Project Module in OSE Data Science

(Summer Semester 2021)

July 23, 2021

Minh-Tri Hoang | M.Sc. Economics, University of Bonn

This notebook contains my replication for the following paper: (Abramitzky, Delavande, & Vasconcelos, 2011). "Marrying Up: The Role of Sex Ratio in Assortative Matching". American Economic Journal: Applied Economics, 3(3), 124-57.

Information about replication and transparency:

- For the replication, I try to remain the original structure of the paper so that readers can easily follow. Tables and figures are labeled and named as those in (Abramitzky et al., 2011).

- My Python scripts, which are used to replicate main results of the paper, can also be found on the auxiliary folder of my repository here.

- The original data and Stata do files used in the paper can be accessed here.

Contents

```
[1]:  import pandas as pd
      import numpy as np
      import math
      import geopandas as gpd
      from geopandas import GeoDataFrame as gdf
      import matplotlib.pyplot as plt
```

```
import matplotlib.lines as mlines
import matplotlib.transforms as mtransforms
from pandas.core.common import flatten
import statsmodels.formula.api as smf
from statsmodels.formula.api import ols
from statsmodels.iolib.summary2 import summary_col
from linearmodels.iv import IV2SLS
from IPython.display import Image
from IPython.display import display, HTML
```

```
[2]: geo = gpd.read_file("data/FRA_adm/FRA_adm2.shp")
     df = pd.read_stata("data/marriage_level.dta")
     df1 = pd.read_stata("data/departement_level.dta")
     pd.options.mode.chained_assignment = None
```

1 Introduction

Understanding how sex ratio, the proportion of men to women in the population, affects the position of men (women) in marriage markets can help economists grasp insights into social inequality, labor supply, income redistribution, and education. To shed some light on the issue, (Abramitzky et al., 2011) investigate the impact of male scarcity on marital assortative matching and marriage market outcomes. The authors take into account World War I (WWI) as an exogenous shock that gives rise to military casualties in France. They adopt the difference-in-difference approach to investigate the impact of war on marriage outcome patterns, considered as the first difference. Meanwhile, the second difference is described as how marriage outcomes differed between levels of military mortality. Using census data, they find that in regions with higher mortality rates, women aged 20-29 were more likely to remain single postwar. Moreover, a 10-percentage-point increase in military mortality rates led to a 2.7 percentage point increase in the proportion of single women aged 20-29. In contrast, it was less likely for men aged 20-29 to remain single postwar.

(Abramitzky et al., 2011) emphasize that military mortality negatively affected divorce rates for young adults while it had a positive impact on the proportion of widowed women aged 30-39. In addition, the proportion of out-of-wedlock births increased while divorce rates fell due to the effect of military mortality. In particular, in regions where the military mortality was 20 percent, the proportion of out-of-wedlock births postwar rose by 1.5 percentage points compared to regions with 10 percent military mortality rate. Finally, they find that men and women married at old ages, resulting in a plummet in the age gap between brides and grooms.

Moreover, (Abramitzky et al., 2011) examine the impact of sex ratio, a proxy of male scarcity, on assortative matching by social class. Using the marriage-level data, the authors find that a decrease in sex ratio due to war casualties raised the status of men in a marriage market postwar, compared to prewar. The phenomenon is called marrying up where a man can marry a woman in a higher social class. The empirical findings illustrate that in French departments with 20 percent military mortality rates, the probability that a given groom married down postwar declined by ten percentage points, compared to regions with 10 percent military mortality rates.

In this notebook, I focus on replicating the main results of the paper and grasp insights about the economic mechanism behind the exogenous shock attributed to WWI. My notebook is structured as follows. Section 2 describes the geographic variation due to WWI. Section 3 and section 4 then indicates the identification and empirical strategy of the authors, respectively. Section 5 focuses on replicating and summarizing the main results of the paper, relying on the authors' theoretical approaches and Stata do files.

2 Historical Context

WWI was a global military conflict that arose during the 1914-1918 period. It induced an enormous number of casualties, which gave rise to a considerable decline in sex ratio, measured by the proportion of men to women in the population. During the great war, France witnessed universal mobilization where millions of Frenchmen were drafted and voluntarily enrolled in the army. While young Frenchmen suffered the high mortality rates (19 to 24 percent), the older ones (aged 40 and above) witnessed the lowest mortality rates (less than 10 percent). (Abramitzky et al., 2011) show that in addition to WWI, there was an exogenous reason emerging during the war period well known as the Spanish flu. However, the authors only focus on measuring war-related mortality under the research scope. Figure 1 shows the geographic variation by the war mortality.

```
[3]:  from auxiliary.auxiliary_m import *
      frmap(df1, geo)
```
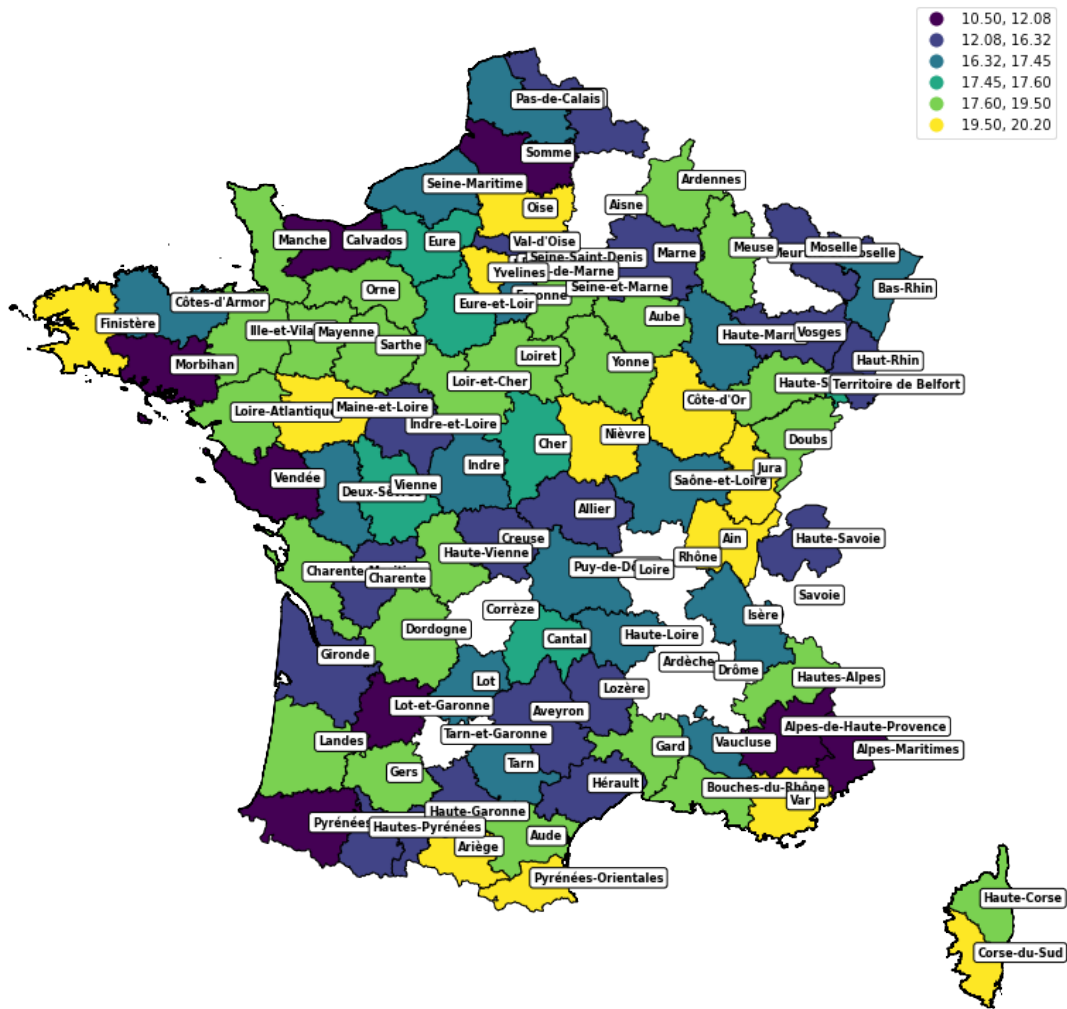

Figure 1. The Geographic Variation in Military Mortality Rates

3 Identification

In the paper, (Abramitzky et al., 2011) apply the difference-in-difference approach (DID) to investigate how the exogenous shock caused by WWI (the first difference) and levels of military

mortality (the second difference) influence marriage outcomes, which are measured by marital status (single, divorced, and widowed) and the proportion of out-of-wedlock births. In addition, the authors adopt the similar method to study the impact of male scarcity, caused by war casualties, on the assortative matching by social class and the age gap between brides and grooms.

**The Difference-in-Difference Approach**

Outcomes	First difference	Second difference	DID
Marital Status (single/divorced/widowed)	Marriage outcome changed following the war in regions	Marriage outcome changed with different levels of military mortality	Impact of military mortality on marriage outcome
The difference between the class of the bride and the groom	Social class changed following the war in regions	Social class changed with different levels of military mortality	Impact of military mortality on assortative matching by social class
The difference between the age of the bride and the groom	Age gap changed following the war in regions	Age gap changed with different levels of military mortality	Impact of military mortality on age gap

Instrumental Variables (IV) Design

As pointed out in the related literature, the study on the effect of the male scarcity (sex ratio) on the marriage market can face a problem called omitted variable bias and reverse causality. For instance, (Abramitzky et al., 2011) emphasize that a low sex ratio could be explained by strong immigration. To avoid this issue, the authors adopt the instrumental variables approach (IV), using military mortality as an instrument for the male scarcity since it demonstrates the exogenous geographical variation that affects the sex ratio while the mortality does not directly affect marriage by social class. The below DAG illustrates the mechanism of the IV approach.

```
[4]:  Image(filename='files/IV_causal_graph.PNG', width=800)
```

[4]:
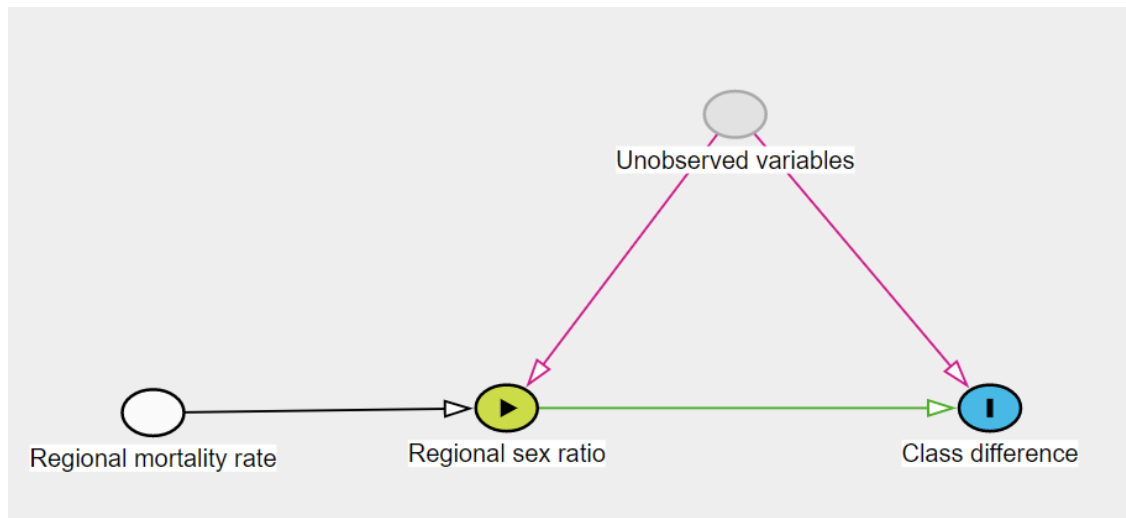
Note: in the causal graph, the pink arrow expresses the biasing path while the green arrow shows the causal path.

4

4 Empirical Strategy

4.1 The Impact of Military Mortality on Marital Status

Using department-level data from the 1911 and 1921 censuses, (Abramitzky et al., 2011) adopt the difference-in-difference method to estimate the impact of male scarcity on marital status and out-of-wedlock births, expressed by the following equation:

$$Y_{dt} = \delta_d + \lambda PW_t + \alpha M_d x PW_t + \epsilon_{dt} \tag{1}$$

where Y_{dt} is the proportion of men/women of a corresponding marital status (single, divorced, or widowed) in department d at time t; δ_d is a department dummy that captures time-invariant department-level factors; PW_t is a postwar dummy that equals one if the observation Y_{dt} comes from the 1921 census and zero otherwise; and M_d is the military mortality rate in department d. The coefficient $M_d x PW_t$ captures the treatment effect. The regression results will be indicated in table 2, section 5.2.1. The Impact of Male Scarcity on Marital Status.

4.2 The Impact of Military Mortality on Assortative Matching by Social Class

To test whether men married up postwar or not, (Abramitzky et al., 2011) continue to apply the difference-in-difference regressions where military mortality is the treatment.

$$Y_{dt} = \delta_d + \lambda PW_t + \alpha M_d x PW_t + \gamma Z_{idt} + \epsilon_{idt} \tag{2}$$

where Y_{dt} is the difference between the class of bride and that of groom; δ_d is a department dummy that captures time-invariant department-level factors, such as local traditions related to marriage; PW_t is a postwar dummy variable that equals one if the wedding took place after the war and zero otherwise; and M_d is the military mortality rate in department d; Z_{idt} are individual-level controls, such as groom class dummies, age, and whether the marriage took place in a rural area, (Abramitzky et al., 2011).

As expressed in section 5.2.2 C. Male Scarcity and Assortative Matching by Social Class, there are two different definitions of the discrepancy between the social classes of bride and groom. In particular, in table 3, (Abramitzky et al., 2011) use their own occupation to allocate individuals into classes. Meanwhile, table 4 defines social class based on father occupation (columns 1-3) or background characteristics (columns 5-6).

4.3 The Impact of Sex Ratio on Marriage Outcomes

The following equation demonstrates the IV approach, which is used for evaluating the impacts of male scarcity (sex ratio) on marriage outcomes. Military mortality is taken into account as an instrument for the sex ratio to avoid the problem of omitted variable bias. The empirical results for equation 3 will be shown in table 5, section 5.2.2 C. Male Scarcity and Assortative Matching by Social Class.

$$Y_{dt} = \delta_d + \lambda PW_t + \alpha SR_{td} + \gamma Z_{idt} + \epsilon_{idt} \tag{3}$$

where Y_{dt} is the difference between the class of bride and that of groom; SR_{td} is the department-level sex ratio, instrumented with military mortality interacted with postwar. Other variables are defined as in equations (1) and (2).

Note: In equations (1), (2), and (3), the authors cluster standard errors at the department level.

5 Replication of Abramitzky et al. (2011)

5.1 Data & Descriptive Statistics

(Abramitzky et al., 2011) use three different databases from several sources including the marriage-level data, the French censuses of population, and the regional data related to military casualties.

- **The marriage-level data** is extracted from the TRA dataset of the survey "L'enquête des 3000 familles", which collected information of 3,000 French couples' offspring during the 1803-1832 period. The survey data includes birth, marriage, and death certificates. The sample size in each region was proportional to its population from the 1806 census to ensure geographical representativeness. Remarkably, in each department, a random sample was drawn among couples whose names started with "TRA" to include names from various local dialects in the sample. The sampling procedure ensures the representativeness of all the social classes (Abramitzky et al., 2011).

- **The French census data and Vital Statistics** are available for the years 1906, 1911, and 1921. In this study, (Abramitzky et al., 2011) use the 1911 and 1921 censuses to obtain department-level variables on marital status (single, divorced, and widowed). In the data, the sex ratio was defined as the proportion of males aged 18–59, to the number of females aged 15–49. In addition, the department-level share of out-of-wedlock births in 1911 and 1921 was collected from the Vital Statistics in France.

- **The military mortality data** provides the proportion of military casualties in 22 military regions, following Huber's study (1931). There were 78 departments falling entirely into a military region, allocated the same mortality rate. For nine departments containing more than one military region, the authors impose a weighted average of the mortality rates of the military regions to be able to construct the mortality rate in each department.

Table 1. Summary statistics

```
[5]: from auxiliary.auxiliary_table import *
     summary_stats(df)
```

[5]:

Variables	Description	Count	Mean	Std
mortality	Military mortality rate	5,878	15.04	3.42
sr	Sex ratio (#18-59 males/ #15-49 females)	6,113	1.01	0.08
sr_39	Sex ratio (#males/ #females, aged 15-39)	6,113	0.89	0.09
sr_49	Sex ratio (#males/ #females, aged 15-49)	6,113	0.90	0.08
agegr	Groom's age	6,174	28.56	7.96
agebr	Bride's age	6,165	25.16	7.27

Variables	Min	25%	50%	75%	Max
mortality	10.50	11.90	15.50	17.60	20.20
sr	0.86	0.96	1.00	1.06	1.60
sr_39	0.75	0.82	0.87	0.95	1.64
sr_49	0.78	0.84	0.89	0.96	1.47
agegr	10.00	24.00	26.00	31.00	77.00
agebr	11.00	21.00	23.00	27.00	78.00

Note: Table 1 is different from table A1 of summary statistics in the original paper, which is not divided into prewar and postwar subgroups.

```
[6]: varlist = ['clbr', 'clfbr', 'clmbr', 'clgr', 'clfgr', 'clmgr']
     mdist_hue(df, varlist, nrows=2, ncols = 3)
```
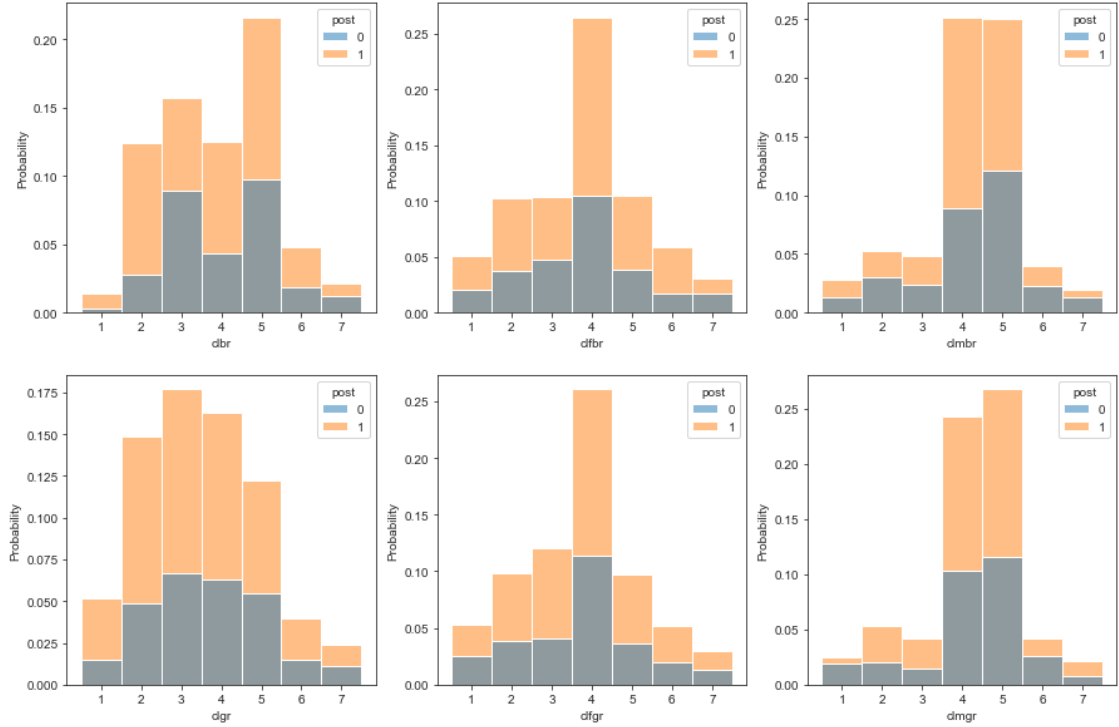
Note: prewar (post = 0) and postwar (post = 1)

Variable name	Variable label
clbr	Bride's class
clfbr	(Bride's) father's class
clmbr	(Bride's) mother's class
clgr	(Groom's) class
clfgr	(Groom's) father's class
clmgr	(Groom's) mother's class

Figure 2. Distribution of social class (prewar and postwar)

5.2 Empirical Results

5.2.1 The Impact of Male Scarcity on Marital Status

The estimates from equation 1 in section 4. Empirical Strategy illustrate that in regions with higher mortality rates, women were likely to remain single postwar. In particular, a ten percentage point increase in mortality led to a 2.7 percentage point increase in the proportion of single women aged 20-29 postwar (see table 2). Meanwhile, the department with a higher casualty ratio witnessed a decrease in the share of single men aged 20-29 postwar. This finding implies that male scarcity gave French males aged 20-29 more incentive to marry.

In addition, mortality rates had a negative impact on divorce. The shares of divorced men and women reduced postwar in departments with higher mortality rates. Specifically, in the 20-29 cohort, the proportion of divorced men and women declined by 0.04 and 0.06 percentage points postwar, respectively, in a department with the mortality rate of 20 percent compared to the 10 percent rate. The effect for the 30-39 female was lower, only 0.13 percentage point. (Abramitzky et al., 2011) explain that in departments with higher mortality rates, women had a tendency to stay single and were less likely to marry. Therefore, divorce did not occur in this situation. In addition, marriages ending in divorce broke down by husbands' deaths in regions with higher mortality. Moreover, male scarcity ensured the men-women matching became stronger and stable,

7

which were hard to dissolve.

Table 2. The Impact of Male Scarcity on Marital Status by Age Group and Gender

```
[7]: table_2(df1)
```

```
[7]:            Single males Single females Divorced males Divorced females
     Age group:
     15-19             0.0097         0.1091**            NaN              NaN
                      (0.0059)        (0.0430)            NaN              NaN
     20-29          -0.3124***        0.2697***       -0.0043**          -0.0057
                      (0.1081)        (0.0954)        (0.0021)          (0.0043)
     30-39             0.0443         0.0366           -0.0125          -0.0130*
                      (0.0840)        (0.0467)        (0.0079)          (0.0070)
     40-49            0.1195**        0.0649**          0.0026          -0.0069
                      (0.0472)        (0.0326)        (0.0066)          (0.0056)


                Widowed males Widowed females
     Age group:
     15-19              NaN             NaN
                        NaN             NaN
     20-29           0.0031          0.0525*
                    (0.0031)        (0.0302)
     30-39          -0.0011          0.2447***
                    (0.0081)        (0.0496)
     40-49           0.0016          0.1380***
                    (0.0129)        (0.0320)
```

Note:

- Coefficient presented is expressed by the interaction between postwar and military mortality. All regressions consist dummy variables for prewar/postwar and departments. Standard errors are presented in parentheses.

- $***$ Significant at the 1 percent level, $**$ Significant at the 5 percent level, $*$ Significant at the 10 percent level.

5.2.2 The Impact of Male Scarcity on Assortative Matching by Social Class

A. Assigning Social Classes Based on the Historical International Social Class Scheme

Using the marriage-level data, (Abramitzky et al., 2011) investigate whether men could have a better matching in regions with higher levels of male scarcity during WWI. The authors first describe how they allocate inhabitants into social classes. Then they make an effort to verify whether there existed assortative matching by social class before WWI. Finally, the authors adopt difference-in-differences and instrumental variables methods in studying how male scarcity affects assortative matching by social class.

For the first step, (Abramitzky et al., 2011) use the marriage certificate data to specify occupations of brides, grooms, and their parents to match them with corresponding social classes. Firstly, the authors adopt the Historical International Standard Classification of Occupations (HISCO) in allocating each occupation into one of seven sectors: (1) professional; (2) technical and related workers, administrative and managerial workers; (3) clerical and related workers; (4) sales workers; (5) service workers; (6) agricultural, animal husbandry and forest workers; fishermen, and hunters;

and (7) production and related workers, transport equipment operators, and laborers. The following table describes an example of common occupations within each class:

Class	Women	Men
1	Accountant, professor	Accountant, engineer
2	Store employee, bank employee	Railway company employee, store employee
3	Seamstress, cook	Mechanic, constructor
4	Farmer, winegrower	Farmer, winegrower
5	Domestic, linen maid	Driver, domestic
6	Factory worker, worker	Factory worker, worker
7	Day laborer, farm worker	Day laborer, farm worker

Secondly, (Abramitzky et al., 2011) use the Historical International Social Class Scheme (HIS-CLASS) to map occupational codes into social classes. The HISCLASS system merges each of the 1,600 HISCO occupations into one of 12 social classes, shortened to the following seven social classes: (1) Class 1: Higher managers and professionals (highest); (2) Class 2: Lower managers and professionals, clerical and sales personnel; (3) Class 3: Foremen and skilled workers; (4) Class 4: Farmers and fishermen; (5) Class 5: Lower skilled workers, (6) Class 6: Unskilled workers, (7) Class 7: Lower-skilled and unskilled farm workers (lowest).

B. Prewar Assortative Matching by Social Class (Bootstrap Simulation)

The section describes the authors' motivation in investigating how robust marriage matching by social class was before WWI. To verify whether individuals married those within their social classes or randomly, (Abramitzky et al., 2011) come up with usage of prewar data to investigate the distribution of social distance, which is defined as the class of the bride minus the class of the groom (classdiff). The authors then construct the well-known null hypothesis, the random matching, which specifies that prewar couples married randomly. To do so, (Abramitzky et al., 2011) adopt the bootstrapping approach to create 95 percent bootstrap confidence intervals for the distribution of social difference under the null hypothesis.

Bootstrap simulation

In this section, I sketch the algorithm of the bootstrapping method applied by (Abramitzky et al., 2011). Specifically, suppose that **N** is the number of prewar marriages. From the realized distribution of groom classes, they draw **N** grooms randomly with replacement. Similarly, they draw **N** brides randomly with replacement from the realized distribution of bride classes. Then they match the list of grooms with the list of brides to derive the distribution of social distances corresponding to random samples. Finally, (Abramitzky et al., 2011) repeat this process **1,000 times**, hence, successfully construct the 95 percent confidence interval. The simulation results are represented in figures within subsections Panel A and Panel B.

```
[8]: from auxiliary.auxiliary_bootstrap_sm import *
```

Panel A. Class defined by own occupation

```
[9]: bootstrap_sm(df, 'clgr', 'clbr')
```
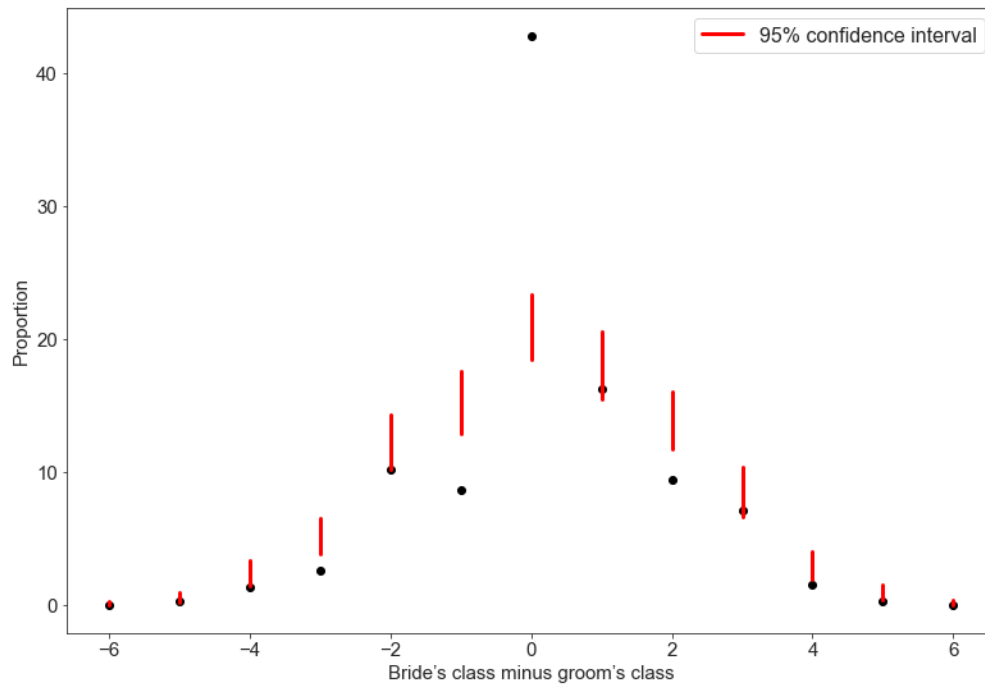
Panel B. Class defned by father's occupation

```
[10]: bootstrap_sm(df, 'clfgr', 'clfbr')
```
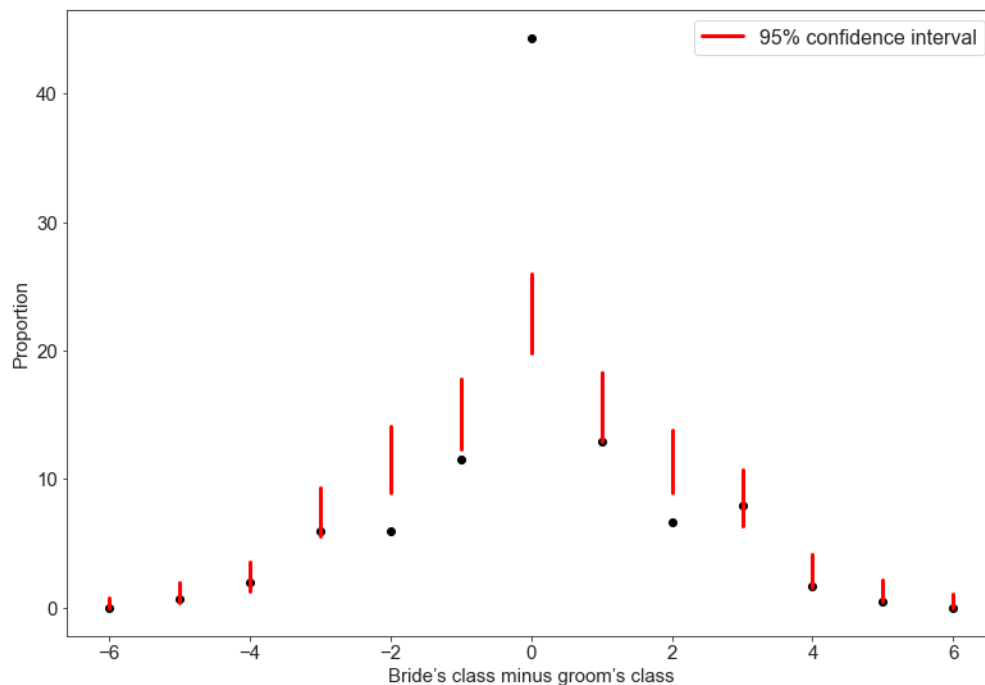

Note: The simulation results from my work in Python are consistent with those in Mata Stata, demonstrated by (Abramitzky et al., 2011).

In the original paper, (Abramitzky et al., 2011) reject the null hypothesis associated with the random matching for social classes. In particular, the figures illustrate that the probability that

men married women within the same social classes was too high (44 percent for own occupation and 45 percent for father's occupation), compared with the possible probabilities in the random matching, 18-24 percent.

C. Male Scarcity and Assortative Matching by Social Class

Table 3 shows the estimation results from equation (2) in section 4. Empirical Strategy. Remarkably, the own occupations of men and women are taken into account to allocate couples into social classes. In addition, "class difference" in table 3 means the difference between the social class of the bride and the class of the groom. A lower value of the class difference means the man "married up" while "married down" is a dummy for whether the groom married a bride of a lower class than his.

(Abramitzky et al., 2011) assure that compared with prewar, the possibility that men had a better matching in regions with higher mortality rates was high. The coefficient of interest in column "married down" can be expressed as follows. Suppose there were two identical regions prewar. After WWI, the probability that a given groom married down declined by approximately 10 percentage points in the region that its mortality rate was 10 percent higher than another region (see table 3).

Table 3. Impact of Male Scarcity on Assortative Matching by Class (OLS) (Classes defined by own occupation)

```
[11]: sample3 = sample_a(df, 'clbr', 'clgr')
      table_3(sample3)
```

[11]:

	All classes of groom	
Dependent variable	Class difference	Married down
Percent of soldiers killed x postwar	-0.0201*	-0.0095**
	(0.0105)	(0.0045)
Postwar	0.1964	0.0918
	(0.1460)	(0.0648)
Rural	0.1905**	0.0320
	(0.0852)	(0.0291)
Bride's age (/100)	-0.5815	0.0870
	(0.4737)	(0.1597)
Groom's Age (/100)	0.3546	-0.1664
	(0.3769)	(0.1383)
Groom class dummies	Yes	Yes
Département dummies	Yes	Yes
R^{2}	0.3869	0.2712
Observations	3117	2994

	groom |	Excluding
Dependent variable	Low-class bride	Class difference
Percent of soldiers killed x postwar	-0.0166***	-0.0350***
	(0.0048)	(0.0126)
Postwar	0.2231***	0.4908***
	(0.0664)	(0.1873)
Rural	-0.0341	0.1914**
	(0.0299)	(0.0915)
Bride's age (/100)	-0.0952	-0.9234*
	(0.1425)	(0.4887)

Groom's Age (/100)	0.1343	0.4298
	(0.1272)	(0.3638)
Groom class dummies	Yes	Yes
Département dummies	Yes	Yes
R^{2}	0.2321	0.3565
Observations	3117	2432

	grooms of class 1 and 2	
Dependent variable	Married down	Low-class bride
Percent of soldiers killed x postwar	-0.0168***	-0.0231***
	(0.0043)	(0.0056)
Postwar	0.2404***	0.3381***
	(0.0597)	(0.0821)
Rural	0.0319	-0.0357
	(0.0292)	(0.0328)
Bride's age (/100)	-0.1639	-0.0844
	(0.1851)	(0.1877)
Groom's Age (/100)	0.0332	0.0485
	(0.1301)	(0.1513)
Groom class dummies	Yes	Yes
Département dummies	Yes	Yes
R^{2}	0.2403	0.2607
Observations	2309	2432

Note:

- ∗ ∗ ∗ Significant at the 1 percent level, ∗∗ Significant at the 5 percent level, ∗ Significant at the 10 percent level.

- Standard errors, clustered at the department level, are presented in parentheses.

Table 4. Impact of Male Scarcity on Assortative Matching by Class (OLS) (alternative definitions of class)

```
[12]: sample4_1 = sample_a(df, 'clfbr', 'clfgr')
      sample4 = sample_a(df, 'clbr', 'clgr') # use the function 'sample_a' to convert␣
       ↪clbr and 'clgr' to numeric
      sample4_2 = sample_2_b(sample4)
      table_4_t(sample4_1, sample4_2)
```

[12]:

	Class	defined by
Dependent variable	Class difference	Married down
Percent of soldiers killed x postwar	0.0092	-0.0025
	(0.0159)	(0.0049)
Postwar	-0.1741	0.0061
	(0.2390)	(0.0732)
Rural	0.1526	0.0156
	(0.0977)	(0.0302)
Bride's age (/100)	-2.4584**	-0.1100
	(1.0504)	(0.2237)
Groom's Age (/100)	-0.3312	-0.4762*
	(0.9524)	(0.2685)
Groom class dummies	Yes	Yes

Groom half-class dummies		
Département dummies	Yes	Yes
R^{2}	0.3552	0.2824
Observations	1966	1882

	father occupation	Classes imputed
Dependent variable	Low-class bride	Class difference
Percent of soldiers killed x postwar	-0.0054	-0.0056**
	(0.0054)	(0.0023)
Postwar	0.0624	0.0878***
	(0.0797)	(0.0298)
Rural	-0.0978***	0.3012***
	(0.0305)	(0.0234)
Bride's age (/100)	-0.5797**	0.3621
	(0.2555)	(0.2854)
Groom's Age (/100)	-0.5145**	0.0342
	(0.2260)	(0.1144)
Groom class dummies	Yes	
Groom half-class dummies		Yes
Département dummies	Yes	Yes
R^{2}	0.2183	0.4609
Observations	1966	5832

	using background	characteristics
Dependent variable	Married down	Low-class bride
Percent of soldiers killed x postwar	-0.0062**	-0.0021
	(0.0025)	(0.0026)
Postwar	0.1076***	0.0325
	(0.0349)	(0.0360)
Rural	0.2094***	0.3972***
	(0.0189)	(0.0263)
Bride's age (/100)	0.3593**	0.0956
	(0.1584)	(0.1171)
Groom's Age (/100)	0.1339	-0.0229
	(0.0857)	(0.0915)
Groom class dummies		
Groom half-class dummies	Yes	Yes
Département dummies	Yes	Yes
R^{2}	0.2435	0.4329
Observations	5832	5832

Note:

- $*\,*\,*$ Significant at the 1 percent level, $*\,*$ Significant at the 5 percent level, $*$ Significant at the 10 percent level.

- Standard errors, clustered at the department level, are presented in parentheses.

Table A5 (Appendix) — Impact of Male Scarcity on Assortative Matching by Class (OLS, No Regional Fixed Effects)

Panel A

```
[13]: samplea5_1 = sample_a(df, 'clbr', 'clgr')
      samplea5_2 = sample_a(df, 'clfbr', 'clfgr')
      table_a5([samplea5_1, samplea5_2])
```

[13]:

	Class	defined by own
Dependent variable	Class difference	Married down
Panel A	NaN	NaN
Percent of soldiers killed x postwar	-0.0181*	-0.0074
	(0.0108)	(0.0045)
Percent of soldiers killed	0.0247	0.0046
	(0.0181)	(0.0080)
Postwar	0.1946	0.0692
	(0.1519)	(0.0653)
Rural	0.3776***	0.0986***
	(0.0936)	(0.0368)
Bride's age (/100)	-0.4282	0.1670
	(0.4157)	(0.1493)
Groom's Age (/100)	0.1153	-0.2458*
	(0.3584)	(0.1334)
Groom class dummies	Yes	Yes
Groom half-class dummies	No	No
R^{2}	0.3437	0.1972
Observations	3117	2994

	occupation	Class
Dependent variable	Low-class bride	Class difference
Panel A	NaN	NaN
Percent of soldiers killed x postwar	-0.0140***	0.0045
	(0.0052)	(0.0159)
Percent of soldiers killed	0.0022	0.0139
	(0.0103)	(0.0152)
Postwar	0.1959***	-0.0927
	(0.0728)	(0.2486)
Rural	0.0638	0.1608*
	(0.0440)	(0.0903)
Bride's age (/100)	0.0274	-2.3663**
	(0.1644)	(0.9532)
Groom's Age (/100)	-0.0013	-0.5865
	(0.1480)	(0.9101)
Groom class dummies	Yes	Yes
Groom half-class dummies	No	No
R^{2}	0.1156	0.3223
Observations	3117	1966

	defined by father	occupation
Dependent variable	Married down	Low-class bride
Panel A	NaN	NaN
Percent of soldiers killed x postwar	-0.0034	-0.0065

14

	(0.0047)	(0.0054)
Percent of soldiers killed	0.0022	0.0024
	(0.0054)	(0.0051)
Postwar	0.0193	0.0810
	(0.0715)	(0.0801)
Rural	0.0205	-0.0847***
	(0.0257)	(0.0304)
Bride's age (/100)	-0.1586	-0.5441**
	(0.2143)	(0.2346)
Groom's Age (/100)	-0.4103	-0.6223***
	(0.2717)	(0.2380)
Groom class dummies	Yes	Yes
Groom half-class dummies	No	No
R^{2}	0.2287	0.1536
Observations	1882	1966

Note:

- ∗ ∗ ∗ Significant at the 1 percent level, ∗∗ Significant at the 5 percent level, ∗ Significant at the 10 percent level.

- Standard errors, clustered at the department level, are presented in parentheses.

Panel B

```
[14]: sample_a5b = sample_a(df, 'clbr', 'clgr') # use the function 'sample_a' to␣
 ↪convert clbr and 'clgr' to numeric
sample_a5b = sample4_2_b(sample_a5b)
table_a5_b(sample_a5b)
```

[14]:
	Class imputed	from background
Dependent variable	Class difference	Married down
panel B	NaN	NaN
Percent of soldiers killed x postwar	-0.0061***	-0.0060**
	(0.0024)	(0.0024)
Percent of soldiers killed	0.0209**	0.0127***
	(0.0084)	(0.0042)
Postwar	0.0979***	0.1053***
	(0.0305)	(0.0329)
Rural	0.3410***	0.2231***
	(0.0347)	(0.0252)
Bride's age (/100)	0.3761	0.3783***
	(0.2590)	(0.1440)
Groom's Age (/100)	-0.0104	0.1266
	(0.1051)	(0.0854)
Groom class dummies	No	No
Département dummies	Yes	Yes
R^{2}	0.4287	0.2218
Observations	5832	5832

	characteristics
Dependent variable	Low-class bride
panel B	NaN

Percent of soldiers killed x postwar	-0.0025
	(0.0026)
Percent of soldiers killed	0.0156***
	(0.0054)
Postwar	0.0379
	(0.0356)
Rural	0.4237***
	(0.0324)
Bride's age (/100)	0.1131
	(0.1238)
Groom's Age (/100)	-0.0420
	(0.0960)
Groom class dummies	No
Département dummies	Yes
R^{2}	0.3969
Observations	5832

Note:

- $***$ Significant at the 1 percent level, $**$ Significant at the 5 percent level, $*$ Significant at the 10 percent level.

- Standard errors, clustered at the department level, are presented in parentheses.

Table 5. Impact of Male Scarcity on Assortative Matching by Class (IV) (equivalent to table 5 in the original paper)

Panel A. Stage 2 regressions

```
[15]: df5_1 = sample_a(df, 'clbr', 'clgr')
      df5_2 = sample_a(df, 'clfbr', 'clfgr')
      table_5_b(df5_1, df5_2)
```

[15]:

	Class	defined by own
Dependent variable	Class difference	Married down
Panel A. Stage 2 regressions	NaN	NaN
Sex ratio (men/women)	4.22*	2.022**
	0.05	0.019
Postwar	0.245	0.116*
	0.124	0.072
Rural	0.197**	0.036
	0.02	0.226
Bride's age (/100)	-0.497	0.126
	0.269	0.372
Groom's age (/100)	0.372	-0.156
	0.3	0.234
Groom class dummies	Yes	Yes
Département dummies	Yes	Yes

	occupation	Class
Dependent variable	Low-class bride	Class difference
Panel A. Stage 2 regressions	NaN	NaN
Sex ratio (men/women)	3.124***	-2.292

	0.002	0.573
Postwar	0.23***	-0.224
	0.001	0.5
Rural	0.008	0.152
	0.803	0.102
Bride's age (/100)	-0.001	-2.523**
	0.995	0.013
Groom's age (/100)	0.114	-0.441
	0.342	0.645
Groom class dummies	Yes	Yes
Département dummies	Yes	Yes

	defined by father	occupation
Dependent variable	Married down	Low-class bride
Panel A. Stage 2 regressions	NaN	NaN
Sex ratio (men/women)	0.632	1.492*
	0.588	0.088
Postwar	0.021	0.116*
	0.824	0.086
Rural	0.016	-0.074***
	0.593	0.001
Bride's age (/100)	-0.091	-0.365*
	0.687	0.093
Groom's age (/100)	-0.447*	-0.565***
	0.083	0.0
Groom class dummies	Yes	Yes
Département dummies	Yes	Yes

Note:

- ∗ ∗ ∗ Significant at the 1 percent level, ∗∗ Significant at the 5 percent level, ∗ Significant at the 10 percent level.

- Different from the original paper, in the table, I use **p-value** instead of clustered standard errors since I have some troubles in extracting standard errors from the **linearmodels** package in Python. However, IV coefficients from my replication are consistent with authors' results.

Panel B. Stage 1 regressions

```
[16]: table_5_c(df5_1, df5_2)
```

[16]:

	Class	defined by own
Dependent variable	Class difference	Married down
Panel B. Stage 1 regressions	NaN	NaN
Percent of soldiers killed x postwar	-0.0045***	-0.0045***
	(0.0010)	(0.0010)
Postwar	-0.0135	-0.0135
	(0.0168)	(0.0168)
Rural	-0.0002	-0.0002
	(0.0011)	(0.0011)
Bride's age (/100)	-0.0107	-0.0107
	(0.0104)	(0.0104)

Groom's age (/100)	-0.0107	-0.0107
	(0.0075)	(0.0075)
Groom class dummies	Yes	Yes
Département dummies	Yes	Yes
R^{2}	0.8855	0.8855
Observations	5498	5498

	occupation	Class
Dependent variable	Low-class bride	Class difference
Panel B. Stage 1 regressions	NaN	NaN
Percent of soldiers killed x postwar	-0.0045***	-0.0045***
	(0.0010)	(0.0012)
Postwar	-0.0135	-0.0135
	(0.0168)	(0.0209)
Rural	-0.0002	0.0003
	(0.0011)	(0.0014)
Bride's age (/100)	-0.0107	-0.0236
	(0.0104)	(0.0195)
Groom's age (/100)	-0.0107	-0.0290
	(0.0075)	(0.0192)
Groom class dummies	Yes	Yes
Département dummies	Yes	Yes
R^{2}	0.8855	0.8874
Observations	5498	2980

	defined by father	occupation
Dependent variable	Married down	Low-class bride
Panel B. Stage 1 regressions	NaN	NaN
Percent of soldiers killed x postwar	-0.0045***	-0.0045***
	(0.0012)	(0.0012)
Postwar	-0.0135	-0.0135
	(0.0209)	(0.0209)
Rural	0.0003	0.0003
	(0.0014)	(0.0014)
Bride's age (/100)	-0.0236	-0.0236
	(0.0195)	(0.0195)
Groom's age (/100)	-0.0290	-0.0290
	(0.0192)	(0.0192)
Groom class dummies	Yes	Yes
Département dummies	Yes	Yes
R^{2}	0.8874	0.8874
Observations	2980	2980

Note:

- ∗∗∗ Significant at the 1 percent level, ∗∗ Significant at the 5 percent level, ∗ Significant at the 10 percent level.

- Standard errors, clustered at the department level, are presented in parentheses.

Table A6 shows estimation results of OLS regression from equation (3) of section 4. Empirical Strategy. The coefficient associated with the sex ratio has the right sign but is much smaller in magnitude than in the IV estimation, 0.74 (OLS) compared to 4.22 (IV). Hence, (Abramitzky et

al., 2011) suggest that taking into account the endogeneity of the sex ratio is crucial.

Table A6 (Appendix). Impact of Male Scarcity on Assortative Matching by Class (OLS equivalent of IV)

```
[17]: sample6_1 = sample_a(df, 'clbr', 'clgr')
      sample6_2 = sample_a(df, 'clfbr', 'clfgr')
      table_a6([sample6_1, sample6_2])
```

[17]:

	Class	defined by own	occupation
Dependent variable	Class difference	Married down	Low-class bride
Sex ratio (males/females)	0.7417	0.4305	0.9622***
	(0.8417)	(0.3012)	(0.3221)
Postwar	-0.0402	-0.0126	0.0577
	(0.1053)	(0.0316)	(0.0379)
Rural	0.2019**	0.0413	-0.0254
	(0.0841)	(0.0292)	(0.0301)
Bride's age (/100)	-0.5328	0.0891	-0.0845
	(0.4513)	(0.1540)	(0.1381)
Groom's Age (/100)	0.2285	-0.1893	0.1018
	(0.3641)	(0.1335)	(0.1275)
Groom class dummies	Yes	Yes	Yes
Groom half-class dummies	No	No	No
Département dummies	Yes	Yes	Yes
R^{2}	0.3883	0.2710	0.2271
Observations	3193	3066	3193

	Class	defined by father	occupation
Dependent variable	Class difference	Married down	Low-class bride
Sex ratio (males/females)	0.4109	0.5336*	0.3625
	(1.1004)	(0.2921)	(0.4397)
Postwar	-0.0036	0.0113	0.0084
	(0.0998)	(0.0282)	(0.0397)
Rural	0.1423	0.0096	-0.0942***
	(0.0980)	(0.0288)	(0.0289)
Bride's age (/100)	-1.9581**	-0.0072	-0.4908**
	(0.9423)	(0.2015)	(0.2210)
Groom's Age (/100)	-0.6016	-0.6160**	-0.5359**
	(0.8952)	(0.2746)	(0.2149)
Groom class dummies	Yes	Yes	Yes
Groom half-class dummies	No	No	No
Département dummies	Yes	Yes	Yes
R^{2}	0.3612	0.2864	0.2177
Observations	2103	2009	2103

Note:

- $***$ Significant at the 1 percent level, $**$ Significant at the 5 percent level, $*$ Significant at the 10 percent level.

- Standard errors, clustered at the department level, are presented in parentheses.

5.2.3　The Impact of Male Scarcity on Spousal Age Gap and Age at Marriage

Table 6 shows the impact of male scarcity on the assortative matching by age. According to (Abramitzky et al., 2011), the average groom was 3.8 years older than his bride prewar. Hence, the average age gap, measured by taking bride's age minus the groom's age, is negative. In column 1 of table 6, the coefficient associated with postwar interacted with mortality is positive and statistically significant, implying that the age gap between brides and grooms shrank (observed in a absolute value) in regions with higher mortality postwar.

Table 6. Impact of Male Scarcity on Assortative Matching by Age (OLS)

```
[18]: table_6(df)
```

```
[18]:
```

	(1)	(2)	(3)
Dependent variable	Age difference	Groom's age	Bride's age
Percent of soldiers killed x postwar	0.0916**	0.0195	0.0827**
	(0.0361)	(0.0384)	(0.0381)
Postwar	-0.6263	-0.9312*	-1.1643*
	(0.5361)	(0.5195)	(0.6155)
Rural	-0.3248***	0.3311	-0.3400*
	(0.1255)	(0.2151)	(0.1842)
Groom class dummies	Yes	Yes	No
Bride class dummies	Yes	No	Yes
Département dummies	Yes	Yes	Yes
R^{2}	0.0630	0.0623	0.0640
Observations	4868	5205	5287

Note:

- $***$ Significant at the 1 percent level, $**$ Significant at the 5 percent level, $*$ Significant at the 10 percent level.

- Standard errors, clustered at the department level, are presented in parentheses.

6　Conclusion

Overall, the results in the notebook strongly support the empirical findings, reported by (Abramitzky et al., 2011). Based on the transparent research methods which are represented in the paper and Stata do-files, I can comprehensively document and (almost) exactly replicate all results of the original paper, including OLS and IV regressions as well as the bootstrapping method.

(Abramitzky et al., 2011) find that a male scarcity in the population attributed to war-related casualty benefits men by giving them better assortative matching in marriage markets. They show that in regions with higher mortality, women were less likely to marry postwar, and if possible, they married at an older age, resulting in a decline in the age gap between brides and grooms.

References

Abramitzky, R., Delavande, A., & Vasconcelos, L. (2011, July). Marrying up: The role of sex ratio in assortative matching. *American Economic Journal: Applied Economics*, *3*(3), 124-57. Retrieved from https://www.aeaweb.org/articles?id=10.1257/app.3.3.124　doi: 10.1257/app.3.3.124