Online Appendix

War-driven permanent emigration, sex ratios, and female labor force participation

Ana Rute Cardoso and Louis-Philippe Morin

August 8, 2023

Appendix A – Institutional setting

Education policies

Aiming to reduce the widespread illiteracy rate in the country, at 75% in 1910 Reis (1993), the First Republic raised in 1911 the compulsory schooling to grade 3; by 1919, it had raised it further to grade 5, even though enforcement was an issue. Education was among its priorities, advocating for new pedagogical methods, mixed-gender schools, and modern academic curricula. However, the regime set in place in 1926, and the New State imposed a different orientation. The reduction of illiteracy remained a priority, but it was now understood as providing basic reading and writing skills. A conservative view pervaded the educational system, which relied on traditional pedagogical methods and a curriculum aimed at instilling nationalism, Catholic moral values, and traditional gender roles.[1] Accordingly, in 1927, compulsory schooling was reduced to four years, a rather unusual direction of change for education policies worldwide. It was further reduced to three years in 1930. Moreover, by 1956, the country had gender-specific compulsory schooling levels—four years for boys and three years for girls.

The law that raised compulsory schooling only for boys also determined a comprehensive set of enforcement conditions: access to jobs in public administration, manufacturing, and the services, was forbidden to individuals without a diploma of grade 4, just like the possibility of getting a driver's license or participate in official sports competitions. The law thus confined women with compulsory education to the agriculture and household spheres. Harmonization of compulsory schooling across gender took place in 1960 when it was raised to grade 4 for girls as well. This change potentially widened the labor market opportunities for women.[2] However, Caillods and Rocha

[1] See, for example, Decree 27:603 dated March 29, 1937.

[2] For a thoughtful analysis of the newspaper coverage of this legal change and the motivations of the government, see Adão and Remédios (2009).

(1983) argue that the gender-specific compulsory schooling law had a long-lasting impact that could justify a high female drop-out rate upon completion of grade 3 as late as the early 1970s.[3]

Reis (1993) provides an insightful discussion of the origins of the low educational achievement in Portugal. He highlights the relevance of one of the explanations by Easterlin (1981) for a State's investment in massive education—the wish to obtain or maintain political, social, or economic power. Several examples worldwide confirm the schooling system as a tool to unify under a single state and language, a set of nations with different social, cultural, and language backgrounds, promoting allegiance to a central government. Portugal, having a common language and stable borders since the 13th century and no social or political tensions comparable to its European counterparts, would have lacked that impetus to educate its population massively.

Laws on voting rights, marriage, divorce, and abortion

The Constitution of 1933 clearly articulated the guiding principle on gender roles in Portugal, as it determined that all citizens were equal before the law, excluding "in the case of the woman, the differences that result from her nature and the good of the family" (art. 5).[4]

The right to vote had been granted in 1911 to citizens over the age of 21 who could read and write and were heads of the family. The omission of any reference to gender opened the possibility for *one* woman to vote, after an intricate process and a court case, on the grounds that she was a widow and mother, thus head of a family. That oversight by the legislator was soon clarified: regulations enacted in 1913, 1918, and 1927 all made explicit that only male heads of a family could vote.[5] Later, in 1931, both gender were recognized the right to vote if fulfilling strict conditions, which were, in any case, different across gender and the scope of the election. In national and regional elections, males were required to know how to read and write or, if they were heads of family, pay over a certain amount in property and income tax; females, instead, were required to have a secondary or tertiary education diploma. In local elections, all males head of a family or living independently could vote; females, only if they were the head of a family.[6] A significant change took place in 1968, when both gender, if literate, were allowed to vote in national elections. However, in the regional and local elections, the constraints on secondary or tertiary education and head of a family for females still applied, such that a tiny fraction was entitled to vote.

Divorce was, rather progressively, regulated for the first time in Portugal in 1910, covering issues such as possible causes, the distinction between mutual consent or

[3]The sequence of laws referred to in the text are the following decrees: March 29, 1911 (no number); 6:137 of September 29, 1919; 13:619 of May 17, 1927; 18:140 of March 22, 1930. They were followed by the decree-laws 40964 of December 31, 1956, and 42994 of May 28, 1960.

[4]For a discussion on the social and political context surrounding the evolution of gender-related regulations in Portugal, see Ferreira (2011).

[5]For an informative overview of the process until 1976, see Souza (2013).

[6]Decree 19:694 of May 5, 1931, updated by Decree 20:073 of July 15, 1931.

litigious divorce, property division, and support orders. However, the 1940 Concordat, a convention with the Holy See, outlawed divorce for those married by the Church, which was the overwhelming majority. It was not until the 1974 revolution that this regulation was reversed. Contraception was outlawed, and so was abortion. The law conferred husbands' rights, such as denouncing an employment contract signed by their wives without their prior consent. During most of the period under analysis, women could not apply for a passport without their husbands' explicit consent.

This entire institutional setting underwent a major turn in the mid-1970s when the juridical order started catching up with the forefront of international regulations on women's rights. Table A.1 lists the major policies with a gender focus during the period under study and their change after the 1974 revolution.

Appendix B – Data

Census for Portugal and France

We rely on the 1981 Census data for Portugal and drop the observations from Madeira and the Azores (4.8% of observations). The microdata include age at the Census date (March 16) but not the birth year. We have thus computed the year of birth as the Census year minus age minus 1. Hence, for individuals whose birthday falls between January 1 and mid-March, the cohort will be measured with error (the true cohort being a year later than coded).The number of children ever born is not reported, and only the number of own children in the household is available for Portugal.

The 1962, 1968, 1975, 1982, and 1990 French Censuses contain information on immigrants in France, their nationality and age, and birth year. France was the overwhelming destination of Portuguese emigrants over the period we cover. We compute the number of Portuguese immigrants in France by birth cohort and gender. For an accurate matching with the Census data of the origin country, the definition of birth cohort relies on age at the Census date. A caveat is that we cannot identify the birth country of naturalized respondents. Hence, we might be underestimating the number of immigrants, especially among older individuals, who may have immigrated a long time ago and obtained host-country citizenship in the meantime.

Computation of Portuguese emigration rates and sex ratios based of the French censuses

We have checked the potential implications of using age instead of birth year to define the cohorts in our analysis. The French IPUMS data contain both the age and the respondent's year of birth. Comparing both methods, we find that it does not materially affect the computation of the sex ratio and emigration rates by cohort.

A second issue relates to the weights used: IPUMS weights all of its observations equally, whereas the French Institut national de la statistique et des études économiques (INSEE) does not. This discrepancy could impact the measurement of sex ratios for

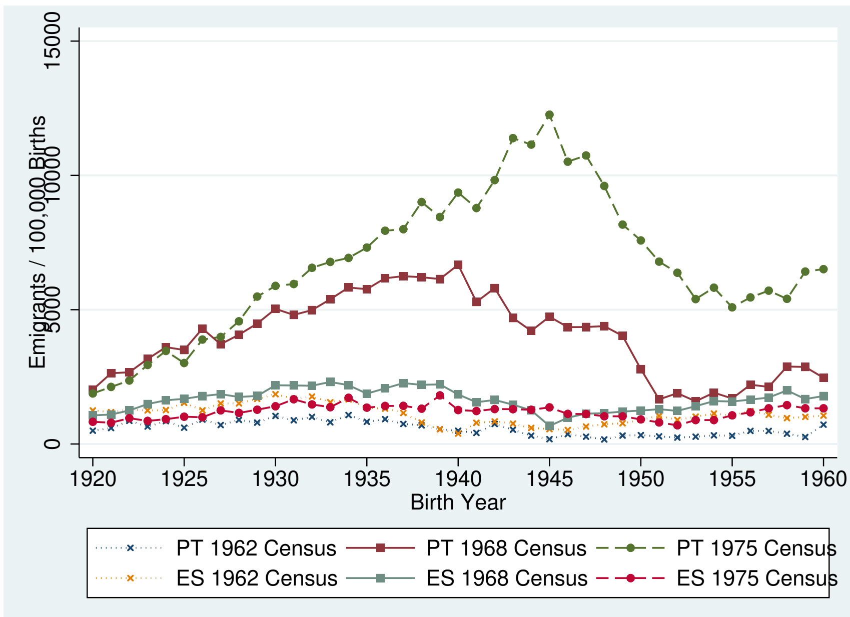
Table A.1: Institutional setting with a gender component, Portugal, 1926-1974.

Year	Policy	Source
	A. Equality before the law	
1933	All citizens are equal before the law, excluding "in the case of the woman, the differences that result from her nature and the good of the family".	Constitution
1971	All citizens are equal before the law, excluding "in the case of gender, the differences in treatment justified by nature".	Constitution
	B. Right to vote	
1927	Only male citizens, if literate or heads of family or living independently.	DL 14802
1931	In national or regional elections: males, if literate or paying over a certain amount in property or income tax; females, if holding a secondary or tertiary education diploma. In local elections: males if heads of a family or living independently; females if heads of a family.	Decrees 19694,20073
1946	In national elections: males if a) literate or b) paying over a certain amount in property or income tax; females if holding a secondary or tertiary education degree, or heads of a family fulfilling condition a) or b), or married (thus not heads of family) and fulfilling condition a) and paying twice the amount set in b) as tax on real estate property (own or joint).	Law 2015
1968	In national elections: all citizens, if literate, as well as those previously registered to vote under Law 2015 of 1946.	Law 2137
	C. Education	
1956	Compulsory schooling: boys four years; girls three years. Starting in 1959, illegal for those without a diploma of 4 years of education to get a job in public admin., a driver's license, or participate in sports competitions; if younger than 21, also illegal to get a job in the services or manufacturing.	DL 40964
1960	Compulsory schooling four years for boys and girls.	DL 42994
	D. Divorce, contraception, abortion	
(1886)	Abortion forbidden.	Penal Code
1940	Divorce forbidden for those married by the Church.	Concordat
1942	Abortion and contraception forbidden.	DL 32171
	E. Post-1974 revolution changes	
1974	All citizens aged 18 or above entitled to vote in all elections.	DL 621A
1975	Divorce legalized.	Concordat amend.
1976	Equality of all citizens before the law.	Constitution
1976	State has duty to promote dissemination of family planning methods.	Constitution
1984	Abortion allowed under specific circumstances (such as serious risk to the health of the woman, rape, fetus malformation).	Law 6/84
2007	Abortion during first 10 weeks of pregnancy legalized.	Law 16/2007

cells with a small number of individuals, such as older cohorts. Using the age of the individual to compute the cohort in either dataset, we acknowledge that both data sets show the same patterns in sex ratios. If anything, the evolution of the sex ratio is slightly smoother in the INSEE data for individuals born in the 1920 and early 1930s.

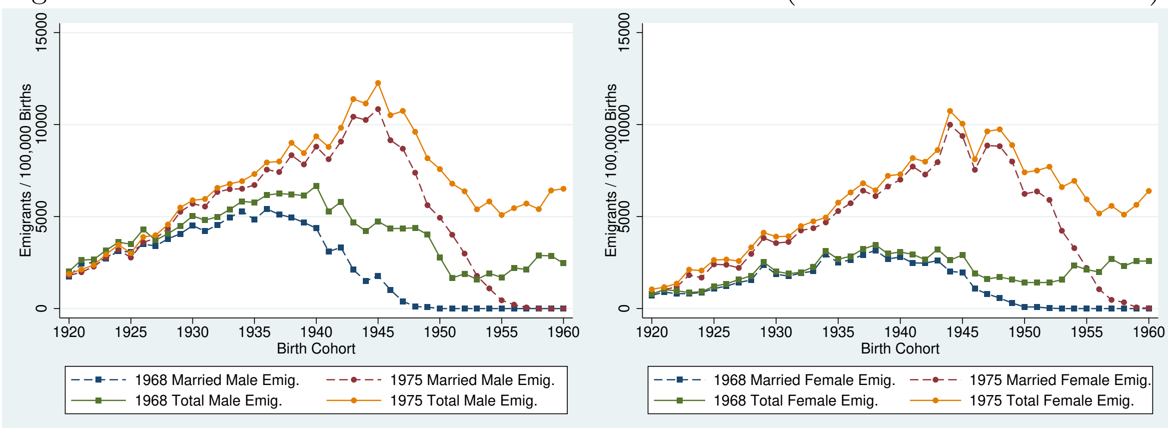
Additional emigration figures based on the French censuses

Figure B.1: Male Emigration Rates, Portugal vs. Spain (based on FR censuses).



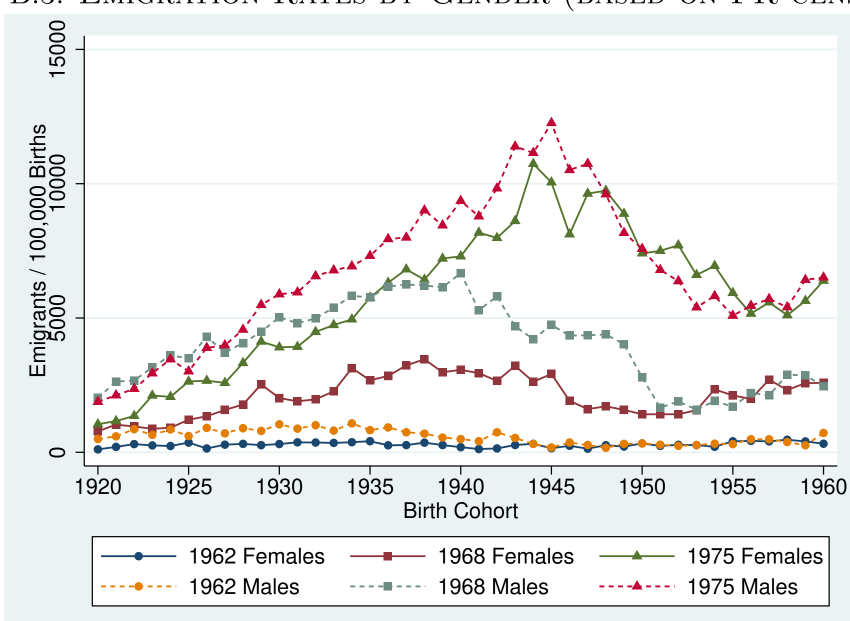
Sources: IPUMS-International (Minnesota Population Center, 2015), Portugal INE (2016), and Spain INE (2017).

Figure B.2: Emigration Rates by Marital Status (based on FR censuses).



Sources: IPUMS-International (Minnesota Population Center, 2015) and INE (2016).

Figure B.3: Emigration Rates by Gender (based on FR censuses).



Sources: IPUMS-International (Minnesota Population Center, 2015) and INE (2016).

Conversion of Portuguese districts into regions

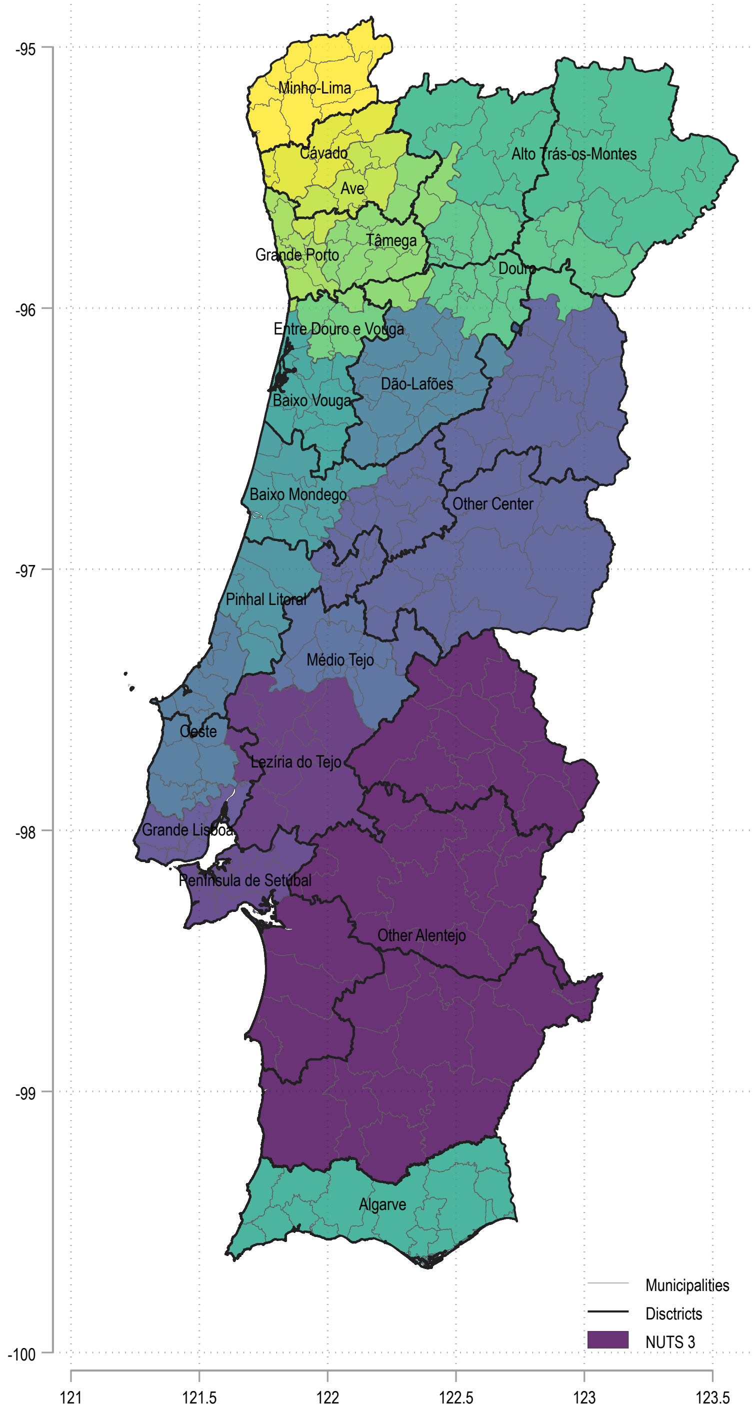
Two main multi-level territorial-unit nomenclatures are used in this paper: administrative and statistical divisions. The 1981 Portuguese Census (provided by IPUMS-International) and QP provide statistical territorial divisions known as NUTS 3.[7] In 1981, there were 20 NUTS-3 regions in mainland Portugal, corresponding to our analysis' geographical unit of observation. Given we observe 18 birth cohorts (born between 1937 and 1954) for each NUTS-3 region, our main regressions contain 360 cohort-region observations.

Some of the original information necessary for constructing our emigration instruments is unavailable at the NUTS 3 level. In particular, the 1950s legal emigration shares from Baganha (1994), and the number of births by cohort are available by districts only. Portuguese districts are administrative divisions of the territory. In the case of mainland Portugal, there are three administrative-division (i.e., local-authority) levels: districts, municipalities, and communes. Mainland Portugal has 278 municipalities distributed in 18 districts.

Figure B.4 shows a map of Portugal's NUTS 3 and districts, along with their municipalities. The 20 NUTS-3 regions are identified with different colors (with their names in the center of the region), while the 18 districts are delimited with thick black lines. Thinner grey lines identify municipalities. As shown in Figure B.4, a district may split across different NUTS. However, municipalities do not cover multiple districts or NUTS. Therefore, in cases where a district covers more than one NUTS, we first allocate district-level values (e.g., the number of births) to each of their municipalities based on their population weights within the district (we consider the 1961 population). We then aggregate municipality values to their respective NUTS-3 region using the municipalities allocation to NUTS, given that municipalities, unlike districts, never cross NUTS boundaries. To perform this exercise, we rely on the following sources. The municipality–district correspondence comes from Portugal, Direção-Geral do Território (2016), while the municipality-NUTS correspondence comes from Portugal, INE (2016). The 1961 municipality population information was obtained from Pordata (2017).

[7]The Nomenclature of Territorial Units for Statistics (NUTS) code has three nested levels of aggregation: NUTS 1, NUTS 2, and NUTS 3, NUTS 3 being the most detailed division level.

Figure B.4: District-Region Correspondence.



Sources: IPUMS-International (Minnesota Population Center, 2015), Portugal, Direção-Geral do Território (2016), INE (2016) and European Environment Agency (2022).

Conversion of occupational codes in the original 1960 Portugal Census to IPUMS codes

The 1960 Portuguese Census reports the occupational classification ISCO-58, and the 1981 one reports ISCO-68. IPUMS, in turn, starts in 1981 and converts whichever occupational classification the country's statistical office used into a 1-digit harmonized classification, standard across countries. Crucially, it reports both the country's original 2-digit classification and the IPUMS harmonized 1-digit classification.

Therefore, to identify whether an occupation reported in IPUMS 1981 was male- or female-dominated as of 1960, we first converted the ISCO-58 classification into ISCO-68, relying on ILO (1969). Having the 1960 employment coded into ISCO-68, which is reported in IPUMS 1981, it was straightforward to aggregate it into the 1-digit harmonized classification, as IPUMS itself reports both variables.

The data available for 1960 by gender and occupation refer to the active population. For comparability, we have thus considered for later years the active population and imposed no constraints on cohort or age (in Table 4).

Individual data on fatal war casualties

The webpage TerraWeb (2016), managed by war veterans, aims at an exhaustive coverage of the fatal casualties, reporting in particular: the specific place of birth (commune and municipality), military rank, military base in mainland Portugal and destination unit in Africa, year of death, and its cause. For a share of the dataset, approximately 7%, the exact birth date is reported as well. For soldiers, corporals, and non-commissioned sergeants[8], the age distribution at death, though not degenerate into one single value, exhibits an extremely narrow range: for soldiers, over 90% of the fatalities occur at ages 21 to 23; for corporals, over 85% occur at ages 22 to 24; and for non-commissioned sergeants, over 80% at ages 22 to 25. This fact can be easily understood. The age of exposure to death risk was almost deterministic, given the fixed starting moment for the military service (the year the individual turned 21) and the two-year fixed length of service in Africa. Therefore, we have relied on the military rank and year of death for the population of deceased men from each region to infer their age and thus their year of birth, by replicating the overall distribution of age at death by military rank. To do so, we imposed a set of constraints on the original dataset and kept the following observations: soldiers, corporals, and non-commissioned sergeants (dropping 899 observations), as other ranks have both fewer casualties and a more dispersed distribution of age at death; individuals originating from mainland Portugal (dropping 372 observations from Madeira and Azores); whose death occurred in 1961-1974 (dropping 68 obs.); and cells rank - age at death with at least three elements (dropping 16 obs.).

[8]In the original terminology, *soldado*, *primeiro-cabo*, and *furriel*, the latter meaning a sergeant who is not part of the permanent staff of the Armed Forces.

Appendix C – The endogeneity problem

To illustrate the potential endogeneity issues when regressing the labor market activity on the sex ratio and the size of the labor market, we present a simple accounting exercise that will also suggest potential instrument-variable candidates to deal with these issues.[9] Let P_{rt}^{gc} be the population of gender $g = f, m$ (female and male, respectively) and cohort c observed in region r in year t (or roughly at age $t-c$). We can decompose this number into:

$$P_{rt}^{gc} = B_r^{gc} + J_{rt}^{gc} - L_{rt}^{gc} = B_r^{gc} + N_{rt}^{gc} \tag{1}$$

where, for gender g, B_r^{gc} is the number of individuals born in region r and cohort (year) c, J_{rt}^{gc} is the number of individuals from cohort c who have joined region r between years c and t and, conversely, L_{rt}^{gc} is the number of individuals who have left the region. The corresponding net flow is expressed as $N_{rt}^{gc} = J_{rt}^{gc} - L_{rt}^{gc}$ (negative, in case of net outflow). These flows include both natural population movements (death of individuals from cohort c) and migratory population movements, whether across regions in the country or international.

Hence, at any point in time, the sex ratio for cohort c is given by

$$R_{rt}^{c} = \frac{B_r^{mc} + J_{rt}^{mc} - L_{rt}^{mc}}{B_r^{fc} + J_{rt}^{fc} - L_{rt}^{fc}}. \tag{2}$$

Dividing and multiplying through the numerator and denominator by the number of male and female births, respectively, we get:

$$R_{rt}^{c} = R_{rc}^{c}\left[\frac{1 + J_{rt}^{mc}/B_r^{mc} - L_{rt}^{mc}/B_r^{mc}}{1 + J_{rt}^{fc}/B_r^{fc} - L_{rt}^{fc}/B_r^{fc}}\right] = R_{rc}^{c}\left[\frac{1 + j_{rt}^{mc} - l_{rt}^{mc}}{1 + j_{rt}^{fc} - l_{rt}^{fc}}\right] \tag{3}$$

where R_{rc}^{c} is the sex ratio at birth for cohort c; the terms j_{rt}^{gc} and l_{rt}^{gc} represent, respectively, the inflow and outflow rates between years c and t for cohort c, region r, and gender g. Taking logs, we can rewrite equation (3) as:

$$\ln R_{rt}^{c} = \ln R_{rc}^{c} + \ln\left[1 + j_{rt}^{mc} - l_{rt}^{mc}\right] - \ln\left[1 + j_{rt}^{fc} - l_{rt}^{fc}\right]. \tag{4}$$

If the flow rates are not too large and the sex ratios not too far from one, we can approximate this relationship using ($\ln X \approx X - 1$):

$$R_{rt}^{c} \approx R_{rc}^{c} + \left[j_{rt}^{mc} - l_{rt}^{mc}\right] - \left[j_{rt}^{fc} - l_{rt}^{fc}\right]. \tag{5}$$

It is evident that if the net flow rates $j - l$ were equal across gender, the sex ratio observed at any point in time would be equal to the sex ratio at birth. Likewise, if men were the only gender migrating or subject to fatalities at the prime ages under analysis, such that $j_{rt}^{fc} = l_{rt}^{fc} = 0$, the difference between the observed sex ratios at time t and at birth would be given by the difference in male inflow and outflow rates.

[9] A similar exercise is undertaken in Card (2001), when estimating the impact of immigrants on native workers.

More generally, given that the sex ratio at birth is relatively homogeneous,[10] we can infer that heterogeneity in sex ratios at any moment in time results from the gender imbalance in population movements. Clearly, net population movements and their gender composition are endogenous. We thus need to find an instrument for the gender composition of population movements. Over the cohorts and period under analysis, out-migration largely predominated over immigration and natural population movements, as we show in more detail in the text. Therefore, the instrument should mostly aim at capturing the change in sex ratio driven by the gender composition of emigration. Instruments should as well tackle the endogeneity of the market size.

[10] We will assume that it is also exogenous (see James (1987) for a review of the literature on the causes of the slight variation in the human sex ratio at birth and Parazzini et al. (1998) for details on its range).

Appendix D – First-stage, reduced-form and OLS regressions

Table D.1: First-stage regressions for FLFP and marriage rates.

	own cohort	own cohort +-1 year	own cohort +-2 years	own cohort +-3 years
	(1)	(2)	(3)	(4)
A. Log Sex Ratio				
Emig. Gender Gap	-5.00***	-5.27***	-5.31***	-5.05***
	(0.997)	(1.21)	(1.28)	(1.39)
Emig. Rate	1.40***	1.42***	1.33***	0.850*
	(0.475)	(0.477)	(0.463)	(0.472)
B. Log Population				
Emig. Gender Gap	0.618	1.51	1.44	1.07
	(1.25)	(2.04)	(2.14)	(2.05)
Emig. Rate	-3.64**	-4.76**	-4.99**	-5.27**
	(1.58)	(2.06)	(2.11)	(2.03)
First-stage Sanderson-Windmeijer F-stats for participation model				
Sex Ratio	29.96	20.69	17.13	11.02
ln Pop	10.45	12.69	12.41	13.56
Kleibergen-Paap Wald rk LM Stat p-value				
	0.009	0.016	0.022	0.024
N	360	360	360	360

Notes: See notes to Table 3 in the main text.

Table D.2: First-stage regressions for occupational allocation.

	own cohort	own cohort +-1 year	own cohort +-2 years	own cohort +-3 years
	(1)	(2)	(3)	(4)
A. Log Sex Ratio				
Emig. Gender Gap	-4.95***	-5.32***	-5.37***	-5.20***
	(0.962)	(1.16)	(1.25)	(1.32)
Emig. Rate	1.25**	1.35***	1.27***	0.845*
	(0.540)	(0.463)	(0.456)	(0.480)
Pop Educ Gap	-0.001	0.007	0.007	0.008
	(0.014)	(0.011)	(0.006)	(0.007)
B. Log Employment				
Emig. Gender Gap	0.764	1.34	1.14	0.710
	(1.35)	(2.13)	(2.23)	(2.20)
Emig. Rate	-4.59***	-5.67***	-5.90***	-6.28***
	(1.60)	(2.17)	(2.25)	(2.27)
Pop Educ Gap	-0.048***	-0.038***	-0.035**	-0.027*
	(0.015)	(0.013)	(0.014)	(0.014)
C. Employment Education Gap				
Emig. Gender Gap	8.81**	11.88**	14.75**	16.90***
	(3.48)	(5.95)	(6.17)	(5.99)
Emig. Rate	-8.14***	-10.43***	-12.75***	-14.51***
	(2.58)	(3.71)	(3.79)	(3.64)
Pop Educ Gap	1.15***	1.15***	1.15***	1.16***
	(0.054)	(0.055)	(0.055)	(0.055)
First-stage Sanderson-Windmeijer F-stats for participation model				
Sex Ratio	36.46	26.23	25.97	24.24
ln Emp	10.90	11.51	12.07	13.78
Educ Gap	156.5	248.6	256.4	278.0
Kleibergen-Paap Wald rk LM Stat p-value				
	0.021	0.017	0.014	0.009
N	356	356	356	356

Notes: See notes to Table 5 in the main text.

Table D.3: First-stage regressions for gender pay gap.

	own cohort	own cohort +-1 year	own cohort +-2 years	own cohort +-3 years
	(1)	(2)	(3)	(4)
A. Log Sex Ratio				
Emig. Gender Gap	-5.50***	-5.64***	-5.73***	-5.55***
	(0.718)	(0.874)	(1.02)	(1.13)
Emig. Gender Gap x Female	0.411*	0.391	0.354	0.439*
	(0.248)	(0.282)	(0.287)	(0.260)
Emig. Rate	1.22**	1.26***	1.28***	0.797*
	(0.489)	(0.447)	(0.406)	(0.419)
Pop Educ	0.018	0.015*	0.012	0.013**
	(0.011)	(0.009)	(0.008)	(0.007)
Pop Educ x Female	-0.003	-0.003**	-0.002*	-0.002***
	(0.003)	(0.001)	(0.001)	(0.001)
B. Log Sex Ratio × Female				
Emig. Gender Gap	-1.35***	-1.02***	-0.968***	-0.719*
	(0.399)	(0.357)	(0.371)	(0.381)
Emig. Gender Gap x Female	-2.45***	-3.44***	-3.64***	-3.90***
	(0.489)	(0.574)	(0.706)	(0.788)
Emig. Rate	0.580**	0.638***	0.638***	0.393*
	(0.243)	(0.234)	(0.215)	(0.219)
Pop Educ	0.020**	0.011	0.009	0.010
	(0.009)	(0.010)	(0.009)	(0.010)
Pop Educ x Female	-0.010*	-0.010**	-0.009**	-0.010**
	(0.006)	(0.005)	(0.004)	(0.004)
C. Log Employment				
Emig. Gender Gap	-6.63***	-10.24***	-11.24***	-11.91***
	(2.14)	(3.08)	(3.32)	(3.44)
Emig. Gender Gap x Female	0.759	0.784	0.848	0.924
	(0.536)	(0.654)	(0.694)	(0.738)
Emig. Rate	-4.75***	-2.95*	-2.90*	-3.34*
	(1.52)	(1.62)	(1.67)	(1.73)
Pop Educ	0.032	0.024	0.023	0.023
	(0.021)	(0.019)	(0.018)	(0.018)
Pop Educ x Female	-0.005	-0.003	-0.003	-0.002
	(0.004)	(0.003)	(0.003)	(0.003)
D. Employment Education				
Emig. Gender Gap	9.21***	10.32***	10.98***	12.32***
	(2.07)	(2.86)	(3.00)	(3.28)
Emig. Gender Gap x Female	-6.59***	-8.53**	-10.28**	-11.84**
	(2.24)	(3.46)	(4.15)	(5.07)
Emig. Rate	2.12	2.44	3.17	3.71
	(2.26)	(2.95)	(3.39)	(3.60)
Pop Educ	0.262***	0.250***	0.246***	0.248***
	(0.063)	(0.059)	(0.057)	(0.056)
Pop Educ x Female	0.072	0.069	0.065	0.059
	(0.058)	(0.060)	(0.061)	(0.063)
E. Employment Education × Female				
Emig. Gender Gap	2.22	3.28	2.83	2.52
	(1.43)	(2.62)	(3.27)	(3.72)
Emig. Gender Gap x Female	4.44*	4.79	4.49	4.56
	(2.37)	(3.48)	(3.97)	(4.65)
Emig. Rate	0.614	0.074	0.915	1.66
	(1.29)	(1.74)	(2.07)	(2.18)
Pop Educ	-0.302***	-0.301***	-0.307***	-0.310***
	(0.056)	(0.053)	(0.053)	(0.053)
Pop Educ x Female	0.892***	0.893***	0.893***	0.893***
	(0.079)	(0.081)	(0.083)	(0.085)
First-stage Sanderson-Windmeijer F-stats for participation model				
Sex Ratio	88.19	76.74	43.00	28.39
Sex Ratio x Female	35.07	29.14	24.95	19.74
ln Emp	43.06	52.88	39.56	19.90
Emp Educ	34.09	42.91	37.22	32.09
Emp Educ x Female	71.88	96.99	95.89	82.36
Kleibergen-Paap Wald rk LM Stat p-value				
	0.017	0.012	0.019	0.034
N	720	720	720	720

Notes: See notes to Table 6 in the main text.

Validity of the instruments

In general, formally and credibly testing the exclusion restriction is a challenge. However, in the case of the marriage rate, we can use Stock-Wright S statistics to jointly test that our endogenous-variable parameters are equal to 0 *and* that our instruments are valid. The intuition behind tests based on Stock-Wright S statistics is that (like the Anderson-Rubin test) if the instruments are valid and the endogenous-variable parameters are 0, the reduced-form parameters should also be 0 (Andrews and Stock, 2005; Chernozhukov and Hansen, 2008). One of the benefits of these tests is that their statistics are weak-identification robust—they are valid even if the instruments are weak.[11]

Panel B of Table D.4 shows that the Stock-Wright S statistics (and Anderson-Rubin tests) do not reject the null hypothesis that the instruments are valid and the endogenous-variable parameters are 0 at standard significance levels for any of our specifications. One could argue that we failed to reject the null hypothesis because our instruments are weak. Indeed, if the instruments were irrelevant, Stock-Wright S statistics should never reject the null hypothesis. However, when we look at the FLFP, the Stock-Wright S statistics reject the null hypothesis for all specifications (Panel A of Table D.4). Since the first-stage regressions for the marriage market and the FLFP are the same, failing to reject the null hypothesis in the case of the marriage market is unlikely to be due to weak instruments.

In contrast with our marriage-rate findings, the sex-ratio and market-size parameters are statistically different from zero in regressions looking at the FLFP, occupational upgrading, and the gender wage gap. Therefore, Stock-Wright S statistics are expected to be uninformative regarding the validity of the instruments for these outcomes.[12]

[11] The main drawback of these tests is related to the interpretation of rejecting the null hypothesis. One can reject the null if the instruments do not satisfy the exclusion restriction or if the endogenous-variable parameters differ from zero.

[12] Recall that rejecting the null hypothesis of these tests would only suggest that our endogenous-variable parameters are different from 0 *or* that our instruments are invalid. Nevertheless, Stock-Wright S statistics and Anderson-Rubin test p-values are presented in Tables D.4 to D.6.

Table D.4: Reduced-form regressions for FLFP and marriage rates.

	RF own cohort	RF own cohort +-1 year	RF own cohort +-2 years	RF own cohort +-3 years
	(1)	(2)	(3)	(4)
A. Female participation				
Emig. Gender Gap	2.12***	3.06***	2.83**	2.93**
	(0.592)	(1.02)	(1.24)	(1.26)
Emig. Rate	-1.55***	-2.16***	-1.99**	-2.07**
	(0.496)	(0.690)	(0.831)	(0.857)
Anderson-Rubin Wald test χ^2_2 p-value				
	0.000	0.001	0.022	0.017
Stock-Wright LM S statistic χ^2_2 p-value				
	0.014	0.027	0.071	0.065
B. (Ever) Marriage rate				
Emig. Gender Gap	-0.200	-0.305	-0.344	-0.274
	(0.231)	(0.354)	(0.391)	(0.418)
Emig. Rate	0.232*	0.264	0.300	0.234
	(0.128)	(0.220)	(0.258)	(0.276)
Anderson-Rubin Wald test χ^2_2 p-value				
	0.138	0.429	0.456	0.652
Stock-Wright LM S statistic χ^2_2 p-value				
	0.228	0.485	0.496	0.660
N	360	360	360	360

Notes: See notes to Table 3 in the main text.

Table D.5: Reduced-form regressions for occupations.

	RF own cohort	RF own cohort +-1 year	RF own cohort +-2 years	RF own cohort +-3 years
	(1)	(2)	(3)	(4)
Female share in top occupations (managers, senior officials, legislators)				
Emig. Gender Gap	1.72	2.66*	2.23	2.31
	(1.07)	(1.33)	(1.41)	(1.55)
Emig. Rate	-0.706	-1.16	-0.818	-0.810
	(1.14)	(1.35)	(1.39)	(1.46)
Pop. Educ. Gap M-F	0.054*	0.055*	0.055*	0.054*
	(0.031)	(0.031)	(0.031)	(0.031)
Anderson-Rubin Wald test χ^2_3 p-value				
	0.106	0.092	0.164	0.195
Stock-Wright LM S statistic χ^2_3 p-value				
	0.255	0.228	0.288	0.319
N	356	356	356	356

Notes: See notes to Table 5 in the main text.

Table D.6: Reduced-form regressions for the wage gap.

	RF own cohort	RF own cohort +-1 year	RF own cohort +-2 years	RF own cohort +-3 years
	(1)	(2)	(3)	(4)
	Log hourly wages			
Emig. Gender Gap	0.713	0.827	0.869	1.01
	(0.435)	(0.552)	(0.541)	(0.641)
Emig. Gender Gap × Female	-0.977	-1.30	-1.49	-1.57
	(0.746)	(1.08)	(1.25)	(1.48)
Emig. Rate	-0.161	-0.151	-0.113	-0.133
	(0.344)	(0.478)	(0.496)	(0.516)
Pop Educ	0.078***	0.077***	0.076***	0.077***
	(0.016)	(0.017)	(0.016)	(0.017)
Pop Educ × Female	-0.006	-0.007	-0.007	-0.007
	(0.009)	(0.010)	(0.010)	(0.010)
Anderson-Rubin Wald test χ^2_5 p-value				
	0.000	0.000	0.000	0.000
Stock-Wright LM S statistic χ^2_5 p-value				
	0.063	0.067	0.067	0.069
N	720	720	720	720

Notes: See notes to Table 6 in the main text.

Table D.7: OLS regressions for FLFP and marriage rates by cohort-window width.

	OLS own cohort	OLS own cohort +-1 year	OLS own cohort +-2 years	OLS own cohort +-3 years
	(1)	(2)	(3)	(4)
		A. Female participation		
ln Sex Ratio	-0.044	-0.117**	-0.176**	-0.265***
	(0.026)	(0.047)	(0.064)	(0.083)
ln Pop	0.143***	0.214***	0.262***	0.276***
	(0.048)	(0.060)	(0.070)	(0.078)
		B. (Ever) Marriage rate		
ln Sex Ratio	0.010	0.024	0.012	0.030
	(0.010)	(0.020)	(0.030)	(0.041)
ln Pop	-0.012	-0.003	-0.011	-0.014
	(0.015)	(0.021)	(0.025)	(0.025)
N	360	360	360	360

Notes: See notes to Table 3 in the main text.

Table D.8: OLS regressions for occupational allocation.

	OLS own cohort	OLS own cohort +-1 year	OLS own cohort +-2 years	OLS own cohort +-3 years
	(1)	(2)	(3)	(4)
Female share in top occupations (managers, senior officials, legislators)				
ln Sex Ratio	-0.162**	-0.288**	-0.120	-0.065
	(0.067)	(0.112)	(0.134)	(0.162)
ln Emp	-0.019	-0.051	-0.122	-0.142
	(0.108)	(0.153)	(0.157)	(0.157)
Education Gap M-F	-0.036*	-0.036*	-0.034*	-0.034*
	(0.018)	(0.019)	(0.019)	(0.019)
N	356	356	356	356

Notes: See notes to Table 5 in the main text.

Table D.9: OLS regressions for the gender pay gap.

	OLS own cohort	OLS own cohort +-1 year	OLS own cohort +-2 years	OLS own cohort +-3 years
	(1)	(2)	(3)	(4)
		Log hourly wages		
ln Sex Ratio	0.047*	0.068	0.088	0.083
	(0.023)	(0.044)	(0.054)	(0.071)
ln Sex Ratio x Female	-0.043	-0.055	-0.080	-0.071
	(0.040)	(0.096)	(0.127)	(0.152)
ln Emp	0.089***	0.088**	0.089**	0.090**
	(0.031)	(0.033)	(0.036)	(0.039)
Educ	0.161***	0.161***	0.162***	0.161***
	(0.019)	(0.020)	(0.020)	(0.020)
Educ x Female	-0.012**	-0.012**	-0.012**	-0.012*
	(0.006)	(0.006)	(0.006)	(0.006)
Female	-0.127***	-0.128***	-0.130***	-0.130***
	(0.036)	(0.038)	(0.038)	(0.040)
N	720	720	720	720

Notes: See notes to Table 6 in the main text.

Appendix E – Constructing a standard shift-share instrument

Adão et al. (2019) show that standard errors based on shift-share instruments can be biased. However, Borusyak et al. (2022) show that the same SSIV parameter estimates can be obtained by running the regressions at the original data level (say, region-cohort) or at the shock level (e.g., the cohort level).[13] Borusyak et al. (2022) show that shock-level regressions will produce appropriate standard errors.

In this section, we illustrate how we can modify our instruments to construct standard shift-share instruments. We use Borusyak et al. (2022)'s notation with some relabeling to make the connection with our setup clearer. In their model, they have industries (n), locations (l that we relabel r), and time periods (t that we relabel c), in the panel version of the model. Following this notation, their panel data shift-share instrument can be written as

$$z_{rc} = \sum_n s_{rnc} g_{nc} \tag{6}$$

where s_{rnc} are the shares and g_{nc} the (national-level) shocks for each industry.

In our case, we have only one industry, and our instrument leverages on the cross-cohort shocks, and our pure shift-share instrument should take the following form:[14]

$$z_{rc} = s_{rc} g_c.$$

We first show that our instruments differ slightly from these typical shift-share instruments. Our leave-out own region emigration gender-gap instrument can be written as :

$$emig\_gender_r^c = \frac{\left[I_{75}^{mc}\left(1 - \frac{E_{r,50s}}{\sum_r E_{r,50s}}\right)\right] \times \frac{E_{r,50s}}{\sum_r E_{r,50s}}}{B_r^{mc}} - \frac{\left[I_{75}^{fc}\left(1 - \frac{E_{r,50s}}{\sum_r E_{r,50s}}\right)\right] \times \frac{E_{r,50s}}{\sum_r E_{r,50s}}}{B_r^{fc}} \tag{7}$$

$$= \left[\left(1 - \frac{E_{r,50s}}{\sum_r E_{r,50s}}\right) \times \frac{E_{r,50s}}{\sum_r E_{r,50s}}\right]\left[\frac{I_{75}^{mc}}{B_r^{mc}} - \frac{I_{75}^{fc}}{B_r^{fc}}\right]. \tag{8}$$

The first term in square brackets varies only across regions and represents the share term of a typical shift-share instrument. However, we can see that the second term in squared brackets (the shocks) varies across cohorts and regions. So, without further restrictions, our instrument cannot be written as a 'pure' shift-share instrument.

[13] This equivalence result does not directly generalize to when one combines SSIVs and other instruments, like in our occupational upgrading and wage gap regressions.

[14] Nunn and Qian (2014) use a similar instrument where they use time-series variation in wheat production as shocks and variation countries' historical tendency to receive food aid from the U.S. as shares.

One possible restriction that would allow us to express our emigration gender-gap instrument as a pure shift-share instrument is to assume that the sex ratio is constant across region (say 1.07, the average sex ratio at birth in our data). In this case, $B_r^{mc} = 1.07 B_r^{fc}$ and we can simplify the expression in equation (8) as

$$emig\_gender_r^c = \left[\frac{1}{1.07 B_r^{fc}} \left(1 - \frac{E_{r,50s}}{\sum\limits_r E_{r,50s}} \right) \times \frac{E_{r,50s}}{\sum\limits_r E_{r,50s}} \right] \left[I_{75}^{mc} - 1.07 I_{75}^{fc} \right]. \quad (9)$$

Doing so, the second term in bracket now varies only across cohorts. This way $emig\_gender_r^c$ can be expressed in the form $z_{rc} = s_{rc} g_c$.

Similarly, the overall emigration rate can be written:

$$Erate_r^c = \frac{\left[I_{75}^{mc} \left(1 - \frac{E_{r,50s}}{\sum_r E_{r,50s}} \right) \right] \times \frac{E_{r,50s}}{\sum_r E_{r,50s}} + \left[I_{75}^{fc} \left(1 - \frac{E_{r,50s}}{\sum_r E_{r,50s}} \right) \right] \times \frac{E_{r,50s}}{\sum_r E_{r,50s}}}{B_r^{mc} + B_r^{fc}} \quad (10)$$

$$= \left[\frac{\left(1 - \frac{E_{r,50s}}{\sum_r E_{r,50s}} \right) \times \frac{E_{r,50s}}{\sum_r E_{r,50s}}}{B_r^{mc} + B_r^{fc}} \right] \times \left[I_{75}^{mc} + I_{75}^{fc} \right]. \quad (11)$$

In the case of $Erate_r^c$, we do not have to make any restriction to write it in the form $z_{rc} = s_{rc} g_c$.

Once we have instruments that can be written as standard SSIVs, we can follow Borusyak et al. (2022)'s methodology to estimate our regressions FLFP and the marriage rate at the shock (i.e., cohort) level to get valid standard errors.

Online Appendix Table F.15 presents the results from estimating our regressions at the cohort (shock) level using our modified instrument on the gender composition of emigration ($emig\_gender_r^c$).[15] Despite using different instruments, the SSIV parameter estimates for the FLFP and marriage rate models are very similar to our main IV estimates. Importantly, our parameters for the FLFP regressions remain significant at 1%. The parameter estimates for the marriage regressions remain close to zero and statistically insignificant.

Emigration rates and region characteristics

Columns (1) to (4) of Table E.1 presents regressions of the 1950s emigrations shares and 1950 region characteristics. Column (5) regresses the average change in sex ratio (between birth and 1981) of each regression on their emigration shares.

[15] Since we have two SSIVs with different share-exposure variables in our regressions, we must choose one of the two for estimation purposes. We have used the share exposure variable for $emig\_gender_r^c$ in Online Appendix Table F.15. However, using the share exposure variable for $Erate_r^c$ yields very similar results (Online Appendix Table F.16).

Table E.1: Emigration Rates and Region Characteristics.

Dependent Var.	Emig. (%) (1)	Emig. (%) (2)	Emig. (%) (3)	Emig. (%) (4)	1981 R Change (5)
Pop. (000s)	-0.004*				
	(0.002)				
Agriculture (%)		0.053			
		(0.032)			
Literacy (%)			0.027		
			(0.119)		
Activity (%)				-0.506**	
				(0.177)	
Emigration (%)					-0.011**
					(0.004)
Constant	5.303***	1.184	2.316	29.143***	-0.096***
	(1.037)	(1.569)	(6.945)	(8.913)	(0.012)
Mean Dep. Var.	3.91	3.91	3.91	3.91	-0.14
N	20	20	20	20	20

Notes: The region populations, agriculture shares, literacy rates, and activity rates are measured in 1950 while the emigration rates are the measured over the 1950s. "1981 R Change" is the difference between the sex ratio at birth and the 1981 sex ratio. Heteroskedasticity-robust standard-errors are in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix F – Robustness checks

Table F.1: Sex ratios and married females (male cohort centered at -2).

	OLS own cohort	IV own cohort	IV own cohort +-1 year	IV own cohort +-2 years	IV own cohort +-3 years
	(1)	(2)	(3)	(4)	(5)
B. (Ever) Marriage rate (male cohort centered at -2)					
ln Sex Ratio	-0.001	-0.012	0.070	0.096	0.105
	(0.009)	(0.120)	(0.069)	(0.118)	(0.173)
ln Pop	-0.019	-0.054*	-0.038	-0.042	-0.036
	(0.018)	(0.029)	(0.038)	(0.040)	(0.040)
First-stage Sanderson-Windmeijer F-stats					
ln Sex Ratio		8.80	11.30	11.77	9.63
ln Pop		8.20	10.11	7.71	8.20
N	360	360	360	360	360

Notes: Under "+- j cohorts", all explanatory variables for cohort c are computed using cohorts c-j to c+j for females, and c-j-2 to c+j-2 for males. See notes to Table 3 in the main text.

Table F.2: Sex ratio and female divorce rate.

	OLS own cohort	IV own cohort	IV own cohort +-1 year	IV own cohort +-2 years	IV own cohort +-3 years
	(1)	(2)	(3)	(4)	(5)
ln Sex Ratio	0.002	0.012	0.021	0.022	0.027
	(0.005)	(0.015)	(0.016)	(0.015)	(0.018)
ln Pop	-0.003	-0.009	-0.009	-0.013	-0.020
	(0.004)	(0.020)	(0.016)	(0.017)	(0.018)
First-stage Sanderson-Windmeijer F-stats					
Sex Ratio		29.96	20.69	17.13	11.02
ln Pop		10.45	12.69	12.41	13.56
N	360	360	360	360	360

Notes: See notes to Table 3 in the main text.

Table F.3: Sex ratios and the gender share gap in top-occupations (relative to population).

	OLS own cohort	IV own cohort	IV own cohort +-1 year	IV own cohort +-2 years	IV own cohort +-3 years
	(1)	(2)	(3)	(4)	(5)
Gender share gap in top occupations (managers, senior officials, legislators)					
ln Sex Ratio	0.012	0.135***	0.197***	0.208***	0.230***
	(0.011)	(0.043)	(0.056)	(0.066)	(0.076)
ln Pop	-0.024	-0.088***	-0.100***	-0.103***	-0.111***
	(0.014)	(0.027)	(0.027)	(0.026)	(0.028)
Education Gap M-F	0.004	0.007**	0.006*	0.006*	0.006*
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
First-stage Sanderson-Windmeijer F-stats					
ln Sex Ratio		40.67	26.37	25.99	23.47
ln Pop		9.93	11.28	11.92	13.83
Education Gap		113.89	206.61	223.77	249.1
N	360	360	360	360	360

Notes: See notes to Table 5 in the main text.

Table F.4: RETORNADOS FROM AFRICA, EMIGRATION RATE AND ITS GENDER COMPOSITION (OLS).

	own cohort	own cohort +-1 year	own cohort +-2 years	own cohort +-3 years
	(1)	(2)	(3)	(4)
Emig. Gender Gap	3.688	5.946	4.040	4.762
	(4.016)	(5.841)	(6.227)	(5.968)
Emig. Rate	1.523	0.133	1.359	0.656
	(4.136)	(3.930)	(3.882)	(3.442)
N	353	360	360	360

Notes: The dependent variable is the sex ratio of the retornados. Under "+- j cohorts", all explanatory variables for cohort c are computed using cohorts c-j to c+j. For the regression on own cohort, seven observations are missing since seven cohort-region cells have no female returnees. The gender composition of emigration is the rate of male emigration minus the rate of female emigration, either one over the respective birth cohort. The rate of emigration is the number of emigrants over the respective birth cohort. All regressions include sets of dummy variables for cohort and region. The cell-size analytic weight is the number of females in the population. Standard-errors clustered at the region level, in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table F.5: Sex ratio, employment and unemployment

	OLS own cohort	IV own cohort	IV own cohort +-1 year	IV own cohort +-2 years	IV own cohort +-3 years
	(1)	(2)	(3)	(4)	(5)
			A. Female employment rate		
ln Sex Ratio	-0.030	-0.357***	-0.460***	-0.433**	-0.499**
	(0.024)	(0.112)	(0.144)	(0.176)	(0.203)
ln Pop	0.150***	0.341**	0.366***	0.335***	0.367***
	(0.044)	(0.134)	(0.119)	(0.121)	(0.128)
First-stage Sanderson-Windmeijer F-stats					
Sex Ratio		29.96	20.69	17.13	11.02
ln Pop		10.45	12.69	12.41	13.56
			B. Female unemployment rate		
ln Sex Ratio	-0.022	0.040	0.095	0.105	0.134
	(0.018)	(0.068)	(0.089)	(0.114)	(0.141)
ln Pop	-0.004	-0.098*	-0.105*	-0.098	-0.104
	(0.025)	(0.057)	(0.058)	(0.064)	(0.068)
First-stage Sanderson-Windmeijer F-stats					
Sex Ratio		29.96	20.69	17.13	11.02
ln Pop		10.45	12.69	12.41	13.56
N	360	360	360	360	360

Notes: See notes to Table 3 in the main text.

Table F.6: Sex ratio, gender gaps in labor force participation and marriage rates.

	OLS own cohort	IV own cohort	IV own cohort +-1 year	IV own cohort +-2 years	IV own cohort +-3 years
	(1)	(2)	(3)	(4)	(5)
			A. Labor force participation gap		
ln Sex Ratio	0.041	0.319***	0.439***	0.417**	0.448**
	(0.031)	(0.096)	(0.133)	(0.168)	(0.185)
ln Pop	-0.110**	-0.191	-0.237*	-0.217*	-0.246*
	(0.049)	(0.135)	(0.126)	(0.131)	(0.137)
First-stage Sanderson-Windmeijer F-stats					
Sex Ratio		33.66	21.27	17.47	11.33
ln Pop		9.23	11.45	11.57	12.45
			B. (Ever) Marriage rate gap		
ln Sex Ratio	0.007	-0.027	-0.048	-0.059	-0.054
	(0.012)	(0.060)	(0.063)	(0.063)	(0.070)
ln Pop	0.013	0.075	0.076	0.080	0.067
	(0.018)	(0.064)	(0.068)	(0.071)	(0.070)
First-stage Sanderson-Windmeijer F-stats					
Sex Ratio		33.66	21.27	17.47	11.33
ln Pop		9.23	11.45	11.57	12.45
N	360	360	360	360	360

Notes: See notes to Table 3 in the main text.

Table F.7: Sex ratio, female marriage and labor market outcomes, controlling for regional-specific cohort trends (own-cohort).

	FLFP		Ever Married		Top Occupation		Wages	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln Sex Ratio	-0.390***	-0.378*	0.034	-0.054	-0.405*	-0.504	0.167	0.240*
	(0.114)	(0.211)	(0.050)	(0.082)	(0.210)	(0.832)	(0.127)	(0.140)
ln Sex Ratio x Female	-	-	-	-	-	-	-0.267	-0.406**
	-	-	-	-	-	-	(0.190)	(0.168)
First-stage Sanderson-Windmeijer F-stats								
ln Sex Ratio	29.96	8.18	29.96	8.18	36.46	6.22	88.19	54.84
ln Sex Ratio x Female	-	-	-	-	-	-	35.07	35.00
ln Pop	10.45	2.29	10.45	2.29	-	-		
ln Emp	-	-	-	-	10.90	2.06	43.06	27.84
Educ gap	-	-	-	-	156.5	7.51	-	-
Educ	-	-	-	-	-	-	34.09	21.87
Educ x Female	-	-	-	-	-	-	71.88	31.87
Region-specific cohort trends	No	Yes	No	Yes	No	Yes	No	Yes
N	360	360	360	360	356	356	720	720

Notes: We treat as endogenous: the sex ratio, the population/employment size, and the education of employed workers (when we control for the education of employed workers). Instruments: gender composition of emigration, rate of emigration, and education of the population in the cell. The gender composition of emigration is the rate of male emigration minus the rate of female emigration, either one over the respective birth cohort. The rate of emigration is the number of emigrants over the respective birth cohort. All regressions include sets of dummy variables for cohort and region, and the interactions of region fixed effects with a cohort trend. The cell-size analytic weight are the female population (in columns 1-4) and the employment level (in columns 5-8). Standard-errors clustered at the region level, in parenthesis.
* significant at 10%; ** significant at 5%; *** significant at 1%.

Table F.8: Sex ratio, female marriage and labor market outcomes, controlling for regional-characteristics interacted with cohort fixed effects (own-cohort).

	FLFP		Ever Married		Top Occupation		Wages	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln Sex Ratio	-0.390***	-0.228*	0.034	-0.021	-0.405*	-0.676***	0.167	0.209*
	(0.114)	(0.122)	(0.050)	(0.054)	(0.210)	(0.229)	(0.127)	(0.110)
ln Sex Ratio x Female	-	-	-	-	-	-	-0.267	-0.326*
	-	-	-	-	-	-	(0.190)	(0.167)
First-stage Sanderson-Windmeijer F-stats								
ln Sex Ratio	29.96	15.71	29.96	15.71	36.46	16.50	88.19	17.84
ln Sex Ratio x Female	-	-	-	-	-	-	35.07	30.32
ln Pop	10.45	4.41	10.45	4.41	-	-		
ln Emp	-	-	-	-	10.90	7.97	43.06	50.45
Educ gap	-	-	-	-	156.5	123.4	-	-
Educ	-	-	-	-	-	-	34.09	25.64
Educ x Female	-	-	-	-	-	-	71.88	48.63
Regional characteristics interacted with cohort F.E.	No	Yes	No	Yes	No	Yes	No	Yes
N	360	360	360	360	356	356	720	720

Notes: We treat as endogenous: the sex ratio, the population/employment size, and the education of employed workers (when we control for the education of employed workers). Instruments: gender composition of emigration, rate of emigration, and education of the population in the cell. The gender composition of emigration is the rate of male emigration minus the rate of female emigration, either one over the respective birth cohort. The rate of emigration is the number of emigrants over the respective birth cohort. All regressions include sets of dummy variables for cohort and region, and the interactions of cohort fixed effects with the following region attributes in 1960: share of employment in agriculture, the activity rate, and unemployment rate. The cell-size analytic weight are the female population (in columns 1-4) and the employment level (in columns 5-8). Standard-errors clustered at the region level, in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%.

Alternative standard-error computation methods

Table F.9: FLFP and marriage rates results, bootstrapping.

	OLS own cohort	IV own cohort	IV own cohort +-1 year	IV own cohort +-2 years	IV own cohort +-3 years
	(1)	(2)	(3)	(4)	(5)
			A. Female participation		
ln Sex Ratio	-0.044	-0.390***	-0.492***	-0.457***	-0.515***
	[0.102]	[0.001]	[0.001]	[0.009]	[0.007]
ln Pop	0.143***	0.275**	0.308***	0.278**	0.310***
	[0.008]	[0.028]	[0.005]	[0.014]	[0.009]
Wild-bootstrap Wald test (Prob> $\|t\|$)					
Sex Ratio	0.088	0.007	0.004	0.021	0.021
ln Pop	0.032	0.034	0.014	0.036	0.019
			B. (Ever) Marriage rate		
ln Sex Ratio	0.010	0.034	0.046	0.052	0.046
	[0.306]	[0.502]	[0.452]	[0.424]	[0.528]
ln Pop	-0.012	-0.051	-0.042	-0.046	-0.037
	[0.403]	[0.110]	[0.203]	[0.201]	[0.342]
Wild-bootstrap Wald test (Prob> $\|t\|$)					
Sex Ratio	0.285	0.539	0.551	0.535	0.620
ln Pop	0.375	0.107	0.232	0.242	0.384
N	360	360	360	360	360

Notes: See notes to Table 3 in the main text. P-values based on clustered standard errors at the region level (our main estimation strategy) are in brackets. Wild-bootstrap Wald test are based on t_{19} obtained from 999 bootstrap replications, imposing the null hypothesis, clustering at the region level and using Rademacher weights.

Table F.10: Occupation results, bootstrapping.

	OLS own cohort	IV own cohort	IV own cohort +-1 year	IV own cohort +-2 years	IV own cohort +-3 years
	(1)	(2)	(3)	(4)	(5)
Female share in top occupations (managers, senior officials, legislators)					
ln Sex Ratio	-0.162**	-0.405*	-0.550**	-0.502*	-0.552
	[0.026]	[0.054]	[0.042]	[0.083]	[0.106]
ln Emp	-0.019	0.119	0.146	0.118	0.146
	[0.865]	[0.567]	[0.478]	[0.560]	[0.472]
Education Gap M-F	-0.036*	-0.042*	-0.039	-0.041*	-0.040*
	[0.063]	[0.087]	[0.123]	[0.094]	[0.094]
Wild-bootstrap Wald test (Prob> $\|t\|$)					
ln Sex Ratio	0.019	0.112	0.074	0.082	0.142
ln Emp	0.898	0.600	0.483	0.539	0.479
N	356	356	356	356	356

Notes: See notes to Table 5 in the main text. P-values based on clustered standard errors at the region level (our main estimation strategy) are in brackets. Wild-bootstrap Wald test are based on t_{19} obtained from 999 bootstrap replications, imposing the null hypothesis, clustering at the region level and using Rademacher weights.

Table F.11: Wages results, bootstrapping.

	OLS own cohort	IV own cohort	IV own cohort +-1 year	IV own cohort +-2 years	IV own cohort +-3 years
	(1)	(2)	(3)	(4)	(5)
			ln hourly wages		
ln Sex Ratio	0.047*	0.167	-0.009	-0.046	-0.092
	[0.052]	[0.187]	[0.970]	[0.862]	[0.794]
ln Sex Ratio x Female	-0.043	-0.267	-0.268	-0.331*	-0.405*
	[0.304]	[0.160]	[0.146]	[0.096]	[0.064]
ln Emp	0.089***	0.155***	0.206**	0.227**	0.257**
	[0.009]	[0.009]	[0.022]	[0.013]	[0.033]
Educ	0.161***	0.257***	0.263***	0.266***	0.271***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Educ x Female	-0.012**	-0.029***	-0.030***	-0.030***	-0.030***
	[0.049]	[0.001]	[0.001]	[0.001]	[0.001]
Female	-0.127***	-0.017	-0.008	-0.010	-0.010
	[0.002]	[0.749]	[0.879]	[0.855]	[0.856]
Wild-bootstrap Wald test (Prob> $\|t\|$)					
ln Sex Ratio	0.120	0.239	0.972	0.896	0.867
ln Sex Ratio x Female	0.470	0.222	0.200	0.121	0.088
ln Emp	0.041	0.026	0.103	0.072	0.110
N	720	720	720	720	720

Notes: See notes to Table 6 in the main text. P-values based on clustered standard errors at the region level (our main estimation strategy) are in brackets. Wild-bootstrap Wald test are based on t_{19} obtained from 999 bootstrap replications, imposing the null hypothesis, clustering at the region level and using Rademacher weights.

Table F.12: Sex ratio, female labor force participation, and marriage rate (two-way clustering).

	OLS	IV own coh	IV own+-1	IV own+-2	IV own+-3
	(1)	(2)	(3)	(4)	(5)
			A. Female participation		
ln Sex Ratio	-0.044*	-0.390***	-0.492***	-0.457***	-0.515***
	(0.022)	(0.124)	(0.153)	(0.170)	(0.184)
ln Pop	0.143**	0.275**	0.308***	0.278**	0.310***
	(0.049)	(0.132)	(0.106)	(0.117)	(0.119)
First-stage Sanderson-Windmeijer F-stats					
Sex Ratio		12.38	14.41	16.82	12.49
ln Pop		7.72	12.28	12.54	13.79
			B. (Ever) Marriage rate		
ln Sex Ratio	0.010	0.034	0.046	0.052	0.046
	(0.012)	(0.058)	(0.064)	(0.072)	(0.081)
ln Pop	-0.012	-0.051	-0.042	-0.046	-0.037
	(0.012)	(0.043)	(0.046)	(0.050)	(0.050)
First-stage Sanderson-Windmeijer F-stats					
Sex Ratio		12.38	14.41	16.82	12.49
ln Pop		7.72	12.28	12.54	13.79
N	360	360	360	360	360

Notes: Standard-errors two-way clustered at the region and cohort levels, in parenthesis. See notes to Table 3 in the main text.

Table F.13: Sex ratios and occupational allocation (two-way clustering).

	OLS	IV own coh	IV own+-1	IV own+-2	IV own+-3
	(1)	(2)	(3)	(4)	(5)
Female share in top occupations (managers, senior officials, legislators)					
ln Sex Ratio	-0.162**	-0.405**	-0.550***	-0.502**	-0.552**
	(0.060)	(0.179)	(0.212)	(0.210)	(0.241)
ln Emp	-0.019	0.119	0.146	0.118	0.146
	(0.113)	(0.179)	(0.184)	(0.177)	(0.164)
Education Gap M-F	-0.036**	-0.042*	-0.039	-0.041*	-0.040*
	(0.014)	(0.024)	(0.024)	(0.023)	(0.022)
First-stage Sanderson-Windmeijer F-stats					
ln Sex Ratio		15.97	20.73	25.25	19.78
ln Emp		9.90	11.75	12.52	14.45
Education Gap		246.7	344.7	335.7	304.6
N	356	356	356	356	356

Notes: Standard-errors two-way clustered at the region and cohort levels, in parenthesis. See notes to Table 5 in the main text.

Table F.14: Sex ratio and the gender pay gap (two-way clustering).

	OLS	IV own coh	IV own+-1	IV own+-2	IV own+-3
	(1)	(2)	(3)	(4)	(5)
Log hourly wages					
ln Sex Ratio	0.047**	0.167	-0.009	-0.046	-0.092
	(0.022)	(0.160)	(0.281)	(0.304)	(0.386)
ln Sex Ratio x Female	-0.043	-0.267	-0.268	-0.331	-0.405
	(0.036)	(0.243)	(0.228)	(0.242)	(0.247)
ln Emp	0.089***	0.155***	0.206**	0.227**	0.257**
	(0.032)	(0.059)	(0.098)	(0.097)	(0.126)
Educ	0.161***	0.257***	0.263***	0.266***	0.271***
	(0.017)	(0.033)	(0.033)	(0.030)	(0.027)
Educ x Female	-0.012**	-0.029***	-0.030***	-0.030***	-0.030***
	(0.005)	(0.009)	(0.009)	(0.009)	(0.009)
Female	-0.127***	-0.017	-0.008	-0.010	-0.010
	(0.031)	(0.051)	(0.055)	(0.055)	(0.057)
First-stage Sanderson-Windmeijer F-stats					
ln Sex Ratio		26.03	66.53	48.59	32.66
ln Sex Ratio x Female		9.58	.	81.38	28.16
ln Emp		9.35	.	21.58	14.81
Educ		27.43	.	115.1	25.35
Educ x Female		71.23	.	187.5	84.51
N	720	720	720	720	720

Notes: Standard-errors two-way clustered at the region and cohort levels, in parenthesis. See notes to Table 6 in the main text.

Table F.15: Sex ratio, female labor force participation, and marriage rate using Borusyak et al. (2022)'s Methodology.

	OLS own cohort	IV own cohort	IV own cohort +-1 year	IV own cohort +-2 years	IV own cohort +-3 years
	(1)	(2)	(3)	(4)	(5)
			A. Female participation		
ln Sex Ratio	-0.201***	-0.414***	-0.530***	-0.505***	-0.558***
	(0.040)	(0.157)	(0.134)	(0.098)	(0.098)
ln Pop	0.142	0.273	0.256**	0.204*	0.244**
	(0.108)	(0.188)	(0.126)	(0.110)	(0.109)
First-stage Sanderson-Windmeijer F-stats					
Sex Ratio		9.82	30.78	50.39	72.73
ln Pop		21.47	116.62	257.91	212.57
			B. (Ever) Marriage rate		
ln Sex Ratio	0.027	0.032	0.050	0.051	0.044
	(0.031)	(0.047)	(0.065)	(0.072)	(0.078)
ln Pop	0.007	-0.056	-0.048	-0.053	-0.044
	(0.046)	(0.062)	(0.058)	(0.056)	(0.052)
First-stage Sanderson-Windmeijer F-stats					
Sex Ratio		9.82	30.78	50.39	72.73
ln Pop		21.47	116.62	257.91	212.57
N	18	18	18	18	18

Notes: Under "+- j years", all explanatory variables for cohort c are computed using cohorts c-j to c+j. We treat as endogenous: the sex ratio and the population size. Instruments: modified gender composition of emigration and rate of emigration (both excluding own region) so that they are standard SSIVs, as in Borusyak et al. (2022). The exposure-share variable is based on Online Appendix equation (9). All regressions control for the average number of years of education as well as region and cohort fixed effects. The cell-size analytic weight is the number of females in the population. Heteroskedasticity-robust standard-errors in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table F.16: Sex ratio, female labor force participation, and marriage rate using Borusyak et al. (2022)'s Methodology (Alternative Exposure-Share Variable).

	OLS own cohort	IV own cohort	IV own cohort +-1 year	IV own cohort +-2 years	IV own cohort +-3 years
	(1)	(2)	(3)	(4)	(5)
			A. Female participation		
ln Sex Ratio	-0.200***	-0.419***	-0.531***	-0.506***	-0.558***
	(0.039)	(0.161)	(0.137)	(0.099)	(0.099)
ln Pop	0.140	0.283	0.264**	0.213**	0.250**
	(0.105)	(0.190)	(0.124)	(0.107)	(0.107)
First-stage Sanderson-Windmeijer F-stats					
Sex Ratio		9.40	29.92	48.73	71.02
ln Pop		24.41	122.22	323.84	220.52
			B. (Ever) Marriage rate		
ln Sex Ratio	0.028	0.034	0.054	0.054	0.046
	(0.031)	(0.048)	(0.066)	(0.073)	(0.079)
ln Pop	0.003	-0.055	-0.049	-0.053	-0.043
	(0.046)	(0.059)	(0.056)	(0.055)	(0.050)
First-stage Sanderson-Windmeijer F-stats					
Sex Ratio		9.40	29.92	48.73	71.02
ln Pop		24.41	122.22	323.84	220.52
N	18	18	18	18	18

Notes: Under "+- j years", all explanatory variables for cohort c are computed using cohorts c-j to c+j. We treat as endogenous: the sex ratio and the population size. Instruments: modified gender composition of emigration and rate of emigration (both excluding own region) so that they are standard SSIVs, as in Borusyak et al. (2022). The exposure-share variable is based on Online Appendix equation (11). All regressions control for the average number of years of education as well as region and cohort fixed effects. The cell-size analytic weight is the number of females in the population. Heteroskedasticity-robust standard-errors in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%.

References

Adão, R., Kolesár, M., and Morales, E. (2019). Shift-share designs: Theory and inference. *Quarterly Journal of Economics*, 134(4):1949–2010.

Adão, Á. and Remédios, M. J. (2009). O alargamento da escolaridade obrigatória para as meninas portuguesas (1960), uma medida legislativa envergonhada: Sua representação nos jornais. *Revista HISTEDBR On-Line*, 9(36).

Andrews, D. W. and Stock, J. H. (2005). Inference with weak instruments. Technical Working Paper 313, National Bureau of Economic Research.

Baganha, M. I. B. (1994). As correntes emigratórias portuguesas no século XX e o seu impacto na economia nacional. *Análise Social*, 29(128):959–980.

Borusyak, K., Hull, P., and Jaravel, X. (2022). Quasi-experimental shift-share research designs. *Review of Economic Studies*, 89(1):181–213.

Caillods, F. and Rocha, A. (1983). *A Mulher no Ensino e no Emprego: Análise dos Modos Evolutivos da Sua Situação*. Ministério da Educação, GEP, Lisboa, Portugal.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64.

Chernozhukov, V. and Hansen, C. (2008). The reduced form: A simple approach to inference with weak instruments. *Economics Letters*, 100(1):68–71.

Easterlin, R. A. (1981). Why isn't the whole world developed? *Journal of Economic History*, 41(1):1–19.

European Environment Agency (2022). Portugal shapefile. `www.eea.europa.eu`. Accessed: December 20, 2022.

Ferreira, V. (2011). Engendering Portugal: Social change, state politics and women's social mobilization. In Pinto, A. C., editor, *Contemporary Portugal: Politics, Society, and Culture*, pages 153–192. Columbia University Press, New York, NY, 2nd edition.

International Labour Organization (1969). *International Standard Classification of Occupations*. International Labour Organization, Geneva, Switzerland, revised 1968 edition.

James, W. H. (1987). The human sex ratio. Part 1: A review of the literature. *Human Biology*, 59(5):721–752.

Minnesota Population Center (2015). *Integrated Public Use Microdata Series, International: Version 6.4 [Machine-readable database]*. University of Minnesota, Minneapolis, MN.

Nunn, N. and Qian, N. (2014). US food aid and civil conflict. *American Economic Review*, 104(6):1630–66.

Parazzini, F., La Vecchia, C., Levi, F., and Franceschi, S. (1998). Trends in male: female ratio among newborn infants in 29 countries from five continents. *Human Reproduction*, 13(5):1394–1396.

Pordata (2017). Base de dados Portugal contemporâneo. www.pordata.pt. Accessed: April 14, 2017.

Portugal, Direção-Geral do Território (2016). Áreas das freguesias, concelhos e distritos. Available from http://www.dgterritorio.pt. Accessed October 11, 2016.

Portugal, Instituto Nacional de Estatística (2016). Sistema de metainformação. Available from www.ine.pt. Accessed October 6, 2016.

Reis, J. (1993). *O Atraso Económico Português em Perspectiva Histórica: Estudos sobre a Economia Portuguesa na Segunda Metade do Século XIX, 1850-1930*. Imprensa Nacional Casa da Moeda, Lisboa.

Souza, M. R. d. (2013). *A Concessão do Voto às Portuguesas: Breve Apontamento*. Comissão para a Cidadania e Igualdade de Género, Lisboa, Portugal, 2nd edition.

Spain, Instituto Nacional de Estadística (2017). Number of births by gender and province, 1900-1980. In Demographic Statistics (INEBase), 1941 to 1982; Statistical Yearbook, 1912 to 1948. Available from www.ine.es digital library. Accessed May 1-5, 2017.

TerraWeb (2016). Dos veteranos da guerra do ultramar. Available from http://ultramar.terraweb.biz/index_MortosGuerraUltramar_Portugal.htm. Accessed November 18, 2016.