Goldin and Katz, 2002), technological advances in housework (Greenwood et al., 2005), and medical progress in pregnancy-related conditions (Albanesi and Olivetti, 2016). The latter, on the other hand, includes changes to divorce laws (Fernández and Wong, 2014) and greater acceptance of working wives by men (Fernández et al., 2004). I add to the second branch by zooming into gender roles that are associated with marriage, and *quantifying* its effect on married women's labor force participation.

My reduced form analysis around WW2 casualties also speaks to a growing literature on how gender roles change. Kuziemko et al. (2018) explore the birth of the first child as a factor that changes individual's preferences, and Fogli and Veldkmap (2011) and Fernández (2013) model gender roles changing as a result of social learning about the uncertain costs of working. My contribution is to tie the structural model and the reduced form results together to estimate how female norms wedges change in response to past female labor

[5] Conservativeness in gender attitudes is measured by the fraction agreeing to "When jobs are scarce, men have more right to a job than women," asked in the World Values Survey.

force participation. In addition, I augment the model with this estimated relationship to illustrate how one-off policies can have long-lasting consequences through the dynamic evolution of norms.

Roadmap The remainder of the paper is organized as follows. Section 1.2 describes empirical facts that motivate my focus on the distinction between married and single women's labor supply decisions. Section 1.3 sets up the structural model, making explicit this difference. Section 1.4 describes the data and how model parameters are calibrated to fit the model to the data. It then discusses the calibration results. Section 1.5 quantifies the effect of changes in gender norms through the lens of the model and benchmarks the results to the effects of other comparable counterfactuals. Section 1.6 presents reduced form results for model validation and a dynamic extension to the model. Section 3.8 concludes.

1.2 Motivating Facts

This section presents descriptive facts that motivate my investigation of the aggregate output effects of gender roles in marriage.

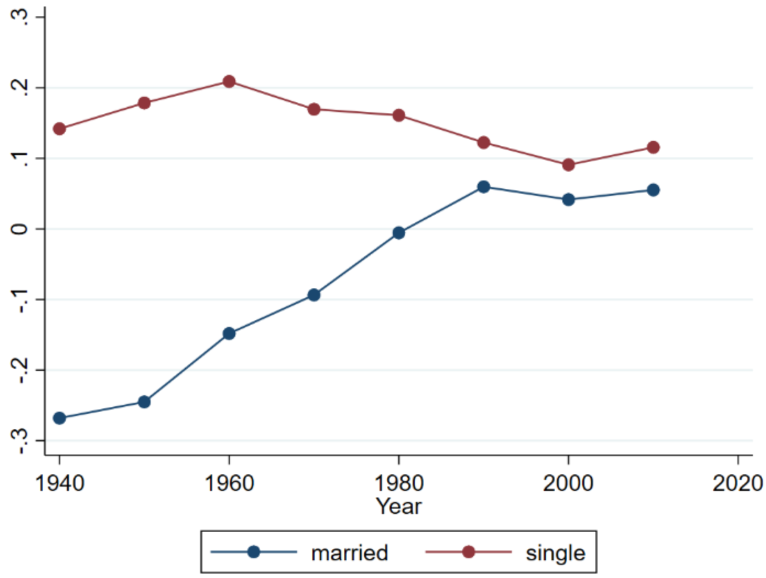
1.2.1 Married vs. single "unexplained" female labor force participation

Figure 1.1 compares the path of the "unexplained" labor force participation (LFP) of married women and single women over time. By "unexplained" LFP, I refer to residuals from the regression of labor force participation status indicator Lab on standard, commonly observed individual characteristics X: age, education, race, and the number of children dummies.

$$Lab_{it} = X_{it}\beta + \epsilon_{it}.$$

The regression sample is all females aged 25-54 between 1940 and 2010 in the U.S. decennial census. Then I take the weighted average of the residuals by marital status and decade.

Figure 1.1: Residualized female labor force participation, by marital status and decade



Notes: This figure compares the labor force participation rates of unmarried (never married, separated, divorced, widowed) to married women between the ages of 25 and 54 in 1940 and 2010, in the United States. The participation rates are residualized for age, education, race, and the number of children dummies. See text for the residualization procedure.

Figure 1.1 shows that the "unexplained" LFP rose for married women but not for single women. It therefore highlights, firstly, that the labor force participation choices of married women are very different from single women, and secondly, that this difference shrinks over time. In addition, this difference exists even when LFP status is residualized for the number of children. Therefore, it shows that the distinction between the labor supply behaviors of married and single women extends beyond the presence of children, which has been the dominant factor setting marrieds apart from singles in the literature. The figure also suggests that technological change around child-bearing or child-rearing cannot explain all of this catch-up, leaving room for cultural change around gender roles within marriage as a potential contributor.

1.2.2 Shift in work patterns upon marriage

To further corroborate the observation that married women's labor supply choice is disparate from single women's, I study how individuals shift their time use immediately upon marriage, and how this shift changes over time.

I follow the event-study approach of Kleven et al. (2019). For this exercise, I use the Panel Survey of Income Dynamics, which is an individual-level panel data where nationally representative individuals of the United States record their weekly paid hours and housework hours.[6] I define event times to be years relative to marriage, such that event time 0 refers to the first year in which an individual's marital status switches to being married

[6]The exact survey question is "About how much time does (he/she) (do you) spend on this housework in an average week–I mean time spent cooking, cleaning, and other work around the house?"

from being single. I run the regression

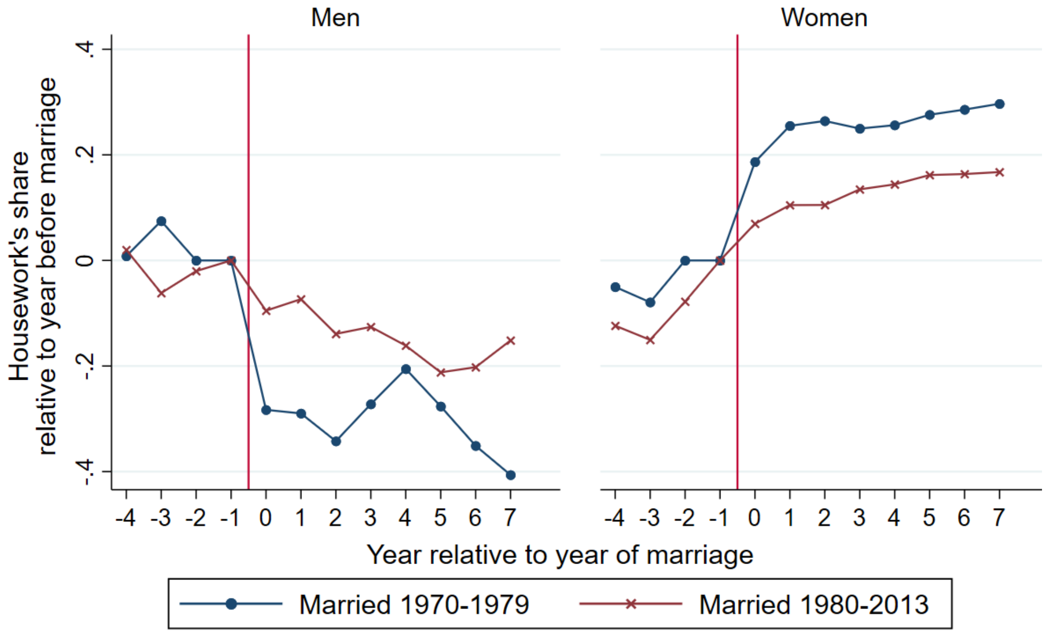
$$housework_{ist}^{g} = \sum_{j \neq -1} \alpha_j^g \cdot \mathbb{1}(j = t) + \sum_k \beta_k^g \cdot \mathbb{1}(k = age_{is}) + \sum_y \gamma_y^g \cdot \mathbb{1}(y = s) + \nu_{ist}^{g}$$

where $housework_{ist}^{g}$ denotes the housework's share of housework and market work hours of individual i of gender g in year s at event time t. This regression tracks how *housework* changes as a function of event time, while controlling for age dynamics via the age dummies and time trends via the year dummies. The event time coefficients ($\hat{\alpha}_t^g$) are then normalized by $\mathbb{E}[\tilde{Y}_{ist}^{g}|t]$, where $\tilde{Y}_{ist}^{g} \equiv \sum_k \hat{\beta}_k^g \cdot \mathbb{1}(k = age_{is}) + \sum_y \hat{\gamma}_y^g \cdot \mathbb{1}(y = s)$ is the level of the predicted outcome when excluding the effect of the event time. $\hat{\alpha}_t^g / \mathbb{E}[\tilde{Y}_{ist}^{g}|t]$ are plotted in Figure 1.2[7].

The blue lines of Figure 1.2 illustrate a sharp jump in women's share of housework hours among total work (market work and housework) hours and a sharp drop in men's, immediately upon transition from singlehood to marriage, for marriages in the 1970s. This finding highlights that being married shifts the responsibility of house chores to women. The magnitude of the time use shift is also not trivial. The jump in the housework's share for women upon marriage amounts to about half of the jump associated with the birth of the first child for the same sample of women. As the sample consists of couples who had no childbirths in the first three years of marriage, the figure additionally suggests that marriage itself - independent of the presence of children - subjects women to the gender role constraints. This idea resonates with the catch-up of married women's LFP with single women's even when residualized for the number of children, shown in Figure 1.1.

[7] The actual event time coefficients (α_j^g) are plotted in Figure 1.A2 of the Appendix. They are statistically insignificant at the 5% level prior to marriage, and are significantly negative for men and significantly positive for women post-marriage

Figure 1.2: Housework's share of housework and market hours, among couples whose first child is born ≥ 4 years after marriage



Notes: This figure plots the share of housework among the sum of housework and market work hours by gender around the year of marriage. The red vertical line plots the timing of marriage. Individuals are unmarried household heads without any live-in partners in the years to the left of the red line, and they are married with live-in spouses in the years to the right of the red line.

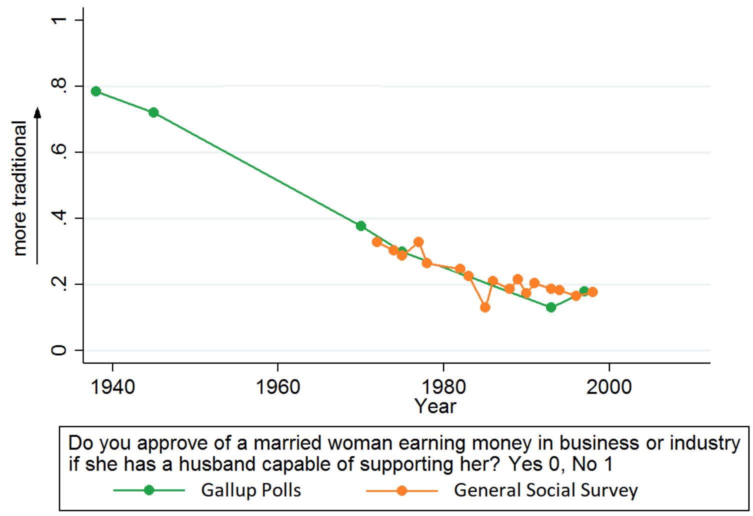
The red lines of Figure 1.2, on the other hand, demonstrate that there are no sharp changes in time allocation for marriages that take place later. The event study coefficients are also statistically insignificant around the year of marriage. I take this null effect for later marriages as suggestive of the decline in the division of labor according to traditional gender roles over time.

1.2.3 Attitudinal survey trends

The last motivating fact supports the notion of weakening gender roles over time.

Figure 1.3 illustrates that attitudes on the gender role of married women have become less traditional over time. Among various survey questions on gender attitudes, the question that was asked for by far the longest period was whether one approved of a married woman working if she had a husband capable of supporting her. While close to 80% answered 'No' to this question in 1938, in 1998 less than 20% did so. The survey question ceased to be asked afterward, which itself could be suggestive of the question being less controversial and thus of less interest than before.

Figure 1.3: Trend in attitudinal survey answers



Notes: This figure plots the fraction of respondents disapproving of a married woman working if she has a husband capable of supporting her, according to the Gallup Polls and the General Social Survey (GSS). Although the two surveys asked almost identical questions, there is a slight difference. The Gallup Polls' specific question was "Should a married woman earn money if she has a husband capable of supporting her?" with possible answers "Yes" (0) and "No" (1), while the GSS asked, "Do you approve of a married women earning money in business or industry if she has a husband capable of supporting her?" with possible answers "Yes" (0) and "No" (1).

Other survey questions on gender roles of married individuals have been asked, however, with answers confirming a continual trend towards less traditional attitudes. These trends are shown in Figure 1.A1 of the Appendix.

1.3 Model

The motivating facts of Section 1.2 suggest that married individuals' labor supply decisions are different from single individuals', that this gap is shrinking over time, and that gender roles could be one of the factors driving this gap.

I proceed by building a structural model, for two purposes. I first use it as a measurement tool to quantify by how much gender roles affect married individuals' labor supply choice. Then I use the model to conduct counterfactuals to gauge their importance for various aggregate measures related to labor supply, earnings, marriage, and education.

Timing In my model, the decisions of individuals are divided into three stages. In the first stage, they choose their level of education as a forward-looking investment decision, balancing the returns to education in the labor and marriage markets and the cost of education (Chiappori, Iyigun, and Weiss, 2009). In the second stage, they enter a frictionless transferable utility (TU) marriage market (Becker, 1973; Shapley and Shubik, 1971), where "types" of individuals equals their education levels chosen previously, and decide on which spousal type to get married to or to stay single. The resulting match is thus characterized by the education levels of the husband and wife for married individuals and by one's own education level for single individuals. Households then exogenously get assigned family composition categories based on the number of children under the age of 5 and under the age of 18 in the household, according to match-specific empirical probabilities. Individuals subsequently enter the third and last stage, each characterized by group: the tuple of (gender, marriage match, family composition). Individuals then draw idiosyncratic market and home abilities. Households make the dichotomous labor supply choice of whether to work in the market or in home production for each individual, taking as given a) the group-specific market wages that the representative firm of the economy pays, b) the group-specific value of home production, and c) the group-specific disutilities that a married couple gets upon deviation from traditional gender roles, i.e. when the wife works in the market and when the husband works at home. After the labor supply decisions are made, households consume and realize utilities.

Since I solve the model backwards, I describe each stage in greater detail starting from the last.

1.3.1 Economic utilities and optimal labor supply choices

Individual utilities consist of an economic and a predetermined noneconomic component. The *economic* component is characterized by the utility functions below, adapted from Chiappori, Costa-Dias, and Meghir (2018)[8]. In this model, the economic gains from marriage arise from two sources. Firstly, there are economies of scale generated by the consumption of public goods. Secondly, marriage enables risk sharing between the two spouses against uncertain future public and private consumption. A more general formulation of the utility function that is consistent with my model is described in Appendix 1.C.1.

Married Individuals

Consider a married household composed of husband m and wife f. Individual $i \in \{m, f\}$ gets the following utility:

$$u_i(Q, C_i, L_f) = \ln(Q) + \ln(C_i - \tau_i w_f L_f) \tag{1.1}$$

where C is the consumption of the private good, and Q is the consumption of the public good (e.g. expenditures on housing, children, heating). $L_f \in \{0, 1\}$ denotes the labor market participation of wife f, and w_i refers to the market wage of individual i. Critically, married individual i gets disutility when labor allocations deviate from traditional gender norms. The disutilities, parameterized by τ, occur proportionally to the value of market income brought home by the wife.

From an ordinal perspective, this utility belongs to Bergstrom and Cornes' Generalized Quasi Linear (GQL) family. Hence, at any period and for any realization of family income, it satisfies the transferable utility (TU) property. Under TU, utility can be transferred

[8]This paper formulates an equilibrium lifecycle model of education, marriage and labor supply and consumption in a transferable utility context. A key innovation of this paper is that labor supply decisions are made over the lifecycle, while maintaining equilibrium in the educational choice and marriage matching.

between spouses at a fixed rate of exchange, and so for Pareto-efficiency, a couple acts as a single decision unit that maximizes the joint marital output. As the set of Pareto efficient allocations is an ordinal concept,[9] any cardinalization of u can be used for the definition of joint marital output. I use $\exp u_i$ as the cardinalization of i's preferences. Then, conditional on the couple's labor choices, any Pareto efficient allocation maximizes the sum of the spouses' exponential utilities, $\exp u_m + \exp u_f$.[10]

Therefore, conditional on labor market participation choices, a married couple solves

$$\max_{Q,C} Q(C - \tau w_f L_f) \tag{1.2}$$

$$\text{s.t.} \qquad pQ + C = w_m L_m + w_f L_f + h_m(1 - L_m) + h_f(1 - L_f)$$

$C \equiv C_m + C_f$ denotes total expenditure on private goods, $\tau \equiv \tau_m + \tau_f$ is the couple's joint disutility from the wife working in the market, and p is the price of the public good relative to the private good (the numeraire). Moreover, $L_m \in \{0, 1\}$ denotes the labor market participation of husband m, and h_i refers to the home productivity of individual i.[11] The solutions to the maximization problem given by (1.2) are

$$Q = \frac{w_m L_m + (1 - \tau) w_f L_f + h_m(1 - L_m) + h_f(1 - L_f)}{2p} \tag{1.3}$$

$$C = pQ + \tau w_f L_f$$

Let us describe the intra-household allocation, before market earnings and home production values are realized. Efficient sharing of the risks against the uncertainty of earnings and home productivities implies that the ratio of marginal utilities of private consumption is constant and equal to the Pareto weight (μ),[12] which is endogenously determined in the marriage market:

$$\frac{\partial u_m}{\partial C_m} = \mu \frac{\partial u_f}{\partial C_f}$$

The resulting indirect utilities are:

$$v_m = 2 \ln Q + \ln p + \ln \frac{1}{1 + \mu}, \qquad v_f = 2 \ln Q + \ln p + \ln \frac{\mu}{1 + \mu} \tag{1.4}$$

---

[9] The set of Pareto efficient allocations remains unchanged when u is replaced with $f(u)$, for a strictly increasing mapping f.

[10] Proof: An allocation is Pareto efficient if it maximizes $\exp u_m$ subject to (a) the budget constraint, and (b) $\exp u_f \geq \bar{u}$. This program is equivalent to a second program that maximizes $\exp u_m + \zeta \exp u_f$ subject to the budget constraint alone, where ζ is the Lagrange multiplier on constraint (b). The first order conditions of the second program with respect to private consumptions yield

$$Q = \lambda = \zeta Q$$

where λ is the Lagrange multiplier on the budget constraint. Thus, it must be that $\zeta = 1$; any Pareto efficient allocation maximizes the sum of exponential utilities.

[11] It might seem non-standard that according to the budget constraint, private and public goods can be bought with "income" from home production. However, see Appendix section 1.C.2 for how the utility maximization problem is identical if I divide goods into market goods and home-produced goods, and have two separate budget constraints for each.

[12] See section 6.3.2 of Browning, Chiappori, and Weiss (2014) for more detail on the characterization of intra-household allocations under efficient risk-sharing.

The total economic utility generated from this marriage then is

$$v = v_m + v_f = \breve{v} + \ln \frac{\mu}{(1+\mu)^2}. \tag{1.5}$$

where $\breve{v} \equiv 4 \ln Q + 2 \ln p$. It is straightforward that the farther the wife's Pareto weight μ is from 1, the husband's, the smaller the total economic utility of the couple.

Clearly, the couple makes labor choices to maximize Q.[13] Hence, from (1.3), the optimal labor choices are

$$L_m^* = \mathbb{1}\left[w_m \geq h_m\right] \tag{1.6}$$

$$L_f^* = \mathbb{1}\left[(1-\tau)w_f \geq h_f\right] \tag{1.7}$$

In the optimal labor choice of married women, τ enters as a "norms wedge". In deciding her labor supply, a married woman values her market wage lower than its face value, as if it is taxed. Another feature that stands out in equation (1.6) is the independence of the husband's and wife's labor supply choices. This feature makes studying the selection into working in the labor market by either the husband or the wife easy.[14] Moreover, in Appendix sections 1.C.5 and 1.C.6, respectively, I consider two extensions to the model: a) norms wedges applying also to married *men*, where the nontraditional activity for them is home production, and b) assuming men always work, enriching the labor supply decisions of women.

Single Individuals

To distinguish from the married case, I use the hat symbol for singles. The economic utilities of singles follow the same formulation as for married couples, except they are not subject to gender roles and hence do not receive disutilities from non-traditional behavior.

A single individual i maximizes the following utility:

$$\hat{u}_i(\hat{Q}_i, \hat{C}_i) = \ln(\hat{Q}_i) + \ln(\hat{C}_i) \tag{1.8}$$

$$\text{s.t.} \qquad p\hat{Q}_i + \hat{C}_i = w_i \hat{L}_i + h_i(1 - \hat{L}_i)$$

The resulting indirect utility is

$$\hat{v}_i = 2 \ln \hat{Q}_i + \ln p \tag{1.9}$$

and the optimal labor choice that maximizes $\hat{Q}_i$ is

$$\hat{L}_i^* = \mathbb{1}\left[w_i \geq h_i\right] \tag{1.10}$$

[13] Moreover, as $\frac{\partial^2 \ln Q}{\partial w_m \partial w_f} = 0$, the model does not predict any assortative matching on market earnings.

[14] Appendix 1.C.4 discusses further whether this feature is an acceptable simplification.

Market Income and Home Production Value As optimal labor supply choices for married and single individuals depend on the comparison of market earnings to home production value, it is imperative to discuss how they are determined. Market income and home production value depend on idiosyncratic market and home abilities, as well as components common to groups defined by gender, marriage match, and family composition. Gender, marriage match, and family composition are all determined before the labor choice stage. After the marriage matching stage, family composition is given exogenously according to match-specific empirical probabilities.[15] Hence, the probability that a (husband type q, wife type r) match has a family composition $\mathcal{K}$, denoted as $d^{qr}(\mathcal{K})$, is simply found from the data.

An individual i of gender g in a (q, r) match, with family composition $\mathcal{K}$, receives income

$$w_i = \bar{w}_g^{qr}(\mathcal{K})\epsilon_i^w, \quad g \in \{M, F\}$$

where $\bar{w}_g^{qr}(\mathcal{K})$ is the market income per unit of effective labor for i's group, and ϵ_i^w is i's market ability. The reason why group wages differ can be thought of as a combination of selection and treatment effects. For instance, I am flexibly letting married women have different market productivity from single women because individuals who get married might be different from those who are single (selection), and marriage might causally affect market productivity (treatment effect). Similarly, college-educated women married to high-school dropout husbands are allowed to have different wages from college-educated women married to college-educated husbands as a result of both selection and treatment effects.

Where i's type equals s, i's home production value is given by

$$h_i = \bar{h}_g^s(\mathcal{K})\epsilon_i^h, \quad g \in \{M, F\}$$

The group component of home production value, $\bar{h}_g^s(\mathcal{K})$, varies by gender, own education level, and family composition. Inherent in this assumption is that marital status and spousal type does not matter for home productivity, which is necessary for me to be able to later disentangle norm tax parameters from home productivity parameters.

I assume that market abilities ϵ^w and home production abilites ϵ^b are drawn independently and identically from the Fréchet distribution with shape parameter θ, after the marriage matching stage.[16] The cumulative distribution functions for these abilities are

$$F(\epsilon^w) = F(\epsilon^b) = F(x) = \exp\left\{-x^{-\theta}\right\}.$$

[15] It is possible for singles to have children in my model, because singles include never-married, divorced, separated, and widowed individuals. This grouping of singles is equivalent to assuming that divorces, separations, and widowhoods occur via shocks exogenous to the schooling years of the couple.

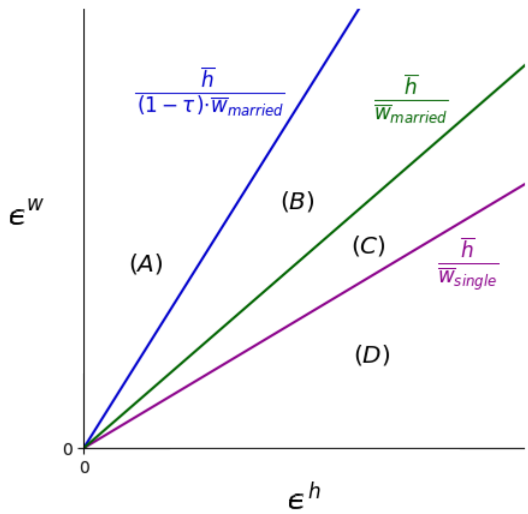
[16] The extensive literature on returns to schooling highlights the correlation between schooling and unobserved abilities. Hence, it might be more plausible that the market and home production abilities are drawn from education-specific distributions. However, in incorporating the correlation between schooling and unobserved abilities into the model, I take a shortcut by assuming that different education levels result in abilities being drawn from the same distribution scaled by different constants. In other words, where ϵ_s^w is the market ability drawn from a distribution specific to education level s, $\epsilon_s^w = c^s\epsilon^w$. Then, it is possible to take the scaling constants (c^s) out of intrinsic abilities and have schooling-specific wages incorporate the scaling constants.

From the convenient property of Fréchet distributions (Eaton and Kortum, 2004), the probability that a woman in a (q, r) match with family composition $\mathcal{K}$ works in the labor market is:

$$P_F^{qr}(\mathcal{K}) \equiv \mathbb{P}\left((1-\tau^{qr})\bar{w}_F^{qr}(\mathcal{K})\epsilon^w > \bar{h}_F^r(\mathcal{K})\epsilon^b\right) = \frac{[(1-\tau^{qr})\bar{w}_F^{qr}(\mathcal{K})]^\theta}{[(1-\tau^{qr})\bar{w}_F^{qr}(\mathcal{K})]^\theta + [\bar{h}_F^r(\mathcal{K})]^\theta} \tag{1.11}$$

The maximum likelihood estimator for this probability is the labor force participation rate of the women in this group. Equation (1.11) is useful for calibrating parameters later in section 1.4. Moreover, Figure 1.4 illustrates how sorting across market work and home production by market and home abilities occurs for married and single women.

Figure 1.4: Sorting across market work and home production of married and single women



Notes: This figure plots how labor allocation between market work and home production is determined for different combinations of market abilities ϵ^w and home abilities ϵ^b, for married and single women with the same education level and family composition (in simplified notation).

(A): For a married woman to work in the market, she must be very talented in market work.

(B): If the female norms wedge was removed, more married women would be engaging in market work.

(C): Along with (A) and (B), the market and home ability combinations of single women doing market work.

(D): Single women who work at home are very talented in home production.

Another implication of Fréchet abilities useful for calibration later on is that the average wage of the women working in the market is[17] :

$$\begin{aligned} \text{avgwage}_F^{qr}(\mathcal{K}) &= \bar{w}_F^{qr}(\mathcal{K})\mathbb{E}\left[\epsilon^w|(1-\tau^{qr})\bar{w}_F^{qr}(\mathcal{K})\epsilon^w > \bar{h}_F^r(\mathcal{K})\epsilon^b\right] \\ &= \bar{w}_F^{qr}(\mathcal{K})\left(\frac{1}{P_F^{qr}(\mathcal{K})}\right)^{\frac{1}{\theta}}\Gamma\left(1-\frac{1}{\theta}\right) \end{aligned} \tag{1.12}$$

1.3.2 Marriage market

Marriage matching occurs based on the expected value of the economic utilities delineated in the previous subsection, together with the noneconomic utilities, of each match.

Following Choo and Siow (2006)'s matching under transferable utility (TU) with random preferences, consider an economy consisting of S types of men and women. These types are defined by the level of education determined prior to the matching stage. Denote n^{qr} as the number of marriages between type-q men and type-r women, n^{q0} as the number of single type-q men, and n^{0r} as the number of single type-r women. Also, M^q is the number of type-q men, and F^r the number of type-r women. The following accounting identities must hold:

$$n^{q0} + \sum_{r=1}^{S} n^{qr} = M^q \qquad \forall\, q = 1, ..., S \tag{1.13}$$

$$n^{0r} + \sum_{q=1}^{S} n^{qr} = F^r \qquad \forall\, r = 1, ..., S \tag{1.14}$$

In this TU model, a type-q man must transfer τ^{qr} amount of utility to a type-r woman to marry her. The utility of type-q man m marrying a type-r woman at time t is

$$V_m^{qr} = \mathbb{E}(v_m^{qr}) - \tau^{qr} + \psi^{qr} + \varepsilon_m^{qr}$$

where $\mathbb{E}(v_m^{qr})$ is the expected economic utility of man m married to a type-r woman, ψ^{qr} is the noneconomic utility ("marital bliss") enjoyed by the couple, and ε_m^{qr} is m's random preference for the match drawn independently and identically from the type I extreme-value distribution. Let $r = 0$ denote the case of singlehood, with $v_m^{q0} \equiv \hat{v}_m^q$, $\tau^{q0} = 0$, and $\psi^{q0} = 0$.

Similarly, the utility of type-r woman f marrying a type-q man is

$$V_f^{qr} = \mathbb{E}(v_f^{qr}) + \tau^{qr} + \psi^{qr} + \varepsilon_f^{qr}$$

where ε_f^{qr} is f's random preference for the match drawn independently and identically from the type I extreme-value distribution.

[17] Where ϵ^{w*} is the market ability ϵ^w conditional on working in the market, its cumulative distribution function is $F^*(x) = \exp\left\{-\frac{1}{P}x^{-\theta}\right\}$ where P is the fraction working in the market. In other words, F^* follows the Fréchet distribution with shape parameter θ and scale parameter $\left(\frac{1}{P}\right)^{\frac{1}{\theta}}$.

The marriage market-clearing equilibrium transfers τ^{qr} are determined such that

$$n^{qr,\mathcal{D}} = n^{qr,\mathcal{S}} = n^{qr}.$$[18]

In equilibrium,[19]

$$\frac{n^{qr}}{\sqrt{n^{q0}n^{0r}}} = \frac{\mathbb{E}(v_m^{qr}) + \mathbb{E}(v_f^{qr}) - \mathbb{E}(\hat{v}_m^q) - \mathbb{E}(\hat{v}_f^r)}{2} + \psi^{qr}$$

Using equation (1.5),

$$\frac{n^{qr}}{\sqrt{n^{q0}n^{0r}}} = \frac{\mathbb{E}(\check{v}^{qr}) - \mathbb{E}(\hat{v}_m^q) - \mathbb{E}(\hat{v}_f^r)}{2} + \Psi^{qr} \tag{1.15}$$

where $\Psi^{qr} \equiv \psi^{qr} + \frac{\ln \mu^{qr} - 2\ln(1+\mu^{qr})}{2}$. The first term on the right-hand side of equation (1.15) is the gain to marriage relative to singlehood from the couple being able to enjoy a greater consumption of the public good together. Ψ^{qr} signifies the utility from marital bliss and the utility from the intra-household allocation of resources based on Pareto weights.[20]

1.3.3 Education

In this section, I describe the women's educational choice problem, without loss of generality. Woman f chooses the education level with the maximum expected utility:

$$\max_{r=1,\ldots,S} U_F^r$$

where

$$U_F^r = \sum_{q=0}^{S} \left[\frac{n^{qr}}{F^r} \Big(\mathbb{E}(v_f^{qr}) + \tau^{qr} + \psi^{qr} \Big) \right] - c_F^r - \xi^r$$

Individuals are forward-looking. The expected utility from schooling level r depends on the consequent matching probabilities in the marriage market and the expected utilities in each type of match. The costs, on the other hand, consist of the gender-specific direct utility cost c_F^r and idiosyncratic cost ξ^r, drawn independently and identically from the Type I extreme value distribution.

[18] There is a 1:1 relationship between τ^{qr} and the Pareto weight μ^{qr}. From the marriage market clearing condition and equation (1.4),

$$\tau^{qr} = \frac{\ln n^{q0} - \ln n^{0r} - \ln \mu^{qr} - \mathbb{E}(\hat{v}_m^q) + \mathbb{E}(\hat{v}_f^r)}{2}$$

[19] See Appendix section 1.C.7 for greater details on the derivation of the marriage market equilibrium.

[20] ψ^{qr} and μ^{qr} cannot be separately identified. As can be seen in footnote 18, μ^{qr} would only be identified if the equilibrium transfers in the marriage market were observable, but they are not. Hence, I seek to identify Ψ^{qr}. Identifying Ψ^{qr} is sufficient for running counterfactuals, conditional on the behavioral assumption of limited foresight, described in section 1.4.2.

From equation (1.4),

$$U_F^r = 2\sum_{q=0}^{S}\left[\frac{n^{qr}}{F^r}\mathbb{E}(\ln Q^{qr})\right] + \ln p - C_F^r - \xi^r \tag{1.16}$$

where $C_F^r \equiv c_F^r - \sum_{q=0}^{S}\frac{n^{qr}}{F^r}\left(\ln\frac{\mu^{qr}}{1+\mu^{qr}} + \tau^{qr} + \psi^{qr}\right)$. C_F^r is the direct cost of getting schooling level r, minus a) the expected utility from intra-household resource allocation, b) the marriage market utility transfer, and c) the noneconomic gains in a match. In the parameter inference section (section 1.4), I will back out the values of C_F^r, and not c_F^r.[21]

The distribution of the idiosyncratic schooling costs imply that the probability an individual of gender g chooses schooling level s is

$$\mathbb{P}(s = \arg\max_{s'=1,..,S} U_g^{s'}) = \frac{\exp\{U_g^s\}}{\sum_{s'=1}^{S}\exp\{U_g^{s'}\}}$$

The maximum likelihood estimator of this probability is $\frac{F^s}{\sum_{s'=1}^{S}F^{s'}}$ for women and $\frac{M^s}{\sum_{s'=1}^{S}M^{s'}}$ for men, i.e. the shares of individuals with education level s for each gender.

1.3.4 Firms

A representative firm produces market output Y^{mkt}. Although there are two market goods in this model, the private good and the public good, I assume that they are derived from the same market output. The relative price p merely measures how much more market output is needed for 1 unit of public good, relative to 1 unit of private good. This simplification is innocuous, given that the value of p has no consequence for equilibrium education, marriage, and labor decisions.

I assume the most simplistic set-up on the firm's side. The firm's production function is linear in male and female effective labor, $\mathcal{M}$ and $\mathcal{F}$:

$$Y^{mkt} = B(\mathcal{M} + \mathcal{F}) \tag{1.17}$$

Normalize, as 1 unit of effective labor, the labor provided by single males with schooling level of 1, market ability of 1, and zero children ($\mathcal{K}_0$), i.e. $\bar{w}_M^{10}(\mathcal{K}_0)$.

$$\mathcal{M} = \sum_{q=1}^{S}\sum_{r=0}^{S}\sum_{\mathcal{K}} n^{qr}d^{qr}(\mathcal{K})\left(\frac{\bar{w}_M^{qr}(\mathcal{K})}{\bar{w}_M^{10}(\mathcal{K}_0)}\right)\left(P_M^{qr}(\mathcal{K})\right)^{1-\frac{1}{\theta}}\Gamma\left(1-\frac{1}{\theta}\right)$$

$$\mathcal{F} = \sum_{r=1}^{S}\sum_{q=0}^{S}\sum_{\mathcal{K}} n^{qr}d^{qr}(\mathcal{K})\left(\frac{\bar{w}_F^{qr}(\mathcal{K})}{\bar{w}_M^{10}(\mathcal{K}_0)}\right)\left(P_F^{qr}(\mathcal{K})\right)^{1-\frac{1}{\theta}}\Gamma\left(1-\frac{1}{\theta}\right)$$

[21] As with Ψ^{qr}, identifying C_F^r is sufficient for running counterfactuals, conditional on the behavioral assumption of limited foresight, described in section 1.4.2.

1.3.5 Aggregate output

Aggregate output is a combination of market output and home production:

$$Y = Y^{mkt} + Y^{home}$$

where

$$Y^{home} = B \sum_{g\in\{M,F\}} \sum_{(q,r)} \sum_{\mathcal{K}} n^{qr} d^{qr}(\mathcal{K}) \left(\frac{\bar{h}_g^r(\mathcal{K})}{\bar{w}_M^{10}(\mathcal{K}_0)} \right) \left(1 - P_g^{qr}(\mathcal{K})\right)^{1-\frac{1}{\theta}} \Gamma\left(1 - \frac{1}{\theta}\right) \quad (1.18)$$

1.3.6 Equilibrium

An equilibrium in this economy consists of schooling choice q for a man, schooling choice r for a woman, marital transfers τ^{qr}, marriage matches (q, r), public consumption Q, private consumption C_i, labor market participation L_i, total efficient male labor $\mathcal{M}$, total efficient female labor $\mathcal{F}$, market wage, market output Y^{mkt}, total home production Y^{home}, and aggregate output Y, such that

1. Individuals choose the schooling level offering the greatest expected utility, taking as given the probability of resulting in a particular match and the expected utility from that match.

2. After schooling choices are made, equilibrium marital transfers $\{\tau^{qr}\}$ equate the supply and demand for each marriage match (q, r) based on the expected utility from each match.

3. After the matching stage and exogenous determination of family composition, each individual chooses public good consumption Q, private good consumption C_i, and labor supply L_i to maximize their utility function. The individual maximizes equation (1.1) jointly with their spouse if married and maximizes equation (1.8) independently if single.

4. A representative firm hires effective male labor $\mathcal{M}$ and effective female labor $\mathcal{F}$, and pays wage equal to the technology parameter B in equation (1.17).

5. Market output Y^{mkt} is given by equation (1.17), and total home production Y^{home} by (1.18).

6. Aggregate output of the economy Y is given by the sum of Y^{mkt} and Y^{home}.

1.3.7 Intuition for aggregate output effects of norms wedges

In the model, how is aggregate output affected by changes in norms wedges? When the norms wedge on market wage for married women decreases, there can be aggregate output effects arising from each of the three stages (in reverse order) of labor supply, marriage, and education choices for women.

First, at the labor supply stage, sorting across market work and home production of *married* women is more aligned with productivity. This channel increases aggregate output.

Second, at the marriage matching stage, marriage becomes more attractive as the disutility from the non-traditional working arrangement when married is lower. Then some of the women who would otherwise have been single would now be married and therefore be newly subject to the norms wedge. As the norms wedge prevents some of these women from pursuing their comparative advantage, aggregate output is lower. There is another effect occurring at the matching stage. The women who are newly induced to be married now receive married wages. As whether the married wages are higher or lower than single wages is an empirical question, this channel has an ambiguous effect on aggregate output.

Third, at the educational choice stage, young women realize that they are more likely to be married and to work in the labor market in the future. Then if there is positive assortative matching on education in the marriage market, the greater likelihood of marriage increases their incentive for higher education. This effect comprises the marriage-market returns to education. In addition, if education is more effective in increasing market productivity than home productivity, the greater likelihood of market work increases young women's incentive for higher education. This effect, on the other hand, comprises the labor market returns to education. Either case, higher education for women would then increase aggregate output through higher market wages and home productivities.

Overall, the effect on aggregate output would depend on the parameter values.

1.4 Data and Parameter Inference

1.4.1 Data

To simulate the U.S. economy within the model framework, I use the U.S. decennial census, consisting of 1-in-100 national random sample of individuals. The nice feature of the U.S. census is that data is collected on all household members so that labor market information is available for both spouses among married couples. Because the presence of other income-earning household members may perturb individual labor decisions, I restrict the sample to either household heads or spouses of heads. I further restrict the sample to individuals aged between 25 and 54, after education is complete and when individuals are the most economically active.[22]

The model in section 1.3 is fitted to the census data every decade, assuming that the data is a reflection of the model steady state. By calibrating the model separately by decade, I am allowing almost all model parameters to change flexibly over time, including family composition probabilities $\{d^{qr}(\mathcal{K})\}$, group market wages $\{\bar{w}_M^{qr}(\mathcal{K}), \bar{w}_F^{qr}(\mathcal{K})\}$ and home productivities $\{\bar{h}_M^{qr}(\mathcal{K}), \bar{h}_F^{r}(\mathcal{K})\}$, gender norms wedges $\{\tau^{qr}(\mathcal{K})\}$, the expected utility enjoyed by a (q, r) match $\{\Psi^{qr}\}$, and the cost of each schooling level $\{C_M^q, C_F^r\}$. The only

[22]Appendix figure 1.A3 shows that the age range of 25-54 is appropriate as the most economically active 30-year window. Hsieh et al. (2019) also use this age range.

parameter that I leave to be constant over time is θ, the inverse measure of the dispersion of market and home abilities.[23]

The practical advantage of my model set-up is that the model is not demanding on the data; the only variables needed for these parameters to be inferred are market wage, labor force participation status, marital status, education, and children. As the earliest decade in which all these variables are observed is 1940, I use the decennial census from 1940 to 2010. The census in 1950 is not used, however, because the 1950 data does not include spousal information.

1.4.2 Assumption on behavior under counterfactual scenarios

Identifying parameter values is necessary to conduct counterfactual simulations. Under counterfactual situations, the marriage matching pattern will be different and the match-specific Pareto weights may change as a result. However, it is impossible to figure out what the counterfactual Pareto weights would be, without imposing further structure on how they are determined. Therefore, I assume that individuals are naive with limited foresight; they expect future marriage market outcomes[24] to remain the same under counterfactual scenarios. This assumption makes identifying Ψ^{qr} and C^s_g, rather than separately identifying μ^{qr}, ψ^{qr} and c^s_g, sufficient for deriving counterfactual marriage and education patterns.

1.4.3 Steps for Parameter Inference

1. $d^{qr}(\mathcal{K})$: probability of a (q, r) match having family composition $\mathcal{K}$

   Set at the empirical probabilities.

2. θ: inverse measure of dispersion of market and home abilities

   Making use of the fact that wages of individuals working in the market follow a Fréchet distribution, I estimate θ through maximum likelihood. Where x_n is the market ability of observation n and P_n denotes the fraction of workers in observation n's group, the maximum likelihood estimator for θ is:[25]

$$\tilde{\theta}_{MLE} = \arg\max_{\theta\in(0,\infty)} \sum_{n=1}^{Nobs} \left[\ln\theta - \ln P_n - x_n^{-\theta} P_n^{-1} - (\theta+1)\ln x_n\right]$$

3. $\bar{w}^{qr}_g(\mathcal{K})$: group market productivity per unit of effective labor

   Using the estimate of θ found in step 2 and the average wage and proportion of market-workers in each group in the data, I can back out $\bar{w}^{qr}_g(\mathcal{K})$ from equation (1.12).

$$\bar{w}^{qr}_g(\mathcal{K}) = \text{avgwage}^{qr}_g(\mathcal{K}) \left(P^{qr}_g(\mathcal{K})\right)^{\frac{1}{\theta}} \frac{1}{\Gamma(1-1/\theta)}$$

[23]See section 1.4.4 for more discussion and for results when θ is estimated for each decade.

[24]Specifically, the exact objects that individuals need to expect unchanged under counterfactual scenarios are the a) Pareto weights $\{\mu^{qr}\}$, b) probabilities of matches $\left\{\frac{n^{qr}}{F^r}, \frac{n^{qr}}{M^q}\right\}$, and c) marriage market equilibrium transfers $\{\tau^{qr}\}$.

[25]How I derive the likelihood function and how I isolate market abilities from observed market wages are detailed in Appendix sections 1.D.1 and 1.D.2, respectively.

4. $\bar{h}_g^{qr}(\mathcal{K})$: group home productivity per unit of effective labor

   $\bar{h}_F^r(\mathcal{K})$ (similarly, $\bar{h}_M^{qr}(\mathcal{K})$) can be backed out from equation (1.11), armed with θ found in step 2 and the average wage and proportion of market-workers among *single* women with r years of schooling and family composition $\mathcal{K}$.

$$\bar{h}_F^r(\mathcal{K}) = \text{avgwage}_F^{0r}(\mathcal{K})\left(1 - P_F^{0r}(\mathcal{K})\right)^{\frac{1}{\theta}} \frac{1}{\Gamma(1-1/\theta)}$$

5. $\tau^{qr}(\mathcal{K})$: group norms wedges

   $\tau^{qr}(\mathcal{K})$ is backed out from equation (1.11) using $\bar{w}_F^{qr}(\mathcal{K})$, $\bar{h}_F^r(\mathcal{K})$, and the fraction of market workers in a group of married women. The idea is that norms wedges are high if the fraction working in the market is much lower than is predicted from market and home productivities.

$$\tau^{qr}(\mathcal{K}) = 1 - \frac{\text{avgwage}_F^{0r}(\mathcal{K})}{\text{avgwage}_F^{qr}(\mathcal{K})}\left(\frac{1 - P_F^{0r}(\mathcal{K})}{1 - P_F^{qr}(\mathcal{K})}\right)^{\frac{1}{\theta}} \tag{1.19}$$

   Intuitively, the disutility from wives working is inferred by comparing the labor choices of married and single women sharing the same level of education and the same family composition $\mathcal{K}$. The difference in their labor market participation rates that cannot be explained by wage differentials is attributed to gender norms.

6. Ψ^{qr}: utility from marital bliss and intra-household resource allocation, in a (q, r) match

   From equation (1.15),

$$\Psi^{qr} = \frac{n^{qr}}{\sqrt{n^{q0}n^{0r}}} - 2A^{qr} + \hat{A}_M^q + \hat{A}_F^r. \tag{1.20}$$

   where

$$A^{qr} = \sum_{\mathcal{K}} d^{qr}(\mathcal{K})\mathbb{E}\Big[\ln\Big(\bar{w}_M^{qr}(\mathcal{K})\epsilon_m^w L_m^* + \bar{h}_M^{qr}(\mathcal{K})\epsilon_m^h(1 - L_m^*) + \\ \big[1 - \tau^{qr}(\mathcal{K})\big]\bar{w}_F^{qr}(\mathcal{K})\epsilon_f^w L_f^* + \bar{h}_F^r(\mathcal{K})\epsilon_f^h(1 - L_f^*)\Big)\Big]$$

   and

$$\hat{A}_g^s = \sum_{\mathcal{K}} d^{qr}(\mathcal{K})\mathbb{E}\Big[\ln\Big(\bar{w}_g^{qr}(\mathcal{K})\epsilon_i^w \hat{L}_i^* + \bar{h}_g^s(\mathcal{K})\epsilon_i^h(1 - \hat{L}_i^*)\Big)\Big]$$

$$(q = 0, r = s \text{ if } g = F, \text{ and } q = s, r = 0 \text{ if } g = M)$$

   There are no closed-form expressions for A^{qr} and $\hat{A}_g^s$ so I simulate them to back out Ψ^{qr}.[26]

[26] Given all the parameter values found in steps 1-5, the simulation is straightforward.

7. C_g^s: direct cost minus expected utility from intra-household resource allocation, of education s

   Use equation (1.16). $\{C_F^r\}_{r=1,...,S}$ are found as the solution to the system of equations

$$\frac{F^r}{\sum_{r'=1}^S F^{r'}} = \frac{\exp\left\{2\sum_{q=0}^S \frac{n^{qr}}{F^r}A^{qr} - C_F^r\right\}}{\sum_{r'=1}^S \exp\left\{2\sum_{q=0}^S \frac{n^{qr'}}{F^{r'}}A^{qr'} - C_F^{r'}\right\}} \qquad \forall\, r = 1, ..., S$$

   where $A^{0r} = \hat{A}_F^r$ and $A^{q0} = \hat{A}_M^q$. $\{C_M^q\}_{q=1,...,S}$ are found similarly.

1.4.4 Calibration Results and Discussion

I now provide a discussion of the calibrated parameter values, and how they match similar estimates in the literature, related measures from external data sources, or well-documented stylized facts.

$\boldsymbol{\theta}$: inverse measure of dispersion of market and home abilities

As shown in Table 1.1, the estimate of θ is 1.837, which is similar to Hsieh et al. (2019)'s estimate of 1.52 for the Fréchet shape parameter dictating the dispersion of abilities across occupations. It is also close to their choice to use 2 for conducting counterfactuals. θ is estimated for the entire sample from 1940 to 2010, under the assumption that the distribution market and home ability endowments remains fixed over time. If θ is estimated decade by decade, the estimates are quite stable although there is a slight non-monotonic upward trend. They range from 1.549 to 1.999, with 1.844 as the median value.[27] Moreover, Appendix figure 1.A4 visually shows that the distribution of inferred market abilities closely resembles the probability distribution of the Fréchet distribution, supporting the assumption of Fréchet-distributed abilities.

Table 1.1: Maximum likelihood estimate of θ

$\hat{\theta}$	1.837***
	(18.31)
N	3570573

Notes: t statistics based on standard errors clustered by sex in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

When the standard errors are not clustered, following the model assumption of independently drawn market abilities, the t statistic is incredibly large at 4579.11 due to the large sample size.

See step 2 of section 1.4.3 for the maximum likelihood estimation strategy.

[27]The θ estimates by decade are 1.549 in 1940, 1.844 in 1960, 1.845 in 1970, 1.718 in 1980, 1.975 in 1990, 1.999 in 2000, and 1.826 in 2010.

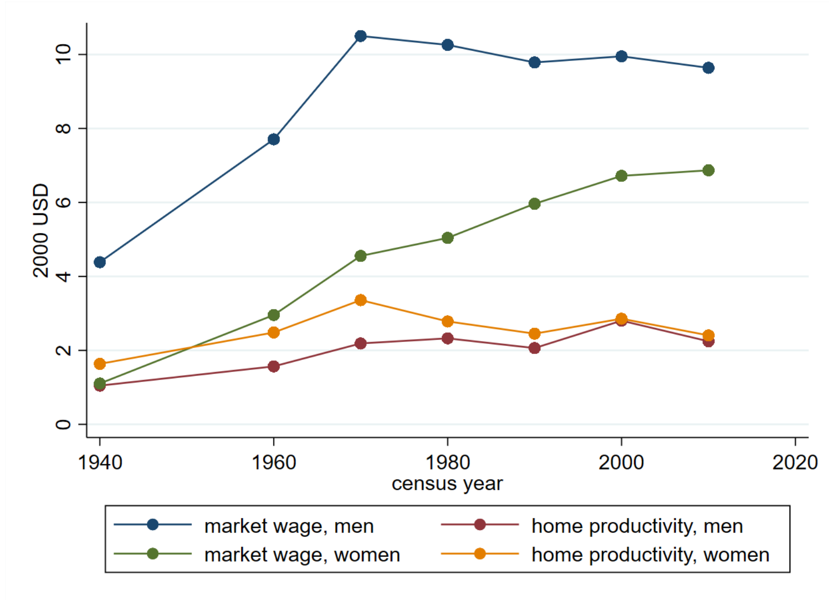
$\bar{w}, \bar{h}$: group market and home productivity

To calibrate the group-specific market and home productivities, I first need to specify the groups. Each group is defined by gender, schooling pair, and family composition. I must ensure that each group is large enough since I match population moments to sample analogs within each group.[28] On the other hand, since I treat all individuals within a group as similar individuals that share the same values of norms wedges, group market productivities, and group home productivities, it must also be that the categorization of the group is specific enough. Of the variables defining each group, average schooling has undergone drastic increases in the sample period. I therefore adjust for the fact that the commonly completed levels of schooling differ by decade. I construct 5 or 6 schooling levels every decade, with at least 5% of the sample belonging to each level. This categorization is in Table 1.A2 of the Appendix. A similar rationale holds for family composition categories; I construct these to have group sizes that are large enough and group categories specific enough. As the largest differences in home production duties relating to children occur for the first child and whether the child is young, these factors formed the basis of the categorization. The family composition categories, detailed in Table 1.A3 of the Appendix, are kept fixed over the entire period.

Figure 1.5 plots the weighted average of group market wage and home productivity by sex and decade. The weight equals the empirical probability of each group. For example, if the share of college-college couples with no child among the entire sample is high in 2010, the group market wage received by the wives of such couples get a greater weight in the computation of the average group market wage for women in 2010.

[28] For example, see equations (1.11) and (1.12), applied in the parameter inference steps 3 and 4 of section 1.4.3.

Figure 1.5: Weighted average of group market productivity ($\bar{w}$) and home productivity ($\bar{h}$) by sex



Notes: This figure plots the weighted average of the group components of market productivity and home productivity by sex and decade. These productivities for each group are inferred using the model structure as outlined in steps 3 and 4 of section 1.4.3.

The group market productivity $\bar{w}$ is increasing in the average wage of the workers in that group as well as in the LFP rate of that group. The reason $\bar{w}$ increases in group LFP rate is that the group LFP rate encompasses the selection effects. The higher the LFP rate, the lower the average idiosyncratic market ability, as the set of workers are less selected on market ability. Then for the same empirically observed average wage of those who work in a group, less of it is accounted for by the average idiosyncratic market ability, so the higher the group market productivity must be. The group home productivity $\bar{h}$ is also increasing in the average wage of the workers in that group, but is decreasing in the labor force participation rate of that group.

How do $\bar{w}$ and $\bar{h}$ vary by education? As mentioned in section 1.3.7 on the intuition for aggregate productivity effects of decreases in norms wedges, whether an increase in education increases $\bar{w}$ or $\bar{h}$ by more matters for the educational choice in counterfactual scenarios. Specifically, if a young woman anticipates a higher likelihood of market work due to a fall in norms wedges, she will increase her education if education increases market productivity by more than home productivity. Table 1.A4 of the Appendix confirms that for both sexes in every year, education increases market productivity by more than home productivity. In fact, while market productivities significantly increase with education every decade, home productivities are either not affected or *decreasing* in education other than for women in the early decades of 1940 and 1960. Thus for home productivity, the effect of education on group LFP rate often dominates that on the average wages of the group's workers. This finding contrasts with the literature on the positive returns to education on childcare (Leibowitz, 1974). It could also be that $\bar{h}$ is underestimated, and more so at higher levels of education, in the later decades. $\bar{h}$ is inferred from singles' labor force participation behavior. If with greater marketization over time[29] it becomes more important for a household to have a wage income, then singles will be more likely to be in the labor force than married individuals as they do not have spouses that can bring in the wage income. Then using the inferred $\bar{h}$ from singles for the $\bar{h}$ for marrieds will underestimate the home productivity for marrieds. Moreover, this underestimation may be more pronounced at higher levels of education where a higher wage income is at stake in the singles' labor supply decision.

τ: gender norms wedges

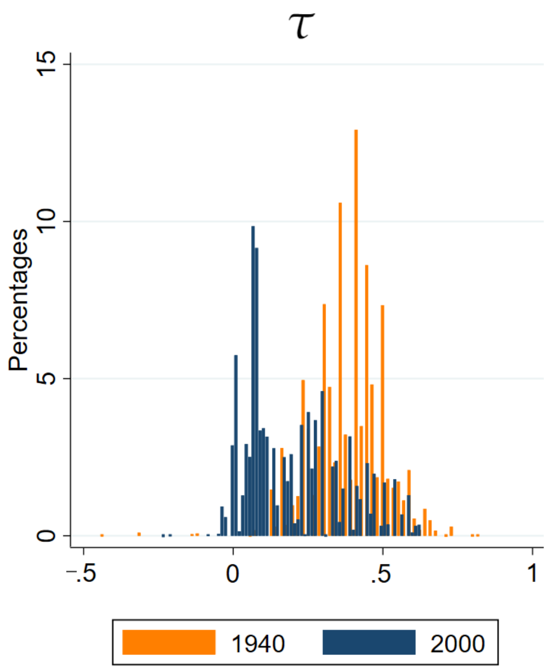
I next calibrate the values of τ, the norms wedges on the market wage of married women. Though labeled as "norms wedges", τ encapsulate any reason that brings married women's LFP to diverge from single women's LFP besides wage differentials. For instance, it includes the differential valuation of staying home between marrieds and singles.

To illustrate the interpretation of τ through the lens of the model, let us take an example. If τ equals 0.6, the interpretation is that the worth of a \$10 market wage to a married woman is only \$4 when she is making her labor supply decision. Thus, $\tau = 0.6$ corresponds to a norms wedge of 60% on the market wage of that woman.

Figure 1.6 plots the histograms of τ, calibrated by group, for the years 1940 and 2000. The height of the bar for each group equals the group's empirical probability. It is very noticeable that the histogram of τ for 2000 sit to the left of that for 1940, signifying a decrease in the norms wedges.

[29] e.g. Ngai and Petrongolo (2017)

Figure 1.6: Histogram of norms wedges on married women's market wages (τ)



Notes: This figure plots the histogram of τ, calibrated by group, for the years 1940 and 2000. The height of the histogram bars is scaled to percentages so that it indicates the empirical probability of the corresponding group in each year. The norm wedges for each group are inferred using the model structure as outlined in step 5 of section 1.4.3.

I take the inferred group-specific τ as noisy estimates of norms wedges, as there are many selection effects unaccounted for in the model, such as the correlation between education choice and market abilities, or the correlation between taste for spousal type and market abilities. Therefore, I consider the weighted median of τ by decade.[30] Figure 1.7 plots these values. τ generally decreases over time, except from 2000 to 2010. The reason for this rise is that while married women's wages have increased relative to single women's between the two decades, their LFP has not.[31,32] To reconcile these two observations, τ must increase since τ encapsulates any reason for which the LFP rates of married women and single women diverge besides wage differentials. A possible interpretation is that the value of home production of married women has increased.[33]

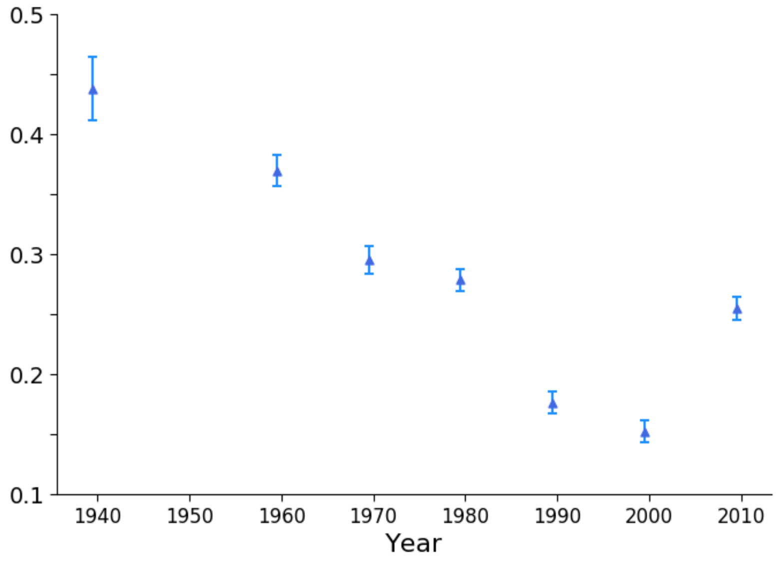
[30]The results are very similar to weighted average, unweighted average, or unweighted median values.

[31]Figure 1.1 illustrates that female labor force participation has plateaued since 1990 regardless of marital status.

[32]Even when I take into account the market work *hours* of married women relative to single women, the rise persists.

[33]The IPUMS Time Use survey indicates a rise in child care hours of married women relative to single women recently.

Figure 1.7: Evolution of inverse norms wedges τ



Notes: This figure plots the weighted median of τ, inferred for each group, by decade. The weight equals the empirical probability of each group. The error bars indicate 95% confidence intervals based on bootstrapped standard errors with 50 replications.

To show that the values of norms wedges imputed are related to directly observable measures of conservativeness, I redo the parameter inference procedure at the state level while pooling all the data across the years. As before, to ensure that each group is specific enough but also large enough, group categories are reformulated: defined by state, schooling pair, and age cohorts.[34] Then I take for each state either the weighted average or the weighted median of the norms wedges τ. Table 1.2 reports the regression coefficients from regressing the state-level norms wedges on the state-level attitudinal survey answers. The state-level attitudinal survey answers are the weighted average by state of individual survey answers taken across multiple periods. Two attitudinal measures are considered. The first is the fraction disapproving of married women working, using the data on the single attitudinal survey question featured in Figure 1.3. The second is a composite attitudinal index that takes the weighted average by state of all the attitudinal survey questions plotted in Figure 1.A1 of the Appendix. It is reassuring that the states with more conservative gender attitudes are also the ones with higher norms wedges τ.

[34]The age cohorts are ages 25-34, 35-44, and 45-54.

Table 1.2: Correlation between state-level norms wedge and attitudinal survey answers

	Dependent variable τ	
	average	median
Regressed on:		
Fraction disapproving of married women working	0.249**	0.282**
	(2.21)	(2.21)
Regressed on:		
Composite attitudinal index	0.450***	0.439**
	(2.94)	(2.50)
N	51	51

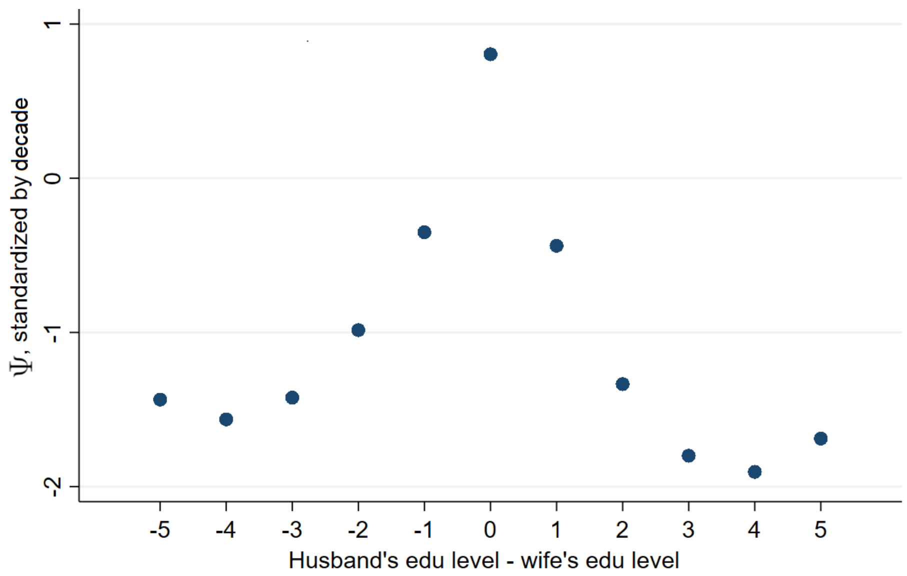
Notes: t statistics based on robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Higher values of the composite attitudinal index correspond to more traditional attitudes. Each group gets weight equal to empirical probability in computing the average and the median. The positive correlation between τ and the attitudinal survey answers survives when the regressions are run at the state-year level with year fixed effects. Adding state fixed effects to this regression, however, renders the coefficients statistically insignificant, due to the fact that τ is a noisy measure of traditional gender norms.

Ψ: utility from marital bliss and intra-household resource allocation

Figure 1.8 plots the average Ψ for each value of the difference in the husband's and wife's education levels. As the values of Ψ differ by decade, driven by changing marriage patterns over time, Ψ are standardized by decade before averages are taken across the decades. The reason for Ψ at the spousal difference of 5 being larger than Ψ at the difference of 4 is that there are 6 schooling categories only in 2 decades, while the other 5 decades have 5 schooling categories. Overall, this plot can be taken as a single-peaked plot, peaking when the husband and wife share the same education level. Therefore, the calibrated Ψ values are congruous with the well-documented fact of assortative matching by education in the U.S. (Greenwood et al., 2016).

To show how the pattern of Ψ has changed *over time*, Figure 1.A5 of the Appendix decomposes the content of Figure 1.8 by decade. The values in Figure 1.A5 are *not* standardized by decade, and hence the fall in the values over time reflect declining attractiveness of marriage.

Figure 1.8: Ψ by spousal education gap



Notes: This figure plots the average Ψ for each value of the difference in the husband's and wife's education levels. As the values of Ψ differ by decade, driven by marriage patterns changing over time, Ψ are standardardized by decade before averages are taken across the decades.

C: direct cost minus expected utility from intra-household resource allocation

The last set of parameters to calibrate relates to the gender-specific cost of schooling, C_g^s. For example, the cost to females of acquiring the schooling level of "high school graduate" is inferred to be small if there are more female high school graduates in the data than is predicted by the expected utility from that level of schooling. The expected utility depends on the marriage market returns and the economic (i.e. wage and home productivity) returns.

The estimates of C, reported in Appendix Table 1.A5, are not comparable across time since the schooling level categories differ by decade.[35] However, the costs can be compared between the two genders within each decade. The cost of attaining the highest schooling level was larger for women from 1940 to1990, but it became *smaller* for women from 2000. This observation matches the stylized fact of women's overtaking of men in educational attainment in the U.S. For instance, the share of 25- to 34-year-old women with at least bachelor's degrees overtook that of men around 1995. The share of women of the same age range with at least some graduate school overtook that of men around 2000, too. Furthermore, a time-series comparison can be made for the decades 1990-2010, as the schooling categories are the same for those decades. The female to male ratio of the cost of the highest schooling level steadily falls from 1.11 in 1990 to 0.96 in 2000, and to 0.94 in 2010.

[35] Refer to Table 1.A2 in the Appendix for the schooling categories.

1.5 Counterfactual Exercises

In order to quantify the contribution to economic growth of changes in gender roles, I will consider how aggregate output Y changes if the norms wedges are the only parameters changing while all others are kept fixed. Moreover, I benchmark the effects of this main counterfactual on the effects of other counterfactuals described below.

1.5.1 Steps for conducting counterfactuals

I denote counterfactual values with underlines.

1. Compute $\underline{A}^{qr}$ and $\underline{P}_g^{qr}(\mathcal{K})$. They differ from the values at the status quo because optimal labor decisions would change under different gender norms. Note that $\hat{A}_g^s$ remains unchanged, as singles' labor decision is unaffected by norms wedges.

2. Compute counterfactual schooling probabilities $\frac{\underline{F}^r}{\sum_{r'=1}^{S} \underline{F}^{r'}}$ and $\frac{\underline{M}^q}{\sum_{q'=1}^{S} \underline{M}^{q'}}$, using $\underline{A}^{qr}$. Then compute $\underline{F}^r$ and $\underline{M}^q$ by assuming that the total population size remains constant.

3. Solve for the $(S \times S + 2S)$ values of marriage matches $\underline{n}^{qr}$, from the $S \times S$ equations given by (1.20) as well as the $2S$ accounting identities given by (1.13) and (1.14).

4. Finally, compute $\underline{Y}^{mkt}$ and $\underline{Y}^{home}$

1.5.2 Counterfactual exercise results

Table 1.3 records the changes in various aggregate variables that would occur in 2010 if gender norms had remained at the level of 1940, holding all other parameter values fixed at the 2010 level. As recognized in section 1.4.4, the calibrated norms wedges for *each group* are most likely noisy estimates. For this reason, I consider the counterfactual of every individual in 2010 being subject to the same, weighted-median norms wedge of 2010 at baseline, and being subject to the same, weighted-median norms wedge of 1940 in the counterfactual scenario.

I consider two adjustment margins: a) when only the labor supply choices are allowed to respond, in column (1), and b) when education, marriage, and labor supply choices are all allowed to respond, in column (2). Because not all variables change in column (1), column (1) clarifies which variables are directly affected by τ. When τ increases from 0.25 in 2000 to 0.44 in 1940, i.e. more traditional gender norms for married women, married women work 14.3% less in the labor market. As a result, the cumulative market output Y^{mkt} of married women falls by 6.9%. However, as fewer married women work in the market, married women's cumulative home production value Y^{home} increases, and so the total output $Y(= Y^{mkt} + Y^{home})$ falls by less, at 2.1%. The dissimilar effects on Y^{mkt} and Y highlight the importance of accounting for nonmarket output, which is almost always excluded from national accounts. It therefore hints that the output gains when women enter the labor market would be overstated in methods that only consider market output.

Table 1.3: Percent changes in various aggregate variables if individuals of 2000 were subject to the female norms wedge of 1940

	Adjustment margins	
	Labor supply (1)	Education, marriage, & labor supply (2)
Education		
Women's years of schooling	-	-1.4
Men's years of schooling	-	-0.8
Selection into marriage		
Marriage rate	-	-32.2
Married women's edu/single women's edu	-	-4.1
Married men's edu/single men's edu	-	-1.2
Labor Force Participation		
Married women's LFP	-14.3	-17.5
Married men's LFP	-	-0.03
Single women's LFP	-	0.6
Single men's LFP	-	0.1
Output per head		
Married women's market output	-7.0	-13.0
Married women's total output	-2.1	-6.5
Married men's market output	-	-0.8
Married men's total output	-	-0.8
Aggregate market output	-2.0	-4.8
Aggregate market & home output	-0.6	-3.5
Within-household gender earnings gap		
Wife's share of household market income	-11.5	-14.9

Notes: This table reports the percentage changes in various aggregate variables that occur when the individuals of 2000 are subject to the female norms wedge of 1940, holding all other parameter values constant at the 2000 level. Column (1) holds the marriage match patterns and educational choices constant at the 2000 level and considers only changes to the married individuals' labor supply decisions. Column (2) additionally allows the (forward-looking) marriage matching and educational choices to change in accordance with the new expected utilities arising from the altered labor supply behavior.

In column (2), the direct effects of higher τ trickle down to indirectly affect education and marriage match choices, too. As norms wedges are modeled as costs to marriage, a higher τ renders marriage less attractive, yielding a fall in the marriage rate of 32.2%. As the norms wedge is more costly for women with higher market ability, the fall in marriage is more pronounced for higher educated women. Because there is assortative marriage matching by education, the fall in marriage is more pronounced also for higher educated *men*. Furthermore, the education of both men and women fall.[36] With a higher τ, the labor and marriage market returns to education are lower for women, and only the marriage market returns to education are lower for men. As the set of married women has a lower market ability on average with higher τ, married women's LFP rate and cumulative market output fall by *more* in column (2) than in column (1). Married women's cumulative total output falls by 6.5%. In aggregate, including the output of single men and women, aggregate market output falls by 4.8% and total output by 3.5%.

In summary, the fall in gender norms wedges over 1940 through 2010 partially accounts for various stylized facts documented in the U.S.: a) rise in married female LFP, b) rise in wife's share of household market income, c) faster growth of educational attainment of women relative to men, and d) increasingly positive selection of men and women into marriage by education (Bar et al., 2018; Juhn and McCue, 2017; Case and Deaton, 2017).

Is the effect of a 4.8% fall in aggregate market output and a 3.5% fall in aggregate total output small or large? The output effects might be viewed as large, as the norms wedge parameters capture quite a narrow concept of gender norms relating to the distinction between married and single women. On the other hand, it might be viewed as small, relative to the output growth that has occurred over 1940-2010. To better benchmark the size of the 2% effect, I conduct additional counterfactuals.

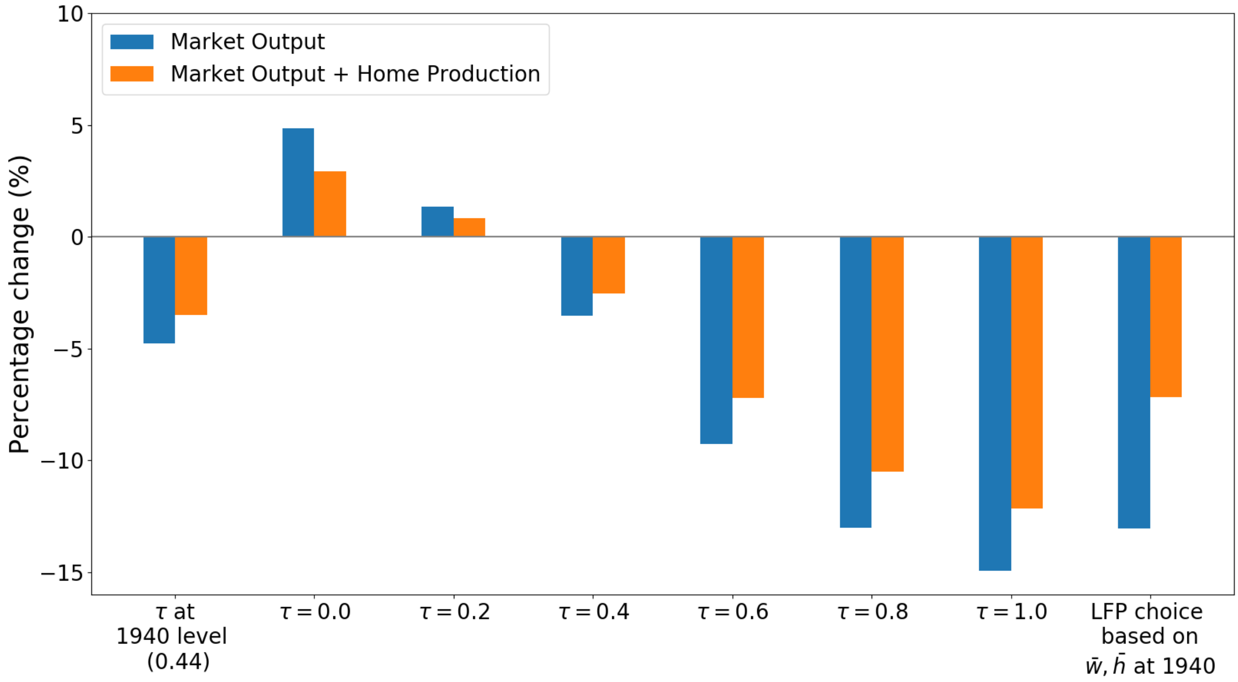
Additional Counterfactuals Figure 1.9 compares the market and total (market and nonmarket) output effects of various counterfactual scenarios, where the baseline year is set at 2010. The other counterfactual scenarios explore the effects of τ changing from 0 to 1, and when labor supply choices are made based on $\bar{w}$ and $\bar{h}$ of 1940. All the three margins of education, marriage, and labor supply are allowed to adjust.

I focus on the last counterfactual, in particular, to benchmark the effects of the main counterfactual. As illustrated in Figure 1.5, the market and home productivities have undergone substantial changes over time, underlying in great part the output growth over 1940-2010. In fact, $\bar{w}$ is lower than $\bar{h}$ for women in 1940, whereas $\bar{w}$ is *more than triple* of $\bar{h}$ in 2010. Not to include the direct effects of productivity changes on output, the counterfactual is only about letting labor supply choices be determined based on $\bar{w}$ and $\bar{h}$ at 1940 levels. Then, the labor supply choice changes enormously, with married women's LFP rate falling by a staggering 66%. Market output consequently falls by 13.1% and total

[36] In an attempt to succinctly express the change in education, I compute for each decade, the weighted average of schooling years of individuals in every education level. The baseline and counterfactual compare the sum of (weighted average of schooling years by education level)×(share of population in each education level). Although the degree of the decline in the years of schooling appears very small, a lot changes with the education *levels*.

output falls by 7.2%. Therefore, the main counterfactual's effects amount to a half of the effect of the last counterfactual. In this sense, the effect of the norms wedge is sizable.

Figure 1.9: Market and total output effects of various counterfactual scenarios



Notes: This figure compares the market and total (market and nonmarket) output effects of various counterfactual scenarios, where the baseline year is set at 2010, with τ at 2010 being 0.25.

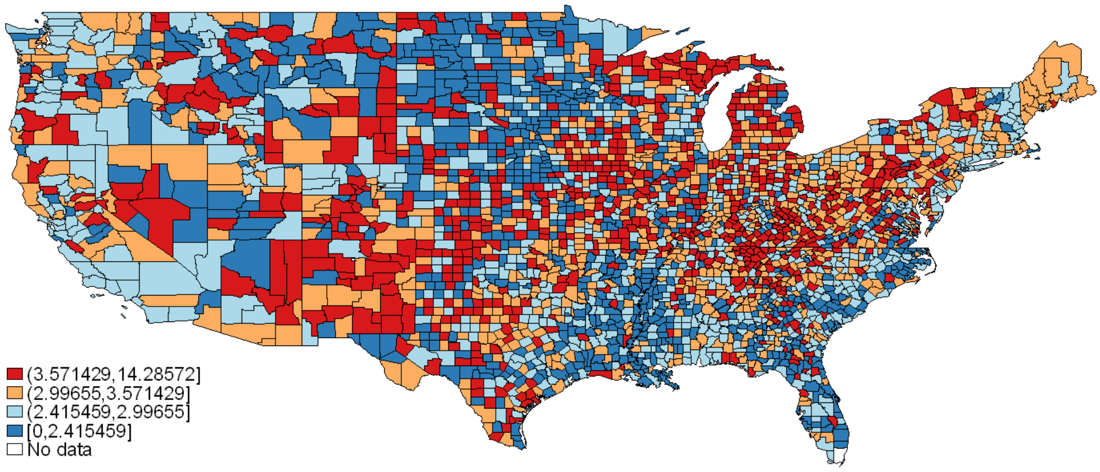
1.6 Reduced Form Exercise

The counterfactual results in the previous section depend on the model structure. The extended discussion on the calibrated parameter values in section 1.4.4 describes how they match similar estimates in the literature, related measures from external data sources, or well-documented stylized facts. Yet, to provide *further* evidence in support of the model, I perform a reduced form exercise. The exercise also allows me to add an extension to the model where *economywide* gender norms respond to *economywide* past female labor force participation, and using this relationship, to conduct dynamic counterfactuals. This model extension is compatible with how I identified norms wedges previously, as long as individuals take norms as given and do not internalize the effect of their labor supply choice on the norms of future generations.

1.6.1 Model validation

For lack of a direct test of model predictions when norms wedges fall, I explore the effects of a shock that *indirectly* affects norms and check that other variables change in the expected direction. Inspired by Fernández et al. (2004),[37] I consider WW2 draftee casualties as a temporary positive shock to female labor force participation that propagates over the long term through weaker gender norms. Underlying this story is the idea of cultural transmission through exposure (Bisin and Verdier, 2000) or social learning (Fernández, 2013; Fogli and Veldkamp, 2011).

Figure 1.10: Map of county-level draftee casualty rates



Notes: This figure color-codes each county into quartiles of draftee casualty rates. From the highest to the lowest quartile, the colors are red, orange, light blue, and blue.

For the reduced form exercise, I match the U.S. decennial census, by county, with the WW2 military casualty records from Ferrara (2019). As a result, every county is characterized by the casualty rates among draftees, as illustrated in Figure 1.10. While earlier studies on the effects of WW2 utilized WW2 mobilization rates by state (e.g. Acemoglu, Autor, and Lyle, 2004; Fernández, Fogli, and Olivetti, 2004), newly digitized data from the National Archives and Records Administration enables the use of *county*-level variation. Moreover, there are two advantages to using casualty rates as opposed to mobilization rates. First, although most women who engaged in wartime work left the labor force upon demobilization (Goldin, 1991), casualties last. Second, casualties are likely to be more random than mobilization rates.

The baseline estimation strategy for the effect of draftee casualties is difference-in-differences with continuous treatment. Hence, I estimate, for individual i in county c at decade t,

$$Y_{ict} = \alpha_c + \lambda_t + \sum_{t \neq 1940} \beta_t \times casualty_c + X_{ict}\gamma + \varepsilon_{ict} \tag{1.21}$$

where Y represents various outcome variables, *casualty* is the county-level draftee casualty rate, and X captures pre-determined individual characteristics, namely dummies for age and race, added for greater precision. In other words, I study the effect of the casualty rate in each decade t relative to 1940, the last decade before the influence of WW2 reached the U.S. With parallel trends, it must be that $\beta_t = 0$ for all $t < 1940$. I later check that the

[37] They argue that WW2 mobilization weakened traditional gender norms over the long run, as sons of women who worked during the war grew up to be more accepting of working wives.

results according to the specification in (1.21) are robust to a) comparing above-median- to below-median-casualty counties in a standard binary difference-in-differences framework, b) controlling for 1940 county characteristics that predict casualty rates, interacted with decade dummies, in order to address the nonrandomness of casualty rates, and c) applying the synthetic difference-in-differences methodology (Arkhangelsky et al., 2019) to further allay concerns over *level* differences in the pre-WW2 period affecting the future trajectory of various outcome variables.

The main results are depicted in Figure 1.11. Plot (A) shows that a 1 percentage point increase in draftee casualty rate induces a 2.5 percentage point increase in female labor force participation rate in 1950. When I dissect the source of this spike, it comes from widows and single women living with their parents increasing their labor force participation, consistent with firms demanding more female labor with lower male labor supply, and new widows increasing their labor supply as a direct consequence of the casualties. The effect of casualties on female labor force participation in 1960 is still positive but a little smaller than in 1950, and then displays a gradual increase over the next decades. Plot (B) shows that for the most part the gradual increase is driven by married women. At the same time, plot (C) shows that gender attitudes become gradually less traditional with higher casualties.[38] Put together, the three panels are consistent with a story of a one-off rise in female labor force participation propagating over the long term through less traditional gender norms that primarily affect married women's labor force participation.

Appendix figure 1.A6 portrays the reduced form results for other variables that buttress this story as well as the model structure. Single women's labor force participation in plot (B) does not mimic the strong, gradual rise observed for married women, reproduced in plot (A). This contrast supports the model assumption where only married individuals' labor force participation decisions are affected by gender norms. Men's employment, as shown in plot (C) does not exhibit a systematic change over time, indicating that the gender norm change is associated with a change in women's behavior, mostly. I therefore think of a change in the *female* norms wedge as the basis of the long-term effect of WW2 draftee casualties. Plot (D) shows that the wife's share of the couple's wage income is also increasing over time, even as real hourly wage for working women is decreasing over time, shown in plot (E). Although not included, the wife's share of the couple's total market hours worked also increases over time, gradually and continually. Furthermore, while wages are equilibrium prices jointly determined by supply and demand, the decrease in female wages points more towards a rise in female labor *supply* than female labor *demand* underlying the rise in female labor force participation. The effect on female wages thus

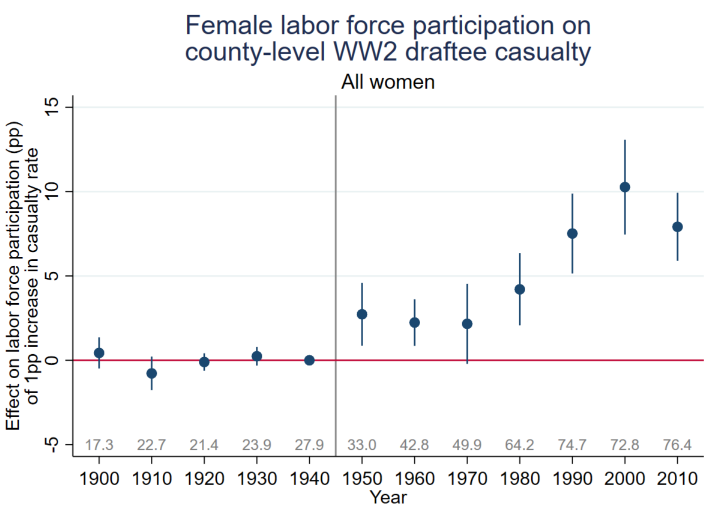
[38]There are a few caveats for plot (C). As the finest geographic variable in the attitudinal survey data is state, the difference-in-differences analysis is performed using state-level draftee casualties. Moreover, the surveys are grouped into five-year intervals. 1945 is counted as post-WW2, since the survey was taken in November, after the official end of WW2 in September. The statistically insignificant drop in 1945 appears to be out of trend, but it is the date in which the sample is by far the smallest; the sample is 1,365 in 1945, while the other dates are based on around 3,000-6,000 observations. 1945 is also the only date in which no survey weights are available. Overall, I take the coefficient plot of plot (C) to indicate that the attitude data is quite noisy, and that the attitudes getting less traditional becomes detectable (statistically significant) from 1985.

supports the model assumption of a decrease in female norms wedge affecting the labor *supply* decisions of married women. In addition, the decrease in female wages is consistent with the model assumption around selection into the labor force, i.e. as more women work, working women are less positively selected. The observation that there are long-term differences in female wages between high- and low-casualty counties indicates that there are labor market frictions precluding the equality of wages across space. In fact, there is no consistent trend of individuals moving to either high- or low-casualty areas to wash out the effects of WW2 casualties, which would have shown up in people migrating to different states from their birth states in plot (F).

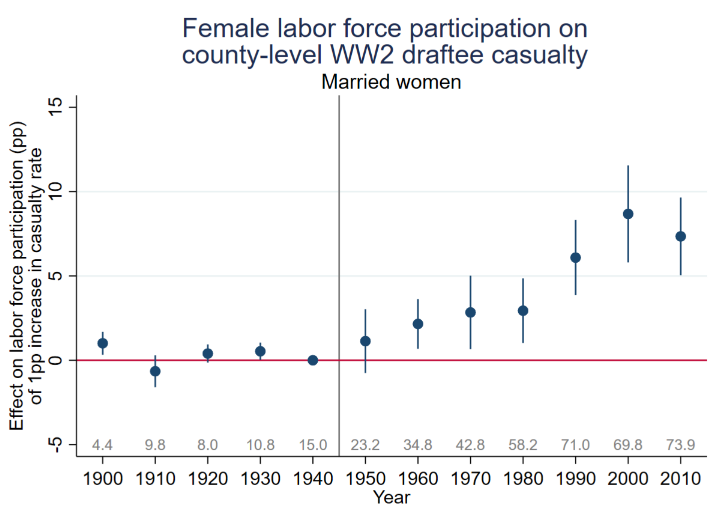
In terms of marriage and education, plot (G) depicts a rise in ever-marriage rates and plot (H) a rise in the education of women overall. Plot (H) uses whether one graduated high school or more as the measure of education, as higher levels of education are very rare to find in 1940, the pre-WW2 benchmark decade. The effect on marriage is consistent with the model assumption where norms wedges are modeled as costs to marriage, and as the female norms wedge falls over time, more people engage in the tradition of marriage. Also, the effect on education supports the model prediction of greater female education due to the greater likelihood of marriage and market work. Lastly, the model assumes that the female norms wedges impose higher costs on women with higher market ability. This assumption leads to the model prediction of increasingly positive selection into marriage by education of women, as the female norms wedge falls. The rise in education of *married* women, in plot (I), but not for single women, in plot (J) supports this prediction.

All in all, it is difficult to reconcile how women are getting married and educated more, married women but not single women are working in the market more, and female wages falling, without a story of changing gender norms. Gender attitudes indeed become less traditional over time in the data. Surely, WW2 casualties can have alternative effects. For instance, the fall in sex ratio can increase husbands' bargaining power. Yet in that case, married women would *not* increase market work, since attitudinal surveys indicate that men hold more traditional views on married women working than women. As another example, casualties might somehow change the industrial structure into one that better enables women to combine work and marriage, such that women with higher market ability get married more. However, while this might explain the rise in married women's market work, the rise in marriage, and the rise in female education, it goes against falling female wages. To generate the sizable rise in married women's work *soley* from a higher market talent of married women, married women's market talent must rise by a great amount, in which case it is unlikely to see a fall in female wages.

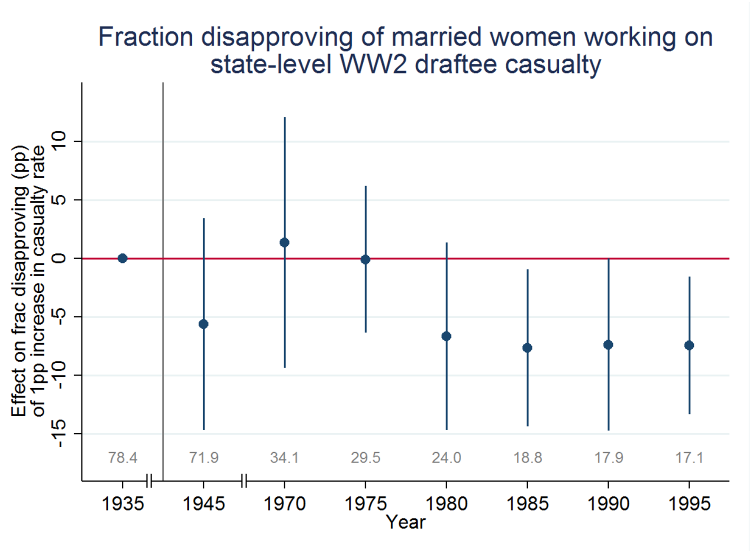
Figure 1.11: The effect of WW2 draftee casualty rates on various outcomes



(A)



(B)



(C)

Notes: This figure plots the difference-in-differences coefficients from estimating equation (1.21) for various outcome variables. Plot (C) uses state-level draftee casualty rates, because state is the finest geographic variable available in attitudes data prior to WW2. In Plot (C), 1945 is counted as post-WW2, since the survey was taken in November, after the official end of WW2 in September.

Robustness I firstly check that the effect of WW2 draftee casualties survive a binary difference-in-differences specification. Column (1) of Appendix table 1.A6 reports the effects of casualties on female labor force participation, pictured in plot (B) of figure 1.11. Column (2), which reports the binary specification results, are very similar to column (1).

Secondly, I control for 1940 county characteristics that predict casualty rates, interacted with decade dummies, to address the nonrandomness of casualty rates. Indeed, the casualty rates are not completely random. Figure 1.10 shows spatial clustering in the casualty rates. During WW2, drafted soldiers were assembled at *state* base camps, and casualties were dictated by outcomes of specific battles, so nearby counties experience similar casualty rates. Moreover, blacks were killed at a lower rate since they were mainly employed in comparatively safer support and supply activities due to racist attitudes that saw them unfit for fighting (Lee, 1965). Appendix table 1.A7 shows that casualty rates were higher in counties with a higher share of whites, a lower share of working-age women, a higher urban resident share, a lower male education, and a lower share of men in agriculture. I therefore control for the effects of these 1940 county characteristics over time in columns (3) and (4) of Appendix table 1.A6. Although the coefficient sizes get smaller, both columns still demonstrate a gradual rise in female labor force participation over time.

Lastly, to further allay concerns over *level* differences in various outcome variables during the pre-WW2 period affecting the future trajectory of those variables, which would bias the usual difference-in-differences coefficients, I apply the synthetic difference-in-differences methodology of Arkhangelsky et al. (2019). Synthetic difference-in-differences estimates weights on control counties and on time periods such that the pre-WW2 path of the doubly-weighted average lies extremely close to the pre-WW2 path of the treatment counties. Hence, any level differences in the pre-WW2 period between treatment and control counties, for any outcome variable, are practically eliminated. Because of the need to divide treatment and control counties, I can only employ a binary specification. Column (5) of Appendix table 1.A6 shows that the synthetic difference-in-differences coefficients also depict a gradual rise in female labor force participation over time.

1.6.2 Dynamic counterfactuals

The reduced form results based on WW2 casualties are useful for validating the assumptions of the structural model. Not only that, but those results also allow me to consider a dynamic extension to the model. Assuming that the model is in steady state each decade, there is no dynamic element linking any two decades in the model. Yet, it is unlikely that gender norms evolve entirely exogenously. In fact, the WW2 reduce form results are congruent with temporarily higher female labor force participation inducing a gradual fall in the female norms wedge.

I estimate how the female norms wedge responds to past female labor force participation, using the reduce form results in plot (A) of figure 1.11. To this end, I impose a function form on how *economywide* female norms wedge in decade t responds to *economywide* female labor force participation in decade $t-1$:

$$\begin{aligned} \underline{\Delta}\tau_t &= f(\underline{\Delta}FLFP_{t-1}, FLFP_{t-1}) + \nu_t \\ \approx\ & \alpha_0 + \alpha_1\underline{\Delta}FLFP_{t-1} + \alpha_2 FLFP_{t-1} + \alpha_3\underline{\Delta}FLFP_{t-1} \cdot FLFP_{t-1} + \nu_t \end{aligned}$$

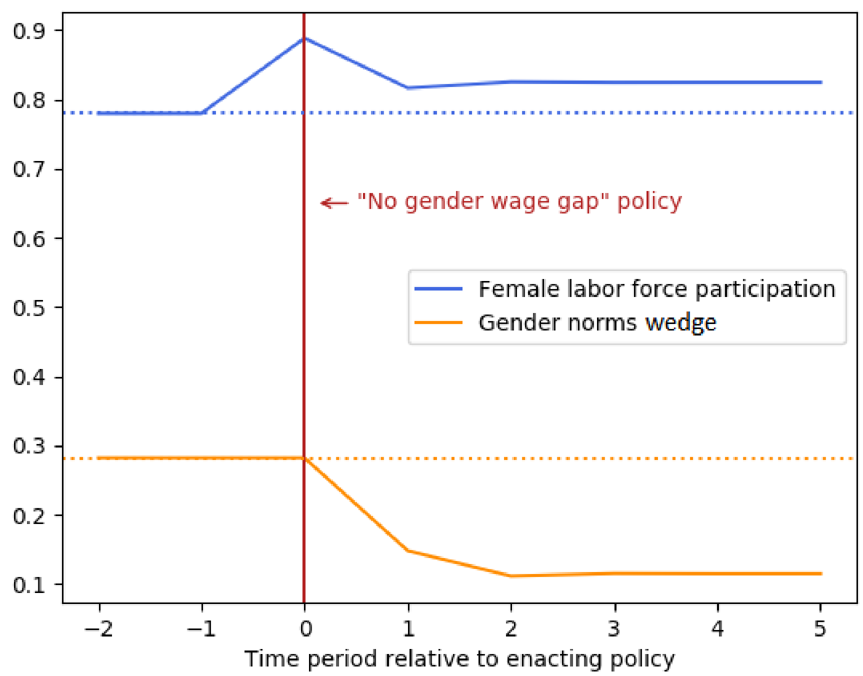
where $\underline{\Delta}$ denotes the gap between treatment and control counties. I also need two additional assumptions on how the long-term effects rise about: a) WW2 draftee casualties affect female labor force participation in 1950 and nothing else, and b) the effect only propagates through a change in the female norms wedge. These assumptions allows me to estimate the relationship between the female norms wedge and the past decade's female labor force participation that would generate the pattern of coefficients in plot (A):

$$\min_{\alpha_0,\alpha_1,\alpha_2,\alpha_3} \sum_t (\text{DID coeff, FLFP}_t - \text{change in FLFP}_t \text{ in model due to } \underline{\Delta} \tau_t)^2$$

The result is $\hat{\alpha}_0 = -0.102$, $\hat{\alpha}_1 = 0.368$, $\hat{\alpha}_2 = 0.242$, $\hat{\alpha}_3 = -1.209$.

Armed with this estimated relationship, I can further conduct *dynamic* counterfactuals. As opposed to the "static" counterfactuals on the effect of a shock on the model steady state in a given decade in section 1.5, I can explore how a shock affects the model steady state *over time*.

Figure 1.12: The effect of paying women male wages, one-off, in 2010



Notes: This figure plots the difference-in-differences coefficients from estimating equation (1.21) for various outcome variables. Plot (C) uses state-level draftee casualty rates, because state is the finest geographic variable available in attitudes data prior to WW2. In Plot (C), 1945 is counted as post-WW2, since the survey was taken in November, after the official end of WW2 in September.

The counterfactual I ask is, what would happen in 2010 if women were paid male wages in a one-off fashion? The counterfactual abstracts from labor demand being affected, as a consequence of the model assumption of firms producing under a linear production function. Forcing firms to pay women male wages, all the more without changing employment, is a far-fetched idea. Yet this thought experiment is illustrative of how a one-off policy can move an economy into a different equilibrium. Figure 1.12 shows that while keeping all other parameters fixed at the 2010 level, paying women male wages for one period induces a contemporaneous spike in female labor force participation, which then induces the female norms wedge to fall a decade later. The female labor force participation that decade is lower than the decade of the policy as the direct effect of the policy is gone, but it is higher than the baseline due to the lower female norms wedge. Consequently, in the *following* decade, the female norms wedge falls even more. The process continues, and from three decades post-policy, the economy stabilizes at a new equilibrium with higher female labor force participation and lower female norms wedge than the baseline.

1.7 Conclusions

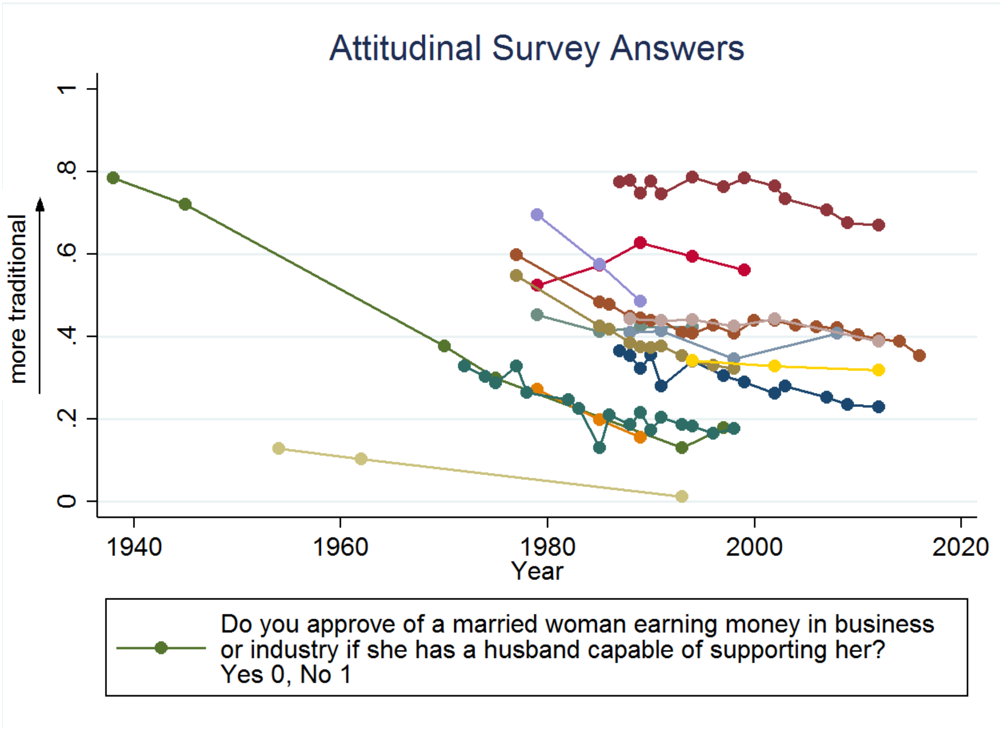
In this paper I measure and study the effects of gender roles associated with marriage on aggregate output, using historical data from the U.S. Gender norms became less constraining on married individuals' labor supply choices. Through direct effects on labor supply choices becoming more aligned with productivity maximization, and through indirect effects on higher education, weaker gender norms increase aggregate market and total output. Moreover, a one-off policy inducing a large rise in female labor force participation can bring an economy to a new equilibrium with higher female labor force participation.

We do not learn about development and growth only from developing countries. Rather, we can also learn from a currently developed country that has undergone large historic changes. In fact, the trend in the U.S. over the last century of gender attitudes becoming less traditional and married women's labor force participation catching up with single women's is resonated in numerous parts of the world. At the same time, one in ten countries of the world still has lower female labor force participation than 1940 U.S. (International Labor Organization, 2019). Thus, this paper can be informative about the potential growth consequences and the underlying channels of cultural change in other countries that currently operate under traditional gender roles or are moving away from it.

A natural extension to the current paper is to take advantage of the fact that parameter identification is straightforward in this model, and apply the model to other countries. There are also other important dimensions that the model does not account for, such as occupations (Hsieh et al., 2019), work flexibility (Goldin, 2014), divorce (Fernández and Wong, 2014; Greenwood et al., 2016), and leisure (Aguiar and Hurst, 2007b). Building these factors into the model will allow a richer understanding of the effects of gender norms in marriage.

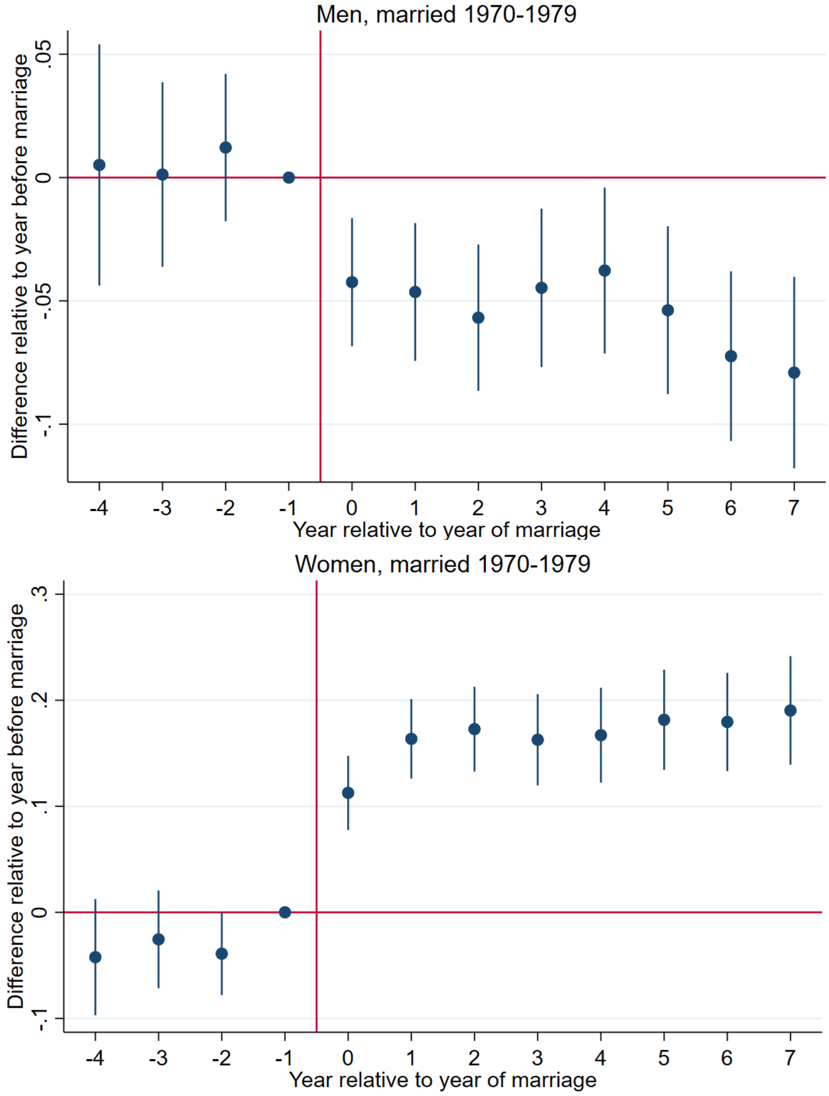
1.A Appendix – Figures

Figure 1.A1: Trends of answers to various attitudinal survey questions relating to gender roles within marriage



Notes: Various survey questions asked multiple times.

Figure 1.A2: Housework's share of housework and market hours, among couples whose first child is born $\geq$ 4 years after marriage

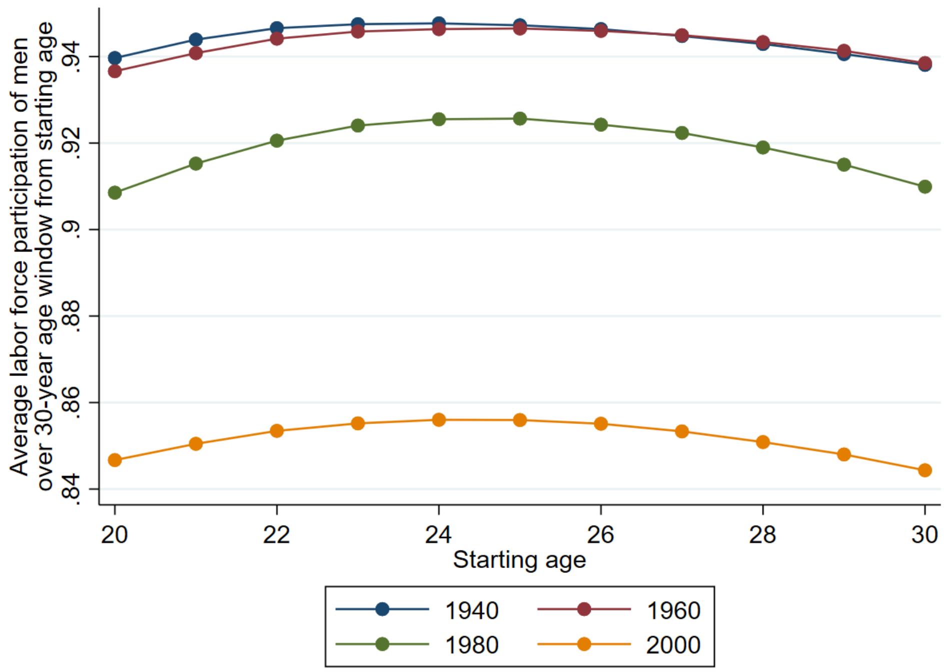


Notes: This figure plots the event-time coefficients (α_j^g) of the regression

$$housework_{ist}^g = \sum_{j \neq -1} \alpha_j^g \cdot \mathbb{1}(j = t) + \sum_k \beta_k^g \cdot \mathbb{1}(k = age_{is}) + \sum_y \gamma_y^g \cdot \mathbb{1}(y = s) + \nu_{ist}^g$$

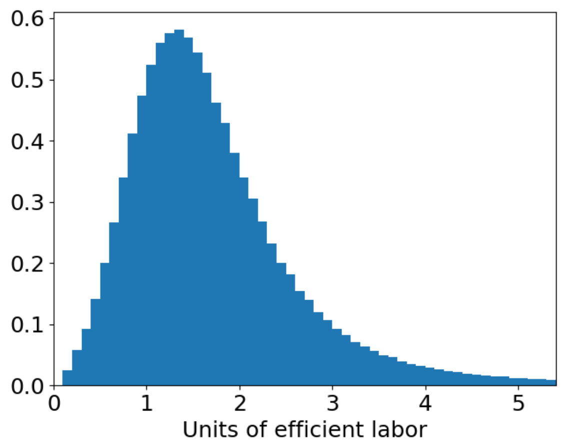
where $housework_{ist}^g$ denotes the housework's share of housework and market work hours of individual i of gender g in year s at event time t. The red vertical line plots the timing of marriage. Individuals are unmarried household heads without any live-in partners in the years to the left of the red line, and they are married with live-in spouses in the years to the right of the red line.

Figure 1.A3: Average labor force participation of men over 30-year window from starting age



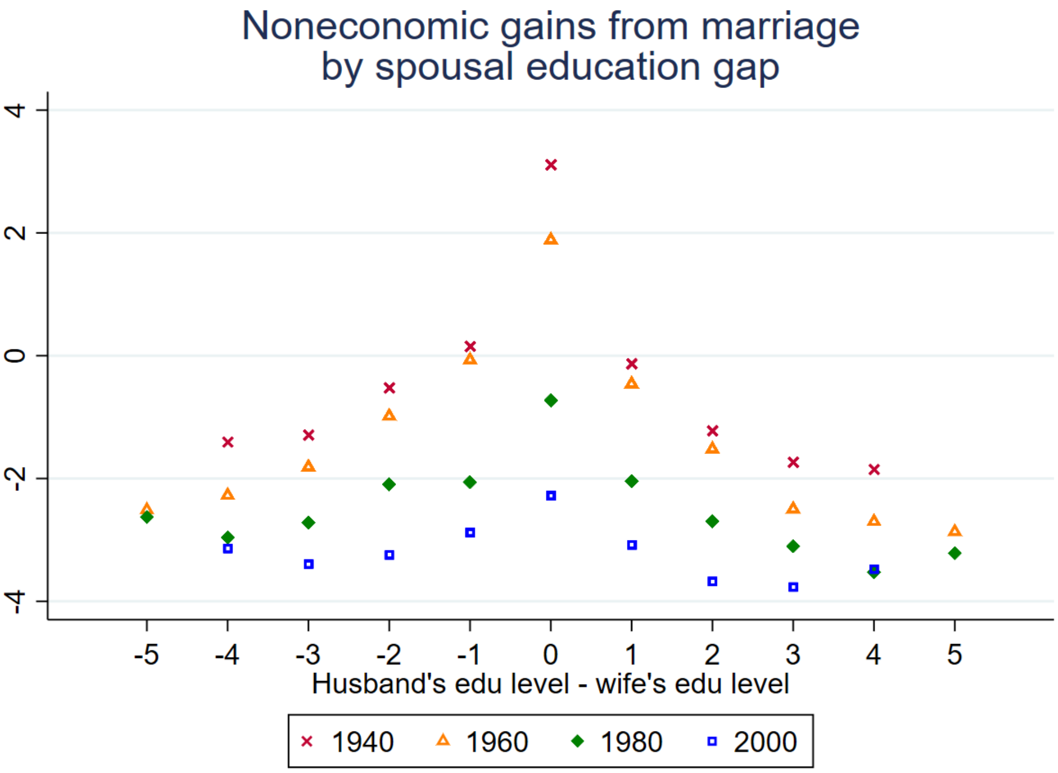
Notes: This figure plots the weighted average of labor force participation among men aged between starting age and (starting age+29). It shows that 25-54 is an appropriate age range for the economically active years of one's life and that this observation is quite stable over time.

Figure 1.A4: Histogram of empirical market abilities



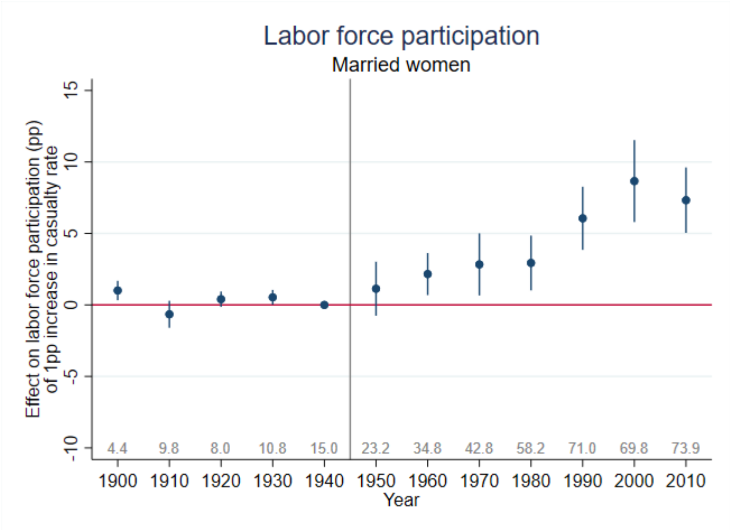
Notes: This figure plots the histogram of the market abilities of all working individuals with wage data, where the market abilities are inferred using the model structure as outlined in step 2 of section 1.4.3.

Figure 1.A5: Ψ by spousal education gap, over time

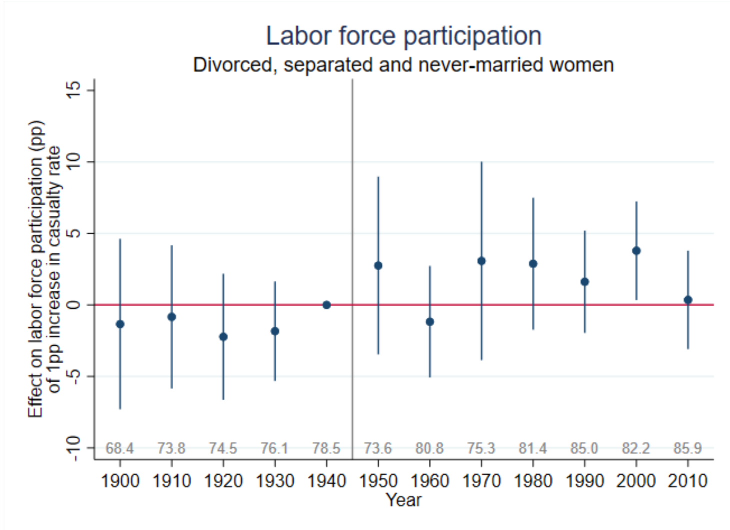


Notes: This figure plots the pattern over time of Ψ^{qr}, the utility from marital bliss in a marriage between a man with education level q and a woman with education level r, when averaged by the spousal educational gap $(q - r)$.

Figure 1.A6: The effect of WW2 draftee casualty rates on various outcomes

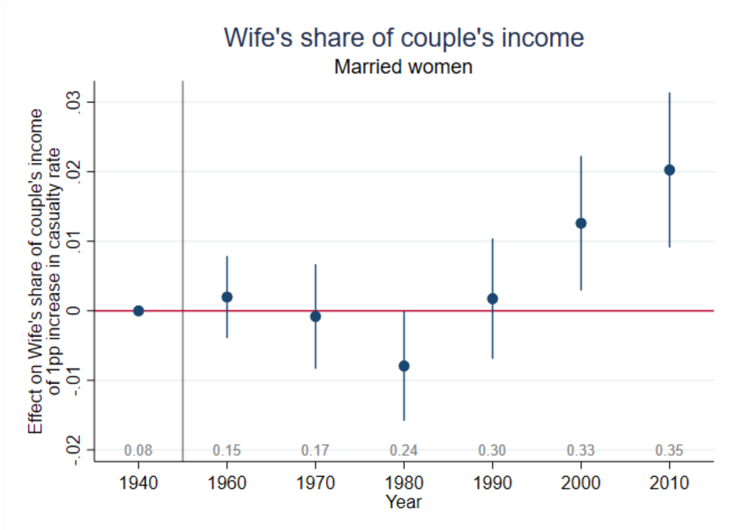


(A)

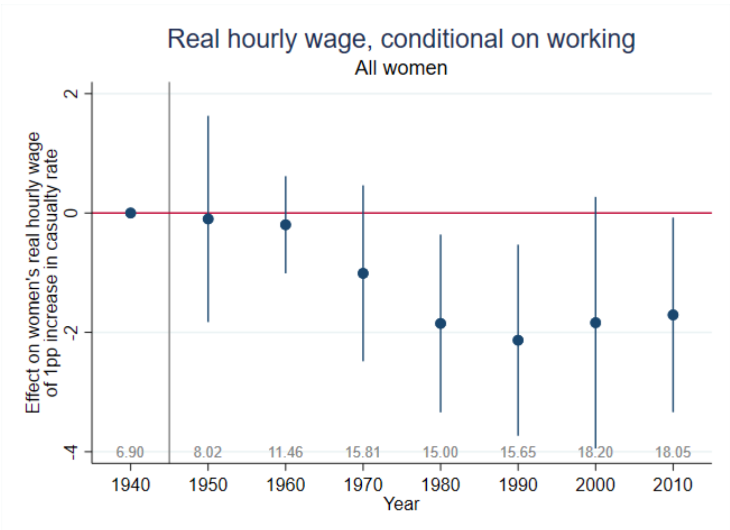


(B)

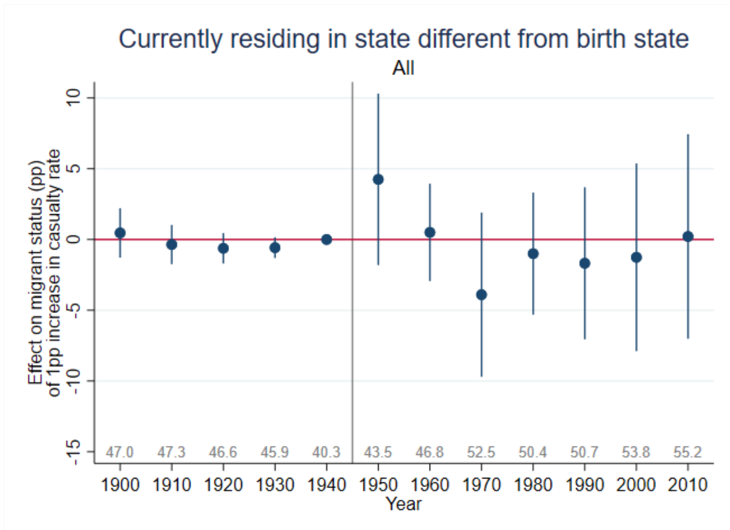
(C)



(D)

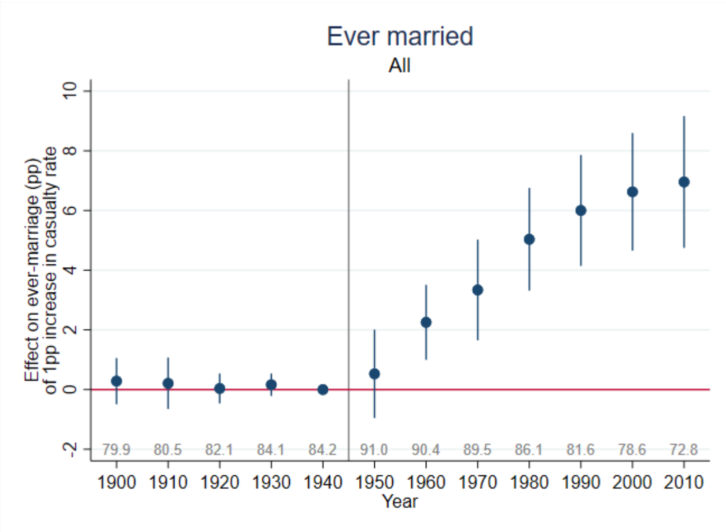


(E)

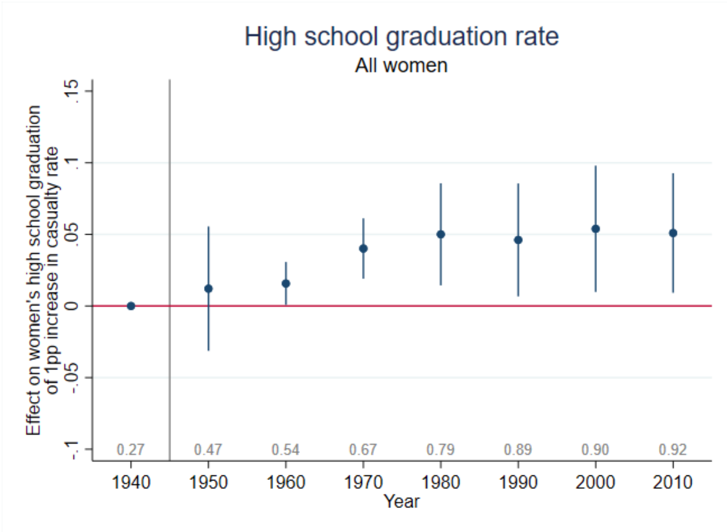


(F)

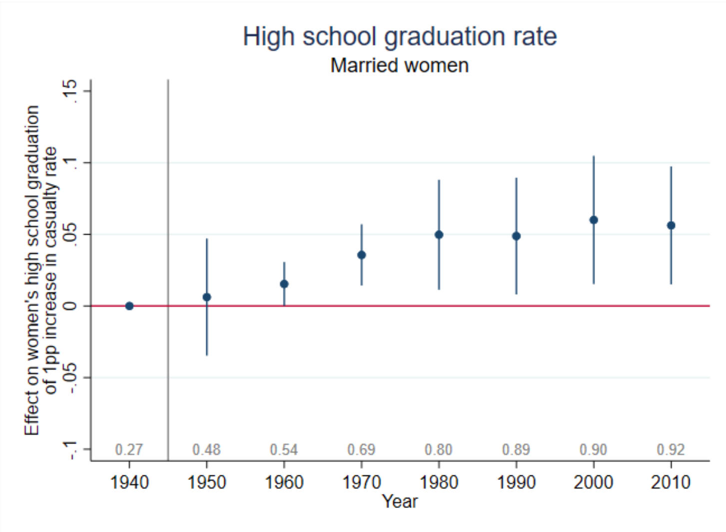
Figure A6 (continued): The effect of WW2 draftee casualty rates on various outcomes



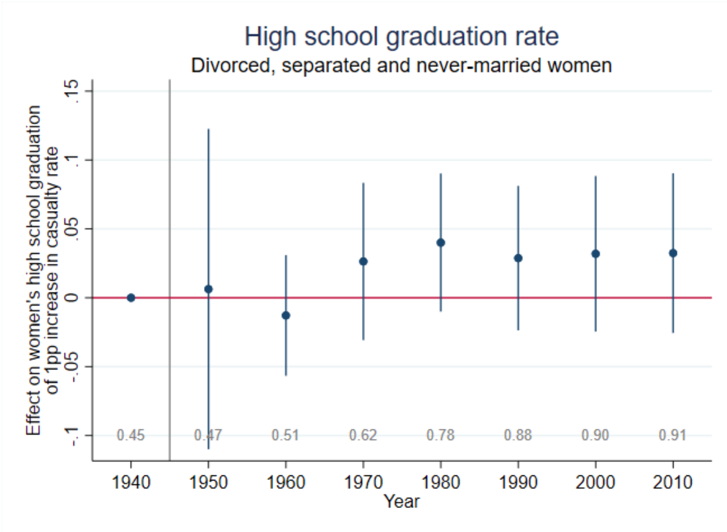
(G)



(H)



(I)



(J)

Notes: This figure plots the difference-in-differences coefficients from estimating equation (1.21) for various outcome variables.

1.B Appendix – Tables

Table 1.A1: Variation in attitudes by individual characteristics

	(1) Average	(2) F-statistic	(3) Shapley decomposition (%)
Year			
1930-1939	0.77		
1940-1949	0.71		
1970-1979	0.26	10.3	19.8
1980-1989	0.16		
1990-1999	0.14		
Marital status			
Married	0.19		
Widowed	0.28		
Divorced	0.18	1.0	2.3
Separated	0.19		
Never married	0.15		
Sex			
Male	0.20	14.2	1.1
Female	0.18		
Race			
White	0.28		
Black	0.25	14.1	3.2
Other	0.21		
Education			
Middle school graduate or lower	0.41		
High school drop-out	0.29		
High school graduate	0.20	153.6	62.6
College drop-out	0.14		
College graduate or higher	0.09		
Age			
20-29	0.23		
30-39	0.26	2.2	11.1
40-49	0.29		
50-59	0.33		

Notes: This table reports by how much various individual characteristics account for the variation in *att*, the indicator variable for an individual's disapproval of a married woman working in the labor market if she has a husband capable of supporting her, in the Gallup Polls and the General Social Survey. The specific attitudinal survey question of interest is in Figure 1.3. Column (1) reports the weighted average disapproval rate by category of each variable, to show the variation across the categories. Columns (2) and (3) are based on the regression of *att* on the dummies for the categories of each variable, on the sample from the General Social Survey. Column (2) reports the F-statistic for the joint significance of the dummies belonging to each variable, while column (3) reports the Shorrocks-Shapely decomposition denoting the relative contribution of each variable to the R squared.

Table 1.A2: Years of completed schooling by schooling category by year

Schooling level	Year						
	1940	1960	1970	1980	1990	2000	2010
1	[0,7]	[0,7]	[0,8]	[0,9]	[0,11]	[0,11]	[0,11]
2	8	8	[9,11]	[10,11]	12	12	12
3	[9,11]	[9,11]	12	12	[13,15]	[13,15]	[13,15]
4	12	12	[13,15]	[13,15]	16	16	16
5	[13,∞)	[13,15]	[16,∞)	16	[17,∞)	[17,∞)	[17,∞)
6		[16,∞)		[17,∞)			

Notes: Years of completed schooling are integers in each interval. For example, individuals with 0,1,...,7 years of completed schooling fall under schooling level 1 in 1940.

Table 1.A3: Description of family composition categories

Family composition categories	Description
1	No child
2	1 child, aged 6-18
3	1 child, aged 0-5
4	2 or more children, all aged 6-18
5	2 or more children, at least one aged 0-5

Notes: All children are one's own children living in the same household.

Table 1.A4: Coefficients from regressing $\bar{w}$ or $\bar{h}$ on schooling level, by sex and year

		1940	1960	1970	1980	1990	2000	2010
Male	Market productivity, $\bar{w}$	0.61***	0.84***	1.53***	1.12***	1.91***	2.48***	2.60***
		(13.98)	(16.22)	(11.97)	(10.56)	(16.75)	(16.74)	(15.40)
	Home productivity, $\bar{h}$	0.021	-0.068**	-0.17***	-0.13***	-0.064	0.029	0.016
		(0.96)	(-3.16)	(-3.80)	(-5.26)	(-1.76)	(0.72)	(0.39)
Female	Market productivity, $\bar{w}$	0.25***	0.46***	1.03***	0.98***	1.71***	1.90***	2.06***
		(5.29)	(6.17)	(7.34)	(13.19)	(20.76)	(30.45)	(38.36)
	Home productivity, $\bar{h}$	0.20***	0.11**	0.085	-0.25***	-0.21***	-0.069*	0.029
		(8.52)	(2.90)	(0.90)	(-7.19)	(-6.02)	(-2.56)	(1.04)
Number of groups		210	210	150	210	150	150	150

Notes: t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
This table reports the coefficients on schooling level in the regression of either market wage or home productivity on schooling level. In this regression, each observation corresponds to each group, weighted by its empirical probability. As detailed in Table 1.A2 of the Appendix, schooling level is a categorical variable with higher values representing greater numbers of completed years of schooling. For the sake of simplicitiy in showing how $\bar{w}$ and $\bar{h}$ vary by education, I treat schooling level as a continuous variable in these regressions. Since the schooling level formulations change over time, the coefficients are not directly comparable across the years.

Table 1.A5: Estimates of the costs of schooling C by sex and decade

		Schooling level					
		1	2	3	4	5	6
1940	Male	-1.59	-0.64	0.00	0.65	1.58	
	Female	-1.70	-0.48	-0.08	0.46	1.81	
1960	Male	-1.07	-0.34	-0.49	-0.48	0.95	1.43
	Female	-1.27	-0.43	-0.75	-0.70	0.95	2.19
1970	Male	-0.88	-0.48	-0.71	0.66	1.41	
	Female	-1.09	-0.96	-0.95	0.83	2.16	
1980	Male	-0.75	0.07	-1.07	-0.33	0.90	1.18
	Female	-0.98	-0.51	-1.30	-0.14	1.19	1.73
1990	Male	-1.09	-1.24	-0.64	0.77	2.20	
	Female	-1.44	-1.33	-0.69	1.00	2.44	
2000	Male	-1.32	-1.63	-0.50	0.91	2.55	
	Female	-1.37	-1.59	-0.53	1.03	2.46	
2010	Male	-1.54	-1.62	-0.64	1.02	2.79	
	Female	-1.58	-1.45	-0.72	1.15	2.61	

Notes: This table reports the estimates of C_g^s, the direct cost minus the expected utility from intra-household resource allocation, by sex and decade. These estimates are not comparable by decade because the schooling level categories differ by decade, except for 1990, 2000, and 2010.

Table 1.A6: The effect of WW2 draftee casualties on female labor force participation (percentage points)

	DID		DID with 1940 county controls		Synthetic DID
	continuous	binary	continuous	binary	binary
	(1)	(2)	(3)	(4)	(5)
1950	2.54***	2.13**	1.01	0.42	3.43***
	(0.98)	(0.86)	(0.83)	(0.87)	(0.15)
1960	2.13***	1.64**	0.66	0.53	4.07***
	(0.75)	(0.67)	(0.75)	(0.69)	(0.13)
1970	2.69**	2.03*	-0.09	-0.07	5.29***
	(1.11)	(1.09)	(-1.38)	(-1.25)	(0.47)
1980	4.03***	3.04***	1.64	0.55	6.81***
	(1.03)	(0.83)	(1.08)	(0.80)	(0.29)
1990	7.11***	5.57***	3.40***	1.76**	7.98***
	(1.15)	(0.94)	(1.16)	(0.87)	(0.33)
2000	9.78***	7.77***	5.08***	2.98***	9.50***
	(1.46)	(1.20)	(1.41)	(1.12)	(0.37)
2010	7.50***	6.23***	3.52***	2.41***	8.41***
	(1.04)	(0.92)	(1.01)	(0.93)	(0.36)
N	2,096,633	2,096,633	2,028,727	2,028,727	2,096,633

Notes: This table reports the decade-specific difference-in-differences coefficients of female labor force participation on WW2 draftee casualties by county. Draftee casualties are considered either as the raw continuous variable or as a binary variable equal to 1 if casualty rates are above the median. When casualties are continuous, the coefficients amount to percentage point changes in married women's labor force participation for every 1 percentage point increase in casualty rates. When casualties are binary, the coefficients amount to percentage point changes in married women's labor force participation for being in above-median counties relative to below-median counties. Columns (1) and (2) reports the result of estimating equation (1.21). Columns (3) and (4) add as controls 1940 county characteristics that predict casualty rates, interacted with decade dummies, in order to address the nonrandomness of casualty rates. Column (5) reports the coefficients from applying the synthetic difference-in-differences methodology (Arkhangelsky et al., 2019) to further allay concerns over *level* differences in the pre-WW2 period affecting the future trajectory of various outcome variables.

Table 1.A7: WW2 casualties and county characteristics in 1940

	Dependent variable: WW2 casualty rate among draftees, county-level			
	(1)	(2)	(3)	(4)
Share white	0.38***	0.32***	0.31***	0.31***
	(13.30)	(9.89)	(9.91)	(9.15)
Share aged 25-54 among women	-0.22***	-0.16***	-0.15***	-0.12***
	(-6.99)	(-3.75)	(-2.85)	(-2.71)
Share city resident	0.22***	0.23***	0.23***	0.18***
	(4.01)	(5.42)	(5.62)	(3.45)
Male avg. schooling	-0.13***	-0.09*	-0.14**	-0.14**
	(-3.07)	(-1.77)	(-2.32)	(-2.41)
Share in agriculture among men	-0.11***	-0.02	-0.03	-0.06
	(-3.88)	(-0.45)	(-0.77)	(-1.25)
Share married			-0.01	-0.01
			(-0.24)	(-0.17)
Female avg. schooling			0.07	0.05
			(1.02)	(0.92)
Avg. no. children in household			0.02	-0.02
			(0.51)	(-0.42)
Female labor force participation			-0.01	-0.02
			(-0.30)	(-0.57)
Additional controls	No	No	No	Yes
State dummies	No	Yes	Yes	Yes
No. counties	2409	2409	2409	2409
R^2	0.226	0.340	0.341	0.345

Notes: t statistics (col (1): robust, col (2)-(4): clustered by state) in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

This table reports the OLS coefficients from regressing county-level WW2 draftee casualty rates on various pre-war county characteristics. The sample excludes outliers, defined as counties having casualty rates or draft rates strictly outside the 2.5^{th} and 97.5^{th} percentiles. Additional controls are county population, share having non-wage income over $50, share aged 25-54 among men, and share in agriculture among women. Average schooling, share in agriculture, share married, average number of children in household, and female labor force participation are computed among 25- to 54-year-olds.

1.C Appendix – Model

1.C.1 General form of the utility function

The general form of the utility function is given by

$$u_i(Q, C_i, L_f, L_m) = H\Big(f(Q)C_i - r(Q)\big[\tau_{Fi}w_f L_f + \tau_{Mi}h_m(1-L_m)\big] + g_i(Q)\Big)$$

where the following conditions hold:

<u>**Conditions**</u>

C1) H is strictly increasing and strictly concave $\Rightarrow$ Individuals are risk-averse

C2) $(H')^{-1}$ is homogeneous or logarithmically homogeneous[39,40,41] $\Rightarrow$ Efficient risk-sharing implies spouses split the sum of individual "inverse-H" utilities[42] in an affine way.

C3) $2p(f')^2 - p{\cdot}f{\cdot}f'' + \big[\tau_F w_f L_f + \tau_M h_m(1-L_m)\big](r''f' - r'f'') - f'g'' + g'f'' > 0$, where $\tau_F \equiv \tau_{Fm} + \tau_{Ff}$, $\tau_M \equiv \tau_{Mm} + \tau_{Mf}$, and $g(Q) \equiv g_m(Q) + g_f(Q)$ $\Rightarrow$ Both spouses' indirect utility functions are increasing in the consumption of the public good *and* the optimal public good consumption is increasing in $M - \tilde{A}$.

This general utility function yields the same result that 1) the optimal labor supply decision of the couple is one that maximizes their pooled income less the disutilities from nontraditional working arrangements, and 2) the optimal labor decisions are made independently based only on individuals' comparisons of the gains from working in the market versus at home.

1.C.2 Model discussion: On how home production features as "income" in the budget constraint

It might seem non-standard that according to the budget constraint, private and public goods can be "bought" with "income" from home production. However, the maximization problem is equivalent to solving

$$\max_{Q,C,Y,B}(Q+Y)\big(C+B-\tau w_f L_f\big)$$

[39] A function is logarithmically homogeneous if it is given by a logarithmic transformation of a homogeneous function. According to Miyake (2015), a function U is logarithmically homogeneous on X if and only if there is a δ-homogeneous function u on X and two parameters $a > 0$ and b such that $U(x) = a\log u(x) + b$ for all $x \in X$. The implication is that $U(\gamma x) = a\delta\log\gamma + U(x)$.

[40] CRRA ($H(x) = \ln(x)$, or $H(x) = \frac{x^{1-\theta}-1}{1-\theta}$ for $\theta > 0, \theta \neq 1$) and CARA ($H(x) = -\exp\{-\theta x + b\}$ for $\theta > 0, b \in R$) utility functions – the most commonly used utility functions for risk-averse individuals – satisfy condition C2).

[41] Mazzocco (2004) shows that a collective household's behavior under uncertainty is equivalent to that of a representative agent if and only if H is of the Identically Shaped Harmonic Absolute Risk Aversion (ISHARA) class:

$$-\frac{H_i''(x)}{H_i'(x)} = \frac{1}{\theta x + a_i}$$

[42] By "inverse-H" utilities, I mean $H^{-1}(u_i)$, or the part of the utility function inside $H(.)$.

$$\text{s.t.} \qquad pQ + C = w_m L_m + w_f L_f$$

$$pY + B = h_m(1 - L_m) + h_f(1 - L_f)$$

where Y is the non-rival, public component of home production (e.g. cleaning of communal area, or food preparation for children) and $B \equiv B_m + B_f$ is the total consumption of the private component of home production (e.g. cleaning of private space, laundry of clothes). Here, market goods and services Q and C can only be financed from market earnings, while consumption of home-produced goods and services occur within the total home production done by the couple. The market value of private home goods is normalized to be the same as that of the private market good (the numeraire), and the market value of public home goods is the same as the price of public market good.

1.C.3 Model discussion: On the perfect substitutability of home-produced goods and market goods

It is difficult to believe that home-produced goods and market goods are perfectly substitutable. Following Gronau (1977), I can categorize home production into work at home, which is perfectly substitutable to work in the market, and leisure (i.e. home consumption time), which has poor market substitutes. Then, the utility function can be formulated as

$$u_i(Q, C_i, L_f, L_m) = H\Big(f(Q)C_i - r(Q)\big[a_i w_i L_i + \tau_i w_f L_f\big] + g_i(Q)\Big)$$

where a_i denotes i's preference for leisure/home consumption time, measured proportionally to i's market wage.[43] Ultimately, the optimal labor supply decisions would come down to

$$L_m^* = \mathbb{1}\left[w_m - a_m \geq h_m\right]$$
$$L_f^* = \mathbb{1}\left[(1 - \tau)w_f - a_f \geq h_f\right]$$

1.C.4 Model discussion: On the independence of husbands' and wives' labor supply decisions

In the real world, there is dependence in husbands' and wives' labor supply decisions (L_m and L_f) for many reasons, including specialization and diminishing marginal utility of household market consumption. The TU nature of the utility function in this model, along with the perfect substitutability between market goods and home produced goods, however, precludes such dependence.

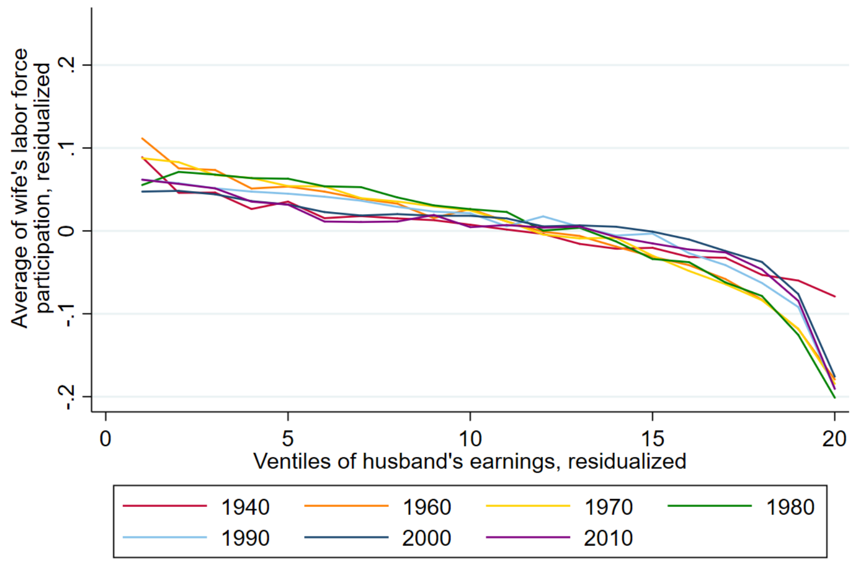
However, even in this model there is some degree of dependence, to the extent that wages and home productivities are modeled to be couple-specific; they depend on the schooling levels of both spouses, as well as their shared family composition characteristics (a vector of family size, number of children under 18, and number of children under 5).[44] For example, let us say a man with a master's degree and a woman who dropped out of

[43] It is easy to also introduce j valuing i's home consumption time, e.g. the husband valuing the wife's play time with their children.

[44] Details about wages and home productivities are in subsection 1.3.1.

high school are married. Empirically, the wife in this household is highly unlikely to work in the labor market, whereas a female high school dropout married to another high school dropout is more likely to work in the labor market. Thus there must be interdependence between husbands' and wives' labor supply decisions. But I can model their labor decisions as independent choices while having *market wages and home productivities to be match-specific.*

Figure 1.A7: Average of wives' labor force participation by ventiles of husbands' earnings



Notes: This figure plots data derived from the sample of married women whose husbands are earning strictly positive wages. The horizontal axis denotes the first to the twentieth ventile of residuals from the regression of the husbands' earnings on age, race, years of schooling, spousal years of schooling, number of children, number of children under 5, family size dummies and county dummies. The reason for adding these particular controls is because these individual characteristics together define a group in the model. On the vertical axis is the weighted average of the residuals from the regression of the wives' labor force participation indicators on the same variables. All regressions are run separately for each decade.

Moreover, the concern of ignoring the dependence between L_m and L_f when evaluating the effects of changing gender norms over time is alleviated if the reasons behind the dependence are fixed over time. Therefore, I plot in Figure 1.A7 the average of wives' labor force participation rates by ventiles of husbands' incomes, decade by decade, since husbands' earnings can be thought to affect households' incentives for division of labor and marginal utilities of market consumption.[45] The pattern is very similar over time, except for the first and last ventile. This result demonstrates that drivers of the dependence of L_m and L_f might be more or less similar over time.

[45]Both husbands' earnings and wives' labor force participation are residualized for race, schoolings of both spouses, number of children, number of children under 5, and family size.

1.C.5 Model extension: adding male norms wedges

The independence in the labor supply decisions of husbands and wives, arising from the perfect substitutability of home-produced goods and market goods, allows for the addition of *male* norms wedges. Then, a married couple receives disutility from not only the wife working in the market, but also the husband working in home production. Therefore, married individual $i \in \{m, f\}$ gets the following utility:

$$u_i(Q, C_i, L_f, L_m) = \ln(Q) + \ln\left(C_i - \tau_i w_f L_f - \tilde{\tau}_i h_m(1 - L_m)\right)$$

where $\tilde{\tau}_i$ represents the disutility that i gets from the husband m working. With $\tilde{\tau} \equiv \tilde{\tau}_m + \tilde{\tau}_f$, the optimal labor supply decisions are:

$$L_m^* = \mathbb{1}\left[w_m \geq (1 - \tilde{\tau})h_m\right]$$
$$L_f^* = \mathbb{1}\left[(1 - \tau)w_f \geq h_f\right]$$

1.C.6 Model extension: enriching women's labor supply decisions

The independence in the labor supply decisions of husbands and wives simplifies the study of these decisions. However, a necessary assumption to generate the independence is the perfect substitutability of home-produced goods and market goods. The perfect substitutability assumption, in turn, implies that there are no incentives for specialization. I can enrich the labor supply decisions of women, however, if I abstract from the labor supply decisions of men, i.e. have men always work in the market.

As in Aguiar and Hurst (2007a), consider the final consumption good being produced according to a CES production function with market goods and home production as inputs:

$$\left((y + wL)^{\rho} + \left(\underline{h} + h(1 - L)\right)^{\rho}\right)^{\frac{1}{\rho}} \tag{1.22}$$

y refers to nonlabor household market income, which includes any remittances from family, government benefits, and, if married, husband's earnings. $\underline{h}$ refers to the basic level of home production conducted in the household, including the amount of home production a woman would have completed regardless of her labor force participation status, and again, if married, any home production completed by the husband.

To make use of the convenient properties of the Fréchet-distributed market and home abilities, a simplification is necessary. A first-order Taylor approximation of equation (1.22) around $(w = 0, h = 0)$ yields

$$(y^{\rho} + \underline{h}^{\rho})^{\frac{1}{\rho}} + (y^{\rho} + \underline{h}^{\rho})^{\frac{1-\rho}{\rho}} y^{\rho-1} wL + (y^{\rho} + \underline{h}^{\rho})^{\frac{1-\rho}{\rho}} \underline{h}^{\rho-1} h(1 - L) \tag{1.23}$$

Hence, the labor supply decision that maximizes equation (1.23), corresponding to the optimal decision for a single woman, is:

$$\hat{L}^* = \mathbb{1}\left[y^{\rho-1} w \geq \underline{h}^{\rho-1} h\right]$$

A married woman, on the other hand, must consider the cost of the norms wedge. Where the cost of working in the market is denominated proportionally to the contribution of market goods in the production of the final good, her optimal labor supply decision is:

$$L^* = \mathbb{1}\left[(1-\tau)y^{\rho-1}w \geq \underline{h}^{\rho-1}h\right]$$

1.C.7 Derivation of marriage market equilibrium

The probability that man m chooses spousal type $r \in \{1, ..., S\}$ or stays single $(r = 0)$ is

$$\mathbb{P}(r = \arg\max_{r'=0,1,..,S} V_m^{qr'}) = \frac{\exp\{\mathbb{E}(v_m^{qr}) - \tau^{qr} + \psi^{qr}\}}{\sum_{r'=0}^{S} \exp\{\mathbb{E}(v_m^{qr'}) - \tau^{qr'} + \psi^{qr'}\}}$$

The maximum likelihood estimator of $\mathbb{P}(r = \arg\max_{r'=0,1,..,S} V_m^{qr'})$ is the fraction of type q men married to r, or $\frac{n^{qr}}{M^q}$.

Hence, in terms of the number of (q, r) marriages *demanded* by type q men,

$$\ln n^{qr,\mathcal{D}} = \ln n^{q0} + \mathbb{E}(v_m^{qr}) - \tau^{qr} + \psi^{qr} - \mathbb{E}(\hat{v}_m^{q})$$

Similarly, woman f of type r choosing her spousal type or remaining single gives the analogue for the number of (q, r) marriages *supplied* by type r women,

$$\ln n^{qr,\mathcal{S}} = \ln n^{0r} + \mathbb{E}(v_f^{qr}) + \tau^{qr} + \psi^{qr} - \mathbb{E}(\hat{v}_f^{r})$$

1.D Appendix – Parameter Inference

1.D.1 Deriving the likelihood function to estimate θ

The probability density function of the Fréchet distribution with shape parameter θ and scale parameter s is:

$$f(x; s, \theta) = \frac{\theta}{s} \exp\left\{-\left(\frac{x}{s}\right)^{-\theta}\right\}\left(\frac{x}{s}\right)^{-\theta-1}$$

The scale parameter in the Fréchet distribution of market abilities among market workers is $\left(\frac{1}{P}\right)^{\frac{1}{\theta}}$, where P is the fraction working in the market among a group defined by gender g, match (q, r), and family composition $\mathcal{K}$. Therefore, where x_n is the market ability of observation n and P_n denotes the fraction of workers in observation n's group, the maximum likelihood estimator for θ is:

$$\tilde{\theta}_{MLE} = \arg \max_{\theta \in (0,\infty)} \sum_{n=1}^{Nobs} \Big[\ln\theta - \ln P_n - x_n^{-\theta} P_n^{-1} - (\theta+1)\ln x_n\Big]$$

1.D.2 Extracting market abilities from market wages

Recall that an individual i of gender g in a (q, r) match with family composition $\mathcal{K}$ receives wage

$$w_{it} = \bar{w}_{gt}^{qr}(\mathcal{K})\epsilon_i^w, \quad g \in \{M, F\}$$

where $\bar{w}_{gt}^{qr}(\mathcal{K})$ is the market wage per unit of effective labor for each group. Therefore the x_n in practice equals ϵ_{it}^{w*} where the asterisk denotes that this it the market ability of those who choose to work in the market. Let us isolate ϵ_i^{w*} from the observed wages:

$$\text{logwage}_{it} = \ln \bar{w}_{gt}^{qr}(\mathcal{K}) + \ln \epsilon_{it}^{w*}$$

To this end, I regress log wages on (decade $\times$ sex $\times$ education pair $\times$ family composition) dummies. For each group, then, the residuals are

$$\begin{aligned} \text{residuals}_{it} &= \text{logwage}_{it} - \overline{\text{logwage}}_{it} \\ &= \left(\ln \bar{w}_{gt}^{qr}(\mathcal{K}) + \ln \epsilon_{it}^{w*}\right) - \left(\ln \bar{w}_{gt}^{qr}(\mathcal{K}) + \overline{\ln \epsilon_{it}^{w*}}\right) \\ &= \ln \epsilon_{it}^{w*} - \overline{\ln \epsilon_{it}^{w*}} \end{aligned}$$

Thus,

$$
\begin{aligned}
x_n &= \exp\{\text{residuals}_{it} + E(\ln \epsilon_{it}^{w*})\} \\
&= \exp\{\text{residuals}_{it} + \tfrac{\gamma}{\theta} + \ln(s_n)\} \\
&= \exp\{\text{residuals}_{it} + \tfrac{\gamma}{\theta} - \tfrac{1}{\theta}\ln P_n\}
\end{aligned}
$$

where γ is the Euler's constant.[46]

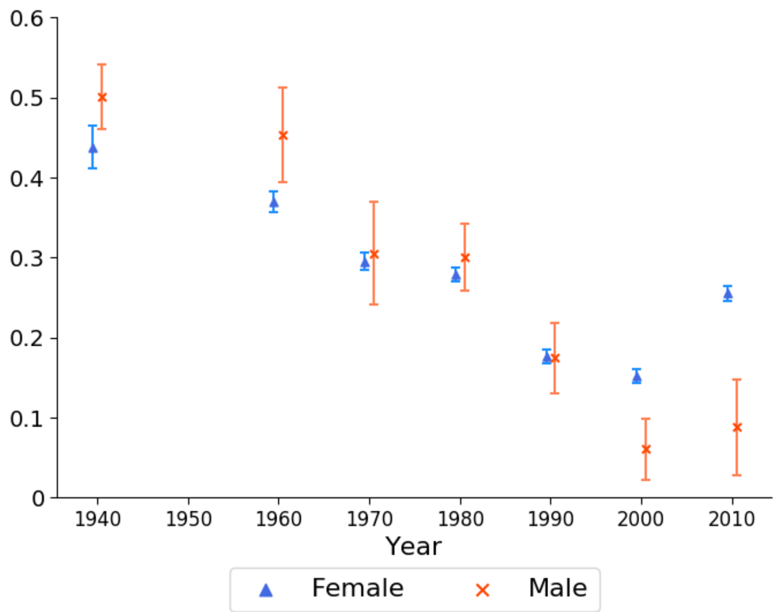
[46] To compute $E(\ln \epsilon_{it}^{w*})$, I need the probability density function of $y = g(x) = \ln(x)$ where x is a Fréchet random variable with scale parameter s and shape parameter θ. $f_Y(y) = f_X(g^{-1}(y))\left|\frac{dx}{dy}\right| = \frac{\theta}{s}\left(\frac{e^y}{s}\right)^{-\theta-1} e^{-\left(\frac{e^y}{s}\right)^{-\theta}} |e^y|$. Thus, $E(y) = \theta s^\theta \int_{-\infty}^{\infty} y e^{-\theta y} e^{-e^{-\theta y} s^\theta} dy = -\frac{1}{\theta}\int_0^\infty e^{-z}(\ln z - \theta \ln s) dz = -\frac{1}{\theta}\Gamma'(1) + \ln s = \frac{\gamma}{\theta} + \ln s$ where γ is the Euler's constant.

1.E Appendix – Norms Wedges under Alternative Models

1.E.1 Male norms wedges

What if married *men* are also subject to norms wedges? The male counterpart to the homemaker gender role would be the breadwinner gender role where married men are expected to earn a market income. This situation corresponds to the model extension in Appendix section 1.C.5. Because the optimal labor supply decisions of husbands and wives are independent, the values of the norms wedges applying to married women are not affected. Nonetheless, it would be interesting to compare the "female" norms wedges to the "male" norms wedges.

Figure 1.A8: Female and male norms wedges by decade



Notes: This figure plots the weighted median of τ and $\tilde{\tau}$ (described in Appendix section 1.C.5), inferred for each group, by decade. The weight equals the empirical probability of each group. The error bars indicate 95% confidence intervals based on bootstrapped standard errors with 50 replications.

Figure 1.A8 shows that the male norms wedge declines over time, resembling the trend of the female norms wedge. The decline indicates that the labor force participation of married men are falling faster than that of comparable single men. In fact, the male norms wedge is even higher than the female counterpart in 1940 and is lower towards the end of the sample period. This result may be counterintuitive; it is difficult to reconcile the consistent and large decline with the fact that the rise of the stay-at-home fathers is only a recent phenomenon.

However, the male norms wedges must be interpreted with a grain of salt. The first reason is an algebraic one. The male norms wedges are proportional to the ratio of *nonparticipation in the labor force* of married and single men.[47] Throughout the sample period, the male labor force participation rate is very close to 1 regardless of marital status, so

[47]The formula for the male norms wedge for a married man in a (q, r) match with family composition

that small changes in the labor force participation rate translate to large changes in norms wedges. For the same reason, the male norms wedge is also much noisier than the female counterpart, as depicted by the wider confidence intervals.[48] The second reason is that nonparticipation in the labor force is less synonymous with home production for men, so $\tilde{\tau}$ would have less to do with actual gender norms. Indeed, state-level male norms wedges are *not* correlated to state-level attitudes towards gender roles, unlike female norms wedges (Table 1.2).

1.E.2 Robustness of norms wedges

This section demonstrates the robustness of the values of the norms wedges to alternative model specifications and sample selection criteria.

Figure 1.A9 plots the norms wedges under four different specifications. "Baseline" refers to the norms wedges in the baseline model, and is a reproduction of the norms wedges in Figure 1.7. "With income taxes" is a modification of the baseline model where individuals compare home productivity to not *before*-tax market wages but *after*-tax market wages. Another way in which a married woman's labor supply decision might differ from a similar single woman's is the aspect of social insurance, where marriage may allow individuals to save on taxes paid. Thus, it is important to check for robustness to after-tax earnings. Because precise after-tax earnings are not recorded in the U.S. Decennial Census, I apply to the reported before-tax earnings, the U.S. federal individual nominal income tax rates from the Tax Foundation (2013). I assume that individuals file for taxes under the most profitable category that they are eligible for, among Married Filing Jointly, Married Filing Separately, Single, and Head of Household.

"With non-labor income" refers to the norms wedges when the women's labor decision is enriched to depend on non-labor income, i.e. the model extension in Appendix section 1.C.6. This specification allows a woman's labor supply decision to depend on the labor and nonlabor earnings of all other members of the household (in particular, her husband if married), as well as the household's social security income and welfare benefits. Therefore, it addresses the concern of interdependence between the labor supply decisions of a husband and a wife. The norms wedge values are computed from 1960, as non-labor income data are not available in 1940.

"Baseline, ages 40-54 only" refers to the norms wedges in the baseline model where the sample excludes 25- to 39-year-olds. This sample selection criteria addresses the concern that a married woman's labor force participation may differ from a similar single woman's because she has different child-bearing prospects. Child-bearing prospects matter for a woman's decision to work in the labor market. For example, with positive returns to experience, a woman would be more likely to work if she expects fewer interruptions during

---

$\mathcal{K}$ is

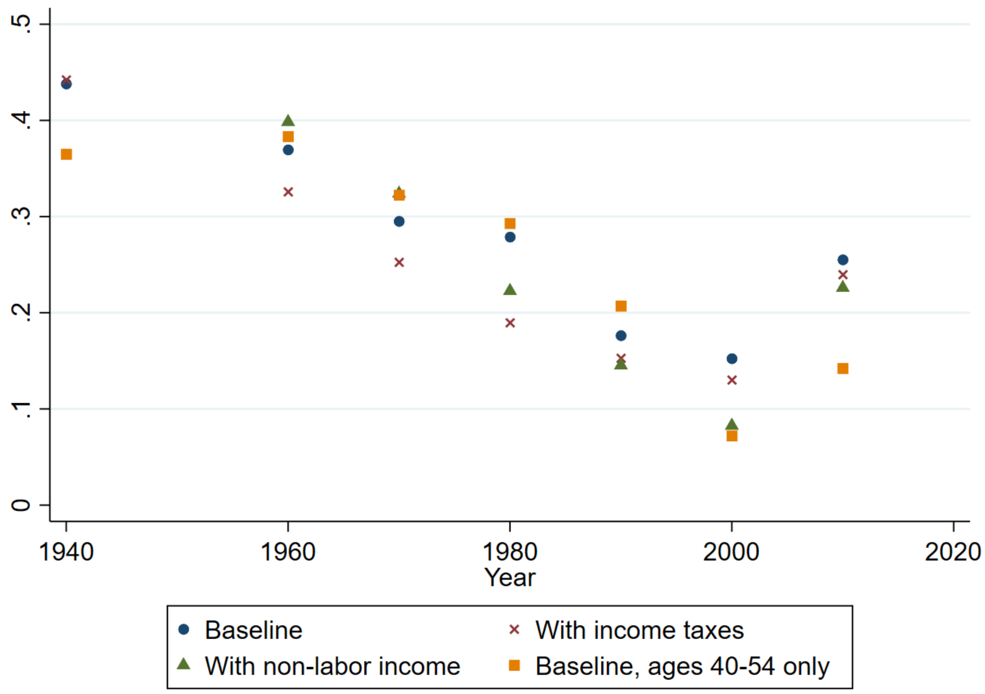
$$\tilde{\tau}^{qr}(\mathcal{K}) = 1 - \frac{\text{avgwage}_M^{qr}(\mathcal{K})}{\text{avgwage}_M^{q0}(\mathcal{K})} \left(\frac{1 - P_M^{qr}(\mathcal{K})}{1 - P_M^{q0}(\mathcal{K})} \right)^{\frac{1}{\theta}}$$

[48] The noisiness can be viewed as a consequence of the difficulty of estimation when the parameter is near the boundary (Andrews, 1999).

her career cycle. Child-bearing prospects also matter from the employer's perspective; an employer may be less willing to hire a woman who they view as likely to leave temporarily or permanently in the future. Hence, I restrict the sample to women who are less likely to conceive more children.

It is reassuring that for all four specifications, the values and trends of the norms wedges are similar.

Figure 1.A9: Norms wedges under alternative specifications or sample criteria



Notes: This figure plots the values of the norms wedges by decade, under alternative model specifications or sample criteria. "Baseline" refers to the norms wedges in the baseline model, and is a reproduction of the norms wedges in Figure 1.7. "With income taxes" is a modification of the baseline model where individuals compare home productivity to not *before*-tax market wages but *after*-tax market wages. The tax rates are based on the U.S. federal individual nominal income tax rates from the Tax Foundation (2013). "With non-labor income" refers to the norms wedges when the women's labor decision is enriched to depend on non-labor income, i.e. the model extension in Appendix section 1.C.6. "Baseline, ages 40-54 only" refers to the norms wedges in the baseline model where the sample excludes 25- to 39-year-olds.

Chapter 2

How do political parties respond to gender quotas? Evidence from South Korea

2.1 Introduction

Gender quotas in politics are used by half of the countries of the world, yet their effectiveness is still an object of debate. Although countries that have adopted quotas do have a higher share of females in national parliaments on average, quotas have not necessarily been accompanied by less conservative attitudes towards women.[1] In such environments, discrimination and hidden barriers against women might still persist, particularly so when the majority of the incumbents are male. Thus, while in some cases quotas may be successful in achieving the policy objective, in other cases they might be rendered ineffective, or even, counterproductive.

In this paper, we study the reaction of highly male-dominated political parties to the introduction of gender quotas in South Korean municipal council elections, over four election cycles. The South Korean setting provides a rare opportunity to study the reactions of affected politicians due to the fact that the gender quota is imposed on only one of the two separate election arms in which a councilor is elected. In South Korea's mixed electoral system, the first group of councilors is elected through a plurality vote in the municipality's constituent wards, and the second group is elected by party-list proportional representation. As the gender quota affected only the proportional representation arm ("PR arm"), we can study how parties strategically respond in the unconstrained ward-level plurality vote arm ("ward arm"). Therefore, the existence of the ward arm enables us to detect what is typically unobservable: the preference of political parties for female candidates, or lack thereof.

[1] Across the world, the countries with gender quotas in national parliaments have an average female share of 26%, relative to 20% for the countries without (World Bank Statistic). In terms of the fraction of respondents agreeing to the statement "Men make better political leaders than women do", asked in the World Values Survey, the countries with quotas rate at 47.5% and those without rate at 47.1%. See Figure 2.A1 in Appendix Section 2.A for the statistics by country.

For our identification strategy, we use the cross-sectional variation in the number of seats reserved for proportional representation as a measure of the intensity of exposure to the gender quota. The number of PR seats increases as a step function of a municipality's council size, creating discontinuities in the intensity of the quota's bite at certain cutoffs of council size. Using a regression discontinuity framework, we study the effect of the quota on how political parties select candidates in the ward election arm.

In the treated municipalities affected more intensely by the quota, parties initially put forth fewer female candidates in the unconstrained ward arm. Hence, fewer female ward councilors were elected, but these municipalities also experience a greater number of female PR councilors as a direct consequence of the quota. Putting the PR and ward arms together, these municipalities experience on net a greater number of female councilors initially. We track the treated and control municipalities over the next three election cycles, and find that the treatment group gradually increase the number of female ward candidates. By the last election cycle, the treated municipalities have completely reversed their initial reaction and have a *greater* number of female ward candidates than the control municipalities.

What is driving the gradual change in the response to the treatment, where treatment municipalities start off with fewer female ward candidates but end with more, than control municipalities? It must be that the treated and control municipalities developed on different paths after the treatment. Two mechanisms are consistent with the finding. The first has to do with dynamically changing attitudes towards women. The political parties might have gradually changed their attitudes towards female candidates from a position of initial aversion to greater favorability, as they learn about the competency of females through exposure to female councilors. The second mechanism, however, has to do with mechanical changes in the pool of female candidates available, with no change in parties' perception of women. Parties in treated municipalities might have simply experienced a shortage of qualified women initially, such that to fill the quota in the PR arm, they had to pull women out of the ward arm. Then over the next election cycles, the treated municipalities experience a faster growth in the female share among their pool of candidates with political experience, resulting in a gradual increase in female candidates.

Although it is difficult to fully disentangle the two mechanisms, our results suggest the former is more important. First of all, even if the latter mechanism was at play, the number of female councilors in the initially treated municipalities were still very small, and so they could not have accounted for all of the rise in the female candidates. Secondly, we see that preferences change also for female candidates who have zero councilor experience. Lastly, the shift in the selection towards female candidates occurs faster and stronger when the first elected female PR councilors are highly educated. Therefore, parties seem to gradually learn about the competency of female councilors, and particularly so upon exposure to more able women.

How, then, does the learning take place? To clarify how the treatment led to lasting effects on the strategies of individual parties, we conduct a supplementary analysis at the *party* level. We use marginal elections in the PR arm, to compare the strategies of parties that marginally won a female councilor in the previous election against the strategies of parties that marginally lost. We find that marginal winners put forth a higher female share among ward candidates in the following election cycle than marginal losers, when the marginal candidate is in the first position in the party list – and therefore must be a woman. However, this finding does not hold when the marginal candidate is in the second position, whether or not the candidate is male or female. Thus, what matters for a party's future selection of candidates is not the *share* of women among the party's councilors but rather whether the party experienced a female councilor or not. Moreover, we find that among marginally winning parties whose marginal candidate was in the first position, it is the parties that listed a man, rather than a woman, in the second position of the party list that increase the female share in the following election. Those parties are the ones with a greater preference for men, choosing to place a man where the gender is not restricted. Thus, it is the parties that had a prior preference towards men that update their beliefs on the competency of women, upon the experience of female councilors.

Gender quotas are often introduced before attitudes have changed sufficiently enough to accommodate the quotas. Hence, it is important to study the effect of the quotas in those contexts. With females holding only 3% of seats before the quota, South Korean politics is definitely one such context, but it is not alone. There are still many countries that currently have a very low female representation in politics, including Haiti, Iran, Kuwait, Lebanon, Mali, Nigeria, Qatar, and Yemen.[2] Not only that, gender quotas are increasingly getting introduced in other settings such as company boards, where the incumbents are similarly, if not more, male-dominated (Ahern and Dittmar, 2012).

Contributions to related literature A large literature exists on the consequences of political gender quotas, studied in numerous contexts. This paper's main contribution lies on the fact that the South Korean setting enables us to study the strategic responses of parties in unusually rich ways. The parallel voting system with the quota only applying to the proportional representation arm implies that there is a whole other arm that is unconstrained. Furthermore, the unconstrained arm of the election system is the way through which around 85% of councilors are elected, and therefore constitutes the more consequential arm. Such a structure of gender quota greatly expands the degree of freedom in which parties can respond, relative to quotas that reserve seats for women (Chattopadhyay and Duflo, 2004; Clayton, 2015), alternate between male and female candidate lists (Besley et al, 2017), or mandate a minimum share of women in candidate lists (De Paola, Scoppa, and Lombardo, 2010; Esteve-Volart and Bagues, 2012; Baltrunaite et al, 2014).[3]

Our second contribution is that we can study learning. Before the quotas, women

[2]See Figure 2.A2 in the Appendix for a visual demonstration of where in the world there is low female representation.

[3]Before describing the richness of the ways in which parties can respond to quotas, it must first be made clear that parties determine the set of candidates running for election. The case is obvious for the PR arm, because one cannot be elected without being a member of a party in that arm. However, also in the ward arm, it is the parties that nominate the candidates to run for election. It is possible that a candidate runs as an independent, but very rarely will he or she win.

held close to no seats in South Korean municipal councils. The introduction of quotas represented a radical shock to the status quo and, as a consequence, an opportunity for parties to learn about female councilors. Other types of quotas, such as those that mandate a minimum share of women in candidate lists, do not ensure that women end up elected. Therefore, these quotas are likely to be limited in increasing female representation to appreciable levels. For example, Bagues and Campa (2020) find that Spain's adoption of such a quota led to only a small increase in female councilors immediately afterwards and that no further gains were achieved in the medium run. Moreover, Dahlerup and Freidenvall (2013) review that among eight European countries with legislated quotas, France, Greece, Ireland, and Slovenia have no rules on the placement of females in winnable seats.

Roadmap The remainder of the paper is organized as follows. Section 2.2 provides a background on the institutional setting of South Korea's municipal council elections. We then describe the data in Section 3.2. Section 2.4 lays out our empirical strategy, and Section 2.5 discusses the results. In Section 2.6, we discuss the pieces of evidence that point towards learning as an explanation for the results in Section 2.5. Finally, Section 2.7 concludes.

2.2 Institutional setting

2.2.1 The role of municipal councils

There are 226 municipal councils in South Korea. Municipal councils represent the legislative branch that works with municipal governments, the executive branch, to oversee local matters. Councils have several legally defined responsibilities, which include reviewing and approving the spending of municipal governments, adopting and revising local bills, monitoring the municipal governments' administrative functions, and examining petitions submitted by residents. Municipal governments administer around a third of South Korea's total public expenditure (Ministry of the Interior and Safety, 2018).

2.2.2 Electoral rules and gender quotas

Municipal councils were established during the mid-1990s, and from then, elections have taken place every four years. Seven elections were held so far, with 2018 being the latest election year. Up to the third election in 2002, all councilors were directly elected through plurality vote in single-member constituent wards. It was extremely rare to find candidates affiliated with a political party.

However, major reforms were made to the electoral rules from the fourth election in 2006. They are summarized in Table 2.1. First, the parallel voting system was introduced, where at least 10% of the councilors needed to be elected through party-list proportional representation. Among a total of 7 to 35 seats in a council, the number of proportional seats increased as a step function of the total council size: 1 for councils with up to 10 seats, 2 for those with 11 to 20 seats, 3 for those with 21 to 30 seats, and so on.

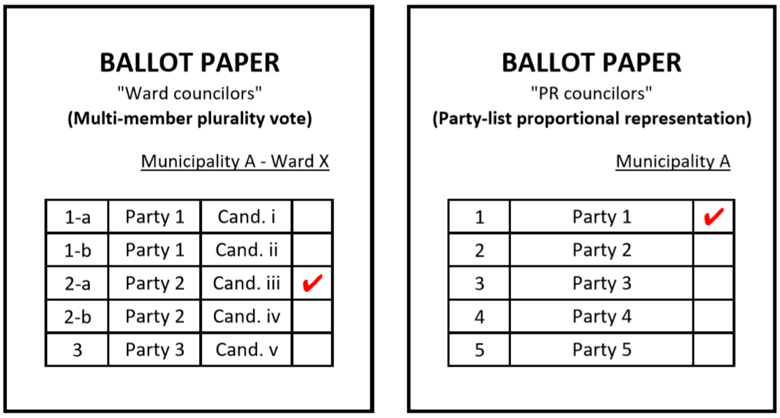
Table 2.1: **Amendments to legislation on municipal council elections**

First applicable election year	Amendment
2006	[PR] Proportional representation introduced [W] Single-member plurality vote → Multi-member plurality vote [PR] Odd-number candidates in party lists must be female (not enforced) [W] Subsidies to parties for nominating female candidates
2010	[PR] Odd-number candidates in party lists must be female (enforced) [W] At least one female candidate per general election district

Notes: Adapted from Lim (2018). [PR] indicates rules relating to propotional representation councilors and [W] to ward councilors.

Second, the remaining seats were reserved for plurality voting in multi-member constituent wards. Each constituency elected between 2 and 4 councilors, and therefore multiple candidates from the same party could run in the same constituency.[4] Figure 2.1 illustrates what the ballot papers look like for the two arms of the municipal council elections.

Figure 2.1: Ballot papers in municipal council elections



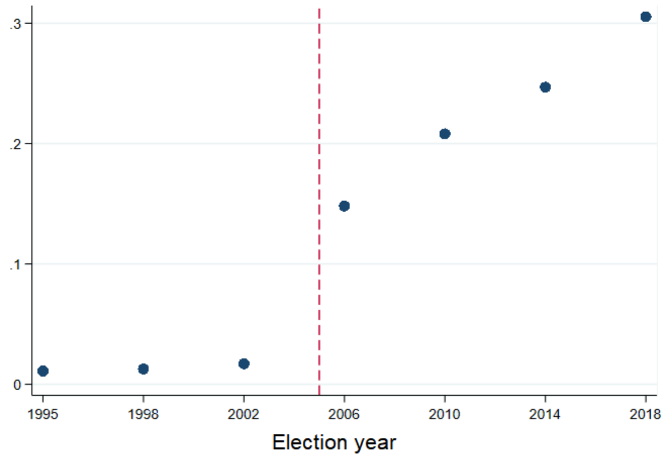
Notes: This figure illustrates the ballot papers that a voter residing in ward X of municipality A receives for the municipal council elections. The one on the left is used to vote for ward councilors and the one on the right is used to vote for PR councilors. The red ticks indicate how the voter might vote.

Third, a gender quota was put in place: all odd-number candidates in the proportional party list needed to be female. As the numbers of seats reserved for proportional representation are small, most elected councilors turned out to be the first candidates in the lists, and therefore female. As a consequence, the introduction of quotas sharply increased the proportion of female councilors. Municipal councils were severely male-dominated prior to the reform, with only 3% of councilors being female. Due to the introduction of quotas in

[4]The maximum number of candidates a party could nominate for a ward equalled the preset number of seats for that ward.

2005, female representation in municipal councils reached more than 30% in the last election in 2018. Figure 2.2 illustrates how the female ratio developed in municipal councils over time. The most striking feature is the sharp rise in the female ratio immediately after the reform.

Figure 2.2: The proportion of females in municipal councils, nationwide average



Notes: This figure illustrates the nationwide average of the gender ratio in municipal councils, for every election cycle since their emergence. The red dotted line indicates the year of the major reform that instituted the gender quota.

Last, subsidies were offered to parties based on the female ratio among the parties' candidates nationwide. However, it is unlikely that the subsidies affected much of the political parties strategies, particularly at the municipality level. The scale of the subsidies have been criticized for being too low to effectively expand female nomination (Jin, 2018; Kim et al, 2003; Lee, 2003). Indeed, they account for only around 5 to 6% of the total value of election subsidies (National Election Commission, 2018). Therefore, the presence of the subsidies are unlikely to have impacted political parties' selection of candidates.

Amendments to electoral rules continued between the 2006 and 2010 elections. It was stipulated that in either the municipal council elections or the higher-up provincial council elections, there must be at least one female candidate in each general election district. As there are around 250 general election districts, compared to 226 municipalities, a general election district approximately compares to a municipality.[5] Legislative Impact Analysis Reports indicate that most parties chose to satisfy this rule in the municipal council elections, due to the larger number of candidates (Lee, 2019). Selecting which ward to place the female candidate would have been a strategic concern for the political parties.

[5]General election districts are divided depending on population size and local representativeness. A large municipality may contain five general election districts, and up to five small municipalities may comprise a general election district.

2.2.3 Background behind the adoption of gender quotas

If some parties had led the move for the reform against opposition from other parties, then we should recognize that parties' strategic responses to the quota might be very heterogeneous in nature. Thus, here we discuss the background behind the adoption of the quota.

Before gender quotas were adopted in the municipal council elections, they were adopted first in the general election for the National Assembly in 2004. The adoption was influenced by increasing demands by women's organizations to raise female representation in politics, which at the time was dramatically behind the international average.[6] As females constitute half the voters, it was in the interest of political parties to put gender quotas forward amongst their election pledges. Moreover, there are views that the adoption of the quota was also a political tactic (Jeon, 2013). Political parties wanted to increase the size of the National Assembly back to what it was before the size cut during the Asian Financial Crisis, and the fact that the majority of the added seats will go to females, with the quota, made for a good excuse to expand the Assembly.

Once the quota was adopted in the general election, it became the natural next step to introduce it in the regional elections. The gender quota in the municipal council election was passed in the National Assembly, led by both major parties. Some argue that there was political motivation behind it, too (Kim, 2005). One new element in the reform was the party nomination system – a ward candidate must be nominated by their party in order to run with the party affiliation – but it was disputed as a ploy to deepen party influence. Political parties used the quota to justify the party nomination system, since the gender quota was embedded in the proportional representation arm where party nomination was essential.

To sum, it is difficult to say that there was a major division among political parties in their support of the gender quota when it was passed.

2.3 Data

Two sources of data are used. First, data related to the execution of the elections are collected by web scraping the website of the National Election Commission. The website posts detailed data on all past elections, including population, candidate information, and vote outcomes. Second, to examine the consequence of the municipal councils' legislative activities, we use the data on municipal governments' expenditures from the Local Finance Disclosure System of the Ministry of the Interior and Safety.

2.3.1 Population

Because ward divisions are centrally determined based on population size, population data is published. The number of residents is available by ward, voting eligibility, gender, and citizenship status. Moreover, the data includes the number of households by ward. This

[6] See Cho and Kim (2010) for a summary of the major activities of women's organizations.

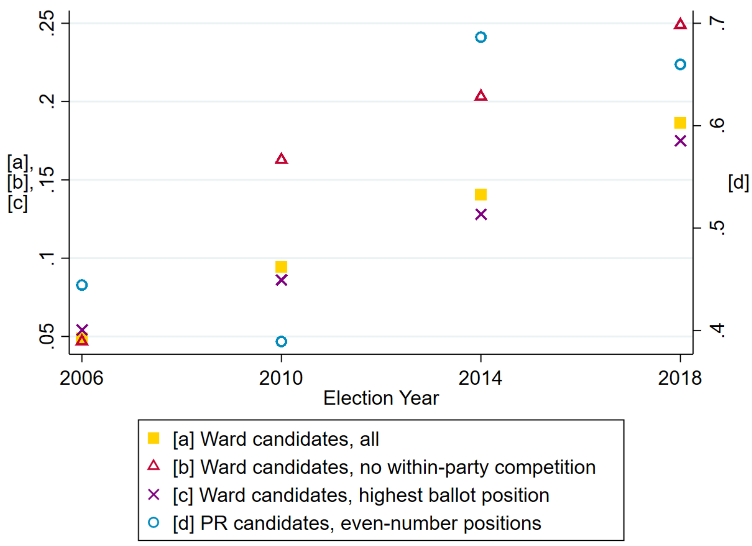
data is used to perform balancing checks in order to validate the identification strategy, which relies on the assumption that municipalities locally around the PR seat thresholds are similar.

2.3.2 Candidates

Various background characteristics of all candidates are also made publicly available by the National Election Commission. These are election arm (ward or PR) classification, election district name, candidate number, party affiliation, name, gender, date of birth, age, occupation, education, and pertinent work experience. Whether a candidate is favored by his or her party is revealed by the election arm and candidate number. Typically, candidates that are deemed less competitive are placed on the PR election arm, and the candidate numbers directly translate to the position on the ballot, in which higher positions attract more votes.

Figure 2.3 illustrates how the female share among candidates have been increasing continually, even when not stipulated by the quota. In particular, plot [b] shows that more females are running in wards as the sole candidates of their parties, and plot [c] shows that more females are taking the highest ballot positions even when multiple same-party candidates are running. Plot [d], on the other hand, shows that more females are taking the even-number party list slots, which would not happen with a strong preference for men.[7]

Figure 2.3: The share of females among non-quota candidates



Notes: This figure plots the share of females among [a] all ward candidates, [b] ward candidates with no within-party competition, [c] ward candidates that have within-party competition but is ranked the highest, and [d] PR candidates in even-number party list positions. The left-hand vertical axis corresponds to [a], [b], and [c], whereas the right-hand one corresponds to [d].

[7]In the latest general election of 2020, where the same gender quota on the PR arm applies, almost all PR candidates in positions 2, 4, and 6 are male.

2.3.3 Votes

The website of the National Election Commission also includes vote counts by ward. These vote counts enable us to see in which wards parties have their strongholds. Therefore, we can categorize wards into safe and contestable ones in the perspective of the political parties. Parties would then allocate their favored and less favored candidates to different wards accordingly.

Moreover, we can learn by which margin the winners won. In the regression discontinuity identification strategy, we rely on the assumption that close victories result in sharp changes in the composition of councilors by party, in an environment where parties enjoy similar degrees of popularity from the voters.

Electoral outcomes determine the gender ratio of the elected councilors. Table 2.2 provides descriptive statistics on the gender composition of councils by election cycle. The table also depicts how the reform in 2005 introduced the PR arm as well as the gender quota in that arm.

Table 2.2: Descriptive statistics on the gender composition of municipal councils

	Election cycle (year)						
	1 (1995)	2 (1998)	3 (2002)	4 (2006)	5 (2010)	6 (2014)	7 (2018)
Total number of councilors							
Min.	7	7	7	7	7	7	7
Mean	19.9	15.0	15.0	12.6	12.6	12.8	12.8
Max.	50	40	41	36	34	43	44
Number of PR councilors							
Min.	-	-	-	1	1	1	1
Mean	-	-	-	1.63	1.63	1.67	1.70
Max.	-	-	-	4	4	5	5
Gender ratio							
Min.	0	0	0	0	0.06	0.08	0.10
Mean	0.01	0.01	0.02	0.15	0.21	0.25	0.29
Max.	0.43	0.22	0.28	0.46	0.57	0.86	0.64
Gender ratio among PR councilors							
Min.	-	-	-	0*	0*	0.50	0.50
Mean	-	-	-	0.87	0.96	0.97	0.98
Max.	-	-	-	1	1	1	1
Minimum number of women required							
Min.	-	-	-	0	1	1	1
Mean	-	-	-	1.12	1.12	1.13	1.13
Max.	-	-	-	2	2	3	3

Notes: *gender quotas were introduced in 2005. However, during the election of 2006, they remained merely a strong recommendation, so it was still legal to place a male in slot 1 of party lists. Most municipalities complied, but 14 of them had no female PR councilors. In election year 2010, the minimum of the gender ratio among PR councilors is 0 because in one council the elected woman was invalidated for being a member of multiple parties.

2.3.4 Municipal budget

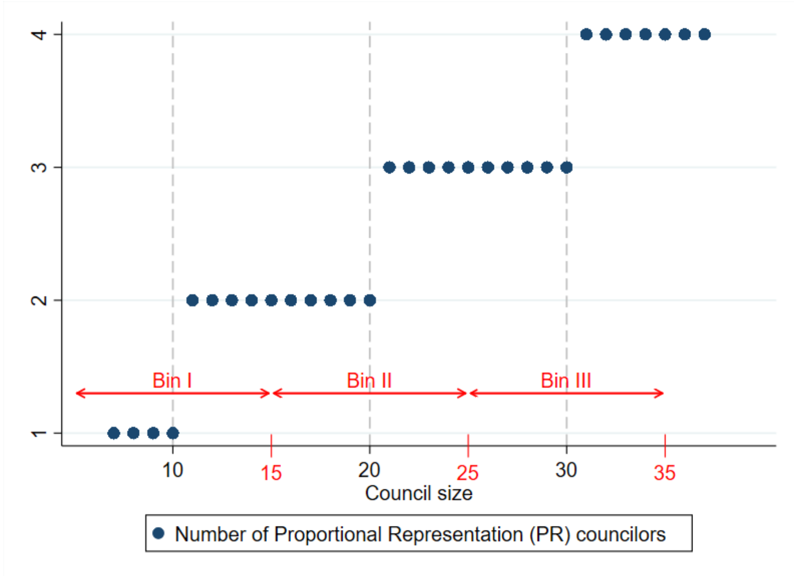
Municipal budget data is used to perform balancing checks, to show that municipalities locally around the PR seat thresholds are similar in terms of economic scale and council performance. The budget of a municipal government reflects the economic prosperity of the municipality, as around a half is sourced from local tax and non-tax revenue. In addition, data is available on the share of the municipality's expenditures spent on running the municipal council (2002-2020). There have been numerous accusations in the past of councilors appropriating large sums of the local budget for their private use (Local Decentralization Bureau, 2019). For instance, they would go on international policy-research trips where the itinerary largely consists of sightseeing. Another example is of councilors ordering member pins made of pure gold. As such, a measure of the performance of a council is the frugality of its operation costs. Newspapers have traditionally included it in their assessments of councils (Jang, 2008).

2.4 Empirical Strategy

2.4.1 Regression discontinuity design around the number of PR seats

To get at the causal effect of the gender quota, we make use of the fact that the gender quota affects municipalities at different intensities depending on the proportion of the PR seats in the council. The number of PR seats increases as a step function of municipal council size, which is pre-determined centrally by the National Election Commission based on population size and regional representativeness. The step function is depicted by the navy dots in Figure 2.4.

Figure 2.4: Councils by bins around each threshold



Notes: This figure depicts how the number of seats reserved for the proportional representation arm increases as a step function of the total number of councilors in a municipality. There are municipalities that do not correspond to the step function, because they are formed by the union of multiple municipalities *after* the election took place. The municipalities pre- and post-union are all excluded from the sample as outliers and are not shown in this figure. Moreover, the figure depicts how a council is categorized into a bin based on its most proximate threshold.

The regression discontinuity design compares the characteristics of ward and PR candidates in municipalities on each side of the step function's thresholds, while controlling for council size. In order to account for the fact that there is not just one but many thresholds, we categorize councils into bins based on the proximity to thresholds, as illustrated in Figure 2.4.

Therefore, this strategy estimates the effect of an additional PR councilor *in general*, rather than an additional *female* PR councilor. Nonetheless, while the gender quota does not necessitate that the second PR councilor be male, in practice almost all PR councilors end up being female.[8] This fact is due to PR candidates even in even-number positions frequently being female, and also due to PR councilors frequently being the number-1 candidates of multiple parties.[9] We also check in Section 2.4.3 that an additional PR councilor strongly implies an increase in the number of female PR councilors.

The regression discontinuity specification is given by:

$$Y_{cbt} = \sum_{s=4}^{7} \beta_s \times Treat_{cbt} + f(x_{cbt}) + \delta_b + \gamma_t + \epsilon_{cbt} \qquad (2.1)$$

where Y_{cbt} denotes the outcome variable for municipal council c belonging to bin b in election cycle t. The running variable is $x_{cbt} \equiv (\text{council size})_{cbt} - \text{threshold}_b$, with $\text{threshold}_b \in \{10, 20, 30\}$. In addition, $Treat_{cbt} \equiv \mathbb{1}(x_{cbt} \geq 0)$, signifying an additional PR councilor.

[8]Table 2.2 shows that among PR councilors, 87% to 98% are female each election.

[9]Appendix Table 2.A1 shows that it is relatively rare to find multiple PR seats getting allocated to the same party.

Therefore, β_s estimates the effect of having an additional PR councilor, pooling all the bins together, in election cycle s. Moreover, the baseline function form of f is linear, and we do not allow for the effect of x_{cbt} to differ to the left and right of the threshhold. The reason for this choice is that making f quadratic or allowing for differential trends on either side of the threshold barely makes a difference.

Another specification, based on treatment status at election cycle 4, is:

$$Y_{cbt} = \sum_{s=4}^{7} \beta_s \times (\textit{Treat at cycle } 4)_{cb} + f(x_{cb4}) + X_{cbt} + \delta_b + \gamma_t + \epsilon_{cbt} \tag{2.2}$$

where $(\textit{Treat at cycle } 4)_{cb} \equiv Treat_{cb4}$, and X_{cbt} denote control variables such as council size or the number of ward seats.

The outcomes we consider are the number of ward and PR candidates by gender. A factor to note is that when the outcome variable relates to the ward elections, we change the running variable to $\tilde{x}_{cbt} \equiv (\text{number of ward councilors})_{cbt} - (\text{number of ward councilors at the threshold})_b$, for ease of interpretation.[10]

2.4.2 Contemporaneous treatment vs. treatment at cycle 4

Equations (2.1) and (2.2) estimate the effects of contemporaneous treatment and initial treatment, respectively. In practice, there is barely any difference which specification we use, because the treatment status changes after election cycle 4 for only 3.7% of the councils. We settle on equation (2.2) as our main specification, though. The first reason is that the initial treatment assignment is more exogenous. Upon the first treatment, the treated and control municipalities may evolve on different paths, which would make them no longer balanced at the subsequent election cycles. Secondly, measuring the effect of the initial treatment maintains the same composition of treated municipalities. If the effect of contemporaneous treatment, specified by equation (2.1), varies over time, then it is unclear whether it is due to the small number of councils that are changing their treatment status, or due to the same councils reacting differently to the treatment over time. With equation (2.2), we can safely conclude that it is the latter.

2.4.3 Did the quota bite?

Because we are interested in the consequence of the change in the gender composition of councilors brought about by the quota, it is important to verify that there is a change in the number of *female* PR councilors at the discontinuity thresholds.

[10] If we keep the running variable based on council size, then the regression estimates the effect of $Treat = 1$, i.e. having one more PR councilor, while controlling for council size. Then in the regression, the councils with $Treat = 1$ effectively have one fewer ward councilor than those with $Treat = 0$. Therefore, it becomes more difficult to interpret the sign of the coefficient on $Treat$ when the outcome variable relates to ward elections, e.g. the number of female ward councilors or candidates. When the running variable is based on the number of ward councilors, however, we are free from this problem. Changing the running variable this way does not change much else. In fact, the coefficients $\hat{\psi}_0$ and $\hat{\psi}_1$ stay the same, as well as the R-squared value.

Table 2.3 reports the results of regressing (2.2) with the number of female PR councilors as the outcome variable, separately for each bin. While having an additional PR councilor at cycle 4 significantly increases the number of female PR councilors over all the cycles at bins 1 and 2, there is no such effect at bin 3. Moreover, there are very few observations at bin 3. The regression results of Table 2.3 are echoed by Figure 2.5, which shows that the average number of female PR councilors sharply increase at the thresholds of bins 1 and 2, but not at bin 3. Therefore, in the reduced-form results that follow, we restrict the sample to bins 1 and 2.

We next focus on the treatment effect over time. In both columns (1) and (2) of Table 2.3, the effect of the treatment at cycle 4 remains similar over the election cycles. Because the vast majority (96.3%) of the initially treated municipalities continue to get treated each cycle, the constancy in the coefficients implies that first-stage effect of the treatment - increasing the number of female PR councilors - is constant, too. This constancy implies that the effects on other outcome variables, i.e. the reduced-form treatment effects, should also be constant over time unless the initial treatment leads treatment and control groups on different paths.

The standard errors are clustered by municipality for two reasons. First, the variation of the initial treatment variable is at the level of the municipality. Second, parties formulate strategies chiefly within a municipality, rather than moving around candidates across municipalities. In fact, there are many factors that tie down a candidate to a certain municipality to be nominated in. A candidate is legally required to have been a resident of the municipality they are running in for at least 60 days prior to the election. In addition, as municipal councilors deal with local grass-roots matters, a candidate familiar with the municipality will win more votes *ceteris paribus*. Hence, a candidate usually runs in the municipality they have a connection with, such as their birthplace, long-term residence, or place of education. Moreover, the final say of a party's nomination lies on the head of the municipal branch of the party, so a candidate typically serves the local activities of the party in the municipality they desire to run in for a long time before getting nominated. Finally, once a candidate is nominated in a municipality, they put on a campaign and become known to the residents. So if they were to run again, they would not start over at a new location. For all these factors, rarely do parties move around candidates across municipalities for strategic reasons.

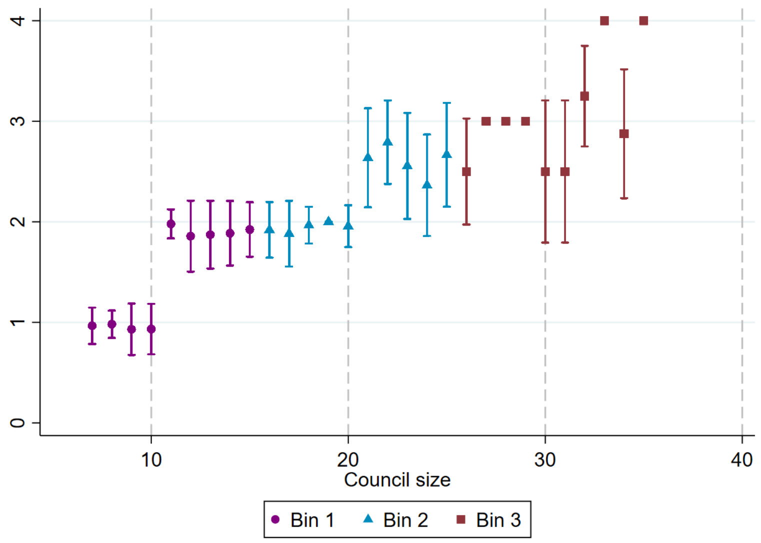
As a way to buttress the validity of the regression discontinuity design, Appendix Section 2.C.1 formally tests and confirms that as council size increases, there is a change in the number of female PR councilors only at the thresholds and at no other point.

Table 2.3: The effect of an additional PR seat at election cycle 4 on the number of female PR councilors

	Number of female PR councilors		
	Bin 1	Bin 2	Bin 3
	(1)	(2)	(3)
Treat at cycle 4× Cycle 4	0.84***	0.52***	-0.23
	(9.35)	(2.82)	(-0.42)
Treat at cycle 4 × Cycle 5	0.84***	0.32*	0.44
	(12.28)	(1.71)	(1.21)
Treat at cycle 4 × Cycle 6	0.77***	0.58***	0.04
	(10.18)	(3.33)	(0.10)
Treat at cycle 4 × Cycle 7	0.77***	0.57***	0.16
	(9.61)	(3.68)	(0.40)
Running variable form	council	council	council
N	670	198	33

Notes: t statistics from standard errors clustered by municipality in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
This table reports the results of regressing (2.2), separately for each bin, with the number of female PR councilors as the outcome variable.

Figure 2.5: The average number of female PR councilors by council size



Notes: The error bars indicate standard deviation of the number of female PR councilors by council size. Where the erroe bars are missing, there is only one municipality for that council size. Therefore, we can tell that there are only a small number of municipal councils belonging to bin 3.

2.4.4 Validity of the regression discontinuity design

Balance tests

A critical part of the identification strategy is that there are no confounders associated with the treatment status at election cycle 4. We regress equation (2.2) for various pre-determined characteristics, to check that they are balanced to the left and right of the threshold. The sample consists of councils at election cycle 4, and the regression results are presented in Table 2.4.

Panel (A) confirms that the population characteristics are balanced. In particular, the voting age population by gender is no different, alleviating the concern that the preference for female councilors among voters may be different between the treated and control municipalities. In Panel B, columns (8) and (9) refer to the vote share received by each main party in the previous election's PR arm. Columns (10) and (11) show that the initial treatment group is balanced in terms of economic prosperity and council performance. Columns (12) and (13) demonstrate that the structure of the ward election arm is balanced, as there is no difference in the number or size of wards between the treatment and control municipalities.

Table 2.4: Balance tests on pre-determined characteristics

***Panel A*: Population characteristics**

	Population		Voting age population			Households	
	Total	Foreign	Total	Male	Female	Total	Foreign
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat at cycle 4	-23.97	0.01	-17.22	-7.84	-9.38	-5.59	0.01
	(-0.78)	(0.78)	(-0.76)	(-0.70)	(-0.82)	(-0.50)	(0.83)
Running variable	31.49***	-0.00	23.22***	11.29***	11.94***	10.66***	-0.00
	(5.50)	(-0.84)	(5.47)	(5.38)	(5.55)	(5.15)	(-0.86)
Running variable form	council	council	council	council	council	council	council
N	219	219	219	219	219	219	219

***Panel B*: Political leaning, economic, and ward division characteristics**

	Past vote share by party		Budget		Ward characteristics	
	Conservative	Progressive	Total	Council expenses	Num of wards	Seats per ward
	(8)	(9)	(10)	(11)	(12)	(13)
Treat at cycle 4	-0.02	-0.00	54.16	0.02	-0.23	0.17
	(-0.26)	(-0.12)	(0.63)	(0.36)	(-1.30)	(1.48)
Running variable	-0.00	0.00	19.04	0.06***	0.45***	-0.06**
	(-0.08)	(0.08)	(1.30)	(7.84)	(11.71)	(-2.53)
Running variable form	council	council	council	council	ward	ward
N	219	219	219	219	219	219

Notes: t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The regression specification follows equation (2.2), and the sample consists of bins 1 and 2 at election cycle 4.

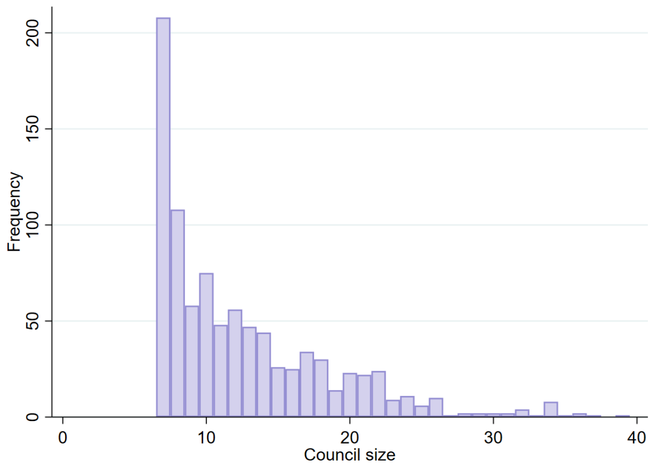
Bunching

Is there a possibility that there is gerrymandering? For example, a council may manipulate its constituent areas to manipulate the council's size and therefore, treatment status.

If there is manipulation, one evidence of it would be bunching at the threshold. Figure 2.6 displays the histogram of the frequency of municipalities by council size. Visually, it is hard to say there is bunching around the thresholds of 11 and 12. In addition, it is difficult to formally test for bunching around the threshold, e.g. the McCray (2008) density test, due to the coarseness in the council size variable.[11] However, there are specific electoral rules against gerrymandering.

The division of election constituencies is determined by the Municipal Council Election Committee. The committee is set up in each district, and it consists of up to 11 members appointed by the district mayor among the individuals nominated by the media, legal community, academic community, civic groups, the district council, and District Election Committee. Municipal councilor or party member cannot be in the committee. The committee determines the council size based on population, administrative districts, topography, transportation, and other conditions. The committee cannot split the smallest administrative district and make it a part of another ward. In sum, there are rules preventing the membership of interested individuals in the committee and also rules circumscribing how the election constituencies are drawn up.

Figure 2.6: Histogram of council size



Notes: The sample includes all municipal councils of election cycles 4, 5, 6, and 7.

2.5 Main Results

2.5.1 The numbers of candidates and councilors by gender

The results of regressing equation (2.2) are reported in Table 2.5. The most interesting result is captured by columns (1) and (2). In response to the treatment at cycle 4, parties initially put up more male ward candidates but gradually decrease the number of male ward candidates. Eventually, at election cycle 7, the parties in the treated municipal councils put up fewer male candidates than those in untreated councils. As for female ward candidates,

[11] No bunching is rejected for randomly selected cutoffs of council size.

the opposite pattern holds: the coefficient sign changes from negative (albeit statistically insignificant) to positive. Thus, the way parties select candidates in reaction to the gender quota is changing over time.

Focusing next on the columns for the councilors, we can see that similarly, the number of female ward councilors in the treated municipalities is lower in the beginning but is higher at the end.[12] Moreover, the higher number of female PR councilors in the treated municipalities at election cycle 4 more than compensates for the lower number of female ward councilors. Consequently, column (10) shows that there are statistically insignificantly more female councilors as a whole at election cycle 4 in the treated municipalities. Then, the coefficients for the later cycles grow in magnitude and become statistically significant.

Table 2.5: The effect of being past the threshold at election cycle 4 on the number of candidates and councilors

	All political parties									
	Candidates				Councilors					
	Ward		PR		Ward		PR		All	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treat at cycle 4 × Cycle 4	3.70***	-0.24	0.94***	0.96***	0.36*	-0.36*	0.09	0.76***	-0.29	0.29
	(3.19)	(-0.69)	(4.27)	(4.39)	(1.65)	(-1.65)	(1.16)	(9.01)	(-1.03)	(1.03)
Treat at cycle 4 × Cycle 5	0.56	0.49	0.63***	1.20***	-0.28	0.28	0.10	0.71***	-0.88***	0.88***
	(0.61)	(1.37)	(3.77)	(5.36)	(-1.24)	(1.24)	(1.52)	(9.23)	(-3.17)	(3.17)
Treat at cycle 4 × Cycle 6	-1.39*	0.91**	0.25*	1.02***	-0.47*	0.47*	0.08	0.70***	-1.06***	1.06***
	(-1.66)	(2.18)	(1.70)	(4.66)	(-1.76)	(1.76)	(1.31)	(8.89)	(-3.45)	(3.45)
Treat at cycle 4 × Cycle 7	-2.23**	1.10**	0.21	1.27***	-0.73**	0.73**	0.04	0.69***	-1.33***	1.33***
	(-2.23)	(2.49)	(1.35)	(5.83)	(-2.50)	(2.50)	(0.61)	(9.04)	(-3.99)	(3.99)
Running variable form	ward	ward	council	council	ward	ward	council	council	council	council
N	868	868	868	868	868	868	868	868	868	868

Notes: t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The regression specification is given by equation (2.2). The sample includes only bins 1 and 2.

What is driving the changing reaction to the gender quota? The parties in municipalities that got the initial treatment are changing their behaviors, becoming more female-friendly in their endorsement of candidates over time. A possible explanation is that that although parties countered the gender quota initially, the quota was not completely undone, as signified by the positive coefficient for cycle 4 in column (10). Then, the consequent experience of female councilors induced parties to become more favorable towards female councilors. Section 2.6 explores deeper into this learning story.

[12]Due to the addition of the running variable in the regression, mechanically the coefficients of columns (5) and (6), as well as those of columns (9) and (10), are of opposite signs. Also mechanically, the coefficients of columns (7) and (8) add up to 1.

2.5.2 Focusing on candidates likely to get elected

Because the analysis is at the municipality level, it is not straightforward to pin down where the effects are coming from. Many parties operate in a municipality, and each party puts forth a large number of ward candidates per ward. The changes in the composition of the electoral body may not mean much if the change in the candidate selection pattern is driven by parties or candidates in positions that have no hope in getting elected. To explain the source of the changing candidate selection with greater clarity, we point to Table 2.6.

Table 2.6 shows that even when we restrict our attention to candidates for whom election is probable, we see the same patterns of (a) the initial preference for males, and (b) the shift in the preference for females. In this table, the sample is restricted to main parties only: the Conservative party and Progressive party.[13] Columns (1) and (2) reproduce Columns (1) and (2) of Table 2.5 but for the main parties only. Columns (5) and (6) are even more selective; these are the candidates of the main parties, running as the candidates in the high-up positions on the ballot for the party in a ward (position 1 if the ward elects 1-2 councilors, and positions 1 and 2 if the ward elects 3-4 councilors). These candidates have a great chance of getting elected.

Because political parties can choose how many candidates to champion, it is difficult to interpret the results in columns (1) and (2) together.[14] Column (4) reports how the number of female ward candidates compares between treatment and control municipalities, when the total number of ward candidates is no different.[15] Hence, column (4) looks at how the *substitution* between the two genders occurs, when the total number of candidates is pre-set. Again, there are fewer females in cycle 4, and gradually more females afterwards. The pattern is mirrored when we restrict the candidates to those in useful positions.

[13]As it can be seen in Appendix Table 2.A2, the majority of candidates and elected councilors is affiliated with these two main parties – 54% and 83% on average, respectively – and their importance increases over time.

[14]The maximum number of ward candidates for a party is the total number of ward seats in the council, but there is no minimum.

[15]One may argue that the total number of ward candidates is a "bad control" in the regression, because it is an outcome of the treatment (Angrist and Pischke, 2009). However, as the total number of ward candidates equals the sum of the number of male and female ward candidates, the controlled regression in column (4) can be simply interpreted as a "summarization" of columns (1) and (2) together; if we were to regress the number of male ward candidates while controlling for the total number of candidates, the regression coefficients would equal exactly the negative of the coefficients in column (4).

Table 2.6: The effect of being past the threshold at election cycle 4 on the number of ward candidates who are likely to get elected

	Main political parties only							
	All positions				Useful positions			
	Male	Female	All	Female	Male	Female	All	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treat at cycle 4 × Cycle 4	1.56*	-0.25	1.31	-0.43*	1.51***	-0.36*	1.15**	-0.54***
	(1.92)	(-1.09)	(1.52)	(-1.95)	(2.85)	(-1.91)	(2.11)	(-2.92)
Treat at cycle 4 × Cycle 5	0.59	0.52**	1.10	0.37*	0.38	0.43**	0.81	0.31*
	(0.79)	(2.09)	(1.31)	(1.73)	(0.75)	(2.32)	(1.49)	(1.81)
Treat at cycle 4 × Cycle 6	0.90	0.83***	1.72**	0.60**	0.50	0.60**	1.09**	0.43*
	(1.22)	(2.82)	(2.11)	(2.25)	(1.01)	(2.44)	(2.00)	(1.94)
Treat at cycle 4 × Cycle 7	0.06	1.29***	1.35*	1.11***	0.37	0.73***	1.10**	0.56**
	(0.08)	(4.00)	(1.74)	(3.66)	(0.74)	(2.90)	(2.11)	(2.37)
Total ward candidates				0.13***				0.15***
				(8.37)				(7.88)
N	867	867	867	867	867	867	867	867

Notes: t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The regression specification is given by equation (2.2). The sample includes only bins 1 and 2 and is restricted to the two main parties. The number of observations is 867 instead of 868 since in one municipality main parties only have proportional candidates. Useful positions refer to candidates in the high-up positions on the ballot for the party in a ward (position 1 if the ward elects 1-2 councilors, and positions 1 and 2 if the ward elects 3-4 councilors).

Table 2.7 uses a different way of gauging the election probability of candidates. It shows how parties select candidates for different types of wards. We categorize wards based on whether the party had a stronghold in the previous election, in which case we call the ward "safe."[16] These are wards where we can assume the party candidates have a very high probability of being elected. Since whether a ward is safe is dependent on the party at hand, the regressions in Table 2.7 are at the ward×party level. We can see in column (5) and (7) that parties had a strong preference for placing male candidates in safe wards in cycle 4, especially in useful seats. The preference for men, however, disappears from the cycle 5 onward. Furthermore, we can see in columns (6) and (8) that, from cycle 5, parties in treatment municipalities started placing more women among candidates in unsafe wards. This is what is driving the overall increase the number of female candidates we observe at the municipality level. Therefore, although the number of female ward candidates increases faster over time in treatment than in control municipalities, the gains for women still remain bounded to the less-preferred wards with lower likelihood of election.

[16] A party is considered to have a stronghold in a ward if the party wins the greatest vote share in the PR arm in the ward, and it got over 10 percentage points more vote share than the next popular party.

Table 2.7: The effect of being past the threshold at election cycle 4 on the number of ward candidates in safe and unsafe wards

	Main political parties only, (ward×party)-level regressions							
	All positions						Useful positions	
	Male	Female	All	Female	Female Safe	Female Unsafe	Female Safe	Female Unsafe
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treat at cycle 4 × Cycle 4	0.06	0.00	0.06	-0.00	-0.01	0.00	-0.05**	-0.01
	(0.68)	(0.11)	(0.74)	(-0.13)	(-0.40)	(0.09)	(-2.02)	(-0.21)
Treat at cycle 4 × Cycle 5	-0.05	0.07**	0.01	0.06**	0.04	0.10***	0.01	0.10***
	(-0.62)	(2.34)	(0.18)	(2.26)	(1.12)	(2.60)	(0.33)	(3.15)
Treat at cycle 4 × Cycle 6	-0.03	0.06*	0.03	0.06*	-0.04	0.10***	-0.01	0.06**
	(-0.33)	(1.95)	(0.39)	(1.76)	(-0.74)	(2.64)	(-0.18)	(2.05)
Treat at cycle 4 × Cycle 7	-0.08	0.07**	-0.01	0.07**	0.04	0.09**	-0.01	0.05
	(-0.96)	(2.18)	(-0.15)	(2.16)	(0.70)	(2.55)	(-0.21)	(1.55)
Total ward candidates				0.11***	0.10***	0.13***	0.09***	0.10***
				(10.17)	(7.05)	(9.10)	(4.55)	(4.22)
N	6035	6035	6035	6035	2337	3698	2337	3698

Notes: t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The regression specification is given by equation (2.2). The sample includes only bins 1 and 2 and is restricted to the two main parties. The level of observation is party ward. A ward is considered as safe for a party if the party wins the greatest vote share in the PR arm in the ward, and it got over 10 percentage points more vote share than the next popular party. Unsafe wards are all the others.

2.5.3 Placebo test

In the last sections, we found that parties reacted to the introduction of quotas in the PR arm by reducing the number of women among ward candidates immediately after the reform. The substitution away from women was stronger in ballot positions and wards where party candidates had higher chances at getting elected. However, from cycle 5, parties in municipalities above the threshold started placing more women among ward candidates.

Before going into the potential mechanisms that can explain the effects we observe, we report the result of a placebo test where we check that the thresholds are meaningful only after and not before the reform to the election system. This test provides supportive evidence that we are estimating the effect of the introduction of the quotas, and the results are not driven by treated municipalities being different from control municipalities ex-ante.

We test whether the number of male and female ward candidates changed at the threshold before and after the reform. We want to make sure that the probability of getting an additional PR seat upon the reform is not correlated with other factors that affect the number of male and female candidates, prior to the reform. Table 2.8 shows that up to election cycle 3, the effect of being past the threshold is not statistically significantly distinguishable from zero. It is at election cycle 4 that the treatment induces an effect, as expected.

Table 2.8: The effect of being past the threshold on the number of male ward candidates

	Number of ward candidates			
	Male	Female	Male	Female
	(1)	(2)	(3)	(4)
Treat × Cycle 1	0.92	0.01	0.03	0.04
	(1.32)	(0.26)	(1.33)	(0.31)
Treat × Cycle 2	0.40	-0.05	-0.61	0.03
	(1.12)	(0.24)	(1.01)	(0.30)
Treat × Cycle 3	1.06	0.06	0.01	0.15
	(1.02)	(0.27)	(0.90)	(0.32)
Treat × Cycle 4	3.22***	0.72**	2.93**	0.59**
	(1.21)	(0.28)	(1.15)	(0.27)
Treat × Cycle 5			-0.71	1.29***
			(0.83)	(0.31)
Treat × Cycle 6			-2.29***	1.58***
			(0.79)	(0.42)
Treat × Cycle 7			-2.79***	1.58***
			(0.98)	(0.44)
Running variable form	ward	ward	ward	ward
N	899	899	1577	1577

Notes: t-statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The running variable and treatment status are defined contemporaneously. While municipality size and divisions remained almost unchanged from cycle 4 onward, they changed dramatically during the first three election cycles. Thus, it would be inaccurate to define treatment for the first three cycles using cycle 4 municipality characteristics. The regression specification is given by equation (2.1).

2.6 Learning about women's competence

There are two broad mechanisms that can explain the initial substitution towards men and the following gradual increase in women among the candidates of parties in treatment municipalities. In the first mechanism, parties dynamically change their attitudes towards women. The initial observed substitution towards men could be due to parties preferring men as candidates and politicians, while the gradual rise in the number of women could be explained by parties' attitudes towards female candidates changing as they learn about women's competence through the exposure to female councilors. However, the dynamic treatment effects can also occur with absolutely no change in parties' perception of women. In the second mechanism, the effects are driven by mechanical changes in the pool of female candidates available. There might have been a shortage of women among the pool of potential candidates when quotas were introduced. If very few women were willing or were qualified enough to become councilors upon the introduction of the quotas, parties that were obliged to have more women among PR candidates would have had fewer women left to place among ward candidates. The overall increase in women in the following cycles then could be explained by the pool of experienced women increasing more in treated municipalities with respect to control ones thanks to the quotas.

We argue in the following sections that the first mechanism on learning appears to be more important.

2.6.1 Alternative mechanism: no learning, but mere change in the size of the pool of qualified women

Even if there is absolutely no learning by parties, the patterns that we see in the number of female candidates and councilors could be consistent with the initial shortage and gradual growth of qualified or experienced women. We provide evidence that this mechanism cannot be the whole story.

First of all, for the treatment effect of fewer female ward candidates in cycle 4 to be explained by a shortage of women, we should find that parties in treatment municipalities have greater trouble finding female candidates. We test this hypothesis. A party is defined as unconstrained if the number of female candidates in its party list is greater to the number of women it needs to list as candidates due to quotas.[17] Table 2.9 shows that parties b below the threshold are not more unconstrained in the number of women they can list as candidates, particularly in cycle 4.

Table 2.9: Probability of being unconstrained in the number of female candidates

	Pr(unconstrained)		
	(1)	(2)	(3)
Treat at cycle 4 × Cycle 4	0.02	-0.09	-0.09
	(0.46)	(-1.19)	(-1.27)
Treat at cycle 5 × Cycle 4	0.02	0.09	0.12*
	(0.49)	(1.19)	(1.65)
Treat at cycle 6 × Cycle 4	0.04	0.02	0.04
	(0.89)	(0.33)	(0.69)
Treat at cycle 7 × Cycle 4	0.00	0.08	0.10
	(0.05)	(1.18)	(1.38)
Running variable from	council	council	council
Sample	all parties	main parties	main parties participating in ward arm
N	3795	1557	1520

Notes: t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The regression specification is given by equation (2.2). The sample includes only bins 1 and 2. A party is defined as unconstrained if the number of female candidates in the party's list is greater than the number of women the party is obliged to place in its list due to quotas. In column (1), all parties are included; column (2) includes only the two main parties; in column (3) the sample is restricted to only the main parties in municipalities where the parties have at least one ward candidate.

[17] For example, the number of women a party needs to include in the party list is 1 if the number of PR seats for the municipality is 1 or 2, and 2 if the number of PR seats is 3 or 4.

Secondly, if the only reason for the treatment effect of rising number of female ward candidates over cycles 5 through 7 was the greater availability of experienced female candidates, then there should be no shift in the gender preference among candidates who have zero councilor experience. However, Table 2.10 shows that the initial withdrawal of female candidates and the gradual reversal is present even for ward candidates who have never been elected before. Hence, the greater availability of experienced women in treated municipalities after the introduction of the quotas cannot be the only reason driving the changing reaction. In addition, even if every municipality had women in its parliament in cycle 4 thanks to the introduction of the quota, the number of female councilors was still very small and so it cannot account for all the rise in the female candidates over time.

Table 2.10: The effect of being past the threshold at election cycle 4 on the number of rookie ward candidates

	Rookie ward candidates	
	Male	Female
	(9)	(10)
Treat at cycle 4 × Cycle 4	0.82	-0.27*
	(1.33)	(-1.66)
Treat at cycle 4 × Cycle 5	-0.63	0.01
	(-1.05)	(0.05)
Treat at cycle 4 × Cycle 6	-0.29	0.15
	(-0.57)	(0.84)
Treat at cycle 4 × Cycle 7	-0.45	0.38*
	(-0.86)	(1.82)
N	868	868

Notes: t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The regression specification is given by equation (2.2). The sample includes only bins 1 and 2.

2.6.2 Is there learning about women's competence?

Then, are parties really learning about women's competence? If parties are learning about the competency of females, then the learning can be expected to take place at a faster rate if parties are exposed to more competent females. We next examine how the treatment effects are different by the education level of the first female councilors of cycle 4. Panels A and B of Table 2.11 reproduce Table 2.5 for councils in which the female PR councilors of cycle 4 have below-median and above-median education levels, respectively. It is clear to see that the shift towards female candidates is more apparent and stronger when the first women are more educated, pointing towards parties learning about the competency of females over time.

Table 2.11: **The effect of being past the threshold at election cycle 4 by education level of the first PR female councilors**

	Panel A: Below-median education level									
	Candidates				Councilors					
	Ward		PR		Ward		PR		All	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treat at cycle 4 × Cycle 4	2.46	-0.55	1.33***	0.67**	0.53	-0.53	0.21*	0.61***	0.06	-0.06
	(1.28)	(-0.94)	(4.36)	(2.06)	(1.33)	(-1.33)	(1.66)	(4.39)	(0.12)	(-0.12)
Treat at cycle 4 × Cycle 5	0.33	0.47	0.94***	1.22***	0.07	-0.07	0.11	0.67***	-0.46	0.46
	(0.22)	(0.75)	(3.49)	(3.34)	(0.16)	(-0.16)	(0.96)	(4.84)	(-0.84)	(0.84)
Treat at cycle 4 × Cycle 6	-1.04	0.63	0.54**	0.44	-0.17	0.17	0.18	0.56***	-0.59	0.59
	(-0.75)	(0.83)	(2.35)	(1.47)	(-0.36)	(0.36)	(1.51)	(3.77)	(-1.03)	(1.03)
Treat at cycle 4 × Cycle 7	-1.90	1.26	0.58*	1.08***	-0.57	0.57	0.09	0.61***	-1.05	1.05
	(-0.99)	(1.58)	(1.96)	(3.32)	(-0.94)	(0.94)	(0.80)	(4.33)	(-1.50)	(1.50)
N	387	387	387	387	387	387	387	387	387	387
	Panel B: Above-median education level									[()]35pt
Treat at cycle 4 × Cycle 4	4.03***	0.16	0.75**	1.12***	0.11	-0.11	0.01	0.88***	-0.71**	0.71**
	(2.67)	(0.35)	(2.55)	(3.80)	(0.41)	(-0.41)	(0.06)	(8.06)	(-2.08)	(2.08)
Treat at cycle 4 × Cycle 5	0.49	0.72	0.56***	1.17***	-0.64**	0.64**	0.13*	0.71***	-1.30***	1.30***
	(0.39)	(1.56)	(2.72)	(4.09)	(-2.26)	(2.26)	(1.69)	(7.83)	(-3.87)	(3.87)
Treat at cycle 4 × Cycle 6	-1.75	1.27**	0.22	1.38***	-0.77**	0.77**	0.06	0.77***	-1.49***	1.49***
	(-1.48)	(2.40)	(1.17)	(4.55)	(-2.24)	(2.24)	(0.78)	(8.66)	(-3.87)	(3.87)
Treat at cycle 4 × Cycle 7	-2.44*	1.07*	0.14	1.28***	-0.88***	0.88***	0.04	0.71***	-1.57***	1.57***
	(-1.93)	(1.97)	(0.79)	(4.38)	(-2.66)	(2.66)	(0.55)	(7.68)	(-4.20)	(4.20)
N	481	481	481	481	481	481	481	481	481	481

Notes: t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The regression specification is given by equation (2.2). The sample includes only bins 1 and 2.

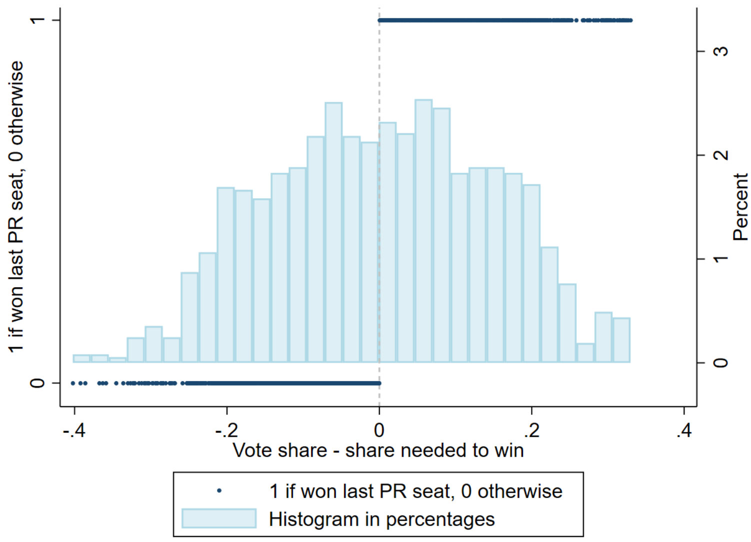
2.6.3 Which parties are learning and how?

The fact that the treatment status at cycle 4 has lasting effects through to cycle 7 shows that the treatment and control municipalities evolve on different paths from cycle 4, through some dynamic linkages across election cycles. A dynamic linkage of particular interest is whether a party learns about the competency of women as a result of the previous election outcome, and then changes its strategy in the current election.

We compare the strategies of parties that marginally won a PR councilor to those that marginally lost a PR councilor in the previous election. This comparison gives us the causal effect of having previously won a PR seat. In close electoral races in which the outcome of the election is uncertain, the winner is typically determined by factors that are beyond the control of parties and candidates, so which party wins the seat can be considered as random (Lee, 2008).

We take marginal parties to be the *two* parties that either marginally won or lost the *last* PR seat for the municipality. In order to differentiate marginal winners from losers, we measure how far off the vote share received by a party was, from the share it needed to win that seat. For party p in municipal council c at election cycle t, this value is given by $v_{cpt} \equiv (\text{vote share})_{cpt} - \bar{v}_{cpt}$, where $\bar{v}$ denotes the verdict-determining vote share. As the simplest example of $\bar{v}$, when two parties are competing for one PR seat, $\bar{v} = 0.5$ for both parties. The precise way we compute $\bar{v}$ for all possible contest scenarios is detailed in Appendix Section 2.D.1.

Figure 2.7: Marginal winners and losers of the last PR seat



Notes: This figure shows the distribution of the vote shares received by the two marginal parties competing for the last PR seat in a municipality. The vote share is computed to be the share of votes received among qualifying parties, i.e. parties that received more than 5% of the raw votes in the PR election arm.

Figure 2.7 shows the distribution of the vote shares received by the two marginal parties competing for the last PR seat in a municipality. The histogram shows that there are plenty of parties that received a vote share close to the share needed to win that seat. We employ a regression discontinuity design of the following form:

$$Y_{cpt} = \beta \times Winner_{cp,t-1} + f(v_{cp,t-1}) + X_{cpt} + \delta_n + \gamma_t + \epsilon_{cpt} \tag{2.3}$$

where $Winner_{cp,t-1} \equiv \mathbb{1}(v_{cp,t-1} \geq 0)$. We denote by $n \in \{1,2\}$, whether the marginal candidate that won – or nearly won – the last PR seat corresponds to the first or second PR candidate in a party's list.[18] $f(v_{cp,t-1})$ is linear and allows for different slopes to the left and right of the cutoff $v_{cp,t-1} = 0$. X_{cpt} represents the control variables, including the number of ward seats and the total council size for the *contemporaneous* election, i.e. election cycle t. A further factor to note is that the sample includes only the two major parties in South Korea, in order to track the parties over time. Due to frequent changes to party names, as well as frequent dissolutions and merges of small parties, parties other

[18]There are only five parties that won or nearly won a third PR councilor, so we exclude these parties. There is no party that won four PR councilors.

than the two major ones are difficult to follow over time. Therefore, X_{cpt} also includes a dummy that indicates which of the two major parties party p is.

When we consider marginal victories in the PR arm, we need to differentiate between the cases when the marginal candidate is in the first or second position in the party list.[19] If the marginal candidate is first on the list, then the candidate is necessarily a female, as enforced by the quota. However, if the marginal candidate is second on the list, the candidate might be male or female. Thus, when we consider close elections around a number-1 PR candidate, we are comparing parties that marginally won a *woman* in the previous election cycle with parties that didn't win any candidate. On the other hand, for close elections around a number-2 PR candidate, we are comparing parties that marginally won a second PR councilor to parties that only won one PR councilor.

We also distinguish between two different types of parties: those that placed men as the number-2 candidate in the party list and those that did not, in the previous election. As parties can place a candidate of any gender in even positions in their lists, the parties that place women in even positions are expected to have more gender-equal attitudes ex-ante.

Table 2.12 reports the result of regressing equation (2.3), when the marginal candidate is in position 1 or position 2 of the party list, and when the party lists a male or female second. We find positive coefficients on $Winner$ in columns (1)-(4) of Panel A. Marginal winners put forth a higher female share among ward candidates in the following election cycle than marginal losers, when the marginal candidate is in the first position and when the party has a male as the number-2 PR candidate. Therefore, parties that had a prior preference towards men that update their beliefs on the competency of women, upon experiencing a female councilor. In addition, the positive coefficients imply that a party's strategy is affected by having a female councilor get elected from within the party. For example, we cannot say from the results that the Conservative party do not learn from experiencing a Progressive female councilor, but we can say that the degree of learning is greater for the Progressive party.

In contrast, the null effect in columns (5)-(8) of Panel B – marginally winning a female number-2 candidate – implies that once a party wins a female councilor, an additional female councilor does not further impact the party's strategy in the following election. Moreover, the null effect in columns (5)-(8) of Panel A – marginally winning a female number-1 candidate for parties with a more gender-equal prior – implies that winning the first woman does not affect a party's future gender preference if the party had a more gender-equal attitude already.

[19] We do not consider the cases of the marginal candidate being third or lower on the list. It is extremely rare that a single party wins that many PR candidates, as can be seen in Appendix Table 2.A1.

Table 2.12: The effect of marginally winning a PR councilor in the previous election

***Panel A*: Position of marginal candidate: 1**

	Female share among party's ward candidates							
	Gender of second position: male				Gender of second position: female			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Winner	0.07*	0.08*	0.10*	0.21***	0.00	0.03	-0.02	-0.04
	(1.73)	(1.73)	(1.81)	(3.09)	(0.04)	(0.33)	(-0.20)	(-0.23)
Bandwidth ($\|v_{cpt}\|$)	0.20	0.15	0.10	0.05	0.20	0.15	0.10	0.05
N	297	231	157	84	104	70	49	24

***Panel B*: Position of marginal candidate: 2**

	Female share among party's ward candidates							
	Gender of second position: male				Gender of second position: female			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Winner	0.01	0.03	0.04	-0.05	-0.01	-0.01	-0.04	-0.05
	(0.26)	(0.68)	(0.92)	(-1.17)	(-0.31)	(-0.21)	(-0.84)	(-0.72)
Bandwidth ($\|v_{cpt}\|$)	0.20	0.15	0.10	0.05	0.20	0.15	0.10	0.05
N	113	84	69	33	98	75	62	33

Notes: t statistics from standard errors clustered by municipality×party in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The regression specification is given by equation (2.3). The Mean Square Error-optimal bandwidth is selected (Calonico, Cattaneo and Titiunik, 2014). The standard errors remain very similar when they are clustered at the municipality level.

2.7 Conclusion

This paper highlights that with time, affirmative action policies can still be effective despite an initial backlash, as long as the policies are not completely undone. Moreover, such is the case even in settings where the target group consists a very small minority among the incumbents. Through exposure to the minority group, the policies provide incumbents an opportunity to learn about the competency of the minority group. Once the learning takes off, the policy itself might be unneeded.

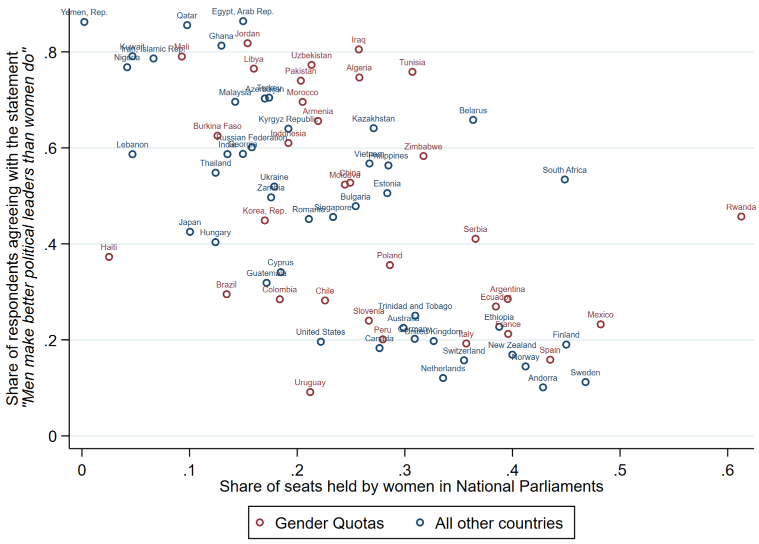
Although gender quotas in parliaments have been adopted broadly worldwide, there are still many countries that have none in place, such as Egypt, India, Liberia, Mauritius, Sao Tome and Principe, Sierra Leone, and Sri Lanka. Unsurprisingly, these countries also suffer from low levels of female representation in national parliaments. The South Korean setting of this paper is unique in that it studies the effect of a gender quota in the legislative body from a starting point of practically zero women. Therefore, this paper is informative about the effect of gender quotas where they are most needed.

What remains to be crystallized is exactly which aspect of women's competence parties are learning about. As of yet, we do not know whether it is their election campaign skills, their loyalty to the party, their keenness as legislators, or their ability to meet the demands of the electorate, that the parties update their beliefs on. Further evidence is needed in this direction.

This paper is a part of a bigger agenda that attempts to study how a gender quota might trigger a gradual process of learning in favor of women. To tackle the precise mechanisms through which the learning takes place, we plan to study in future work the specific interactions among councilors recorded in the transcripts of council meetings.

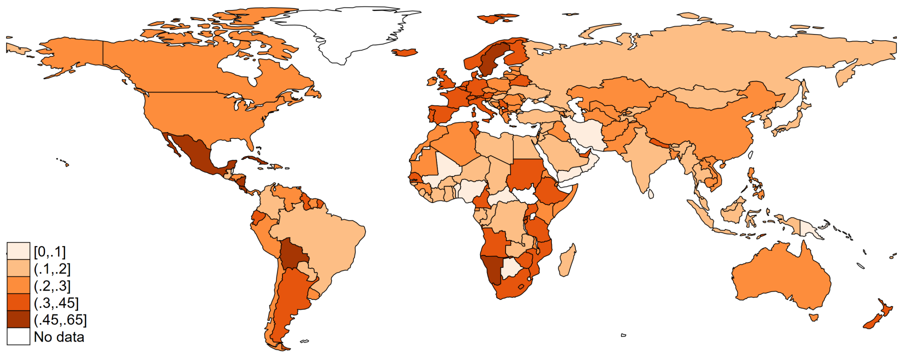
2.A Appendix – Figures

Figure 2.A1: Female share in national parliaments and attitudes towards women



Notes: Sources: Attitudes towards women: World Values Survey waves 5 (2005-2009) and 6 (2010-2014); Share of seats held by women in national parliaments: World Bank Gender Statistics, average of years 2018, 2019, 2020; Gender quotas in national parliaments: International IDEA Institute for Democratic and Electoral Assistance Gender Quotas Database, 2020

Figure 2.A2: Proportion of seats held by women in national parliaments (%)



Notes: Source: World Bank Gender Statistics, average of years 2018, 2019, 2020

2.B Appendix – Tables

Table 2.A1: The allocation of Proportional Representation seats across parties

	Councils by the number of PR seats					
	1 PR seat		**2 PR seats**		**3 PR seats**	
	N.	Percent.	N.	Percent.	N.	Percent.
Election Cycle 4						
1 Party	117	100%	15	17.86%	0	0
2 Parties	0	0	69	82.14%	15	83.33%
3 Parties	0	0	0	0	3	16.67%
Election Cycle 5						
1 Party	117	100%	5	6.02%	0	0
2 Parties	0	0	78	93.98%	13	72.22%
3 Parties	0	0	0	0	7	27.78%
Election Cycle 6						
1 Party	110	100%	18	20.22%	0	0
2 Parties	0	0	71	79.78%	17	100%
3 Parties	0	0	0	0	0	0%
Election Cycle 7						
1 Party	105	100%	9	9.89%	1	4.35%
2 Parties	0	0	82	90.11%	18	94.74%
3 Parties	0	0	0	0	1	5.26%
Total	**449**		**347**		**72**	

Notes: The sample is restricted to bins 1 and 2, i.e. to municipal councils with up to 25 councilors.

Table 2.A2: Candidates and councilors' party affiliation

		Candidates				Councilors			
		Direct		Proportional		Direct		Proportional	
	N	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Election Cycle 1									
Independent	226	1	0			1	0		
Election Cycle 2									
Independent	228	1	0			1	0		
Election Cycle 3									
Independent	228	1	0			1	0		
Election Cycle 4									
Independent	230	0.41	0.15	0	0	0.11	0.14	0	0
Progressive party	230	0.16	0.08	0.29	0.17	0.20	0.18	0.18	0.25
Conservative party	230	0.25	0.13	0.47	0.30	0.54	0.31	0.64	0.38
Election Cycle 5									
Independent	228	0.32	017	0	0	0.14	0.16	0	0
Progressive party	228	0.21	0.16	0.31	0.29	0.33	0.27	0.41	0.39
Conservative party	228	0.33	0.18	0.49	0.31	0.43	0.25	0.44	0.40
Election Cycle 6									
Independent	227	0.34	0.16	0	0	0.13	0.15	0	0
Progressive party	227	0.24	0.16	0.38	0.27	0.37	0.25	0.40	0.37
Conservative party	227	0.36	0.18	0.55	0.30	0.49	0.26	0.59	0.38
Election Cycle 7									
Independent	226	0.20	0.16	0	0	0.09	0.14	0	0
Progressive party	226	0.32	0.12	0.43	0.20	0.54	0.21	0.66	0.33
Conservative party	226	0.29	0.16	0.38	0.25	0.35	0.20	0.33	0.33

2.C Appendix – Identification

2.C.1 Confirming that the number of female PR councilors changes only at the thresholds

In order to buttress the regression discontinuity design, we test whether there is a change in the number of female PR councilors as council size increases, at points *other* than the thresholds. We regress, for each value of $x \in \{-4, -3, ..., 3, 4\}$ i.e. distance from the threshold,

$$\text{(number of female PR councilors)}_{cbt} = \beta \times TreatOne_{cbt} + \delta_b + \gamma_t + \epsilon_{cbt} \tag{2.4}$$

$$\text{where } TreatOne_{cbt} = \begin{cases} 1, & \text{if (council size)}_{cbt} = x \\ 0, & \text{if (council size)}_{cbt} = x - 1 \end{cases}$$

Equation (2.4), therefore, estimates the change in the number of female PR councilors when the council size increases by 1, for all points around the threshold. Table 2.A3 reports the results. It confirms that there is a positive effect only at the threshold.

Table 2.A3: The effect of an increase in council size on the number of female PR councilors

	x value								
	-4	-3	-2	-1	0	1	2	3	4
Coefficient ($\hat{\beta}$)	-0.03	0.03	-0.03	-0.01	0.92***	-0.01	-0.03	-0.03	0.09
Standard error	(-0.36)	(1.32)	(-0.99)	(-0.20)	(15.00)	(-0.14)	(-0.36)	(-0.35)	(1.04)
N	267	380	210	170	168	150	136	111	87

Notes: t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table reports the results of regression equation (2.4). The sample includes only bins 1 and 2.

2.C.2 Robustness to bandwidth choice

Table 2.A4: The effect of being past the threshold on the number of candidates and councilors, for various bandwidths

	Candidates				Councilors					
	Ward		PR		Ward		PR		All	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
***Panel A*: distance[†] ≤ 4**										
Treat	0.37	0.31	0.58***	1.23***	-0.15	0.15	0.09**	0.91***	-0.06	1.06***
	(0.46)	(0.94)	(4.64)	(8.29)	(-0.74)	(0.74)	(1.97)	(19.58)	(-0.26)	(4.78)
N	868	868	868	868	868	868	868	868	868	868
***Panel B*: distance ≤ 3**										
Treat	0.14	0.33	0.51***	1.24***	-0.13	0.13	0.08	0.92***	-0.04	1.04***
	(0.16)	(0.99)	(3.77)	(7.95)	(-0.59)	(0.59)	(1.57)	(17.77)	(-0.19)	(4.48)
N	811	811	811	811	811	811	811	811	811	811
***Panel C*: distance ≤ 2**										
Treat	0.89	0.37	0.54***	1.29***	-0.13	0.13	0.08	0.92***	-0.06	1.06***
	(0.99)	(1.07)	(3.59)	(7.50)	(-0.59)	(0.59)	(1.33)	(16.07)	(-0.23)	(4.16)
N	514	514	514	514	514	514	514	514	514	514
***Panel D*: distance ≤ 1**										
Treat	0.34	0.44	0.61***	1.25***	-0.24	0.24	0.09	0.91***	-0.15	1.15***
	(0.34)	(1.17)	(3.55)	(6.66)	(-0.94)	(0.94)	(1.36)	(14.03)	(-0.53)	(4.02)
N	320	320	320	320	320	320	320	320	320	320
***Panel E*: distance $= 0$**										
Treat	0.38	0.42	0.58***	1.23***	-0.24	0.24	0.08	0.92***	-0.16	1.16***
	(0.37)	(1.10)	(3.26)	(6.53)	(-0.91)	(0.91)	(1.25)	(15.00)	(-0.55)	(3.99)
N	168	168	168	168	168	168	168	168	168	168

Notes: t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
This table reports the results of regression equation (2.1). The sample includes only bins 1 and 2.
[†] Distance refers to the distance to the threshold. To illustrate, the council sizes for which distance equals 0 are 10, 11, 20, and 21, while the council sizes for which distance equals 1 are 9, 12, 19, and 22.

2.D Appendix – Computational details

2.D.1 Computing the running variable in the regression discontinuity design of section 2.6.3

The purpose of the regression discontinuity design of section 2.6.3 is to compare the strategies of parties that marginally won a PR councilor to those that marginally lost a PR councilor in the previous election. Thus, we are interested in the causal effect of having won a female PR councilor. We take marginal parties to be the *two* parties that either marginally won or lost the *last* PR seat. In order to differentiate marginal winners from losers, we measure how far off the vote share received by a party was, from the share it needed to win that seat. The running variable for party p in municipal council c at election cycle t equals $v_{cpt} = voteshare_{cp,t-1} - \bar{v}_{cp,t-1}$, where $\bar{v}$ denotes the verdict-determining vote share.

To compute $\bar{v}$, we first need to describe the rules by which PR seats get allocated:

Rules for allocating PR seats

1. Among parties running for prop rep in a locality, only the parties getting >=5% of votes qualify.
2. Of the qualifying parties, first compute $X =$(number of prop MP seats in the locality)$\times$(vote share of each qualifying party).
3. Allocate to each qualifying party the number of seats equal to the integer part of X.
4. Allocate the remaining seats by the ranking of the decimal part of X.

E.g. Municipal council A has 3 PR seats. There are 3 parties (1, 2, and 3) running for proportional representation. The vote shares of the parties are: party 1: 60%, 2: 38%, and 3: 2%. Party 3 got less than 5%, so it does not qualify. Among the qualifying parties, the vote shares are then party 1: $60/(60+38) \approx$ 61.22%, and 2: $38/(60+38) \approx 38.77\%$. The values of X's are party 1: $3 \times 0.6122 \approx 1.83$, and 2: $3 \times 0.3877 \approx 1.16$. Parties 1 and 2 both have 1 in the integer part of X, so they first get one PR councilor each. The last PR seat goes to party 1, because $0.83 > 0.16$.

Below, we compute $\bar{v}$ for all possible contest scenarios.[20] While doing so, we distinguish whether the marginal candidate that won – or nearly won – the last PR seat corresponds to the first, second, or third PR candidate in a party's list. For notational convenience, we call s the position in the party list of the marginal candidate of a party, and V the sum of the vote shares (among qualifying parties) received by the two marginal parties.

$\boxed{1}$ When there is one PR seat in the municipality

i) The two most popular parties contest over the only PR seat. Marginal parties: ranks 1 and 2

- Rank 1: $\bar{v} = \frac{V}{2}$, $s = 1$
- Rank 2: $\bar{v} = \frac{V}{2}$, $s = 1$

[20] An example is when there are three PR seats in a municipality, and the rank-1 and rank-2 parties contest over the last seat. Let v_n denote the vote share (among qualifying parties) received by the rank-n party. Rank 1 wins if $3v_1 - 2 > 3v_2 \iff v_1 > v_2 + \frac{2}{3}$. Therefore, $\bar{v}$ for the rank-1 party equals $v_2 + \frac{2}{3}$. On the other hand, $\bar{v}$ for the rank-2 party equals $v_1 - \frac{2}{3}$.

$\boxed{2}$ When there are two PR seats in the municipality

i) The contest is over whether the rank-2 party wins the second PR seat. Marginal parties: ranks 1 and 2

- Rank 1: $\bar{v} = \frac{2V+1}{4}$, $s = 2$
- Rank 2: $\bar{v} = \frac{2V-1}{4}$, $s = 1$

$\boxed{3}$ When there are three PR seats in the municipality

i) The contest is over whether the third PR seat goes to the rank-1 party or the rank-2 party. Marginal parties: ranks 1 and 2

- Rank 1: $\bar{v} = \frac{3V+2}{6}$, $s = 3$
- Rank 2: $\bar{v} = \frac{3V-2}{6}$, $s = 1$

ii) Where the rank-2 party wins a seat for sure, the contest is over whether the third PR seat goes to the rank-1 party or the rank-3 party. Marginal parties: ranks 1 and 3

- Rank 1: $\bar{v} = \frac{3V+1}{6}$, $s = 2$
- Rank 3: $\bar{v} = \frac{3V-1}{6}$, $s = 1$

$\boxed{4}$ When there are four PR seats in the municipality

i) The contest is over whether the fourth PR seat goes to the rank-1 party or the rank-2 party. Marginal parties: ranks 1 and 2

- Rank 1: $\bar{v} = \frac{4V+3}{8}$, $s = 4$
- Rank 2: $\bar{v} = \frac{4V-3}{8}$, $s = 1$

ii) Where the rank-1 party wins two seats for sure and the rank-2 party wins a seat for sure, the contest is over whether the fourth PR seat goes to the rank-1 party or the rank-2 party. Marginal parties: ranks 1 and 2

- Rank 1: $\bar{v} = \frac{4V+1}{8}$, $s = 3$
- Rank 3: $\bar{v} = \frac{4V-1}{8}$, $s = 2$

iii) Where the rank-1 party wins two seats for sure and the rank-2 party wins a seat for sure, the contest is over whether the fourth PR seat goes to the rank-2 party or the rank-3 party. Marginal parties: ranks 2 and 3

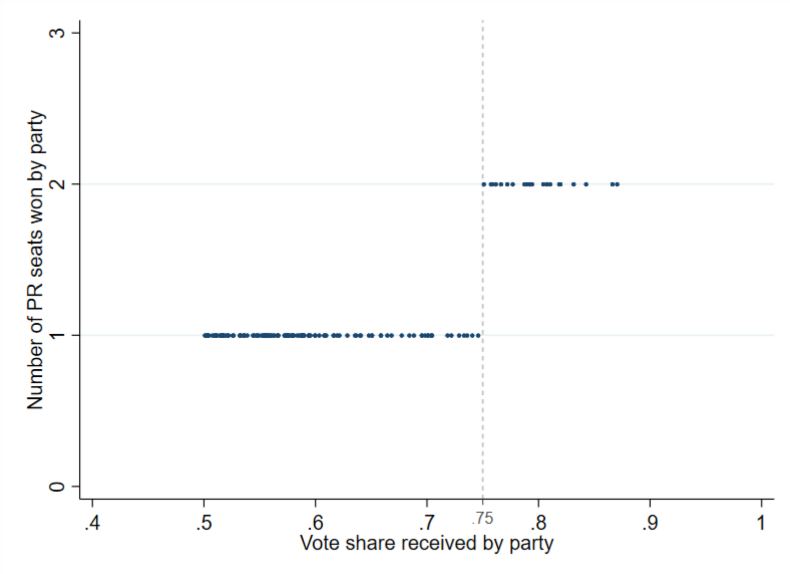
- Rank 2: $\bar{v} = \frac{4V+1}{8}$, $s = 2$
- Rank 3: $\bar{v} = \frac{4V-1}{8}$, $s = 1$

iv) Where the rank-2 and rank-3 parties win a seat each for sure, the contest is over whether the fourth PR seat goes to the rank-1 party or the rank-4 party. Marginal parties: ranks 1 and 4

- Rank 1: $\bar{v} = \frac{4V+1}{8}$, $s = 2$
- Rank 4: $\bar{v} = \frac{4V-1}{8}$, $s = 1$

As an example, take the case of the rank-1 party in a municipality with two PR seats and two qualifying parties. The party's $\bar{v} = 0.75$, according to the computation given above. Indeed, Figure 2.A3 shows that among such rank-1 parties, those receiving a vote share greater than 0.75 win two PR councilors whereas those receiving a vote share below 0.75 win one PR councilor.

Figure 2.A3: Marginal winners and losers of the last PR seat, among rank-1 parties in municipalities with two PR seats and two qualifying parties



Notes: This figure shows that in municipalities with two PR seats and two qualifying parties, the rank-1 parties must receive a vote share greater or equal to 0.75 in order to win both PR seats. The reason the vote share received is always greater than 0.5 is because these parties are the rank-1 parties. Note that the vote share is the share of votes among qualifying parties only.

Chapter 3

Homemaker Norms for Married Women: A Cross-Country Comparison

3.1 Introduction

A low level of female labor force participation implies that a large portion of human talent rests unrealized. Countries around the world are exposed to this issue at widely varying degrees, with female labor force participation ranging from 6% in Yemen to 84% in Rwanda in 2019 (World Bank Statistics). A large literature points to varying gender norms on female employment as a driver of this variation (Alesina, Giuliano, and Nunn, 2013; Jayachandran, 2015, Almond and Edlund, 2008; Abrevaya, 2009; Fernández and Fogli, 2009). Meanwhile, the low rate of female labor force participation is further exacerbated when we consider married women.[1] On top of the differential norm governing the appropriate behavior for men versus women, Lee (2020) suggests that *married* women face a further constraint against participating in the labor market relative to single women.[2]

This paper seeks to highlight the role of marriage as a factor driving international differences in female labor force participation. Adopting Lee (2020)'s computation of the "homemaker norms wedge" applying to married women, I verify that the homemaker norms wedge varies across countries.[3] The norms wedges measure how much a 10-dollar market wage is valued at for married women, when they are deciding between working in the labor market versus working at home. The wedges are computed from the gap in the labor force participation of married women and that of similar single women that is not explained by the wage differentials.

Many theories aiming to explain why the labor force participation of females varies across countries look to differences in economic conditions. First, where men have com-

[1] Among the country censuses of the 2010s compiled in the IPUMS International database, the ratio of married to single women's labor force participation was as low as 0.38 in Iran in 2011. The average ratio is 0.85.

[2] By Lee (2020), I refer to the Chapter 1 of this thesis.

[3] From here on, the homemaker norms wedge is sometimes referred to as plainly, the "norms wedge."

parative advantage in the more physically-intensive agriculture and manufacturing sectors and women in the more mentally-intensive service sector, the sectoral composition in the economy affects relative female labor productivity (Goldin, 1995; Galor and Weil, 1996; Pitt, Rosenzweig, and Hassan, 2012; Ngai and Petrongolo, 2017). Second, there are differences in women's earning capacity arising from gender gaps in educational attainment, due to lower expected employment opportunities or son preferences (Heath and Jayachandran, 2017). Third, the burden of home production and child care or child bearing, which keeps women at home, varies by country. Economic development abates these burdens, through electrification (Dinkelman, 2011), invention and spread of household appliances (Greenwood et al., 2005; Coen-Pirani et al., 2010), lower fertility (Miller, 2010), improved child care facilities, and medical advancement in maternal health care (Albaseni and Olivetti, 2009, Jayachandran and Lleras-Muney, 2009).

On the other hand, the homemaker norms wedge of Lee (2020) focuses on the differences not between men and women, but between married and single women. Therefore, sectoral transitions that favor women in general do not affect the norms wedge. Moreover, the norms wedges are computed by matching a group of married women to a group of single women with the same level of education and family composition – number of children under the age 18 and under the age 5 – so that earning capacity or the burden of home production is more or less kept similar in the two groups of women. Nonetheless, even with education and family composition matched, married women's labor force participation typically falls behind single women's. It is the norms wedge that picks up this married-single gap. Thus, the wedges are summative measures that encapsulate all the reasons besides wages for which the gap arise.

Using the country censuses compiled in the IPUMS International database, I compute the homemaker norms wedges for seventeen different countries and find that they exhibit a wide cross-country variation. The norms wedge is almost negligible in Canada, implying that a 10-dollar market wage is valued at very close to 10 dollars for married women in Canada when they make their labor force participation decisions. In contrast, a 10-dollar market wage is only valued at 5.7 dollars by the married women in Indonesia. Because the IPUMS International database includes censuses taken at different points in time, the country-average of the norms wedges over all available census years reflect not solely the cross-country differences but also the time series differences; the norms wedges do change over time in a given country.[4] However, as the wedges are slow to change over time, the cross-country variation eclipses the within-country variation, and so I maintain the main focus of this paper on the cross-country variation. Furthermore, the homemaker norms wedges are large in magnitude, amounting to around a half of the "gender norms wedge", the men-women counterpart to the homemaker norms wedge; the gender norms wedge rationalizes by how much the labor force participation of women falls behind the level predicted if they behaved like similar men.

Although the homemaker norms wedges compound all the reasons besides wages that

[4]Lee (2020) documents how the homemaker norms wedge changes over time in the United States.

create a gap in the labor force participation between married and single women, I find that culture plays an important part in explaining the cross-country variation in the wedges. The norms wedges are strongly correlated with answers to questions on the appropriate behavioral prescriptions for men and women, asked in the World Values Survey. This relationship persists even within-country, when I exploit the change in the norms wedges and attitudinal survey answers over time for the countries surveyed at multiple time periods. However, the norms wedges are not correlated with a comparable placebo question about attitudes on immigrants.

To isolate the effect of culture through an epidemiological approach, I repeat the exercise for groups of married and single women of different ancestry among the U.S. population. The idea behind this approach is that the effect of culture can be identified through the variation in economic outcomes of individuals who share the same economic and institutional environment, but whose social beliefs are potentially different (Fernández, 2011). Indeed, the norms wedges by ancestry are also found to be correlated with gender attitudes, and they account for around 45% of the cross-country variation in the norms wedges.

Finally, I quantify how much of the cross-country variation in female labor force participation is driven by the variation in the homemaker norms wedges. I compute the counterfactual female labor force participation for the scenario where the norms wedges disappear altogether. The labor force participation of married women increases by 38% on average in the counterfactual. Taking married and single women together, female labor force participation increases by 21% on average. The cross-country variance in female labor force participation decreases by 8%.

The rest of this paper is organized as follows. Section 3.2 explains the data sources used. Section 3.3 then discusses what the homemaker norms wedges mean and how they are computed. The resulting wedge values are discussed in Section 3.4. Next, I check that culture is pertinent for the norms wedges in Section 3.5. Section 3.6 isolates the effect of culture through the epidemiological approach and discusses how much of the cross-country variation in the norms wedges is due to cultural differences. Lastly, I return to the motivation of the large variation in female labor force participation across countries, and I examine how much of it is driven by differences in the norms wedges. Section 3.8 concludes.

3.2 Data

3.2.1 Country censuses

The data used to compute the norms wedges for each country comes from the IPUMS International database, which compiles the censuses of numerous countries. A few countries' household survey data are also included. The main advantage of this dataset is that the variables present in each country's census are harmonized. This feature allows the model to be applied to each country in a consistent manner. Out of the countries in the database, I select the ones which contain all the variables necessary to apply the model of Lee (2020):

market earnings, labor force participation status, marital status, education, children, and the code linking a person to their spouse. Although some 98 countries are listed as part of the IPUMS International database, the earnings variable is not available broadly and therefore is the most limiting factor in the country selection.

The countries used are Brazil, Canada, Colombia, Dominican Republic, India, Indonesia, Israel, Italy, Jamaica, Mexico, Panama, South Africa, Trinidad and Tobago, United States, Uruguay, and Venezuela.[5] Table 3.1 lists the countries and census years for which all the variables needed to compute the norms wedges are available. The main focus of this paper is the cross-country variation in the norms wedges, since it markedly dominates within-country time series variation.[6] Therefore, when I conduct the cross-country analysis, I will work with the average over the years by country. I do make explicit use of the country-year panel structure when I demonstrate the strength of the correlation between norms wedges and gender attitudes by showing that the correlation is robust to the addition of country fixed effects in Section 3.5.1.

Appendix Section 3.A discusses in greater detail how I harmonize the variables are heterogeneous across countries.

Table 3.1: List of census or survey

Country	Census/survey year(s)
Brazil	1980, 1991, 2000, 2010
Canada	2011
Colombia	1973
Dominican Republic	1981, 2002
India†	1983, 1987, 1993, 1999, 2004
Indonesia	1976, 1995
Israel	1972, 1995
Italy†	2011‡
Jamaica	1982, 1991, 2001
Mexico	1970, 1990, 1995, 2000, 2010, 2015
Panama	1970, 1980, 1990, 2000, 2010
Puerto Rico	1990, 2000, 2005, 2010
South Africa	1996, 2001, 2007, 2011
Trinidad and Tobago	1970, 2000
United States	1980, 1990, 2000, 2010
Uruguay	2006
Venezuela	1971, 1981, 1990, 2001

Notes: † Surveys, not censuses

‡ I combined the survey data from years 2011, 2012, ..., 2018, due to the small sample size in each year.

[5] Germany is another country for which earnings data is available in its 1970 census, but the spousal identifier is missing there.

[6] I compare the cross-country and time series variation later in Section 3.5.2.

3.2.2 Attitudinal measures

The second dataset I employ is the World Values Survey (WVS), which asks various attitudinal survey questions to respondents in countries around the world. Of different surveys on attitudes, the WVS is well known to be the one with the greatest coverage globally. The WVS allows me to check whether the norms wedges are in line with conventional measures of attitudes. An example of the ideal attitudinal survey question is, "Do you approve or disapprove of a married woman earning money in business or industry if she has a husband capable of supporting her?", asked in the General Social Survey. The question directly asks about the appropriate role of the married woman. Unfortunately, such a question is not asked in the WVS.[7]

Instead, to measure gender attitudes, I make use of the survey question that asks whether the respondent agrees or disagrees with the statement: "When jobs are scarce, men have more right to a job than women." This particular question is especially appealing for two reasons. Firstly, it is consistently asked for the greatest number of survey waves and therefore allows more country- and year-specific norms wedges to be matched to the attitudinal measure.[8] Secondly, it is asked just alongside a similar yet very different statement: "When jobs are scarce, employers should give priority to [own nationality] people over immigrants." Both statements revolve around job scarcity and are phrased similarly, and the answer categories are the same. Yet, the former asks about gender and the latter about immigrants. Therefore, the latter would be useful as a check for whether any correlation between norms wedges and attitudes on gender, if present, is a fluke. If a similar correlation exists between the wedges and the attitudes on immigrants, then it might merely be that the wedges reflect *general* conservative values.

3.2.3 Ancestry information among U.S. immigrants

To what extent can the variation in norms wedges by country be explained by differences in culture? In order to break apart culture from other economic and institutional factors affecting the norms wedge values, I consider computing the norms wedges by ancestry among the U.S. population. The idea is that groups of workers in the U.S. who operate within the same economic and institutional conditions might display different labor force participation patterns due to different cultural beliefs that stem from different ancestry. This procedure is performed on the United States censuses in the IPUMS International database, using the variable *ancest*. *Ancest* records the respondent's self-reported ancestry or ethnic origin, where typical examples include German, English, Irish, and Mexican. If the respondent gave multiple ancestries, then *ancest* records the first response. Since *ancestr* is self-reported, it captures the ethnic origin that the respondent identifies with. It therefore would sensibly represent the culture that the respondent's actions and choices are influenced by.

[7] The gender-related questions that are asked in the WVS are typically about what is expected from a woman in general, or a (working) mother in general, not distinguishing married and single women.

[8] The question was asked in all WVS waves but the first. Each survey wave covers 3-5 years, and the second wave covered 1990-1994.

3.3 Computation of norms wedges

This section discusses how the homemaker norms wedges are computed for each country and year. Before I proceed on to the computation results, I provide a brief background for what the norms wedge means, how it is derived from the model, and which assumptions need to be made. In particular, I highlight which factors are assumed to be the same across countries, and which factors are allowed to be country-specific, when I take the one model to different country datasets.

3.3.1 Derivation of the homemaker norms wedge

To explain what the norms wedge means and how it is derived, I reproduce the key pertinent parts of the model in Lee (2020). The norms wedge measures how much a 10-dollar market wage is valued at for married women, when they are deciding between working in the labor market versus working at home. A married woman f's labor force participation decision is represented by:

$$L_f^* = \mathbb{1}\left[(1-\tau)w_f \geq h_f\right] \tag{3.1}$$

where w_f denotes her market wage, h_f her home productivity, and τ the norms wedge. For example, with $\tau = 0.4$, a married women who would make \$20 by working in the market behaves as if she would earn \$12. Furthermore, she would only work if her home productivity is lower than \$12.

The labor force participation decision of married women can be contrasted against that of a single woman i:

$$L_i^* = \mathbb{1}\left[w_i \geq h_i\right] \tag{3.2}$$

In order to identify the τ in equation (3.1), I need to know the labor force participation decision and potential market earnings and home productivity of married women. Labor force participation is observed in the data. Also observed in the data are market earnings of those who chose to work, from which potential market earnings can be deduced accounting for selection into the labor market. However, home productivity is *not* observed. Therefore, it is necessary to find a counterfactual group from which home productivity can be extracted. As shown in equation (3.2), single women's labor force participation decision in the model is based on market earnings and home productivity alone, with no wedge involved, so I can back out the potential home productivity of single women. Assuming that single and married women who are similar in terms of education and family composition – number of children under the age 18 and under the age 5 – share the same level of home productivity, I can finally back out the value of τ.

All in all, the computation of the norms wedges relies on figuring out what the norm wedge must have been, to rationalize by how much the labor force participation of married women falls behind the level predicted if they behaved like similar single women. The norms wedges are defined as:

$$\tau = 1 - \frac{avrwage_{\text{ single women}}}{avrwage_{\text{ married women}}} \left(\frac{1 - LFP_{\text{single women}}}{1 - LFP_{\text{married women}}} \right)^{\frac{1}{\theta}} \tag{3.3}$$

where *avrwage* denotes the average wage of the workers among a group of married or single women, and LFP the labor force participation rate of that group. θ is an inverse measure of the dispersion of idiosyncratic abilities in the population, which parameterizes the selection into market work.

3.3.2 Taking the model to different countries

When I compute the norms wedges across countries, I take equation (3.3) and compute it using the data for each country.

Is comparing the norms wedge value of country A and country B tantamount to comparing apples and oranges? From equation (3.3), it is clear that the norms wedge value has a closed form solution. Thus, when I compare the norms wedges of two countries, I are comparing exactly that in each country. The norms wedge can therefore be thought of as a summative measure the encapsulates all the reasons for which the labor force participation rate of a group of married women differs from that of a group of similar single women in a country, apart from wage differentials. The fact that the norms wedge is a summative measure is an advantage in itself, allowing for a simple way to compare different countries.

Moreover, cross-country differences in numerous factors would not confound the norms wedge values, as long as those factors affect married and single women of a country in the same way. For instance, total factor productivity (TFP) and thus the wage levels of the labor force are different by country, but that does not affect the value of the norms wedge, as long as similar married and single women operate in comparable labor markets. This feature also means that differences in data collection methodology by country do not matter for the norms wedge value. Moreover, in matching a group of married women to a group of single women by education and family composition, the model already takes into account that there are differences in education levels and child-bearing propensities by country.

However, which exact ingredients go in to diverging norms wedges by country remains a black box, without tailoring the model in substantial ways to each country. Therefore, the extent to which the norms wedge value can stay true to its name – that it can be interpreted as a measure of the societal norm against married women working – would vary by country. For example, if there are positive returns to experience and the risk of divorce is high in a country, a married woman would be more likely to work in the market *not* because of modern gender role values but as a precautionary action. Depending on the likelihood of divorce by country, how well the norms wedge value truly reflects the societal norm would vary. As another example, if the national welfare system is weak and market earnings are volatile (such as in seasonal or agricultural jobs), a married woman would be more likely to work in order to secure an additional source of income for the rainy days when the husband cannot provide. Then, the norms wedges would also reflect differences in the national welfare system.[9]

3.4 Norms wedge values

3.4.1 Aggregating homemaker norms wedges at the country level

The calibration exercise involves backing out the value of τ for each (education-family composition) match between married and single women. Hence, the value of τ is sensitive to the size of each group, or cell, of married and single women. As it can be seen in equation (3.3), τ is based on group means: the labor force participation rate and the average of the wages earned by the market workers. The larger the cell size of both married and single women in a given (education-family composition) match, the more precise and stable is τ for that match. On the contrary, with small cell sizes, the more likely it is that the τ value is unreliable. In fact, as evident in equation (3.3), τ is not constrained to be between 0 and 1; rather, the range of τ is $(-\infty, 1]$. It is thus possible that an imprecise τ takes a very low negative value that is clearly implausible. In the IPUMS International database, different countries have different numbers of observations, and so the countries with small populations or small census sampling would be more prone to this issue.

Therefore, in putting the τ values by group together at the country-year level, many summary statistics can be considered. The summary statistics I consider include the mean and median that are unweighted or weighted by the cell size – either the raw observation number per cell or the combined person weights in the census per cell. In addition, I consider first truncating the values of the group-specific τ at the top and bottom 1%, 5%, and 10%, and then looking at the unweighted or weighted mean and median. The last summary statistic I consider is the unweighted or weighted mean only among τ values computed with cells in which the number of observations exceeds an arbitrary threshold, 100.

Table 3.2 demonstrates how correlated the various summary statistics are, where each summary statistic is at the country-year level. The mean values tend to be quite sensitive to the weighting, truncation, or restriction criteria, due to the presence of outlier values. The median values on the other hand, are quite robust to the different criteria. Moreover, it can be seen that truncating the top and bottom values of τ within a country barely makes a difference, whereas weighting does. Overall, the median values are highly correlated with one another. The selection of which median value to use consistently throughout the rest of the analysis therefore becomes tricky. I select the one that is the most stable over time within countries, i.e. the one with the lowest variance, within country and across years. This statistic is the unweighted median τ conditional on cell sizes exceeding 100 observations, which corresponds to 10 in Table 3.2.

To arrive at the norms wedge value at the country level, I take the average of the

[9] In Appendix Section 3.A, I discuss the caveat that the earnings variable in the IPUMS International censuses are before-tax numbers.

country-year summary statistic (10) across years for each country.

Table 3.2: Correlation coefficients of various summary statistics for norms wedges

centrality	mean	median	mean	median	mean	median	mean	median	mean	median
weighting	-	-	-	-	obs no.†	obs no.†	census wt.‡	census wt.‡	-	-
truncation	-	-	5%	5%	5%	5%	5%	5%	-	-
restriction	-	-	-	-	-	-	-	-	> 100§	> 100§
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	1.000									
(2)	0.781	1.000								
(3)	0.921	0.948	1.000							
(4)	0.761	0.997	0.938	1.000						
(5)	0.783	0.900	0.914	0.902	1.000					
(6)	0.655	0.859	0.828	0.863	0.965	1.000				
(7)	0.779	0.901	0.912	0.905	0.998	0.967	1.000			
(8)	0.660	0.853	0.825	0.858	0.959	0.995	0.966	1.000		
(9)	0.828	0.951	0.964	0.949	0.955	0.888	0.956	0.884	1.000	
(10)	0.706	0.976	0.893	0.982	0.889	0.860	0.892	0.856	0.932	1.000

Notes: (1), (2), ..., (10) indicate summary statistics that differ by measure of centrality, weighting, sample truncation, and sample restriction.

† obs no.: the number of observations in each cell

‡ census wt.: the total sum of person weights in each cell

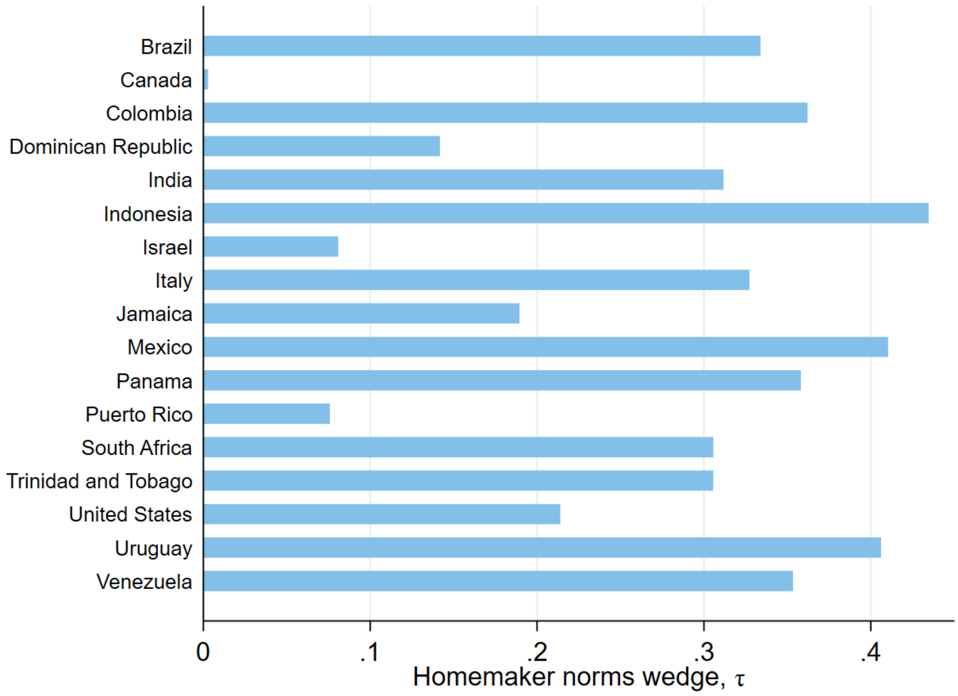
§ >100: only cells with strictly more than 100 observations

3.4.2 Norms wedges across countries

Figure 3.1 depicts the values of the norms wedges by country. There is substantial cross-country variation. As societal norms have changed over time, the fact that the IPUMS International database covers censuses of countries from different points in time contributes to this variation. For instance, the only year in which the data for Canada is available is 2011, whereas the only year in which the data for Colombia is available is 1973.[10] However, differences in years used by country are far from the main drivers of the variation. I show later in Section 3.5.2 that, at least for the sample period of interest, the time series variation within country is eclipsed by the cross-country variation.

[10] Table 3.1 provides information on which years are covered for each country.

Figure 3.1: Homemaker norms wedge by country



Notes: This figure illustrates the values of the homemaker norms wedges by country. For each country, I take the average of the country-year summary statistic (10).

3.4.3 Magnitudes of norms wedges

Are the values of the norms wedges, as seen in Figure 3.1, large or small? It would be useful to have a benchmark to compare the norms wedges to.

To this end, I compute the "men vs. women" counterpart of the norms wedge: what is the "gender norms wedge" that rationalizes by how much the labor force participation of *women* falls behind the level predicted if they behaved like similar *men*? The gender norms wedge can be juxtaposed to the "married women vs. single women" comparison that is the norms wedge.

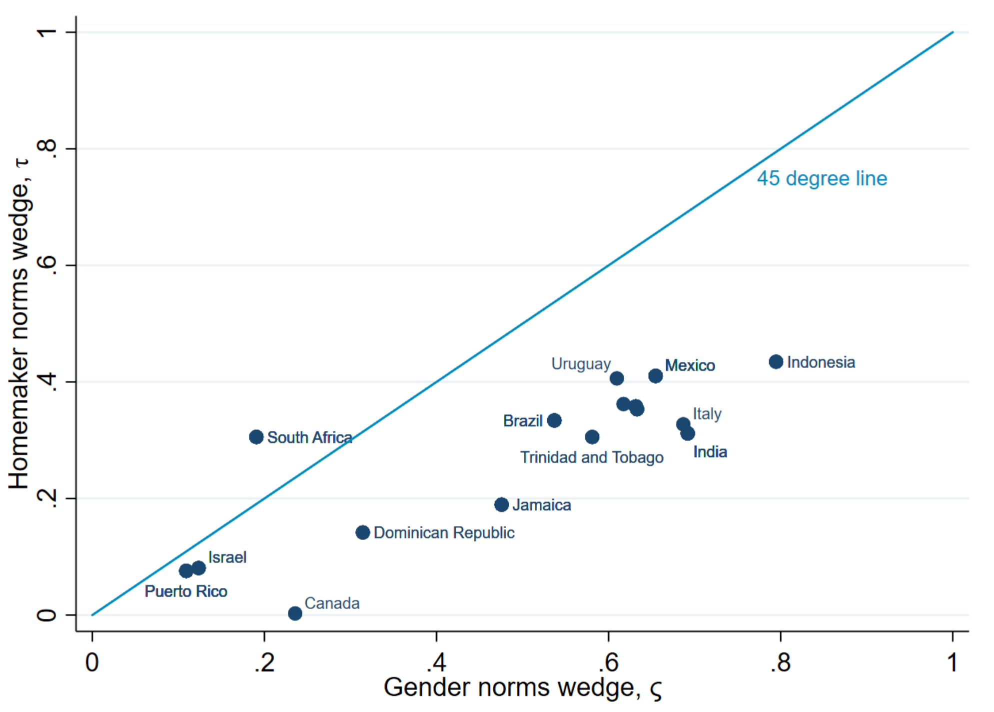
The gender norms wedge is defined as:

$$\varsigma = 1 - \frac{avrwage_{\text{ men}}}{avrwage_{\text{women}}} \left(\frac{1 - LFP_{\text{men}}}{1 - LFP_{\text{women}}} \right)^{\frac{1}{\theta}} \tag{3.4}$$

For the gender norms wedge, men and women are matched on their education and family composition categories, similarly to the norms wedge computation.

Figure 3.2 plots the values of the norms wedges against the gender norms wedges by country. For all countries but South Africa, ς is greater than τ. In fact, the magnitude of τ approximates around half of the magnitude of ς. Moreover, there is greater degree of cross-country variation in ς than in τ.

Figure 3.2: Homemaker norms wedge and gender norms wedge by country



Notes: This figure compares the plots the homemaker norms wedge and the gender norms wedge for each country. The gender norms wedge is the "men vs. women" counterpart of the "married women vs. single women" comparison that is the homemaker norms wedge.

A caveat to keep in mind is that the assumption that men and women share a similar level of home productivity is less plausible than it is for married and single women. When children are small, in particular, biological gender differences contribute to the greater need for women to stay home than men. There is also a vast literature on gender differences in mental or physical tendencies, such as risk-taking (Charness and Gneezy, 2012), competitiveness (Niederle and Vesterlund, 2007), opportunities to rise to the top (Athey, Avery, and Zemsky, 2000), and how their performance is assessed (Mengel, Sauermann, and Zölitz, 2018). Therefore, where these gender differences contribute to why females participate less in the labor market relative to males than what the wage differentials suggest, they would contribute to the large magnitude of the gender norms wedge. It is less likely that the gender norms wedge serves as a measure of a social norm than the homemaker norms wedge.

Nonetheless, the gender norms wedges is still useful as a benchmark value. That there are vast gender differences in numerous dimensions, in fact, functions to highlight the size of the homemaker norms wedge, amounting to a half of the gender norms wedge.

3.5 Is culture pertinent for the norms wedge?

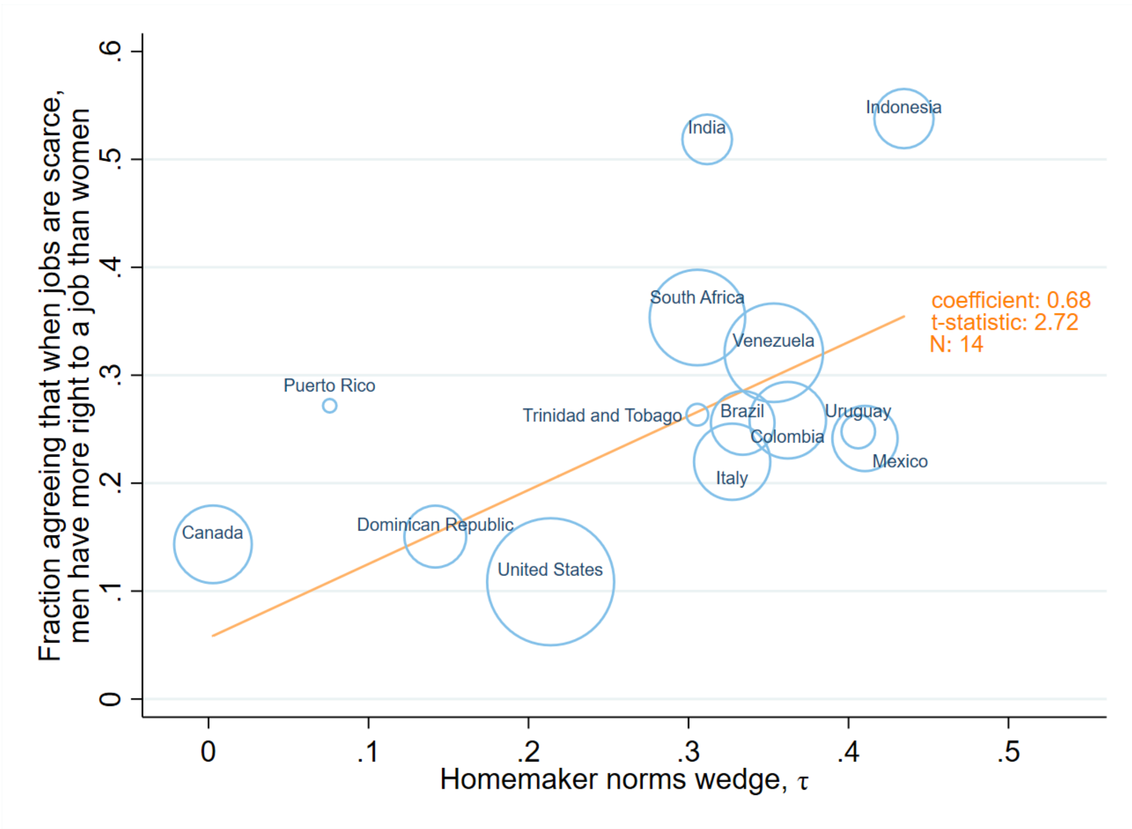
3.5.1 Correlation between norms wedges and attitudes

The norms wedges are summative measures that encapsulate all the reasons besides wages for which the gap in the labor force participation of married women and that of similar single women arises. Therefore, the cross-country variation in the norms wedges naturally reflect differences in economic structure and institutions. However, could differences in culture also drive the cross-cross variation? If the norms wedges are driven by economic or institutional differences alone, then the nomenclature "norms" wedge would be misleading.

Taking attitudinal survey data as proxies for cultural norms, I gauge whether the norms wedges are in line with attitudes regarding gendered behavioral prescriptions. To this end, I correlate the τ's and the answer to the WVS question, "Do you agree with the following statement: men have more right to a job than women." The τ and the attitudinal survey answers are aggregated at the country level.[11] The result is plotted in Figure 3.3. The correlation shows that in spite of all the cross-country differences the norms wedges fail to take into account, the countries with large τ's are indeed the ones that tend to have more conservative gender attitudes.

[11]The values of τ aggrefated at the country level are in Figure 3.1.

Figure 3.3: Cross-country correlation between norms wedges and attitudes



Notes: The attitudinal survey data originate from the World Values Survey, and the data used to compute the norms wedges are from the IPUMS International database. The reported t-statistic is based on robust standard errors. The size of the bubble for each country represents the size of the sample used in the computation of τ.

However, it is difficult to see the correlation in Figure 3.3 as a causal relationship. One of the most representative confounding variable would be GDP.[12] Richer countries might have more progressive attitudes on all fronts in general, which would result in a lower fraction agreeing that men have more right to a job then women when jobs are scarce. Richer countries might also have better transport infrastructure which allows a married woman – who might live near their husband's workplace, unlike a similar single woman who can live near their own workplace – to travel with ease to work, resulting in a lower norms wedge.

Therefore confirm that the relationship between attitudes and norms wedges is robust, I construct a country-year panel dataset. A panel allows me to address country-specific factors that influence the correlation, such as the level of economic development, sectoral composition of the economy, infrastructure, and legal or social institutions. Instead of matching country-level means of attitudes and τ, I match the two variables by the closest year.[13] The advantage of using this particular WVS question becomes apparent here. As explained in Section 3.2.2, it is the question that got asked the most consistently over time

[12] Goldin (1995) documents a U-shaped cross-country relationship between female labor force participation and economic development.

[13] Appendix Table 3.A2 lists how the years of the IPUMS International censuses and the years of the World Values Survey are matched.

in the WVS that relates to gender. It was asked in all five waves after the first wave, covering the years 1990 to 2014. Because the number of countries for which the norms wedge can be computed is quite small, it is critical to include as many time periods as possible.

Moreover, the other advantage of this WVS question is that there is a directly comparable question on the topic of immigrants. Both are similarly phrased and answered, and are on the same issue of job scarcity. The two questions were asked side by side in the survey questionnaire, so a country-year panel can be constructed for the exact same country-year pairs.

Table 3.3 reports the results of regressing the attitudinal measures on the norms wedge, augmented by country fixed effects. Clearly, the sample is still very small in this panel dataset.[14] Nonetheless, even with the country fixed effects, τ and the WVS question on gender are correlated. What is reassuring, too, is that the coefficient value remains similar with the addition of country fixed effects. Without the country fixed effects, the coefficient value is 0.68, as depicted in Figure 3.3. The coefficient value with country fixed effects, reported in column (1) of Table 3.3, is 0.65. When the fixed effects for the World Value Survey waves are added on top, in column (2), the coefficient size decreases but remains statistically significant. Overall, the relationship between the homemaker norms wedges and gender attitudes appear to be robust.

On the other hand, τ and the WVS question on immigrants are found not to be, in columns (3) and (4). Hence, it is not the case that the correlation between norms wedges and attitudes on gender, depicted in Figure 3.3, represents a correlation between the norms wedges and conservative values in general.

[14] The number of countries with multiple years is also very small. Thus, the regression results in Table 3.3 include both the cases when the standard errors are clustered by country and when they are not. The t-statistics are similar.

Table 3.3: Norms wedges against attitudes on female or immigrant workers

	Fraction agreeing that when jobs are scarce, ...			
	men have more right to a job then **women**		employers should prioritize **own nationals** over **immigrants**	
	(1)	(2)	(3)	(4)
Homemaker norms wedge (τ)	0.648**	0.453*	0.639	0.268
	(2.39)	(1.99)	(1.30)	(0.44)
	[2.75]	[1.99]	[1.55]	[0.64]
Country fixed effects	✓	✓	✓	✓
Survey wave fixed effects		✓		✓
No. of countries	14	14	13	13
No. of countries with >1 waves	7	7	7	7
N	29	29	28	28

Notes: t statistics from standard errors clustered by country in parentheses, and t statistics from robust standard errors in square brackets; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
When the sample for columns (1) and (2) are fixed at the sample for columns (3) and (4), the regression results barely change. The reason is that the country fixed effect absorbs the one country that is present in the former sample and missing in the latter sample: Colombia.

3.5.2 Cross-sectional vs. time series variation in norms wedges

Although Table 3.3 demonstrates that there is certainly variation in the norms wedges and attitudes within a country over time, the more interesting variation remains the cross-country one – at least for the relatively short time range covered by the data. Table 3.4 decomposes the variation in the norms wedges and the gender attitudes into cross-country and time series components. In Panel A, the Shapely decomposition provides an additive decomposition of the R squared in the OLS regression of the norms wedge on country and year indicators. It quantifies the relative contribution to the R squared of the group of country dummies and the group of year dummies. Although there are far greater number of year dummies than country dummies, the share of contribution to the R squared lies much more heavily on the country dummies. The country dummies contribute to over three times the contribution of the year dummies. Similarly, in Panel B, the WVS question on gender also varies much more greatly across countries than across years. The Shapley decomposition shows that the country dummies contribute 87% of the variation in the attitudinal survey answer.

The between-within decomposition decomposes a variable (x_{it}, of country i and year t) into between-country ($\bar{x}_i \equiv \frac{1}{T}\sum_{t=1}^{T} x_{it}$) and within-country ($x_{it} - \bar{x}_i + \bar{\bar{x}}$, with the global mean $\bar{\bar{x}} \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}$ added back in to make the results comparable) components. For both the homemaker norms wedge and the gender attitude measure, the between-country variance is over five times the within-country variance.

Therefore, although the panel structure was useful to demonstrate the robustness of the relationship between norms wedges and attitudes, I focus on the cross-country variation in the sections to follow.[15]

Table 3.4: Cross-sectional and time series decomposition of norms wedges and attitudes

***Panel A*: Homemaker norms wedge (τ)** [N=51]

	Shapley decomposition			**Between-within decomposition**		
	Number of categories	Share of contribution to R^2		Standard deviaion	Min.	Max.
Countries	17	76.05%	Between-country	0.130	0.003	0.434
Years	26	23.95%	Within-country	0.057	0.158	0.399

***Panel B*: Fraction agreeing that when jobs are scarce, men have more right to a job than women** [N=152]

	Shapley decomposition			**Between-within decomposition**		
	Number of categories	Share of contribution to R^2		Standard deviaion	Min.	Max.
Countries	59	86.78%	Between-country	0.203	0.040	0.877
Years	22	13.22%	Within-country	0.053	0.222	0.574

Notes: The Shapely decomposition provides an additive decomposition of the R squared of the OLS regression. It quantifies the relative contribution of each group of regressors to the R squared. The between-within decomposition decomposes the standard deviation of a variable into between and within components. The value of N refers to the number of observations for each variable.

3.6 How much of the cross-country variation in norms wedges can culture account for?

It is not sufficient to show that the homemaker norms wedges are correlated with attitudes, to establish that culture plays a role in explaining the cross-country variation in the wedges.

To get closer to causality, I adopt the idea behind the epidemiological approach in the literature assessing whether culture matters for economic decisions, reviewed by Fernández (2011). The epidemiological approach attempts to identify the effect of culture through the variation in economic outcomes of individuals who share the same economic and institutional environment, but whose social beliefs are potentially different. For immigrants in the United States operating in the same labor market, differences in their labor force participation is a reflection of different cultural heritage. Thus I am able to separate culture

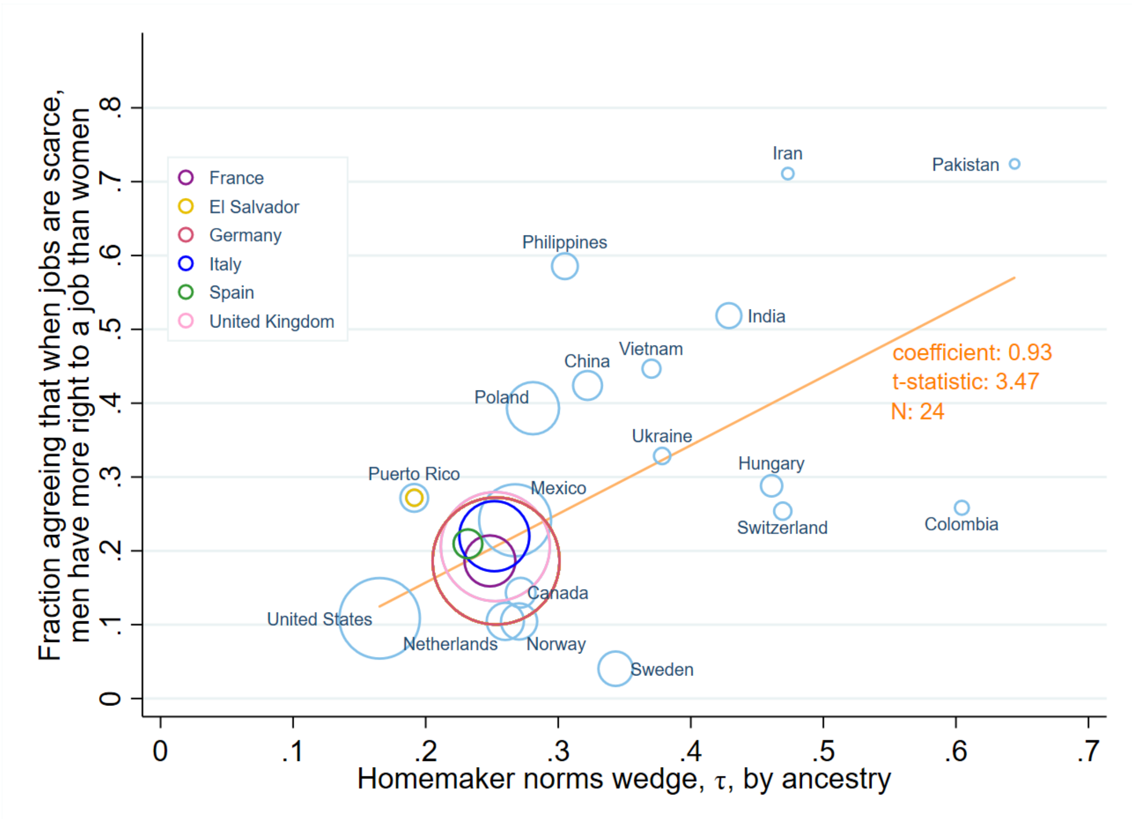
[15]In addition, adding country fixed effects necessitates that 7 out of 14 countries "drop out" from the variation, since they are observed only in one time period.

from differences in economic and institutional environment, by computing the τ separately for U.S. immigrants from different origin countries.

I compute the τ separately by self-reported ancestry in the U.S. censuses of the IPUMS International database. To illustrate, the τ corresponding to the Italian heritage represents the gap in the labor force participation rates of married Italian American women to similar single Italian American women, that is not accounted for by the wage differentials. Groups of married and single Italian American women are matched by education and family composition, as before. After identifying the group-specific τ for each census year, I take all those values of τ, and aggregate them at the country level by taking the unweighted median of the τ's computed with cells of sizes greater than 100. This statistic is the same as the one used to aggregate the norms wedges by country in Section 3.4.

Do these norms wedges by ancestry make sense? I check whether they are correlated with attitudes on gender. Figure 3.4 reproduces Figure 3.3 for the norms wedges by ancestry. I could compute and aggregate the norms wedge by ancestry for 24 origin countries, more than the 14 countries for which I could compute the country-level norms wedge. Similarly to norms wedges at the country level, the norms wedges by ancestry among the U.S. population are strongly correlated with gender attitudes.

Figure 3.4: Correlation between norms wedges by ancestry and gender attitudes

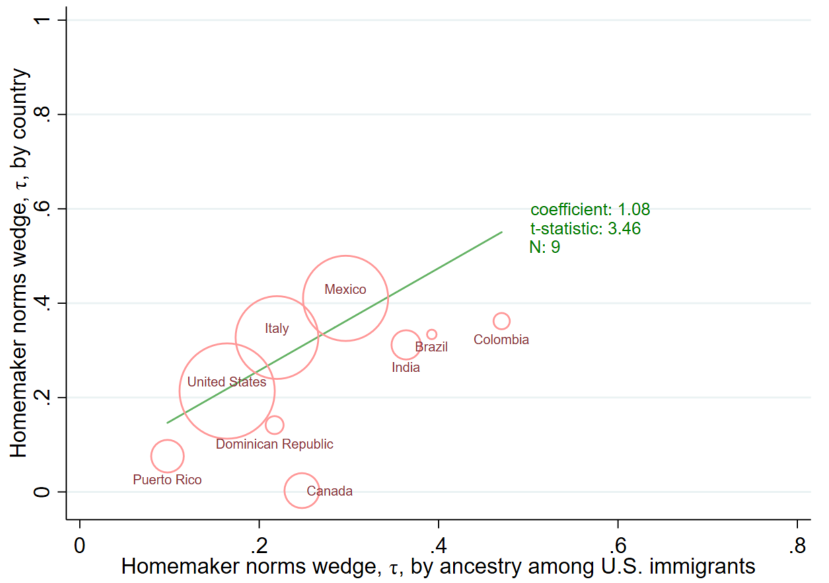


Notes: The attitudinal survey data originate from the World Values Survey, and the data used to compute the norms wedges are from the IPUMS International database. The reported t-statistic is based on robust standard errors. The size of the bubble for each country represents the size of the sample used in the computation of τ by ancestry. Some of the bubbles are colored differently for visual clarity.

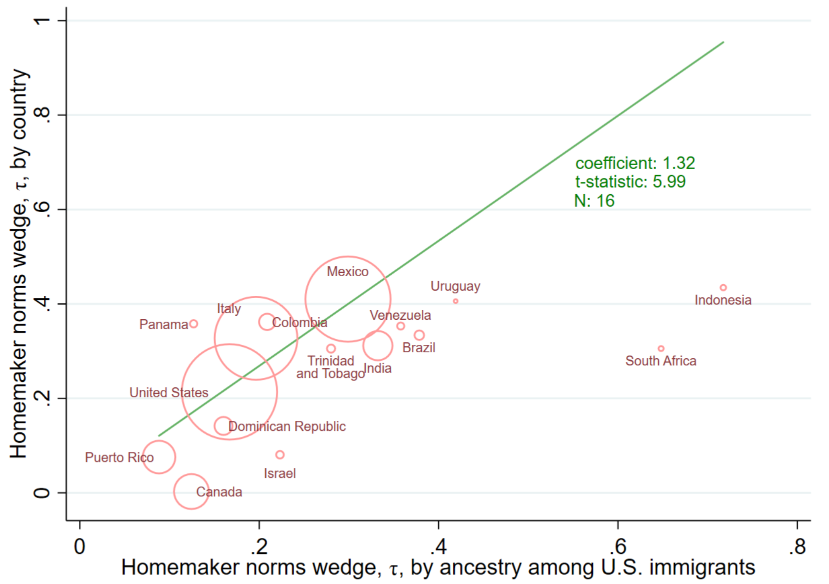
Then, how much of the cross-country variation in norms wedges be explained by culture, as measured through the epidemiological approach? To this end, I plot the norms wedges by country against the norms wedges by ancestry among U.S. immigrants in Figure 3.5.[16] Plot (A) of Figure 3.5 depicts the scatter plot of the two types of norms wedges when the same statistic, unweighted median conditional on cell sizes exceeding 100, is used for both. The number of countries for which the two types of norms wedges are matched is rather small at 9. The reason is that the U.S. immigrant population in the 5% census sample cannot be large for all ancestral backgrounds. For example, where the size of the Indonesian American population is not large, it becomes difficult to meet the restriction that τ is to be calculated only if the cell sizes are large enough. Hence, in Plot (B) I also repeat the scatter plot for when the norms wedges by ancestry are computed with the cell size restriction lifted.

[16] Appendix Table 3.A3 reports the regression results. For greater comparability, Appendix Table 3.A3 also reports the results of the regression when the cell size restriction is lifted for the norms wedges by country, too. The results remain similar.

Figure 3.5: Correlation between norms wedges by country and by ancestry



(A)



(B)

Notes: This figure plots the homemaker norms wedge τ by country against τ by ancestry among the U.S. population. Plot (A) depicts the scatter plot of the two types of norms wedges when the same statistic, unweighted median conditional on cell sizes exceeding 100, is used for both. For Plot (B), the norms wedges by ancestry are computed without the cell size restriction. The size of the bubble for each country represents the size of the sample used in the computation of τ by ancestry.

In both plots (A) and (B) of Figure 3.5, there is clearly a strong positive relationship between the two types of norms wedges. Therefore, taking the norms wedge by ancestry as a good measure of the isolated effect of culture, the norms wedge by country does indeed contain the effects of culture.

To see how much of the cross-country variation in norms wedges can culture account for, I consider two factors: the R squared and the coefficient value. Take a simple linear regression given by

$$y_i = \alpha + \beta x_i + \epsilon_i \tag{3.5}$$

The R squared value of this regression represents the degree of variation in y that the linear projection of x can account for. Visually, it is displayed by how close the observation points are situated to the line of best fit. The coefficient $\hat{\beta}$, on the other hand, is useful for gauging whether the relationship is positive or negative, and also for comparing the magnitude of the change in y that is associated with a one-unit change in x.

If, for example, the norms wedge by country was always exactly double the norms wedge by ancestry, then the R squared would equal 1, and the coefficient estimate $\hat{\beta}$ (as well as the true parameter value β) would equal 2. The conclusion in that scenario would be that the cross-country variation in the norms wedge is entirely explained by cultural differences, and that the effect of culture is attenuated by half among the U.S. immigrants, perhaps by cultural assimilation within the U.S, post-immigration.

By contrast, if the norms wedge by country had absolutely no correlation with the norms wedge by ancestry, then both the R squared and the coefficient estimate would be very close to 0. The conclusion then would be that factors that are completely orthogonal to culture drive all of the cross-country differences in the norms wedges.

The R squared value of plot (A) is 0.44, while for plot (B), the R squared value is larger at 0.69.[17] Hence, 44 to 70% of the variation in the cross-country differences in the norms wedge can be accounted for by differences in culture. Moreover, the coefficient estimates for both plots are greater 1, which implies that the effect of culture is stronger in the country-level norms wedges. The reason could be that the strength of culture is weakened post-immigration. It could also be that other country variables that are correlated with culture affect the norms wedges in the same direction as culture. To illustrate, countries with a stronger homemaker social norm for married women may tend have inferior transport infrastructure which further impedes married women from traveling to work.

3.7 Female labor force participation across countries

The previous sections showed that a) there is a large degree of variation in the country-level norms wedges, b) norms wedges are sizable, relative to gender norms wedges, and c) a hefty portion of the variation can be accounted for by cultural differences.

The cross-country variation in the norms wedges demonstrates itself in the large variation in female labor force participation across countries. Figure 3.6 plots on the horizontal axis the labor force participation rate of all (married and single) women and of married women by country, taking the average across all available IPUMS International census years. The large variation is evident; married women's labor force participation ranges from a low of 10.3% to a high of 80.8%.[18]

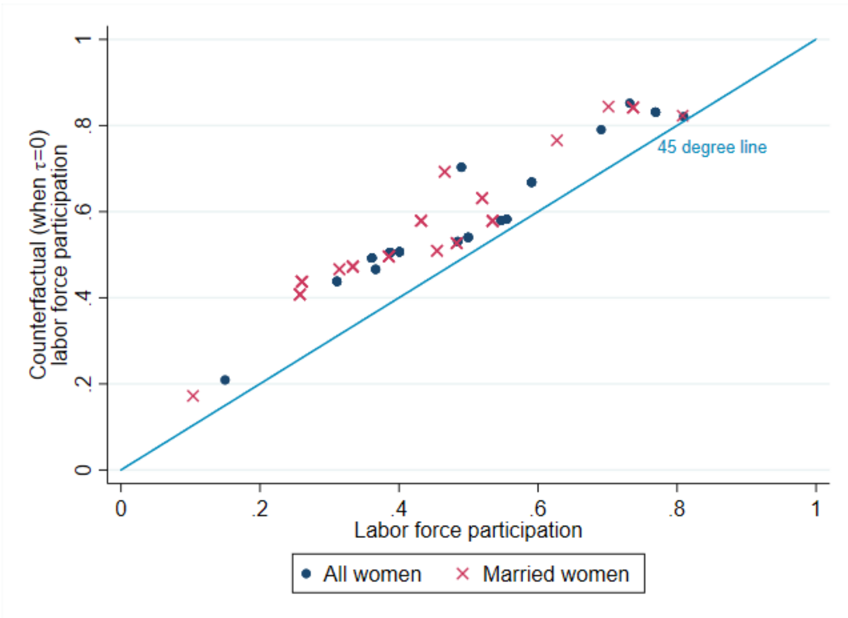
[17] Visually, too, the scatter is tighter around the line of best fit in plot (B) than in plot (A).

[18] As these numbers are averages across census years, it does mix up the time series variation along with the cross-sectional variation. Specifically, the low of 10.3% corresponds to Colombia in 1973 and the

To what extent do the norms wedges contribute to the cross-country variation in female labor force participation? I compute how the pattern of female labor force participation change, had the norms wedges disappeared altogether. Those counterfactual values are plotted on the vertical axis of Figure 3.6. As the scatter points in Figure 3.6 are situated above the 45 degree line, the figure shows that the counterfactual values are larger than the actual values. On average, the labor force participation rate of married women rises by 38% in the counterfactual. Taking married and single women together, the rise is at 21%. The reason the rise is larger for married women is simple: the norms wedges directly impact the labor force participation decisions of married women only. Proportionally speaking, it is the countries that have low female labor force participation that experience greater increases when the norms wedges disappear.

Moreover, the cross-country variance of married female labor force participation drops by 7.8%. Thus, differences in the norms wedge are far from driving all of the differences in female labor force participation across countries, but they do contribute to a non-negligible degree.

Figure 3.6: Female labor force participation with zero norms wedges



Notes: This figure plots the female labor force participation by country. I differentiate between the labor force participation of married women and that of all (married and single) women. The horizontal axis stands for the empirically observed labor force participation rate, whereas the vertical axis stands for the counterfactual rate computed when the homemaker norms wedges disappear for all countries.

3.8 Conclusion

A low level of female labor force participation implies that a large part of human talent rests unrealized. This paper demonstrates that the homemaker norms wedge, which glues

high of 80.8% to Canada in 2011. However, even just taking recent years, there is much divergence across countries. In 2015, the labor force participation rate of Mexico's married women was 34.2%.

married women to the home, exists. Not only that, they are substantial in size in many countries. The homemaker norms wedge that married women experience relative to similar single women amounts to a *half* of the gender norms wedge that women in general experience relative to similar men. Against a backdrop of a large and vigorously growing subfield of economics on various gaps by gender, it calls to attention how important marriage is, in creating gaps amongst individuals of the same gender. Married women do face additional constraints against participating in the labor market relative to single women, and their participation would rise by close to 40% if the homemaker norms wedge were to disappear.

The homemaker norms wedge diverging by ancestry among U.S. immigrants also underscores the fact that culture persists even when economic conditions change. Culture is slow to change, and cultural shifts often lags behind economic progress.[19] Recall that the female labor force participation rate was 6% in Yemen and 84% in Rwanda in 2019. A slow-moving culture implies that it is difficult to expect Yemen to converge to the levels of Rwanda any time soon, even if the two countries were to be placed under the same economic conditions.

Nonetheless, it is not impossible to accelerate cultural change. Cultural change does occur over time within countries (Lee, 2020), and policies aimed at behavioral and attitudinal change have been found to be effectual (La Ferrara, 2016). Policies can indeed work to hasten the adoption of gender-equal attitudes. At the same time, culture is far from explaining all of the cross-country variation in female labor force participation. Even if the homemaker norms wedges are removed everywhere, 92% of the variation would remain. Hence, other factors are also important. Jayachandran (2015) review that policies aimed at improving the earning potential of women, such as school expansion, microfinance, and business skill training, have effectively contributed to raising female labor force participation in developing countries. Therefore, policies that tackle economics and culture should work in tandem in order to achieve higher female labor force participation across the board worldwide.

[19] Although traditional plough cultivation that requires male brute force have long been largely outmoded, descendants of societies that practiced it still have more conservative gender roles today (Boserup, 1970; Alesina, Giuliano, and Nunn, 2013).

3.A Appendix – Data

3.A.1 Harmonizing earnings in the IPUMS International database censuses

The earnings variable is not homogeneous across these countries. Where available, I use the variable *incwage*, which reports the respondent's weekly, monthly, or annual wage and salary income. The next alternative is the variable *incearn*, which reports the respondent's total income from their labor, including wages, business income, and farm income, made in the previous month or year. When neither are available, I use the variable *inctot*, which reports the respondent's total personal income from all sources in the previous month or year. The reason why I utilize the earnings variables in this order is that in the model, the decision to work in the labor market depends on the comparison of potential market earnings and home productivity that one could generate. Therefore, I do not want business profits (particularly negative profits) or other unearned sources of income such as remittances to confound the labor force participation decision. Table 3.A1 lists the census or survey countries and years, and also depicts which earnings variable is used in each sample.

A caveat for all earnings variables for all the countries is that the recorded earnings are before-tax. Of course, countries have different national systems for taxes and benefits, which affects the take-home earnings that really matter for the labor force participation decision. Since to compute the homemaker norms wedge, I compare the wages of married and single women within a country, the fact that some nations have much higher tax rates in general than others would not affect the norms wedge. However, national tax systems would also vary in how differently they treat married individuals from single, which would be included in the norms wedge value.

3.B Appendix – Tables

Table 3.A1: List of census or survey, with earnings variable used

Country	Census/survey year(s)	Earnings variable
Brazil	1980, 1991, 2000, 2010	incearn
Canada	2011	incearn
Colombia	1973	inctot
Dominican Republic	1981	incwage
	2002	inctot
India†	1983, 1987, 1993, 1999, 2004	incwage
Indonesia	1976, 1995	incwage
Israel	1972	incearn
	1995	incwage
Italy†	2011‡	incwage
Jamaica	1982, 1991, 2001	incwage
Mexico	1970	inctot
	1990, 1995, 2000, 2010, 2015	incearn
Panama	1970	incwage
	1980	inctot
	1990, 2000, 2010	incwage
Puerto Rico	1990, 2000, 2005, 2010	incwage
South Africa	1996, 2001, 2007, 2011	inctot
Trinidad and Tobago	1970	inctot
	2000	incwage
United States	1980, 1990, 2000, 2010	incwage
Uruguay	2006	incwage
Venezuela	1971, 1981, 1990, 2001	incearn

Notes: *incwage* reports the respondent's weekly, monthly, or annual wage and salary income; *incearn* reports the respondent's total income from their labor, including wages, business income, and farm income, made in the previous month or year; *inctot* reports the respondent's total personal income from all sources in the previous month or year.

† Surveys

‡ I combined the survey data from years 2011, 2012, ..., 2018, due to the small sample size in each year.

Table 3.A2: Matching IPUMS International censuses to the World Values Survey by the closest year

	IPUMS	WVS		IPUMS	WVS
Brazil	1991	1991	Puerto Rico	1990	1995
	2000	2006		2000	2001
	2010	2014	South Africa	1996	1996
Canada	2011	2006		2001	2001
Colombia	1973	1997		2007	2006
Dominican Republic	2002	1996		2011	2013
India	1987	1990	Trinidad and Tobago	2000	2006
	1993	1995	United States	1990	1995
	1999	2001		2000	1999
	2004	2006		2010	2011
Indonesia	1995	2001	Uruguay	2006	2006
Italy	2011	2005	Venezuela	1990	1996
Mexico	1990	1990		2001	2000
	2000	2000			
	2010	2005			
	2015	2012			

Notes: This table lists how the years of the censuses in the IPUMS International database and the years of the World Values Survey (WVS) are matched. They are matched by the closest available year.

Table 3.A3: Correlation between norms wedges by country and by ancestry

	Norms wedge by country		
	(1)	(2)	(3)
Norms wedge by ancestry	1.084**	1.325***	1.371***
	(3.46)	(5.99)	(6.02)
Cell size restriction, norms wedge by country	>100	>100	-
Cell size restriction, norms wedge by ancestry	>100	-	-
N	9	16	16
R^2	0.444	0.686	0.704

Notes: t statistics from robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ Columns (1) and (2) correspond to the regression results plotted in plots (A) and (B), respectively, of Figure 3.5.

References

Abrevaya, J. (2009). Are there missing girls in the united states? evidence from birth data. *American Economic Journal: Applied Economics 1*(2), 1–34.

Acemoglu, D., D. H. Autor, and D. Lyle (2004). Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury. *Journal of Political Economy 112*(3), 497–551.

Aguiar, M. and E. Hurst (2007a). Life-cycle prices and production. *American Economic Review 97*(5), 1533–1559.

Aguiar, M. and E. Hurst (2007b). Measuring Trends in Leisure: The Allocation of Time Over Five Decades*. *The Quarterly Journal of Economics 122*(3), 969–1006.

Ahern, K. R. and A. K. Dittmar (2012). The Changing of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation *. *The Quarterly Journal of Economics 127*(1), 137–197.

Akerlof, G. A. and R. E. Kranton (2000). Economics and Identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Albanesi, S. and C. Olivetti (2016). Gender roles and medical progress. *Journal of Political Economy 124*(3), 650 – 695.

Alesina, A., P. Giuliano, and N. Nunn (2013). On the Origins of Gender Roles: Women and the Plough. *The Quarterly Journal of Economics 128*(2), 469–530.

Almond, D. and L. Edlund (2008). Son-biased sex ratios in the 2000 united states census. *Proceedings of the National Academy of Sciences 105*(15), 5681–5682.

Andrews, D. W. K. (1999). Estimation when a parameter is on a boundary. *Econometrica 67*(6), 1341–1383.

Angrist, J. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion* (1 ed.). Princeton University Press.

Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager (2019). Synthetic Difference In Differences. Working Paper 25532, National Bureau of Economic Research.

Athey, S., C. Avery, and P. Zemsky (2000). Mentoring and diversity. *American Economic Review 90*(4), 765–786.

Bagues, M. and P. Campa (2017). Can gender quotas in candidate lists empower women? evidence from a regression discontinuity design. IZA Discussion Papers 10888, Institute of Labor Economics (IZA).

Baltrunaite, A., P. Bello, A. Casarico, and P. Profeta (2014). Gender quotas and the quality of politicians. *Journal of Public Economics 118*(C), 62–74.

Bar, M., O. Leukhina, M. Hazan, and H. Zoabi (2018). Why did Rich Families Increase their Fertility? Inequality and Marketization of Child Care. *Journal of Economic Growth 23*, 427–463.

Becker, G. (1981). *A Treatise on the Family.* Cambridge: Harvard University Press. Enlarged Edition, 1991.

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

Bedi, A. S., T. Majilla, and M. Rieger (2018). Gender Norms and the Motherhood Penalty: Experimental Evidence from India. IZA Discussion Papers 11360, Institute of Labor Economics (IZA).

Bertrand, M., E. Kamenica, and J. Pan (2015). Gender Identity and Relative Income within Households. *The Quarterly Journal of Economics 130*(2), 571–614.

Besley, T., O. Folke, T. Persson, and J. Rickne (2017). Gender quotas and the crisis of the mediocre man: Theory and evidence from sweden. *American Economic Review 107*(8), 2204–42.

Bisin, A. and T. Verdier (2000). "Beyond the Melting Pot": Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits*. *The Quarterly Journal of Economics 115*(3), 955–988.

Blau, F. D. and L. M. Kahn (2007). Changes in the labor supply behavior of married women: 1980–2000. *Journal of Labor Economics 25*(3), 393–438.

Boserup, E. (1970). *Women's role in economic development.* London: George Allen and Universe Ltd.

Browning, M., P.-A. Chiappori, and Y. Weiss (2014). *Economics of the Family.* Cambridge University Press.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Case, A. and A. Deaton (2017). Mortality and Morbidity in the 21st Century. *Brookings Papers on Economic Activity, Spring 2017*.

Charness, G. and U. Gneezy (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior Organization 83*(1), 50 – 58. Gender Differences in Risk Aversion and Competition.

Chattopadhyay, R. and E. Duflo (2004). Women as policy makers: Evidence from a randomized policy experiment in india. *Econometrica 72*(5), 1409–1443.

Chiappori, P.-A., M. Costa-Dias, and C. Meghir (2018). The Marriage Market, Labor Supply, and Education Choice. *Journal of Political Economy 126*(S1), 26–72.

Chiappori, P.-A., M. Iyigun, and Y. Weiss (2009). Investment in Schooling and the Marriage Market. *American Economic Review 99*(5), 1689–1713.

Chiappori, P.-A., B. Salanié, and Y. Weiss (2017). Partner Choice, Investment in Children, and the Marital College Premium. *American Economic Review 107*(8), 2109–67.

Cho, H.-O. and E. H. Kim (2010). A Study of Historical Context and Institutionalization Process of Gender Quota System and Women's Political Empowerment Movement in Korea 2000-2010. *Journal of Korean Social Trend and Perspective* (79), 110–139.

Choo, E. and A. Siow (2006). Who Marries Whom and Why. *Journal of Political Economy 114*(1), 175–201.

Clayton, A. (2015). Women's political engagement under quota-mandated female representation: Evidence from a randomized policy experiment. *Comparative Political Studies 48*(3), 333–369.

Coen-Pirani, D., A. León, and S. Lugauer (2010). The effect of household appliances on female labor force participation: Evidence from microdata. *Labour Economics 17*(3), 503–513.

Couprie, H., E. Cudeville, and C. Sofer (2017). Efficiency versus Gender Roles and Stereotypes: An Experiment in Domestic Production. THEMA Working Papers 2017-10, THEMA (THéorie Economique, Modélisation et Applications), Université de Cergy-Pontoise.

Dahlerup, D. and L. Freidenvall (2013). Electoral gender quota systems and their implementation in europe. Technical report, The European Parliament, the Committee on Women's Rights and Gender Equality.

De Paola, M., V. Scoppa, and R. Lombardo (2010). Can gender quotas break down negative stereotypes? evidence from changes in electoral rules. *Journal of Public Economics 94*(5), 344 – 353.

Dinkelman, T. (2011). The effects of rural electrification on employment: New evidence from south africa. *American Economic Review 101*(7), 3078–3108.

Doepke, M., M. Hazan, and Y. D. Maoz (2015). The Baby Boom and World War II: A Macroeconomic Analysis. *Review of Economic Studies 82*(3), 1031–1073.

Eaton, J. and S. Kortum (2004). Technology, Geography, and Trade. *Econometrica 70*(5), 1741–1779.

Erosa, A., L. Fuster, G. Kambourov, and R. Rogerson (2017). Hours, occupations, and gender differences in labor market outcomes. Working Paper 23636, National Bureau of Economic Research.

Esteve-Volart, B. and M. Bagues (2012). Are women pawns in the political game? evidence from elections to the spanish senate. *Journal of Public Economics 96*(3), 387–399.

Fernández, R. (2011). Does culture matter? In J. Benhabib, M. O. Jackson, and A. Bisin (Eds.), *Handbook of Social Economics*, Volume 1A, pp. 481–510. The Netherlands: North-Holland.

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

Fernández, R. and A. Fogli (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American Economic Journal: Macroeconomics 1*(1), 146–77.

Fernández, R., A. Fogli, and C. Olivetti (2004). Mothers and Sons: Preference Formation and Female Labor Force Dynamics*. *The Quarterly Journal of Economics 119*(4), 1249–1299.

Fernández, R. and J. Wong (2014). Unilateral Divorce, the Decreasing Gender Gap, and Married Women's Labor Force Participation. *American Economic Review 104*(5), 342–47.

Ferrara, A. (2019). World War II and African American Socioeconomic Progress. Working paper.

Field, E. M., R. Pande, N. Rigol, S. G. Schaner, and C. T. Moore (2019). On Her Own Account: How Strengthening Women's Financial Control Affects Labor Supply and Gender Norms. Working Paper 26294, National Bureau of Economic Research.

Fogli, A. and L. Veldkamp (2011). Nature or Nurture? Learning and the Geography of Female Labor Force Participation. *Econometrica 79*(4), 1103–1138.

Galor, O. and D. Weil (1996). The gender gap, fertility, and growth. *American Economic Review 86*(3), 374–87.

Goldin, C. (1995). *The U-Shaped Female Labor Force Function in Economic Development and Economic History*, pp. 61–90. University of Chicago Press.

Goldin, C. (2006). The quiet revolution that transformed women's employment, education, and family. *American Economic Review 96*(2), 1–21.

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

Goldin, C. and L. F. Katz (2002). The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions. *Journal of Political Economy 110*(4), 730–770.

Goldin, C. D. (1991). The Role of World War II in the Rise of Women's Employment. *American Economic Review 81*(4), 741–756.

Greenwood, J., N. Guner, G. Kocharkov, and C. Santos (2016). Technology and the Changing Family: A Unified Model of Marriage, Divorce, Educational Attainment, and Married Female Labor-Force Participation, journal = American Economic Journal: Macroeconomics. *8*(1), 1–41.

Greenwood, J., A. Seshadri, and M. Yorukoglu (2005). Engines of Liberation. *The Review of Economic Studies 72*(1), 109–133.

Gronau, R. (1977). Leisure, home production, and work-the theory of the allocation of time revisited. *Journal of Political Economy 85*(6), 1099–1123.

Heath, R. and S. Jayachandran (2017). The causes and consequences of increased female education and labor force participation in developing countries. In S. Averett, L. Argys, and S. Hoffman (Eds.), *The Oxford Handbook of Women and the Economy.*

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019). The Allocation of Talent and U.S. Economic Growth. *Econometrica 87*(5), 1439–1474.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and Manufacturing TFP in China and India*. *The Quarterly Journal of Economics 124*(4), 1403–1448.

International Labor Organization (2019). Labor force participation rate, female (% of female population ages 15+). International Labor Organization Database (ILOSTAT).

Jang, G. S. (2008). Busan Provincial Council gets "Poor Report Card". *"No-cut" News*.

Jayachandran, S. (2015). The roots of gender inequality in developing countries. *Annual Review of Economics 7*(1), 63–88.

Jayachandran, S. and A. Lleras-Muney (2009). Life Expectancy and Human Capital Investments: Evidence from Maternal Mortality Declines*. *The Quarterly Journal of Economics 124*(1), 349–397.

Jeon, J. Y. (2013). Political Dynamics of Adopting Gender Quota in Korea. *Journal of Korean Politics 22*(1), 29–52.

Jin, J. W. (2018). Female nomination subsidy monopolized by two major parties. *Women News*.

Juhn, C. and K. McCue (2016). Selection and Specialization in the Evolution of Marriage Earnings Gaps. *The Russell Sage Journal of the Social Sciences [Working Paper version] 2*(4).

Juhn, C. and K. McCue (2017). Specialization Then and Now: Marriage, Children, and the Gender Earnings Gap across Cohorts. *Journal of Economic Perspectives 31*(1), 183–204.

Kim, S. H. (2005). Local Council Elections - "Reformed" or "Deformed"? *Pressian News*.

Kim, W. H., M. J. Kim, H. C. Lee, and H. K. Kim (2003). *Research on measures to expand female candidate nomination by political parties*. Korea Women's Development Institute.

Kleven, H., C. Landais, and J. E. Søgaard (2019). Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics 11*(4), 181–209.

Kuziemko, I., J. Pan, J. Shen, and E. Washington (2018). The mommy effect: Do women anticipate the employment effects of motherhood? Working Paper 24740, National Bureau of Economic Research.

La Ferrara, E. (2016). Mass Media and Social Change: Can We Use Television to Fight Poverty? *Journal of the European Economic Association 14*(4), 791–827.

Lee, D. S. (2008). Randomized experiments from non-random selection in u.s. house elections. *Journal of Econometrics 142*(2), 675 – 697. The regression discontinuity design: Theory and applications.

Lee, J. E. (2020). Marriage and Misallocation: Evidence from 70 Years of U.S. History. Working paper.

Lee, J. J. (2019). *The Legislative Impact Analysis of Gender Quota in Local Council District Elections*. Number 44. National Assembly Research Service.

Lee, M. S. (2003). Measures to expand female political representation in preparation for the 17^{th} general election. Panel discussion in the Meeting of Female Members of the National Assembly and Municipal Councils To Increase the Participation of Women in Politics.

Lee, U. (1965). The Employment of Negro Troops. In S. Conn (Ed.), *United States Army in World War II*. Washington D.C: Center of Military History U.S. Army.

Lim, H. K. (2018). Women's political representation and the current state of Gyeonggido: Focused on local elections in 2018. (102). Gyeonggido Family and Women Research Institute [Issue Analysis].

Local Decentralization Bureau – Election and Local Council Division (2019). Ministry of the Interior and Safety to Block “Self-Assessment" of Overseas Training for Local Councils and to Redeem Unjust Spending. Ministry of the Interior and Safety.

Lundberg, S. and R. A. Pollak (2007). The american family and family economics. *Journal of Economic Perspectives 21*(2), 3–26.

Mazzocco, M. (2004). Saving, Risk Sharing, and Preferences for Risk. *American Economic Review 94*(4), 1169–1182.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

Mengel, F., J. Sauermann, and U. Zölitz (2018). Gender Bias in Teaching Evaluations. *Journal of the European Economic Association 17*(2), 535–566.

Miller, G. (2010). Contraception as development? new evidence from family planning in colombia. *Economic Journal 120*(545), 709–736.

Ministry of the Interior and Safety (2012). Municipal council operation guide. Publication Registration Number: 11-1311000-000437-01.

Ministry of the Interior and Safety (2018). Municipal councilor's activities guide. Publication Registration Number: 11-1741000-000101-15.

Miyake, M. (2016). Logarithmically homogeneous preferences. *Journal of Mathematical Economics 67*, 1 – 9.

National Election Commission (2018). Election subsidy payments by year (2001-2018). National Election Commission Data Resource Archive.

Ngai, L. R. and B. Petrongolo (2017). Gender Gaps and the Rise of the Service Economy. *American Economic Journal: Macroeconomics 9*(4), 1–44.

Niederle, M. and L. Vesterlund (2007). Do Women Shy Away From Competition? Do Men Compete Too Much?*. *The Quarterly Journal of Economics 122*(3), 1067–1101.

Pitt, M. M., M. R. Rosenzweig, and M. N. Hassan (2012). Human capital investment and the gender division of labor in a brawn-based economy. *American Economic Review 102*(7), 3531–60.

Pollak, R. A. (2013). Allocating Household Time: When Does Efficiency Imply Specialization? NBER Working Papers 19178, National Bureau of Economic Research, Inc.

Restuccia, D. and R. Rogerson (2007). Policy distortions and aggregate productivity with heterogeneous plants. Working Paper 13018, National Bureau of Economic Research.

Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers 3*(2), 135–146.

Shapley, L. and M. Shubik (1971). The Assignment Game I: The Core. *International Journal of Game Theory 1*(1), 111–130.

Tax Foundation (2013). U.S. Federal Individual Income Tax Rates History, 1862-2013 - Nominal Dollars [table].