I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15487

Gender Discrimination and the Sex Ratio of Immigrants

Saul Lach
Nachum Sicherman

AUGUST 2022

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15487

Gender Discrimination and the Sex Ratio of Immigrants

Saul Lach
The Hebrew University and CEPR

Nachum Sicherman
Columbia University and IZA

AUGUST 2022

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Gender Discrimination and the Sex Ratio of Immigrants*

We use data on international migration to study the causal effect of gender discrimination on the sex-ratio of immigrants to the U.S. during the 1970-2019 period. We measure gender discrimination in the countries of origin using the Women, Business, and the Law (WBL) index, which measures legal differences in access to economic opportunities between men and women. Controlling for country fixed effects and regional time trends, as well as for potentially confounding factors, we find that a one standard deviation increase in the WBL index in a country of origin (a decrease in gender discrimination) decreases the share of women immigrating to the U.S. from that country by 1.7 percentage points, on average. This large effect of gender discrimination on the sex ratio of immigrants is robust to specification changes, and is not significant when examining senior citizens.

JEL Classification: F22, J16
Keywords: gender discrimination, sex-ratios, international migration

Corresponding author:
Nachum Sicherman
Graduate School of Business
Columbia University
New York, NY 10027
USA
E-mail: nachum.sicherman@columbia.edu

* We thank Ann Bartel and John Donaldson for comments and suggestions. We also thank Brenden Eum, Jett Pettus, and Saikun Shi for superb research assistance.

1 Introduction

The sex-ratio, the ratio of the number of men to the number of women in a population, has important social and economic implications. For example, Becker (1981) argues that the relative scarcity of women improves their bargaining position in the marriage market, while others have observed that high sex-ratios have an impact on savings, crime and other economic outcomes.[1]

This paper focuses on the sex-ratio of immigrants, which varies over time and across countries.[2] The sex ratio of immigrants has a direct impact on the sex ratio of the population in the countries of origin and of destination. For example, Raphael (2013) finds that the high sex ratio of Mexican emigrants decreased nuptiality rates and increased the proportion of women who never had a child among the women who remained in Mexico, while increasing their educational attainment and employment rates. Conversely, Angrist (2002) shows that the heavily male migration to the U.S. in the first half of the 20th century increased nuptiality rates among the daughters of immigrants, but also diminished their labor force participation.[3]

[1] The dramatic increase of the sex ratio in various East Asian countries prompted Amartya Sen (1990) to warn about the effects of sex-preference in Asia. Such effects were documented in various studies (e.g., Edlund et al. (2013), Wei and Zhang (2011)). See also Abramtizky et al (2011) for an analysis of the shortage of men on marriage market outcomes in post-WWI France.

[2] For example, in the 1860-1920 period, between 60 and 80 percent of immigrants to the U.S., South America, Northwest Europe, and Antipodean Pacific were males (Ferenczi 1929). Also, the percentage of females immigrating to the U.S. varied from 30.4 in the 1900s to 61.2 in the 1940s (Houstoun et.al. 1984, Appendix Table A-1). For more examples, see Donato and Gabaccia (2015).

[3] Panunzio (1942) is an early study showing how the sex-ratio of immigrants affects intermarriage rates among different ethnic groups in Los Angeles during the 1924-33 period. For additional examples and discussion, see Donato and Gabaccia (2015, pages 92-95)

One may speculate as to the factors determining the gender composition of immigrants but, to the best of our knowledge, there is no systematic research into their relevance and importance. For example, the sex-ratio of immigrants may be determined by the population sex-ratio in the country of origin, by the nature of migration itself (e.g., whether it is driven by labor market considerations, war, religious prosecution, famine, etc.), as well as by migration laws that, either explicitly or implicitly, are gender biased.

In this paper we focus on an hitherto understudied factor: gender discrimination in the country of origin. Gender discrimination may not only be a "push" factor in women's decision to emigrate (e.g., through reduced economic opportunities), but it may also restrict their ability to migrate (e.g., through restricted mobility). The net effect of gender discrimination on the sex-ratio of immigrants is, therefore, an empirical question.

We tackle this empirical issue by focusing on immigration to the United States between 1970 and 2019. Specifically, we use the American Community Surveys to identify foreign born individuals along with their gender, country of origin, and year of immigration to the U.S. To measure gender discrimination in the country of origin we rely on an index recently developed by the World Bank, the "Women, Business and the Law" (WBL) index (Hyland, 2020). This index synthesizes legislation relevant to a woman's access to employment and entrepreneurial activity across 190 countries from 1970 until the present.

Controlling for year and country fixed effects, we find a strong positive effect of gender discrimination in a country of origin on the share of women immigrating to the U.S. from that country. This effect is robust to the inclusion of regional trends and other controls, but disappears when analyzing population subgroups where the effect of gender discrimination should not be present (e.g., senior citizens). We argue that the effect is plausibly causal. The estimated effect implies

that a one standard deviation increase in the WBL index, a decrease in gender discrimination, decreases the share of women immigrating to the U.S. by 1.7 percentage points. This large effect can account for the observed decline in this share over the 1970-2019 period. Among the eight components of the WBL index the ones that appear to have the strongest effects are those related to Assets, Entrepreneurship and Marriage legal restrictions.

The paper is organized as follows. The sources of data used in our empirical work are described and analyzed in Section 2. The results of the regression analysis are presented in Section 3. Conclusions close the paper.

2 Data sources and trends

Our empirical analysis relies on two main datasets: the Women, Business and the Law (WBL) dataset produced by the World Bank from which we obtain the index of legal gender discrimination, and the American Community Surveys (ACS) from which we obtain the data on immigration to the U.S. by country of origin, gender and year.

The original set of countries in the WBL dataset comprises 190 countries but the ACSs record immigrants from only 145. Appendix Table A2 lists these 145 countries and the number of surveyed immigrants to the U.S. during the 1970-2019 period. In what follows, we describe the datasets and present descriptive statistics for the subset of 145 countries on which the regression analysis will be based.

2.1 The WBL dataset

The "Women, Business and the Law 2020" dataset measures legal differences in access to economic opportunities between men and women in 190 countries. It is

structured around the life cycle of a working woman.[4] To ensure comparability, a woman in question is assumed to reside in the main business city of her economy and to be employed in the formal sector. Eight indicators were constructed around women's interactions with the law as they begin, progress through and end their careers. The indicators were chosen based on statistically significant associations with outcomes related to women's "economic empowerment", including women's labor force participation rates. The eight indicators are: mobility, workplace, pay, marriage, parenthood, entrepreneurship, assets, and pension. The mobility indicator, for example, assesses laws affecting women's agency and freedom of movement—two factors likely to influence their decision to enter the labor force, as well as their decision to emigrate. Each indicator looks at a specific set of regulations and the ways in which they affect women's economic participation as entrepreneurs and employees.

In total, 35 questions are scored across the eight indicators (see Appendix Table A1). Overall scores are then calculated by taking the simple average of each indicator, with 100 representing the highest possible score. The resulting *WBL index* is an easily replicable way to benchmark the regulatory environment for women as entrepreneurs and employees.[5]

[4]See World Bank (2020). The data can be downloaded from `http://pubdocs.worldbank.org/en/506381582842200909/WBL50YearPanelData2020.xlsx`

[5]There are numerous variables that are likely to be highly correlated with gender discrimination or gender inequality. Such variables include, for example, educational disparities, health disparities (e.g., access to prenatal care), labor market outcomes (e.g., wage gap, presence in managerial positions), access to power (e.g., representation in parliament), and legal disparities.

We are aware of only two other projects that attempt to use a subset of such variables to construct an overall international index of gender inequality or discrimination. The Economist Intelligence Unit (EIU), the research and analysis division of The Economist Group, constructed an index called Women's Economic Opportunity Index (available to download

Hyland et al. (2020) find positive correlations between more equal laws pertaining to women in the workforce – a higher WBL index – and more equal labor market outcomes, such as higher female labor force participation and a smaller wage gap between men and women.

Figure 1 shows that the average WBL index across countries *increases over time* from 48 in 1970 to 77.7 in 2019.[6] This 30 point increase reflects the advances in women's rights achieved during the period, at least as reflected in the law.

---

from https://www.eiu.com/public/topical_report.aspx?campaignid=weoindex2012). The Gender Inequality Index (GII), constructed by the U.N., is another attempt to construct a general measure of gender inequality.

The WBL data that we use in this study is superior to the other alternative measures because it is more general, it is consistent across countries, and covers more countries and years than any of the other indeces.

[6] Weighting the country-specific WBL index by its share of immigrants to the U.S. lowers the level of the index, but this weighted average exhibits the same pattern as the simple average. In fact, the correlation between these two averages of the WBL index is 0.99.

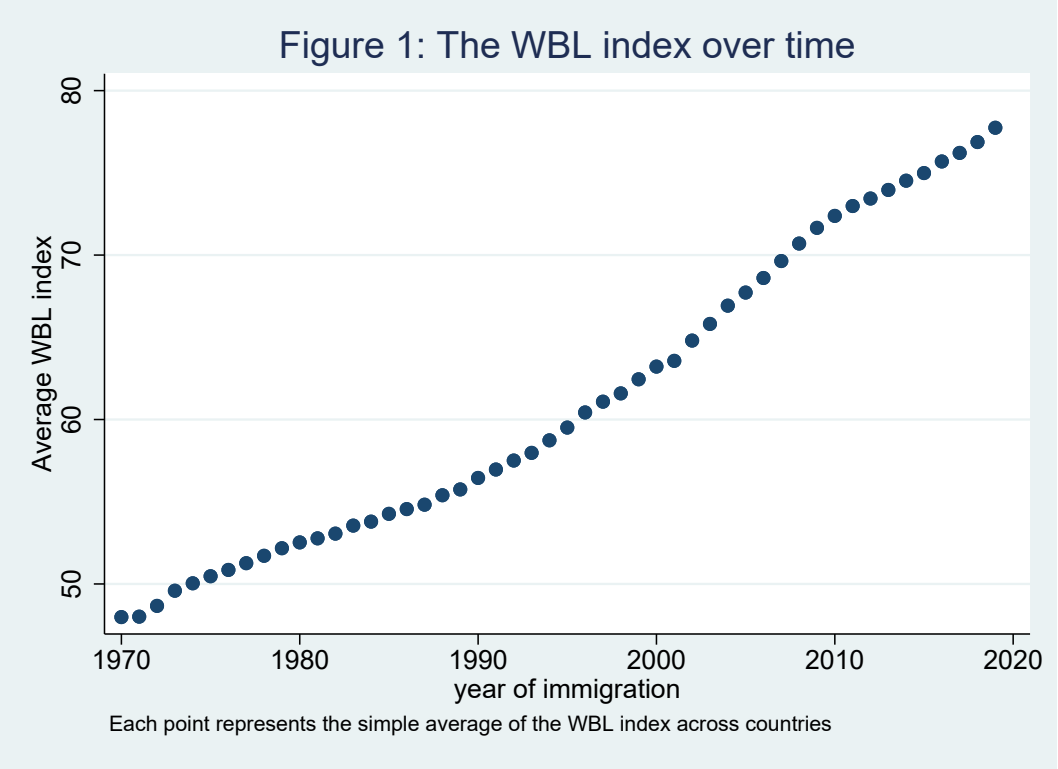
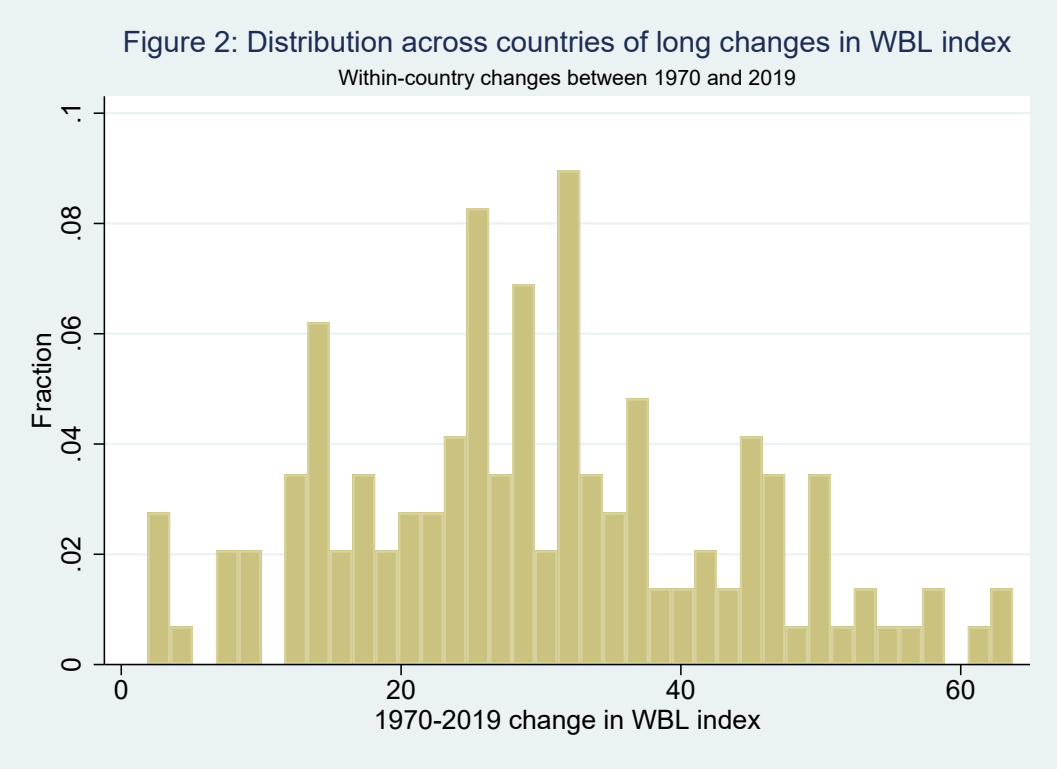


Figure 2 shows that all countries increased their WBL index during the 1970-2019. The average (median) change per country was 29.8 (28.8) points but there is variation across countries in the extent of the change. Some countries increased by just a few points (e.g., by less than 5 points in Iran and Kuwait) while others increased by a considerable amount (e.g., by more that 50 points in Spain, South Africa, Belgium, Togo, among others).



A final caveat is in order. The data set is constructed using laws and regulations that are currently in force. Unless they are codified, religious laws and social conventions are not considered. Because the indicators reflect legal equality of opportunity, they do not necessarily reflect their actual implementation. Moreover, the indicators do not cover legal gender discrimination across all aspects of a woman's life nor do they cover other types of gender discrimination. For brevity, however, we use the term "gender discrimination" to mean discrimination against women of the type captured by the WBL index.

2.2 The American Community Surveys

Ideally, we would like to use the sex-ratio of the outflows of migrants from all origin countries to all destination countries. Unfortunately, the availability of flow data is very limited given the inherent difficulty of defining and measuring immigration

flows directly. This is in stark contrast to the availability of data on the *stock* of immigrants.[7]

The most comprehensive data of international inflows and outflows of migrants is, probably, the "DEMIG C2C database" (DEMIG 2015).[8] We do not use these data mainly because the definition of who is an immigrant and coverage vary across countries and over time. Moreover, in most countries, the assigned timing of immigration does not match the actual time of immigration. This timing inconsistency can potentially obscure the relationship between the decision to migrate and changes in gender discrimination.[9]

We therefore focus on immigration to the U.S. and use readily available survey data to provide consistent estimates of the sex-ratio of immigrants across countries of origin and over time. We note that the U.S. has been the main desti-

[7]By "stock data" we mean estimates (usually based on censuses, population registers, or surveys) of the percentage of foreign born residents at a point in time. The Population Division at the United Nations (UN) publishes data on the *stock* of international migrants by country of origin, gender, and age, for the period 1990-2020, at five year intervals (available at: /urlhttps://www.un.org/development/desa/pd/content/international-migrant-stock).

[8]The UN also publishes data on international migration *flows*, for a limited number of years (per country), over the period 1980-2013 `https://www.un.org/development/desa/pd/data/international-migration-flows`(available at https://www.un.org/development/desa/pd/data/international-migration-flows). Unfortunately, these data are not reported by gender and, therefore, not useful for our purposes.

[9]For example, in the U.S., an immigrant is defined as a foreigner who is admitted for permanent residency ("green card"). Such a definition ignores a large number of immigrants that arrive to the U.S. either illegally or using other types of visas (e.g., employment, students). In addition, the time between arrival to the U.S. and the time of becoming permanent residents can vary dramatically, as evidenced by the IRCA legalization which significantly inflated the inflow numbers from 1989 to 1998.

nation of international migration in recent decades: around 48 million immigrants lived in the U.S. in 2015 which is more than four times the number of immigrants living in the second largest recipient country, the Russian Federation.[10] We are, to some extent, trading off external validity with more accurate data.

The American Community Surveys (ACS) is a nationwide annual survey, conducted by the U.S. Census Bureau, and designed to "provide communities with reliable and timely social, economic, housing, and demographic data every year".[11] It has an annual sample size of about 3.5 million addresses. ACS 1-year estimates are data that have been collected over a 12-month period and are available for geographic areas with at least 65,000 people. The survey collects detailed information about all individuals living in a given address, including country of birth and year of immigration to the U.S.

We use these surveys to identify foreign-born individuals. For each such individual we record his or her gender, age during the survey year, and the year the individual immigrated to the U.S. This allows us to compute the number of women and men immigrating to the U.S. from a given country in a given year (the year of immigration) reported in these surveys. We pool the data over the annual 2000-2019 surveys and compute the share of women immigrating to the U.S. from a given country in a given year. To be precise, let $m_{it\tau}(f_{it\tau})$ be the number of male (female) immigrants from country i in year t obtained from the ACS in calendar year τ. Then $m_{it} = \sum_{\tau} m_{it\tau}$, for $t \leq \tau \leq 2019$ is the number of male immigrants from country i in year t. Similarly for f_{it}.[12] We measure the gender composition

[10] https://www.migrationpolicy.org/programs/data-hub/charts/top-25-destinations-international-migrants

[11] See U.S. Census Bureau (2018) for detailed information about this survey.

[12] For example, the number of women immigrating to the U.S. from Mexico in 1980 is the total number of women born in Mexico and immigrating to the U.S. in 1980 as reported in all

of immigrants (to the U.S.) from country i in year t by the share of women among all immigrants from country i,

$$s_{it} = \frac{f_{it}}{m_{it} + f_{it}}$$

The overall sample size is 6.1 million individuals who migrated to the U.S. between 1911-2019. Because the WBL index started in 1970 we limit our sample to the 5.3 million individuals who reported arriving to the U.S. since 1970.

Figure 3 shows the share of women immigrating to the U.S. between 1970 and 2019 (for various age groups). The figure shows a clear decline over time, from an average of 52-54 percent in the early 1970s to an average of 50-51 percent after 2010.[13] As the graph shows, this trend is not monotonic over time, as there are marked fluctuations during the 1985-2000 period. Overall, this is a significant change in the sex ratio of immigrants to the U.S.

To give some perspective on the magnitude of this decline, consider the concern raised by Nobel Laureate Amartya Sen (1990) about the devastating effects of sex-preference (for males) in Asia. His calculations, that "more than 100 million women are missing", were based on comparing sex-ratios at birth for different countries in Asia to the rest of the world. The differences between the Asian countries and the rest of the world were around 2-3 percentage points.[14]

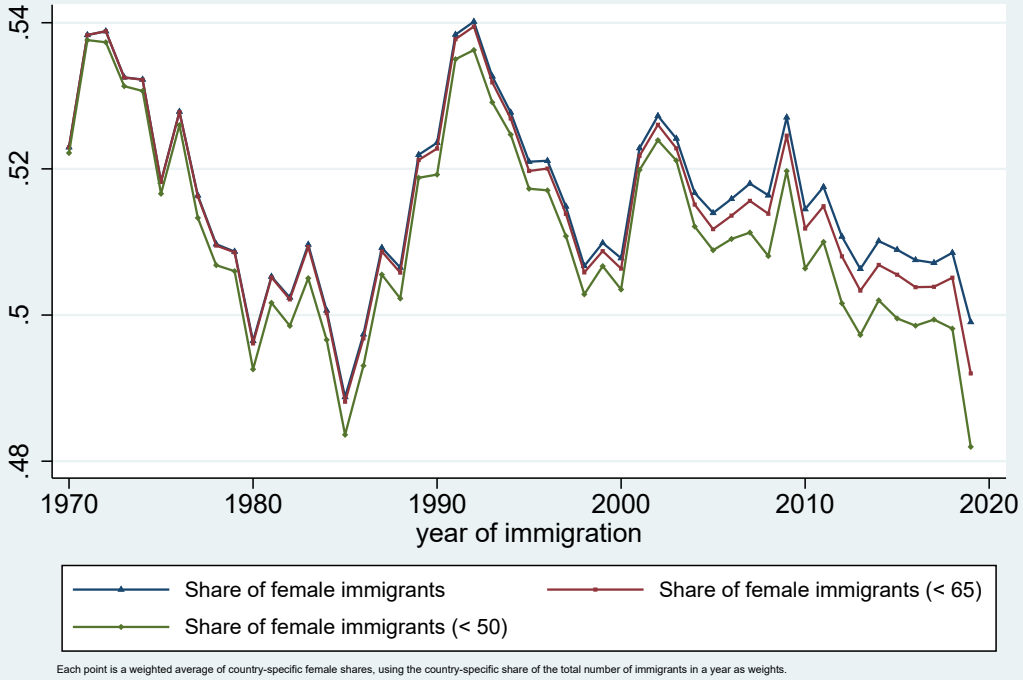
---

post-1980 annual surveys.

[13] The data points are weighted averages across countries, using the share of immigrants as weights. The decline is even stronger if one considers the 1960s when this share hovered around 56 percent.

[14] For example, the sex ratio (males over females) at birth in China peaked in 2005 at 1.186, which translates into a percentage of males out of the total population of 54.2%. https://www.unicef.cn/en/figure-19-sex-ratio-birth-19822017

Figure 3: Share of women immigrating to the US over time



Each point is a weighted average of country-specific female shares, using the country-specific share of the total number of immigrants in a year as weights.

Two remarks are in order. First, because we use annual survey data, we cannot estimate the *total number* of immigrants (by gender), m_{it} and f_{it}. This is the main reason we focus on the sex-ratio of immigrants and not on the level of female (male) immigration. Unfortunately, this does not allow us to determine whether the effect of gender discrimination works through changes in f_{it} or in m_{it} or in both. We remark, however, that our theoretical framework suggests that stronger gender discrimination either increases or decrease f_{it}. We have not considered the case that m_{it} is also affected by gender discrimination, although this is also a possibility. In any case, the estimated effects will show the net effect of gender discrimination on s_{it}, reflecting both women's and men's responses (if any).

Second, pooling data over annual surveys, may introduce a bias towards higher female shares. The earlier the year of immigration t is, the larger the

measured total number of immigrants will be because we pool data from more annual surveys. In contrast, the earlier t is, the likelihood that an immigrant died before the survey year increases, reducing the number of immigrants that are sampled.

The first factor – a larger number of survey years – is likely to apply to both genders equally, while the second factor – increased likelihood of death – may vary by gender. Specifically, if women live on average longer than men, then our estimates of the share of women by year of immigration will be upward biased. This bias, however, is larger the earlier the year of immigration because men are more likely to die by the survey year than women. This could be part of the reason for the observed declining trend in Figure 3.

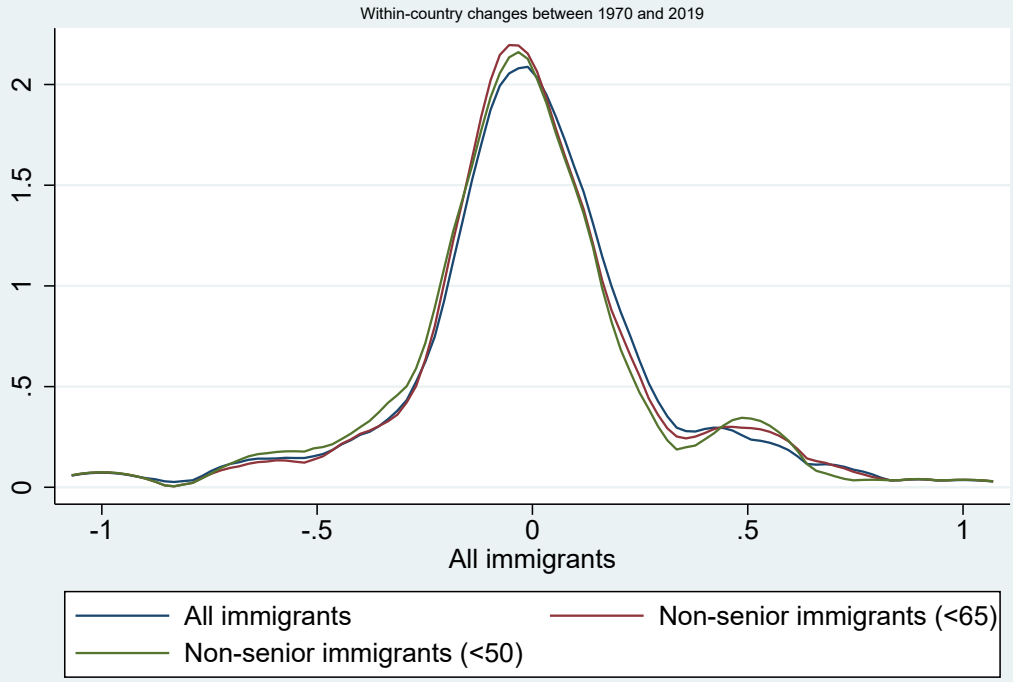
In addition, this bias is likely to increase with age at migration. We therefore compute the share of women immigrants among non-senior immigrants, defined as individuals who were less than 65 or 50 years of age when they arrived in the U.S. If gender differences in mortality rates among senior immigrants are larger than among non-senior ones, we should expect the share of women immigrants to be smaller among non-seniors than among *all* immigrants. This is indeed the case as shown by the red squares (<65) and green rhomboids (<50) in Figure 3: the younger the subpopulation of immigrants considered, the lower the share of women because the effect of their longer longevity is less important. Note, however, that no matter how the shares are computed, they all exhibit the same declining pattern over time.

Moreover, the distribution of the "long" changes (between 1970 and 2019) in the share of women immigrating to the U.S. is not very sensitive to the choice of subpopulation as seen in the similarity of the three density functions in Figure 4. This is important because in the empirical analysis we will be using within-country changes over time. Indeed, the pairwise correlations across countries in these long

changes between the three shares considered is above 0.96 .

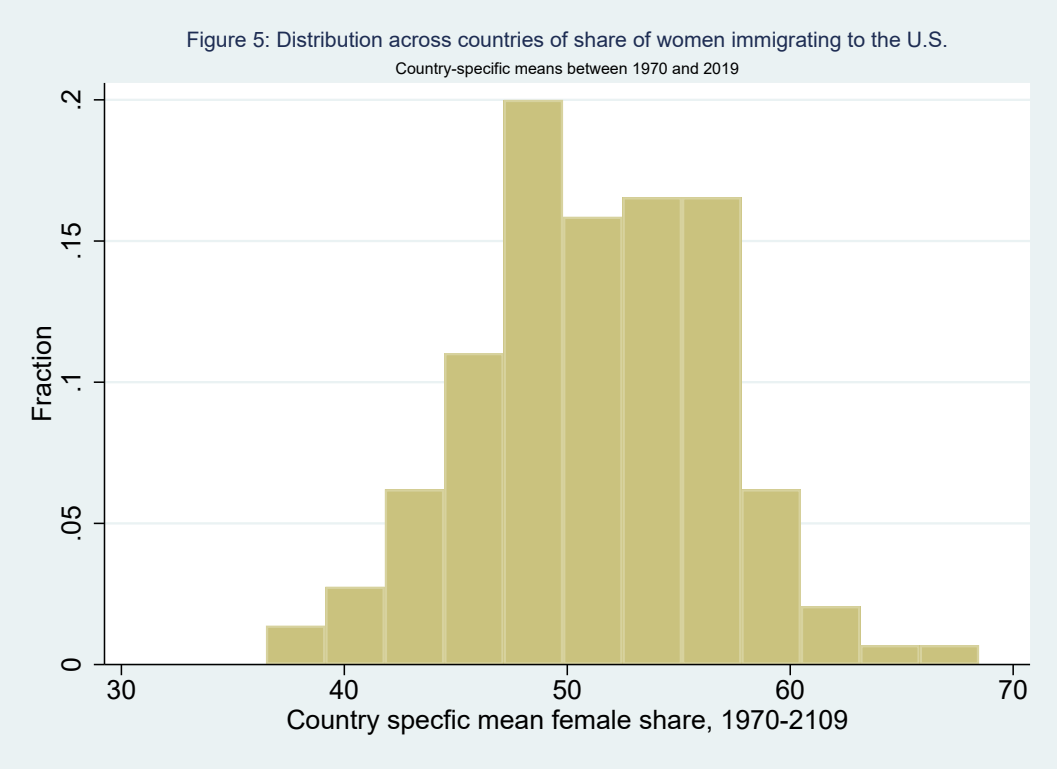
In sum, although we cannot use the total immigration numbers obtained from the ACSs, the female shares, s_{it}, can be consistently estimated from the survey data and changes over time in these share do not seem to be affected by the type of biases discussed above.

Figure 4: Distribution acoss countries of long changes in share of women immigrating to the U.S.



Finally, we observe some variation across countries of origin in the share of women immigrating to the U.S. The country-specific mean shares (computed over the 50 years), $\overline{s}_i = \frac{1}{50}\sum_{t=1}^{50} s_{it}$, averages to 0.514 with a standard deviation of 0.054 across countries. Figure 5 displays the histogram of country-specific average shares.[15]

[15] For countries with very small number of immigrants the observed female share could depart considerably from their "true" female share. We address this issue in the regression analysis



2.3 Other datasets

We use several variables from other datasets as controls in our regression analyses and in various robustness checks. These variables, measured at the country of origin, are: the sex-ratio, GDP per capita, the incidence of armed conflict, divorce rates, and population size. The reasoning for including each of these variables is discussed below, and a detailed description of the variables' construction is available in Appendix A.

---

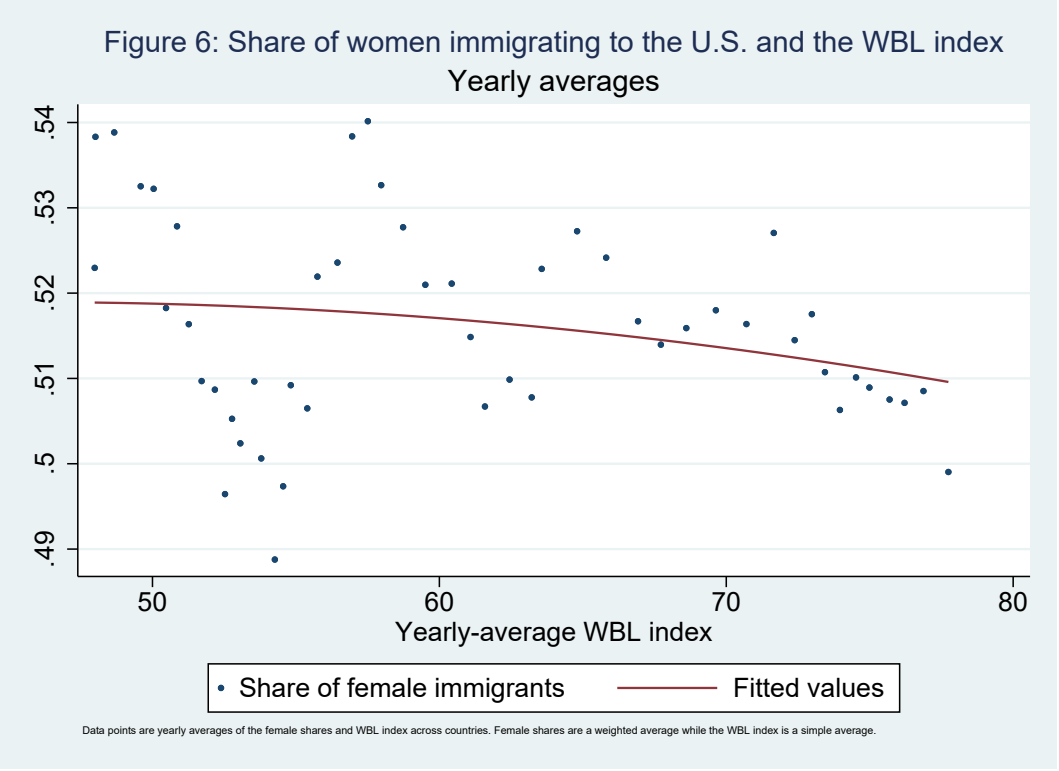
by weighting each country-year observation by its number of immigrants so that countries with lower number of immigrants have a smaller effect on the estimates.

2.4 Trends in the data

We focus on the effect of gender discrimination, as reflected by the legal restrictions that affect the ability of women to enter the labor force and/or to engage in entrepreneurial activity, on the gender composition of migration.

In Figure 6 we plot the yearly average of the share of women immigrating to the U.S. against the yearly average of the WBL index (i.e., averages across countries for each year). Given that the WBL index trends upwards (Figure 1), reflecting the worldwide decrease in gender discrimination, and that the average s_{it} trends downwards (Figure 3) it is not surprising that the average share of women immigrating to the U.S. and the average WBL index are *negatively correlated over time* as shown in Figure 6. This correlation, however, is not very strong, being equal to -0.24 over the 50 years.[16]

[16] Restricting to non-senior immigrants increases this correlation to -0.33 and -0.40 for those above 65 or above 50 years old, respectively.

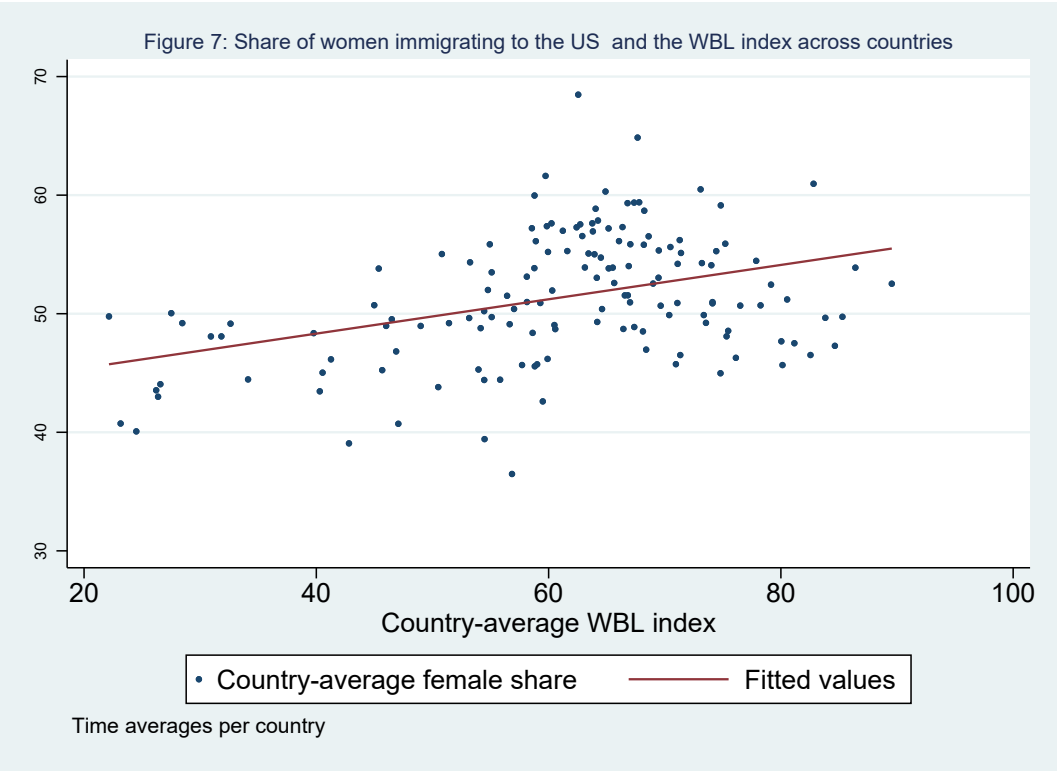


In Figure 7 we plot the country-specific mean shares $\overline{s}_i$, displayed in Figure 5, against the country-specific average WBL index (computed over the 50 years period). Interestingly, these averages are *positively correlated* (correlation coefficient is 0.38). Countries with less discrimination against women, a higher WBL index, are associated with a higher share of women among its emigrants to the U.S.[17]

This positive correlation disappears (correlation coefficient is -0.05), however, when we examine the long changes between 1970 and 2019 in these variables. This suggests that there may be country-specific but time-invariant factors affecting the gender composition of a country's emigrants to the U.S. and the extent to which there is legal discrimination against women (e.g., social norms). In fact, the

[17] The correlation between the country-specific average WBL index and the share of non-senior women is also 0.38.

correlation between s_{it} and WBL_{it} over all countries and years is just 0.15.[18]



In sum, we observe a negative correlation at the aggregate level: over time the WBL index and the share of women immigrating to the U.S. move in opposite directions. On the other hand, we also observe a positive association between the country-specific average WBL index and the share of women emigrating to the U.S. Overall, the raw data do not show strong evidence in support of the hypothesis that gender discrimination is a push factor in the migration decision.

These correlations (or the lack of) may be masking differences across countries in observed and unobserved time-invariant factors affecting the gender composition of immigrants and the extent of gender discrimination in the countries of

[18]The correlations when using the share of women among seniors is 0.14 in both cases (above 50 and 65 years old).

origin. They may also reflect observed and unobserved factors that are trending over time that may generate spurious correlations between the WBL index and the share of women immigrating to the U.S. These considerations are particularly relevant in our data because of the long sample period analyzed and the large heterogeneity across countries. In order to account for these effects we turn to regression analysis.

3 Regression results

The empirical model relates the share of females immigrating to the U.S. from country i in year t, s_{it}, to the WBL index in country i in year t, WBL_{it}, and other controls x_{it}. The basic specification is

$$s_{it} = \beta_0 + \lambda_i + f_i(t) + \beta WBL_{it} + x_{it}\delta + u_{it} \tag{1}$$

where λ_i is a country of origin fixed effect, $f_i(t)$ is a trend function that will be either year fixed effects or year fixed effects interacted with regional dummies.

Neoclassical economic theory assumes that individuals choose their country of residence to maximize their well-being (Borjas, 1989). The migration decision is determined by comparing (expected) opportunities across countries, net of migration costs.[19] Mincer (1978) extended the model to the family, where the net family gain rather than the net personal gain motivates migration.

For women, gender discrimination is likely to reduce the expected benefits of staying in a country, thereby increasing their likelihood of migration relative to men. In other words, gender discrimination "pushes" women to migrate. This

[19] For a model that focus on earning differentials, see Borjas (1994). This framework does not consider "return migration" where migration "may have been planned as part of an optimal life-cycle residential location sequence" (Borjas and Bratsberg 1996). For a theoretical framework of mobility as part of a planned sequence, see Sicherman and Galor (1990).

argument implies $\beta < 0$ in (1). In the case of family migration, when parents consider the future welfare of their children, the gender composition of the children is also likely to affect the family decision to migrate.

On the other hand, gender discrimination may increase the cost of migration for women, thereby decreasing their likelihood of migration relative to men. This could be the case if gender discrimination is associated with financial constraints, lack of freedom and independence, or other explicit and implicit laws, norms and expectations that limit women's ability to migrate. This argument implies $\beta > 0$ in (1).

The net effect of gender discrimination on the sex-ratios of immigrants depends, therefore, on the relative importance of these opposite forces. It is therefore an empirical question. The goal of the empirical analysis is to estimate β.

The main challenge for a causal interpretation of the estimated effects from a regression such as (1) is the existence of omitted variables correlated with the WBL index. It is, however, not obvious how the gender composition of immigrants to the U.S. is determined and, to the best of our knowledge, there is no literature to guide us in selecting the appropriate controls.[20]

We know, however, that gender discrimination is correlated with economic activity and this surely affects the level of migration and *might* also affect its gender composition via changes in the structure of production favoring one gender over the other (e.g., the share of agriculture/manufacturing/services). Because of lack of worldwide consistent data on the structure of production we use log GDP

[20]Reverse causality (simultaneity) is also a possibility if, for example, governments introduce gender-equality legislations in response to increasing relative number of women leaving their countries of origin. We are not aware of evidence supporting this direction of causation but, even if it exists, the implied positive correlation between WBL and u will tend to bias the estimated coefficients upwards. Because, as we will see below, the estimated β is significantly negative simultaneity, if it exists, weakens the estimated effects.

per capita to control for both the level of economic activity and its distribution across sectors.

Because overall sex-ratios vary across countries and over time, we control for the share of women in the country of origin in the regressions. If emigrants are randomly selected, their gender composition will equal the sex ratio in the origin country. If the sex ratio in a country is correlated with gender discrimination, including this control variable avoids the mechanical link between the sex ratio in the country of origin and that of its emigrants to be attributed to the WBL index.

Another variable that could affect the well-being of men and women differently and, therefore, the sex-ratio of emigrants, is the occurrence of armed conflicts. It is not clear, a priori, whether the adverse effect of an armed conflict should be stronger for men or for women. One reason why men should be affected more is that, in almost all countries, soldiers, and especially those involved in combat, are men. However, depending on the nature and location of the conflict, women could be more adversely affected by armed conflicts than men. Ormhaug (2009), for example, finds that, in general, men are more likely to die during conflicts, whereas women die more often of indirect causes after the conflict is over. We therefore include in the regressions an indicator for armed conflict in a country-year, as defined by the Uppsala Conflict Data Program (see Appendix A). Doing this avoids a possible omitted variable bias if armed conflicts do impact the gender composition of migrants and are also correlated with gender discrimination.

The last control variable is population size in the country of origin (in logs). Again, the gender composition of a country's emigrants may be correlated with the country's size and so may be the extent of gender discrimination.

In addition, the presence of fixed country effects goes a long way towards capturing additional permanent features of a country affecting both gender discrimination and the gender composition of its emigrants. For example, long-established

norms about attitudes towards women and their role in society, may affect the decision to emigrate differentially across gender. Moreover, these norms may also be related to the occurrence and prevalence of gender discrimination in the country.[21] Thus, we use within-country changes in gender discrimination to identify its effect on the share of women immigrating to the U.S. from that country.

In sum, for a causal interpretation of our estimates we assume that, conditional on the controls and the presence of country and year fixed effects, the remaining variation in the gender composition of immigrants to the U.S. (across countries and over time) is driven by factors unrelated to gender discrimination. We believe this identifying assumption is plausible in the present context because it is difficult to come up with other potential factors that may affect the gender composition of emigrants and be correlated with gender discrimination that are not outcomes of gender discrimination.[22]

Table 1 starts with a simple OLS regression of s_{it} on the WBL index only.[23]

[21] For example, for many years young Filipino women emigrate to take care of older people (e.g., in Israel) or to attend nursing school. Thus, the gender composition from the Phillipines has a permanent component reflecting these norms.

[22] A case that comes to mind is divorce. The easier women can get divorced, the easier they can emigrate out of the country. But divorce rates can be considered an outcome of gender discrimination. In fact, the WBL index includes two questions about this: "Can a woman obtain a judgment of divorce in the same way as a man?" And, "Does a woman have the same rights to remarry as a man?" For these reasons we do not include divorce rates among the controls. See, however, the robustness checks in Section 3.1.

[23] Because the dependent variable is an average over the total number of immigrants from country i in year t, $m_{it} + f_{it}$, we weight observations by the total number of immigrants from country i. The dependent variable is s_{it} multiplied by 100 for easier interpretation: a one point change in the WBL index changes the share of women immigrating to the U.S. by β percentage points.

The coefficient in column 1 is very small and insignificantly different from zero, reflecting the lack of a measurable relationship mentioned at the end of Section 2. Standard errors account for two-way clustering at the country and year levels, thereby allowing for arbitrary serial correlation within countries as well as contemporaneous correlations across countries. Adding year dummies does not change this picture (column 2) but adding country dummies does (column 3), which is not surprising in light of the findings in Figure 7. The estimate of β increases to -0.13 and is significantly different from zero. The R^2 also increases considerably when country fixed effects are added reflecting the heterogeneity in immigrants' sex ratios across countries of origin observed in Figure 5. Adding the control variable GDP per-capita (column 4), the percentage of women in the country of origin (column 5), an indicator for the occurrence of an armed conflict (column 6), and (log) population in the country of origin (column 7), and all these controls together (column 8) does not affect the point estimate of β. In fact, except for population, these controls are not significantly different from zero either individually or jointly, reinforcing our belief that omitted variable bias is not a serious problem when the goal is estimating the effect of gender discrimination on the gender composition of migration.

A reasonable concern about model (1) is that unobserved trends in a country of origin are driving both the increase in its WBL index and in the gender composition of its emigrants over time. Global trends are flexibly accounted for by the year dummies while the control variables used, especially GDP per-capita and population size, should account for part of the country-specific trends.

Table 1: The effect of gender discrimination on women

	Dep. Var: share of women immigrating to the US, 1970-2019								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9) **baseline**
WBL index	0.000433	-0.00158	-0.130***	-0.134***	-0.129***	-0.129***	-0.119***	-0.121***	-0.167***
	(0.0309)	(0.0461)	(0.0357)	(0.0353)	(0.0359)	(0.0363)	(0.0315)	(0.0320)	(0.0439)
Observations	7,134	7,134	7,134	6,855	7,134	7,134	7,134	6,855	6,855
R-squared	0.000	0.039	0.676	0.680	0.676	0.676	0.682	0.689	0.783
Controls	**None**	**None**	**None**	**GDP**	**% women in origin**	**Conflict**	**Population**	**All**	**All**
Fixed effects	**None**	**Year**	**Country & Year**	**Country & Year**	**Country & Year**	**Country & Year**	**Country & Year**	**Country & Year**	**Country & 16 Regional Trends**

Standard errors in parentheses are two-way clustered at the country and year levels.
All controls inlcude log GDP per-capita, a conflict indicator, the share of women and log population in the country of origin.
Observations weighted by the number of immigrants from each country of origin.
*** p<0.01, ** p<0.05, * p<0.1

We can even be more flexible by allowing for differences in regional trends. We do this by first defining regions and then adding year dummies interacted with the regional indicators to model (1). This is a fairly flexible way of accounting for unobserved regional trends in the gender composition of immigrants to the U.S. Regions are defined by the coding of the countries of origin in the ACS. Specifically, we group countries into 16 regions according to a common first and second digit in the country codes.[24] Results for the model with regional trends appear in column

[24]This procedure results in 20 regions but three of these regions are composed of a single country, Canada, Mexico and Cape Verde, which were reassigned to the UK/Ireland, Central America and Africa regions, respectively. Also the islands in the Pacific were assigned to the

9. The estimated effect of the WBL index becomes even stronger. Accounting for regional trends does not affect the qualitative effect of gender discrimination on the gender composition of migration, thereby lending support to a causal interpretation of this effect. We adopt the specification in column 9 as our baseline model.[25]

A one standard deviation increase in the WBL index (about 10 index points) is associated with a 1.7 percentage point decrease in the share of women immigrating to the U.S.[26] This is a very large effect. Given that the change in the average WBL index over the 50 years between 1970 and 2019 was about 30 points, this effect accounts for all, and more, of the decline in the observed average share of women immigrating to the U.S. If the estimated effect admits a causal interpretation, we conclude that gender discrimination in origin countries is an important determinant of the sex ratio of immigrants to the U.S.

3.1 Robustness checks

In Section 2.2 we remarked that the share of female immigrating to the U.S. may be upward biased and that this bias diminishes as we consider younger immigrants. We therefore start our robustness checks in Table 2 by changing the dependent variable to the share of women among non-seniors: individuals less than 65 years old (column 1) or individuals less than 50 years old (column 2) at the time of immigration to the U.S. The estimated effects are almost identical to that in the baseline model.

Australia/New Zealand regions.

[25] In this baseline specification none of the control regressors, including population, is statistically significant either individually or jointly.

[26] We computed the standard devation of the WBL index for each country (over time). The mean of these standard deviations (across countries) is 10.1 (or 10.8 if weighted by immigration shares).

Table 2: Robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<65	**<50**	**quadratic**	**excl. Mexico**	**excl. former USSR**	**excl. Eastern Europe**	**excl. small countries**
WBL index	-0.169***	-0.171***	-0.500***	-0.149***	-0.169***	-0.168***	-0.167***
	(0.0450)	(0.0465)	(0.179)	(0.0444)	(0.0442)	(0.0452)	(0.0444)
$(\text{WBL index})^2$			0.00275**				
			(0.00127)				
Observations	6,850	6,848	6,855	6,805	6,473	6,213	6,154
R-squared	0.781	0.775	0.785	0.751	0.782	0.791	0.791
Controls	All	All	All	All	All	All	All

All regressions include fixed country effects and years dummies intercated with 16 regional dummies.
Standard errors in parentheses are two-way clustered at the country and year levels.
All controls inlcude log GDP per-capita, a conflict indicator, the share of women and log population in the country of origin.
"Small countries" means country-observations with less that 15 immigrants to the U.S. per year.
Observations weighted by the number of immigrants from each country of origin.
*** p<0.01, ** p<0.05, * p<0.1

Column 3 examines the role of nonlinearities by adding a quadratic of the WBL regressor. Although the quadratic term is significantly different from zero at the 3.6 percent level, the marginal effect of a change in WBL averaged over the data values is 0.151 which is very close to the effect estimated in the baseline linear model (column 9 in Table 1). The linear model delivers the correct effect on average. There is, however, evidence of decreasing effects (in an absolute sense) as the WBL index increases. The marginal effect evaluated at the first quartile of WBL (50) is -0.22 (se 0.06), at the median (61) it is -0.16 (se 0.04), and at the third quartile of WBL (73) it is -0.10 (se 0.04). Adding a cubic term does not

change the estimated marginal effects (results not reported).

Mexico is an outlier in terms of the number of immigrants it sends to the U.S. The average annual flow of immigrants from Mexico is 4.7 times larger than that from the second largest country, the Philippines. Because Mexico receives a very large weight in the estimation procedure one may worry that this country is driving the results. This is a concern if the effect of gender discrimination vary significantly across countries. In column 4 we re-estimate the model excluding Mexico. Although the estimated effect of the WBL index declines slightly it is still of the same order of magnitude as the one obtained when Mexico is included.

The next three columns present estimates over different subsamples of countries. One may suspect that immigration from former communist countries, which are included in the sample only after 1990-91 due to missing GDP data before that, is affected differently by gender discrimination and their inclusion in the sample may be driving the estimated effect. In column 5 we exclude the 13 former USSR countries which became independent around 1991, while in column 6 we exclude 12 Eastern European countries.[27] The estimated effects are essentially the same as in the baseline model allaying concerns about effect heterogeneity.

Countries exhibiting the largest annual changes in the share of women emigrating to the U.S. are usually the smallest countries in terms of the total number of immigrants to the U.S. In fact, all the annual changes in s_{it} above 50 points (in absolute value) occur in country-years with less than 15 immigrants.[28] Because we

[27] In column 5 we exclude Armenia, Azerbaijan, Byelorussia, Estonia, Kazakhstan, Kirghizia, Latvia, Lithuania, Moldavia, Other USSR/Russia, Republic of Ukraine and Uzbekistan. In column 6 we exclude Bosnia, Bulgaria, Croatia, Czech Republic, Hungary, Kosovo, Montenegro, Northern Macedonia, Poland, Romania, Serbia and Slovakia.

[28] Except Estonia which had 20 immigrants in 1989 and a change in s_{it} between 1988 and 1989 of 53.3 points.

use the within-country variation in estimation there may be a concern regarding the effect of these extreme observations. Recall, however, that we weight each country-year observation by its number of immigrants so that countries with lower number of immigrants have a smaller effect on the estimates. For completeness, however, in column 7, we exclude 701 country-year observations with less than 15 immigrants. The estimated effect of gender discrimination remains robust to this change in sample composition.

There are many cases in which the WBL index does not change from one year to the next. In fact, in most years the index does not change. Excluding these "zero observations" from the sample reduces the sample size considerably to 1010 country-year observations over 144 countries. The estimate effect of WBL in this subsample is somewhat smaller at -0.117 (se 0.055) but still significantly different from zero at the 5 percent significant level.

Another variable that could possibly be correlated with women's decision to migrate is the easiness by which a divorce can be obtained in a country, although it is not a-priori obvious in which direction this effect works. Moreover, the easiness of obtaining a divorce is part of the overall WBL index and, as mentioned above, may directly affect divorce rates in a country. In fact, we did not include divorce rates as a control variable precisely because they should be considered an outcome of gender discrimination. There is, nevertheless, an interest in checking whether variation in the WBL index, controlling for variation in divorce rates, still has an effect on the gender composition of migration. There are partial data available on divorce rates over countries and times. In the UN database, we observe at most 5 years of data at roughly 10 year intervals, but this varies a lot across countries and some countries have no data at all (e.g., Argentina). We linearly interpolate the divorce rate data (but do not extrapolate) and end up with a sample of 2,859 observations covering 97 out of the 145 countries. Adding the divorce data to

the baseline specification reduces the estimated effect of the WBL index to -0.149 (se 0.059), while the coefficient of the divorce rate is negative and significantly different from zero. Thus, even when shutting down a channel through which gender discrimination may affect the gender composition of migration, the effect remains of a similar order of magnitude.[29] In subsection 3.5 we address the effect of each of the eight components of the WBL index on the share of women immigrating to the U.S.

The final robustness check addresses the nature of the dependent variable. Because the dependent variable s_{it} represents the average of an underlying gender indicator variable, we used the fractional regression model developed by Papke and Wooldridge (1996) to estimate the parameters of its conditional mean. The estimated marginal effect averaged over the data values is -0.00167 with a standard error of 0.000378 which is exactly the same effect in the baseline model.[30]

Overall, these checks reveal that the estimated effect of gender discrimination in Table 1 is robust to different departures from the baseline model. In the next subsections we examine other extensions.

3.2 Dynamic effects

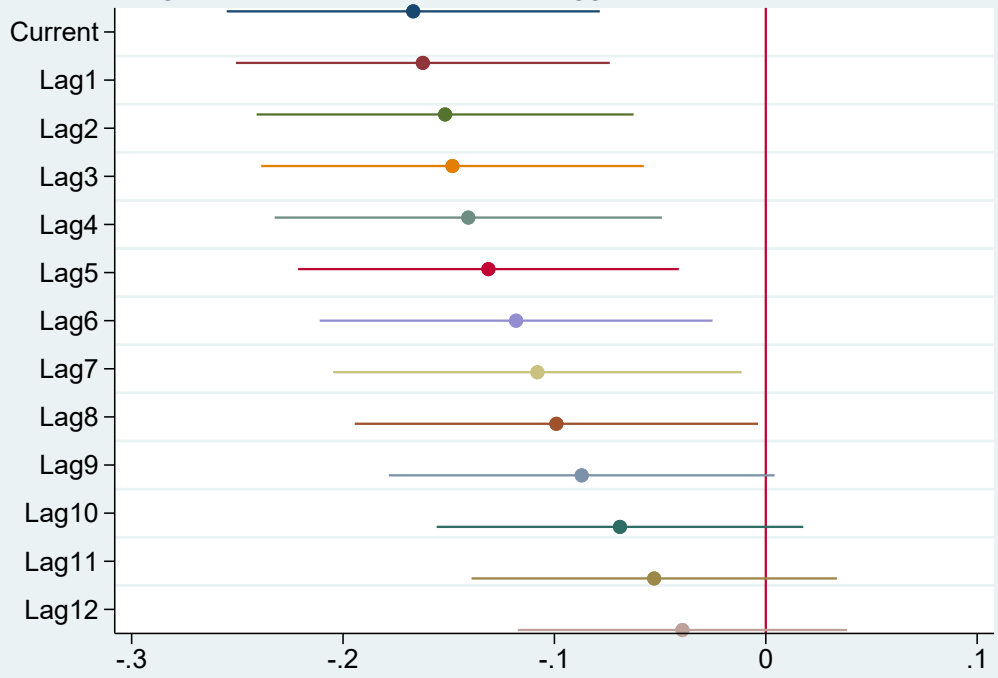
We have been using contemporaneous values of the WBL index but, of course, its effects on the gender composition of migration can be long-lasting. These dynamic effects could be estimated by adding lags of WBL to the baseline model. The problem is that the slow changing nature of the WBL index implies that it

[29] We remark that the estimated effect of the WBL index for this smaller subsample when excluding the divorce rate is -0.186 (se 0.078).

[30] Recall that in the OLS regressions we use $100s_{it}$ as the dependent variable. Results available upon request. The same estimates are obtained using a probit or a logit link function. Standard errors clustered at the country level only.

is very strongly autocorrelated thereby not leaving enough independent variation in the lagged values to estimate their effect.[31] What we can do is to estimate the model using only one lag of WBL at a time. Estimating the baseline model using lags of the WBL index, one at a time, gives the coefficients and 95% confidence intervals plotted in Figure 8.

Figure 8: Coefficients and CI of lagged WBL index, one at a time



Not surprisingly, the effect of WBL is long-lasting over time. This effect declines slowly as time goes by from a current effect of -0.167 to a lagged effect from 8 years ago of -0.10. For longer lags, the WBL effect is not significantly different from zero.

[31] In fact, regressing WBL on lagged WBL, country fixed effects and 16 regional trends gives an autoregressive coefficient of 0.93 and R^2 of 0.99. Adding four additional lags does not (cannot) improve the fit (the additional lagged coefficients are very close to zero).

3.3 Effects by age group

Economic discrimination against women of the type measured by the WBL index should not have much of an effect on women who are not, or not expected to, be engaged in economic activities. This suggests using the gender composition of senior immigrants as a placebo test because we do not expect the WBL index to have a significant effect on the share of senior women immigrating to the U.S.

In Table 3 we show results of estimating equation (1) for various age groups: children (0-17), adults (18+) and seniors (50+ and 65+).

Table 3: The effect of gender discrimination by age group

	(1) children	(2) adults	(3) seniors(50+)	(4) seniors(65+)
WBL index	-0.122**	-0.193***	-0.0661	-0.0179
	(0.0522)	(0.0439)	(0.0463)	(0.0751)
Observations	6,707	6,803	5,544	3,690
R-squared	0.559	0.792	0.520	0.428
Controls	All	All	All	All

All regressions include fixed country effects and years dummies intercated with 16 regional dummies.
Standard errors in parentheses are two-way clustered at the country and year levels.
Observations weighted by the number of immigrants from each country of origin in the appropriate age group.
All controls inlcude GDP per-capita, a conflict indicator and the share of women and population in the country of origin.
Children are immigrants 1-17 years old, adults are 18+ years old, and seniors are either above 50
or above 65 years old.
*** p<0.01, ** p<0.05, * p<0.1

The estimated coefficients for seniors in columns 3 and 4 are small and not significantly different from zero. These results are consistent with the notion that

the factors captured by the WBL index cease to have a gender-differential effect on the decision of senior people to emigrate to the U.S. This placebo test strengthens the case for a causal interpretation of the estimated effect of gender discrimination in the baseline model.[32]

The estimated coefficients for the other age groups, in columns 1 and 2, are in line with the estimated effect in the baseline model. It is interesting to note that gender discrimination also appears to have a significant effect on the sex ratio of immigrant children, albeit a bit smaller than for adults, which is consistent with Mincer's (1978) model of family migration where the future welfare of children affects the family's decision to migrate.[33]

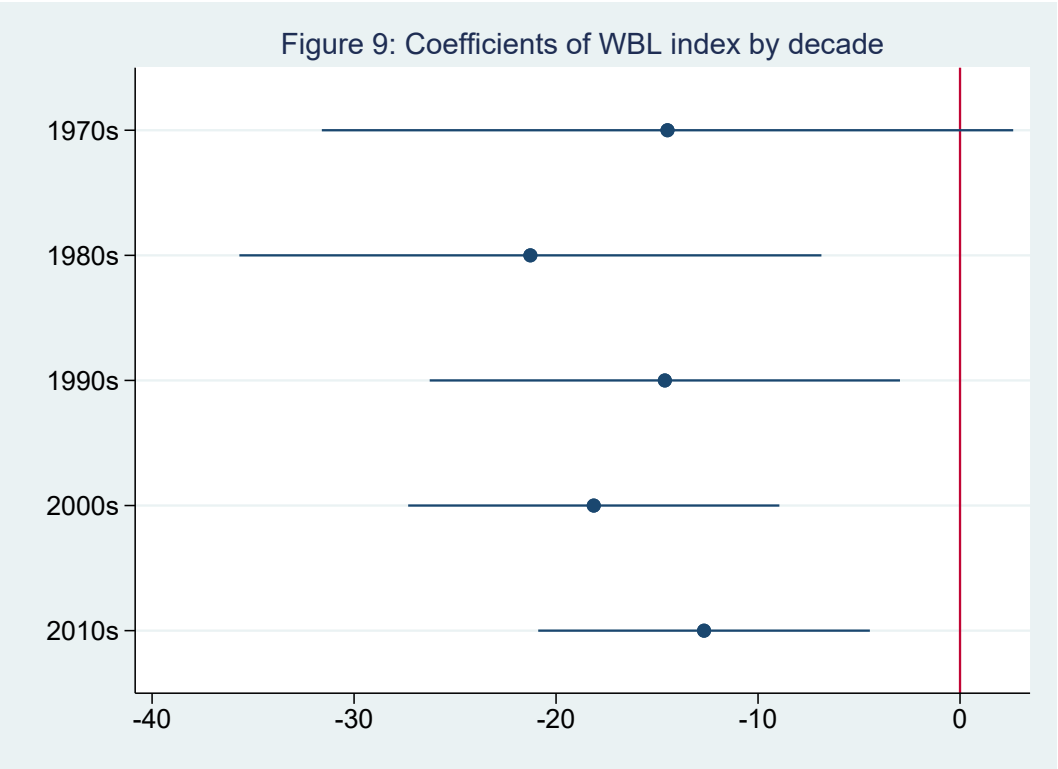
3.4 Effects over time

The relationship between the gender composition of immigrants to the U.S. and the WBL index may change over time. This is a particularly relevant concern when using 50 years of data. To allow for the effect of gender discrimination to vary over time we interacted the WBL index in the baseline specification with a

[32] The number of observations used in the seniors samples is smaller than in the baseline specification because the number of senior immigrants is zero in many years. To check that the differences in the estimates between seniors and non-seniors are not driven by the differences in samples we re-estimated the non-seniors specifications, columns 1 and 2 in Table 2, using the restricted samples in columns 3 and 4 of Table 3, respectively. The estimated coefficient of the WBL index is -0.171 (se 0.052) and -0.173 (se 0.048), respectively. These estimates are of similar order of magnitude to the ones in the full sample (and significantly different from zero), meaning that the small insignificant estimated coefficients in the seniors regressions are not related to the smaller samples. This further validates the placebo test.

[33] The aggregate nature of our data does not allow us to check whether gender discrimination increases the probability of families with (relatively) more daughters to leave their countries of origin.

decade indicator. The decade-specific coefficients are plotted in Figure 9.



While the coefficient is not significantly different from zero in the initial decade, it turns significantly negative from the 1980s onwards. The variation in the estimated effects across decades is not that large: the differences between the four estimated coefficient in the last four decades (1980-2020) is not significantly different from zero.[34] Thus, the estimated effect of gender discrimination on the share of women immigrating to the U.S. appear to be quite stable over time.

3.5 The components of the WBL index

As mentioned in Section 2, the WBL index is a simple average of eight indicators (components) reflecting various aspects of women's interactions with the law dur-

[34] The p value of the test for equality of the four coefficients is 0.014.

ing their lifetime (see Appendix A for details). These components are, of course, correlated but not overly so: the average (median) of the (absolute) 28 pairwise correlation coefficients is 0.33 (0.34). It is therefore reasonable to attempt to identify the effect of the individual indicators.

In this subsection we estimate model (1) using each of these components, jointly and separately, in order to learn about the channels driving the estimated effects presented above. Table 4 shows the estimated coefficients.

Table 4: The components of gender discrimination

	Dep. Var: share of women immigrating to the US, 1970-2019								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MOBILITY	-0.0798								0.00768
	(0.0498)								(0.0369)
WORKPLACE		-0.0214							0.00506
		(0.0167)							(0.0148)
PAY			-0.0453*						-0.0273
			(0.0229)						(0.0204)
MARRIAGE				-0.0717***					-0.0315*
				(0.0209)					(0.0178)
PARENTHOOD					0.0140				0.0148
					(0.0315)				(0.0235)
ENTREPRENEURSH						-0.0456***			-0.0363**
						(0.0155)			(0.0163)
ASSETS							-0.0889***		-0.0704**
							(0.0270)		(0.0285)
PENSION								-0.0364**	-0.0250*
								(0.0137)	(0.0132)
Observations	6,805	6,805	6,805	6,805	6,805	6,805	6,805	6,805	6,805
R-squared	0.775	0.772	0.773	0.777	0.770	0.776	0.783	0.772	0.791
Controls	All	All	All	All	All	All	All	All	All

All regressions include fixed country effects and years dummies intercated with 16 regional dummies.

Standard errors in parentheses are two-way clustered at the country and year levels.

Observations weighted by the number of immigrants from each country of origin.

All controls inlcude GDP per-capita, a conflict indicator and the share of women and population in the country of origin.

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

The Marriage, Entrepreneurship, Assets and Pension components of the

WBL index have significant negative coefficients in the single and multiple regressions. On the other hand, the factors captured by the Mobility, Workplace, Pay and Parenthood components do not seem to have a measurable effect on the gender composition of immigrants to the U.S. Because the components are measured on the same scale (0-100) we can compare the size of their coefficients. The most important component is "Assets" which refers to gender differences in the right to own, administer and inherit family assets. This suggests that restricted access to financial resources in the home country is a key determinant of women's decision to migrate.

4 Conclusions

Human migration is as old as humanity itself. The causes and effects of migration have been studied intensively across various academic disciplines. One dimension of international migration that has received relatively little attention is the sex-ratio of immigrants. There is a large variation in the gender mix of immigrants, both over time and across countries, with important implications for both countries of origin and destination. Nevertheless, we are not aware of studies that attempt to explain such variations on a global scale.

In this paper we estimate the effect of gender discrimination against women on the sex-ratio of immigrants. Gender discrimination can push women to seek better economic opportunities abroad but it can also limit their ability to migrate.

We test the net effect of these opposing factors by regressing the percentage of women that immigrated to the U.S. from country i in year t on an index of gender discrimination in country i and year t. Using both country and year fixed effects, we find that a one standard deviation increase in the WBL index, a decrease in gender discrimination, is associated with a 1.7 percentage point decrease in the

share of women immigrating to the U.S. In order to establish that these findings reflect a casual effect, we control for potential confounding factors and for regional time trends. Our finding is robust to various departures from the baseline model and it also satisfies a placebo test.

In sum, the evidence strongly supports the hypothesis that an increase in gender discrimination in the countries of origin increases the share of women among its emigrants. Future research based on disaggregated data could explore the relationship between the presence of women in families and their propensity to migrate.

References

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

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

[3] Becker, Gary S. 1981. A Treatise on the Family. Cambridge, MA: Harvard University Press.

[4] Borjas, George J. "Economic Theory and International Migration", The International Migration Review, Autumn, 1989, Vol. 23, No. 3, Special Silver Anniversary Issue: International Migration an Assessment for the 90's (Autumn, 1989), pp. 457-485

[5] Borjas, George J. "The Economics of Immigration", Journal of Economic Literature, Dec., 1994, Vol. 32, No. 4 (Dec., 1994), pp. 1667- 1717

[6] Borjas, George J. and Bernt Bratsberg "Who Leaves? The Outmigration of the Foreign-Born", The Review of Economics and Statistics, Feb., 1996, Vol. 78, No. 1 (Feb., 1996), pp. 165-176.

[7] DEMIG (2015) DEMIG C2C, version 1.2, Limited Online Edition. Oxford: International Migration Institute, University of Oxford. www.migrationinstitute.org

[8] Donato, Katharine M., and Amada Armenta. 2011. "What Do We Know About Undocumented Migration." Annual Review of Sociology 37:529-43.

[9] World Bank. 2020. Women, Business and the Law 2020. Women, Business and the Law. Washington, DC: World Bank. doi:10.1596/978-1-4648-1532-4. License: Creative Commons Attribution CC BY 3.0 IGO

[10] L. Edlund, H. Li, J. Yi, J. Zhang, "Sex ratios and crime: Evidence from China." Review of Economics and Statistics, 95 (5) (2013), pp. 1520-1534.

[11] Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand (2002) Armed Conflict 1946-2001: A New Dataset. Journal of Peace Research 39(5).

[12] Houstoun, Marion F., Roger G. Kramer and Joan Mackin Barrett, "Female Predominance in Immigration to the United States Since 1930: A First Look", The International Migration Review, Winter, 1984, Vol. 18, No. 4, Special Issue: Women in Migration (Winter, 1984), pp. 908-963.

[13] Hyland, Marie, Simeon Djankov, and Pinelopi Koujianou Goldberg. "Gendered Laws and Women in the Workforce". AER: Insights 2020, 2(4): 475-490. https://doi.org/10.1257/aeri.20190542

[14] Ormhaug, Christin "Armed Conflict Deaths Disaggregated by Gender", PRIO Paper, November 2009.

[15] Panunzio, Constantine "Intermarriage in Los Angeles, 1924-33" American Journal of Sociology 47(5):690-701, 1942.

[16] Papke, Leslie E. Papke and Jeffrey M. Wooldridge "Econometric methods for fractional response variables with an application to 401(k) plan participation rates", Journal of Applied Econometrics, vol. 11, issue 6, pages 619-632, November 1996.

[17] Pettersson, Therese, Shawn Davis, Amber Deniz, Garoun Engström, Nanar Hawach, Stina Högbladh, Margareta Sollenberg & Magnus Öberg (2021). Organized violence 1989-2020, with a special emphasis on Syria. Journal of Peace Research 58(4).

[18] Raphael, Steven "International Migration, Sex Ratios, and the Socioeconomic Outcomes of Nonmigrant Mexican Women", Demography (2013) 50:971-991.

[19] Ruggles Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. IPUMS USA: Version 8.0 [dataset]. Minneapolis, MN: IPUMS, 2018. https://doi.org/10.18128/D010.V8.0

[20] Sen, Amartya "More than 100 Million Women Are Missing" The New York Review of Books, December 20, 1990.

[21] Sicherman, Nachum and Oded Galor "A Theory of Career Mobility", Journal of Political Economy, Vol. 98, No. 1 (Feb., 1990), pp. 169-192.

[22] United Nations, Department of Economic and Social Affairs, Population Division (2017). World Population Prospects: The 2017 Revision, DVD Edition.

[23] U.S. Census Bureau, Understanding and Using American Community Survey Data: What All Data Users Need to Know, U.S. Government Publishing Office, Washington, DC, 2020 and 2021.

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

A Appendix A: Description of data sources and adjustments

A.1 American Community Survey (ACS)

We use the American Community Survey, conducted by the United States Census Bureau, to collect information on individuals who migrated to the U.S. We use the 2000-2019 ACS annual surveys. Data is limited to households (excluding individuals in "group quarters", such as prison or other institutions). We limit the data to individuals who were born outside the U.S. and utilize information about their gender, age at the year they arrived at the U.S., and their "place of birth". The main variable that we construct, using the ACS, is the percentage of females that arrived at the U.S. from a given country at a given year. For some of the analyses we construct this variable by age groups.

The overall sample size is 6.1 million individuals, who migrated to the U.S. during 1911-2019. Fifty two percent of the sample are females.[35]

A.2 Women, Business, and the Law (WBL)

Women, Business, and the Law (WBL) is a World Bank Group project collecting data on the laws and regulations that affect women's economic opportunities. The "Women, Business and the Law 2021" dataset measures legal differences between men's and women's access to economic opportunities in 190 economies. Thirty-five aspects of the law are scored across eight indicators of four or five binary questions each (see Appendix Table A1 below). These indicators were constructed around

[35] For additional information on the design and methodology of the ACS, including data collection and processing, visit: https://www.census.gov/programs-surveys/acs/methodology.html

women's interactions with the law as they begin, progress through, and end their careers and are used to align different areas of the law with the economic decisions women make at various stages of their lives. The indicators were chosen based on statistically significant associations with outcomes related to women's economic empowerment, including women's labor force participation rates. The eight indicators are: mobility, workplace, pay, marriage, parenthood, entrepreneurship, assets, and pension. The mobility indicator, for example, assesses laws affecting women's agency and freedom of movement, two factors that are likely to influence their decision to enter the labor force. Each subsequent indicator also looks at a specific set of regulations and the ways in which they affect women's economic participation as entrepreneurs and employees.

The methodology was designed as an easily replicable measure of the legal environment for women as entrepreneurs and employees. The data has been updated based on feedback from respondents with expertise in family, labor, and criminal law. Indicator-level scores are obtained by calculating the unweighted average of the questions within that indicator and scaling the result to 100. Overall scores are then calculated by taking the average of each indicator, with 100 representing the highest possible score.

To ensure comparability, the woman in question is assumed to reside in the main business city of her economy and to be employed in the formal sector.

Appendix Table A1: Women, Business and the Law Indicators

Mobility
1. Can a woman choose where to live in the same way as a man?
2. Can a woman travel outside her home in the same way as a man?
3. Can a woman apply for a passport in the same way as a man?
4. Can a woman travel outside the country in the same way as a man?
Workplace
1. Can a woman get a job in the same way as a man?
2. Does the law prohibit discrimination in employment based on gender?
3. Is there legislation on sexual harassment in employment?
4. Are there criminal penalties or civil remedies for sexual harassment in employment?
Pay
1. Does the law mandate equal remuneration for work of equal value?
2. Can a woman work at night in the same way as a man?
3. Can a woman work in a job deemed dangerous in the same way as a man?
4. Can a woman work in an industrial job in the same way as a man?
Marriage
1. Is there no legal provision that requires a married woman to obey her husband?
2. Can a woman be head of household in the same way as a man?
3. Is there legislation specifically addressing domestic violence?
4. Can a woman obtain a judgment of divorce in the same way as a man?
5. Does a woman have the same rights to remarry as a man?
Parenthood
1. Is paid leave of at least 14 weeks available to mothers?
2. Does the government administer 100% of maternity leave benefits?
3. Is paid leave available to fathers?
4. Is there paid parental leave?
5. Is dismissal of pregnant workers prohibited?
Entrepreneurship
1. Does the law prohibit discrimination in access to credit based on gender?
2. Can a woman sign a contract in the same way as a man?
3. Can a woman register a business in the same way as a man?
4. Can a woman open a bank account in the same way as a man?
Assets
1. Do men and women have equal ownership rights to immovable property?
2. Do sons and daughters have equal rights to inherit assets from their parents?
3. Do male and female surviving spouses have equal rights to inherit assets?
4. Does the law grant spouses equal administrative authority over assets during marriage?
5. Does the law provide for the valuation of nonmonetary contributions?
Pension
1. Is the age at which men and women can retire with full pension benefits the same?
2. Is the age at which men and women can retire with partial pension benefits the same?
3. Is the mandatory retirement age for men and women the same?
4. Are periods of absence due to childcare accounted for in pension benefits?

A.3 Matching and merging the ACS and the WBL data sets

For most observations in the ACS data there is a perfect matching between the reported place of birth and the economy listed in the WBL dataset. There are, however, several mismatches that we have to address. Below we report all the adjustments we make in order to best match the two data sets.

1. Macedonia (43330) does not appear in the WBL but appears in the ACS.[36] Macedonia in the ACS is part of Greece. Therefore, Macedonia was dropped, and individuals born in Macedonia were re-classified as being born in Greece (43300).
2. Israel and "West Bank & Gaza" are reported separately in the WBL but combined in ACS as Israel/Palestine (53400). We average the WBL index of both entities and combined into a single country Israel/Palestine (53400). We used 2020 population weights for averaging. The two countries were substituted by the combined Israel/Palestine.
3. Germany and Canada appear in the ACS as having many sub-regions but the actual immigrant data reports only the aggregates for Canada (15000) and Germany (45300).
4. Yemen, or officially the "Republic of Yemen", was created in 1990, uniting north and south Yemen. The ACS, however, has two separate codes, one for "Yemen Arab Republic (North)" (54400) and one for "Yemen, PDR (South)" (54500) for all years (and no separate code for Yemen). Nevertheless, in all survey years only migrants from North Yemen (total of 6039) appear in the ACS surveys. The WBL index exists only for the "Yemen Republic". We,

[36] The five-digit codes in parentheses are the country codes used by the Census for "place of birth (detailed)" (BPLD).

therefore, match the index to North Yemen.

5. Although Korea was divided into two separate countries in 1945, the WBL index exists only for "Korea". We assume it is South Korea. In the ACS we have data for Korea (50200) in all years, for North Korea (50210) in only two years (2003-2004), and for South Korea (50220) in nine years (2003-2011). In a year when more than one code is available, respondents can choose whichever code they prefer. For example, in 2004, 2,035 listed "Korea", 7 listed "North Korea", and 1,544 listed "South Korea", as their place of birth. We added the number of immigrants coming from "Korea" and "South Korea" into a single BPLD code 50200, to which the WBL index of "Korea" was assigned.[37]
6. The WBL has one index for the United Kingdom (U.K.), while the ACS has separate codes for different countries in the U.K. (41000- 41200), plus one code for "United Kingdom, n.s./n.e.c." (41300).[38] We combined the codes 41000-41200, plus Northern Ireland (41410) and Bermuda (16010), with "UK, n.s./n.e.c." into one code, 41300. Ireland (41400), of course, is

[37] The number of immigrants coming from North Korea was ignored because of the country's peculiarities and because it is never more than 10 immigrants per year.

[38] "n.s." stand for "not specified" and refers to cases where more precise information about the place of birth is not available. "Americas, ns", for example, refers to instances when the place of birth was indicated as being in "America" or the "Americas" without any further specification. n.s. is used in cases where a response to a particular census question is relevant (i.e., it is clearly an answer to the question, as opposed to something written in the incorrect column) but general, so that it is impossible to map it to a detailed code. "n.e.c." stands for "not elsewhere classified" and includes all responses for which there is no specific code for the given geography. n.e.c. is used in cases where a response to a particular census question is relevant (i.e., it is clearly an answer to the question, as opposed to something written in the incorrect column) and often quite detailed, but where no specific code exists to match the response.

and independent republic, not part of UK.

7. Azores does not appear in the WBL. We added the immigration data from Azores (43610) to that of Portugal (43600).
8. Czechoslovakia was split into two countries, Slovakia and the Czech Republic, in 1993. The ACS has three separate codes, Czechoslovakia (45200), Slovakia (45212), and the Czech Republic (45213), which are available in all survey years. Respondents in the ACS could choose any of the three codes in 2000-2019 (and only Czechoslovakia up to 1993). This means that after 1993 they could (and did) report coming from either Slovakia or the Czech Republic before they were separate countries, or, alternatively, report coming from Czechoslovakia after 1993, even though it did not exist as a country anymore. The WBL index, on the other hand, is reported separately for Slovakia and the Czech Republic in all years, including years in which the two separate countries did not exist. For all years we split the number of immigrants from "Czechoslovakia "(45200) into Slovakia (45212) and the Czech Republic (45213) using 2020 population weights.
9. The country of Yugoslavia was set up after World War II as a federation of six republics: Bosnia and Herzegovina, Croatia, Macedonia, Montenegro, Serbia, and Slovenia. Following a series of political conflicts during the early 1990s the country was broken up forming eventually the following independent countries: (North) Macedonia (1991), Croatia, Slovenia, Bosnia and Herzegovina (1992), Serbia (2006), Montenegro (2006), and Kosovo (2008). The table below lists the countries and years they are reported in both data

sets:

Country	ACS	WBL
Yugoslavia	all years	none
Croatia	all years	all years
Montenegro	2012-2019	all years
Serbia	2012-2019	all years
Bosnia	all years	none
Bosnia & Herzegovina	none	all years
North Macedonia	none	all years
Slovenia	none	all years
Kosovo	2018-2019	all years

Because the WBL index does not exist for Yugoslavia, we split the immigrants from Yugoslavia to the countries that were part of Yugoslavia, using the same procedure we applied to Czechoslovakia, using 2020 population weights. Slovenia (45780) has a WBL index and, while it appears in the list of ACS countries, it has zero immigrants to the U.S. during the 1970-2019 period. Because of our split of the data for Yugoslavia, Slovenia will appear with positive immigration numbers.

Northern Macedonia does not appear in ACS data and was added manually with BPLD 9999. Similarly to Slovenia, it will get a portion of the data for Yugoslavia.

10. The Russian Federation has a WBL index and was matched to immigrants with BPLD code 46500 corresponding to "Other USSR/Russia".
11. Surprisingly, Cuba does not have a WBL index and therefore it drops from our sample.

There are 2,150 observations in the WBL, corresponding to 42 countries, that could not be matched to ACS data because no immigrants from those countries

are reported in any of the ACS surveys used. This is either because these countries were not listed (in all or certain survey years) or because there are no immigrants from those countries in the years they were listed.

To conclude, we started with 190 countries in the WBL data, 42 of them do not have immigration data, and three belong to the U.S.A. (United States, Puerto Rico, and Northern Mariana Islands). This leaves us with 145 countries and 50 years of data (1970-2109) on the WBL index and immigration. A "zero" is assigned to the total number of immigrants in years in which there are no immigrants from a country of origin. In these years the sex ratio cannot be computed and appears as "missing". The 145 countries and the total number of immigrants during 1970-2019 appear in Table A2.

Appendix Table A2: Number of immigrants in ACS 1970-2019

Country	Immigrants	Country	Immigrants	Country	Immigrants
Mexico	1,369,506	Greece	16,648	Bahamas, The	4,266
Philippines	291,297	Turkey	16,360	Grenada	4,071
India	279,580	Panama	16,218	Lithuania	3,860
China	255,852	Spain	15,590	Kazakhstan	3,824
Vietnam	192,792	South Africa	15,461	Kuwait	3,822
Korea, Rep.	159,932	Australia	14,459	Denmark	3,786
El Salvador	144,279	Ireland	13,741	Austria	3,755
Canada	112,103	Kenya	13,468	Dominica	3,749
Germany	108,443	Myanmar	13,075	Eritrea	3,691
United Kingdom	105,431	Indonesia	13,003	Uganda	3,241
Dominican Republic	103,059	Chile	12,830	Slovak Republic	3,066
Guatemala	98,117	Armenia	11,138	Norway	3,052
Colombia	85,764	Nepal	10,850	Tanzania	2,957
Jamaica	84,484	Costa Rica	10,769	Zimbabwe	2,914
Haiti	69,092	Netherlands	10,495	St. Vincent & the Grenadines	2,764
Japan	64,674	Malaysia	10,239	Azerbaijan	2,747
Taiwan, China	61,997	Syrian Arab Republic	10,041	Finland	2,675
Russian Federation	59,274	Afghanistan	10,032	Paraguay	2,573
Honduras	57,458	Albania	9,947	Algeria	2,565
Poland	54,712	Bolivia	9,620	Bhutan	2,511
Peru	53,912	Bulgaria	9,434	St. Lucia	2,468
Iran, Islamic Rep.	51,492	Jordan	9,348	Georgia	2,333
Brazil	49,456	Morocco	8,762	Senegal	2,315
Ukraine	47,448	Liberia	8,619	Antigua and Barbuda	2,286
Pakistan	46,029	Belarus	8,502	Tonga	2,213
Ecuador	45,772	Saudi Arabia	7,921	Latvia	2,210
Guyana	36,533	Somalia	7,245	Congo, Dem. Rep.	2,114
Hong Kong SAR, China	34,941	Uzbekistan	7,229	United Arab Emirates	1,551
Thailand	34,240	Sri Lanka	6,942	Guinea	1,394
Nigeria	32,633	Sweden	6,757	Togo	1,359
Italy	31,687	Hungary	6,633	Iceland	1,287
Venezuela, RB	31,554	Barbados	6,485	Congo, Rep.	1,242
Nicaragua	31,363	Yemen, Rep.	6,124	Marshall Islands	1,224
Trinidad and Tobago	29,350	Switzerland	6,122	Montenegro	1,173
Lao PDR	26,461	Uruguay	6,117	Zambia	963
Bangladesh	26,100	Croatia	6,030	Libya	940
Romania	24,003	Belize	6,025	Kosovo	940
France	23,356	Fiji	5,915	St. Kitts and Nevis	880
Argentina	22,800	Moldova	5,637	Gambia, The	821
Cambodia	22,593	Singapore	5,610	Slovenia	789
Ethiopia	22,311	Belgium	5,262	North Macedonia	789
Egypt, Arab Rep.	21,996	Serbia	5,167	Côte d'Ivoire	528
Portugal	21,926	Cameroon	5,144	Estonia	507
Israel	20,866	New Zealand	5,097	Cyprus	500
Iraq	20,622	Sudan	4,873	Mongolia	344
Lebanon	17,658	Czech Republic	4,754	Kyrgyz Republic	332
Bosnia & Herzegovina	17,026	Cabo Verde	4,544	Rwanda	298
Ghana	16,714	Sierra Leone	4,428	Tunisia	262
				South Sudan	155

A.4 GDP Per Capita

We use the GDP per capita at current prices in U.S. Dollars, downloaded from https://unstats.un.org/unsd/snaama/Downloads.We made the following adjustment to match the GDP data with the ACS data:

1. Since the GDP data is separated for Israel and the State of Palestine, we averaged them into a single Israel/Palestine observation using 2020 population as weights.
2. Ethiopia: Until 1989 GDP numbers are reported for "Ethiopia (Former)". For the years 1990-1993 numbers are reported for both "Ethiopia" and "Ethiopia (Former)", and after 1993 data is reported only for "Ethiopia". The 1970-1989 data for Ethiopia were taken from "Ethiopia (Former)".
3. Sudan: Until 2007 the reported GDP numbers are for "Sudan (Former)". For the years 2008-2010 data is reported for both "Sudan (Former)" and" Sudan. After 2010 GDP data is reported only for Sudan. Therefore, the 1970-2007 data for Sudan were taken from "Sudan (Former)".
4. GDP data for Slovakia and "Czechia" are reported starting from 1990, while data for "Czechoslovakia (Former)" is reported only until 1990. We assigned the "Czechoslovakia (Former)" GDP per capita to Slovakia and the Czech Republic for the period 1970–1989.
5. GDP data for data for Croatia, Bosnia and Herzegovina, Montenegro, North Macedonia, Kosovo, Serbia, and Slovenia is available only after 1989. Therefore, for the period 1970-1989, we use the GDP per capita values reported for "Yugoslavia (Former)".
6. Yemen (id 887) has data from 1989 onwards. For previous years we took a simple average of the GDP per capita between Yemen Arab Republic and Yemen, Democratic.
7. Taiwan has no GDP data in the UN files. For the years 1980-2019 we

used the International Monetary Fund, World Economic Outlook Database, April 2021 (https://www.imf.org/en/Publications/WEO/weo-database/2021/April). For the years 1961-1979 the source is: https://countryeconomy.com/gdp/taiwan. For 2020, the figure was manually downloaded from the Internet.

In sum, 279 country-year observations have missing GDP data. A few countries have 1 or 2 years of data missing, while the bulk of the missing observations correspond to former U.S.S.R. countries before 1990. This accounts for the slightly smaller number of observations when GDP per capita is used in the regressions (see Table 1).

A.5 International Sex Ratios data

Sex ratio over 5-year periods and across countries were downloaded from : https://population.un.org/wpp/Download/Standard/Population/. The file is named "Sex Ratio of Total Population".

The following adjustments were made to match the sex-ratio data to the WBL/ACS data:

1. Israel and Palestine were averaged into a single Israel/Palestine observation using 2020 population weights.
2. Kosovo does not have data on sex ratios. It was assigned North Macedonia's data.
3. Dominica does not have data on sex ratios. It was assigned a simple average of the sex ratios in the neighboring islands Guadeloupe and Martinique.
4. St. Kitts-Nevis does not have data on sex ratios. It was assigned the sex ratios in neighboring Antigua and Barbuda.
5. Marshall Islands does not have data on sex ratios. It was assigned the sex ratio in Micronesia.

A.6 Armed Conflict data

We use the UCDP/PRIO Armed Conflict Dataset version 21.1, which covers the period 1946-2020. These data can be downloaded from https://ucdp.uu.se/downloads/index.html#armedconflict.

For this project we generated a variable indicating the number of armed conflicts per country-year and the highest intensity among the conflicts in each country-year.

There are 103 countries experiencing at least one conflict between 1970 and 2020, for a total of 1487 country-year observations (excluding the U.S) with at least one conflict. But 19 of them do not have WBL/ACS data. This means that in the WBL/ACS data 84 countries experienced at least one conflict between 1970 and 2019, while 61 countries never experienced one. Countries and years with no conflict were assigned 0 number of conflicts and intensity.

We now describe the data in detail. UCDP defines state-based armed conflict as: "a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in a calendar year" (see codebook for more details).

The main unit of observation in this dataset is an "Armed Conflict" as defined by UCDP. Each conflict is listed in all years where fighting in one or more dyad(s) caused at least 25 battle-related deaths. We define the location of the conflict as the name(s) of the country/countries whose government(s) have a primary claim to the issue in dispute.

There are four types of conflict:

1. Extrasystemic armed conflict occurs between a state and a non-state group outside its own territory. These conflicts are by definition territorial, since the government side is fighting to retain control of a territory outside the

state system.

2. Interstate armed conflict occurs between two or more states.
3. Internal armed conflict occurs between the government of a state and one or more internal opposition group(s) without intervention from other states.
4. Internationalized internal armed conflict occurs between the government of a state and one or more internal opposition group(s) with intervention from other states (secondary parties) on one or both sides.

Intensity level of the conflict: The intensity level in the dyad per calendar year. Two different intensity levels are coded: minor armed conflicts and wars. Minor: between 25 and 999 battle-related deaths in a given year. War: at least 1,000 battle-related deaths in a given year.

A.7 Population data set

Data were downloaded from : https://population.un.org/wpp/Download/Standard/Population/. The file is named "File POP/1-1: Total population (both sexes combined) by region, subregion and country, annually for 1950-2100 (thousands)".

The following adjustments were made to match the population data to the WBL/ACS data:

1. Israel and Palestine were averaged into a single Israel/Palestine observation using the total population of both entities.
2. Kosovo is missing in the UN. Data taken from World Bank https://data.worldbank.org/indicator/SP.POP.TOTL?locations=XK .