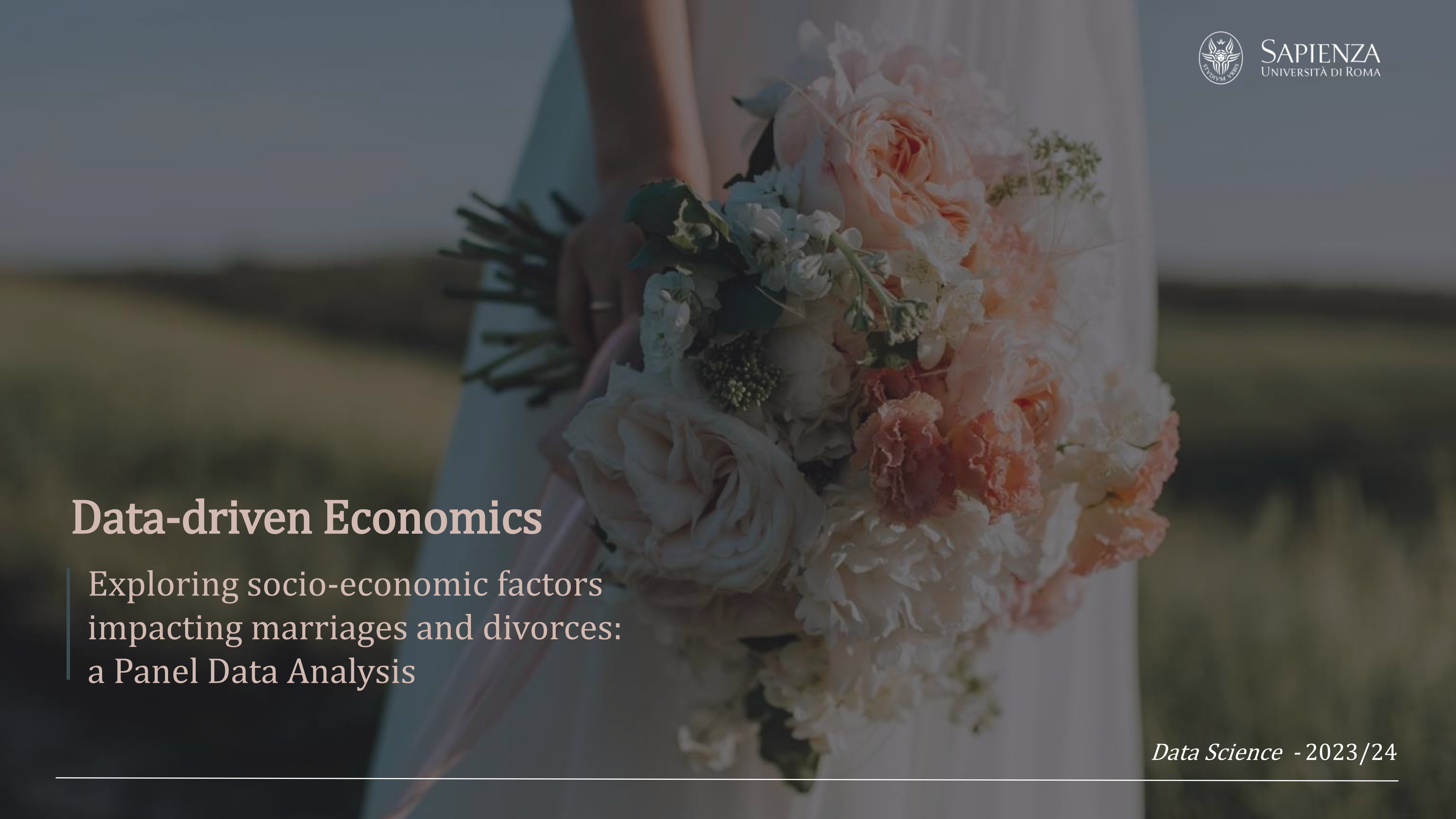


Sapienza Università di Roma

Data-driven Economics

Exploring socio-economic factors impacting marriages and divorces: a Panel Data Analysis

Data Science - 2023/24

Problem Overview - Main Goal

In our study we would like to understand how welfare can influence the western social "micro-structure", based on the concept of **monogamy** and **family**.

We are interested in how the change in stress levels, poverty, competitiveness and female empowerment are determinants in shaping our social rules and habits:

- Are people more prone to divorce over the years?
- How can gender inequality change the social structure?
- Do changes in the gender gap have the same effect on the number of new marriages and new divorces?

Source if inspiration:

From biology research involving primate species, it appears that the role of the female, the level of inter-subject competitiveness, and physical characteristics of the primate are decisive in the imposition of a polygamous or monogamous structure!

Does woman also define the nature of romantic relationships in human society?

Problem Overview – Balanced Panel Data

- **What** :
  we study the causal relationships between socioeconomic factors such as net median income and employment rate, and levels of new marriages and divorces in different European nations. In particular, we analyze how indices of welfare distinctly by sex are influencing social relationships.

- **Where:**
  we focus the analysis across 25 European countries

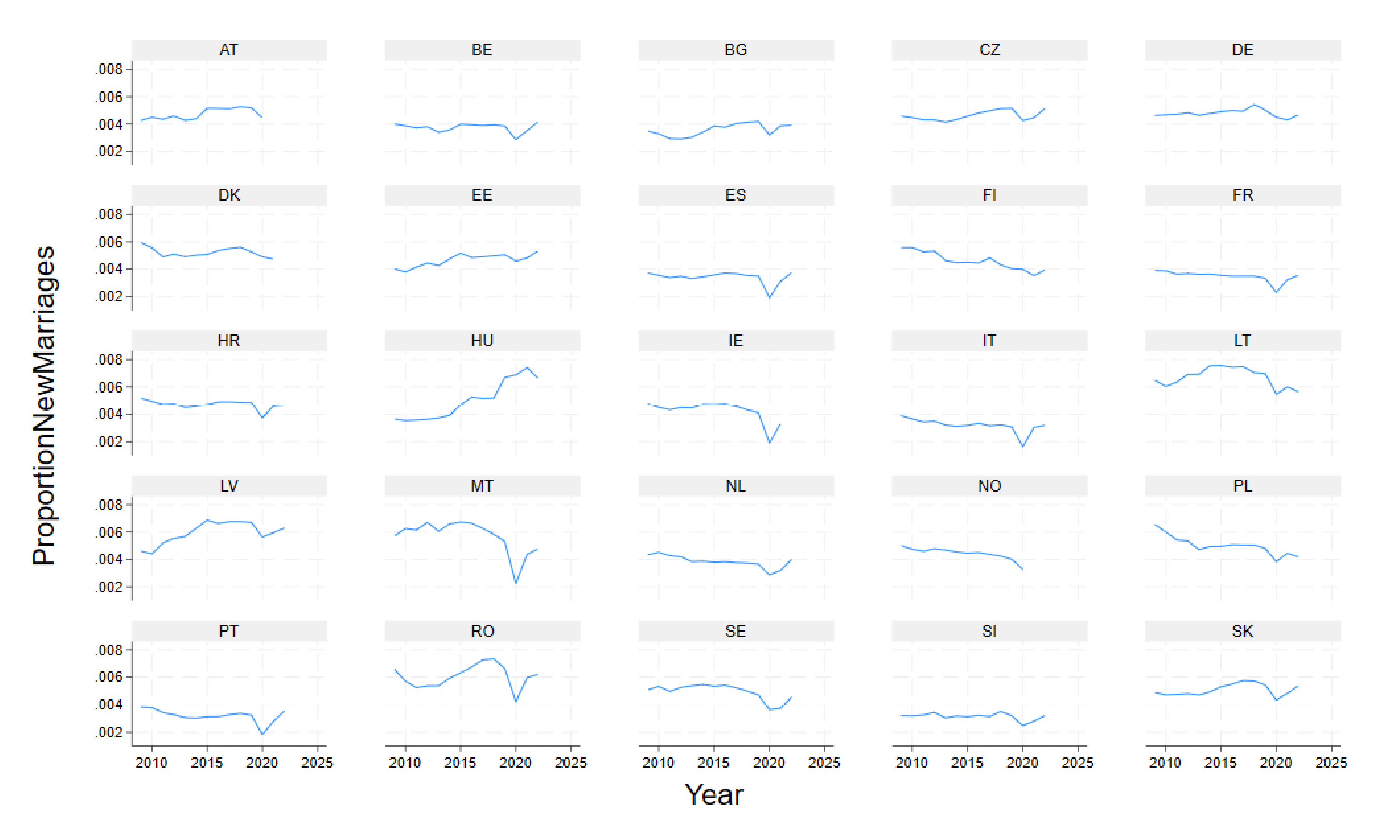
- **When:**
  we restricted the study over 13 years (2009-2022)

- **Why**:
  The findings from this analysis can have significant policy implications for ***governments*** and ***policymakers***. Understanding the factors influencing marriage and divorce rates can inform the design of targeted interventions aimed at promoting stable family structures and social well-being.

Sources: EUROSTAT United Nations

Exploratory Data Analysis

Dependent variable



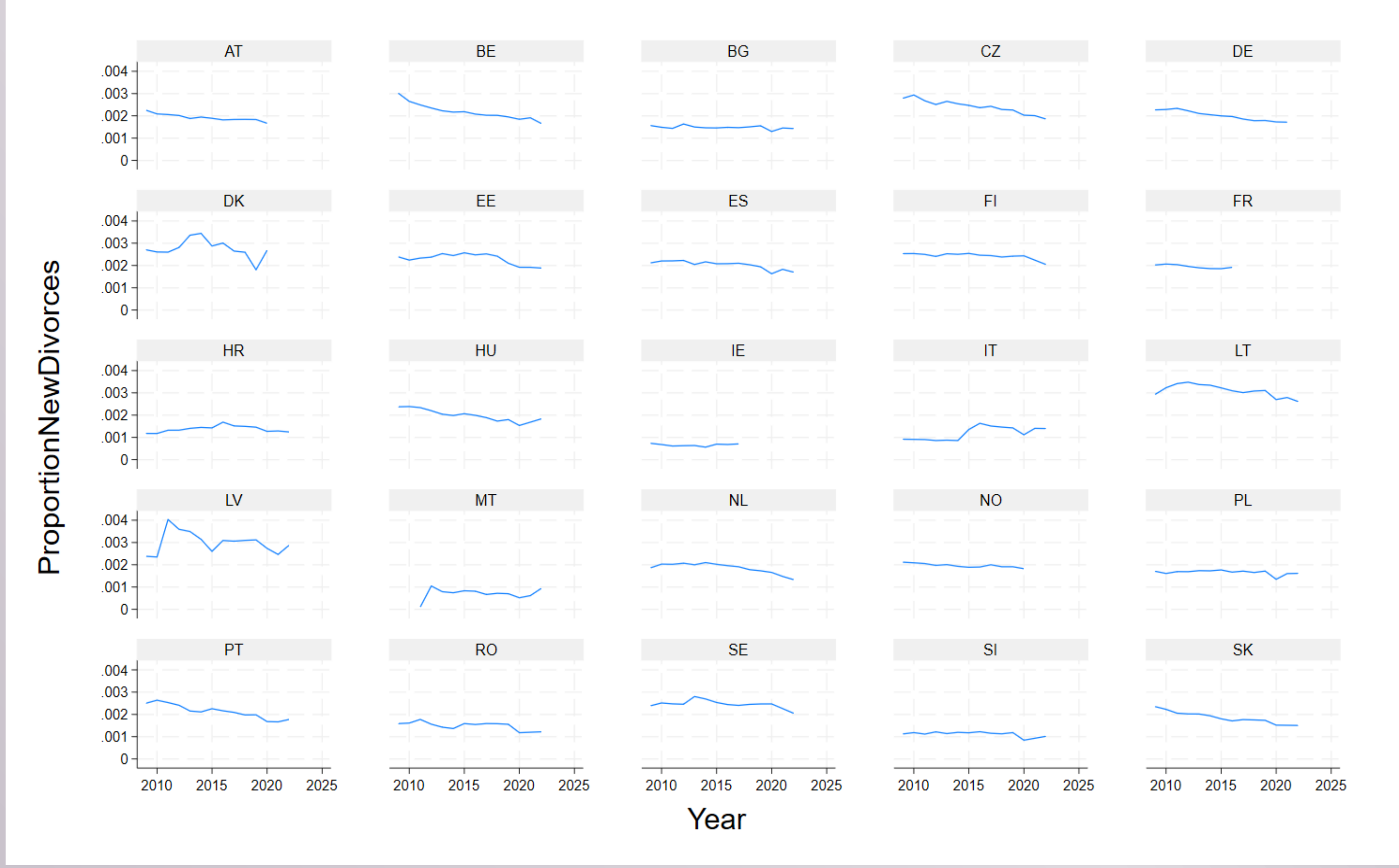
The **number of new marriages** across countries for each year is normalized by the country population.

In this case the curves seem to vary across countries, but it is interesting to note a *common local minimum* corresponding to the pandemic year (2020), in contrast with the number of divorces (which doesn't show any minimum or maximum).

Exploratory Data Analysis

SAPIENZA UNIVERSITÀ DI ROMA

Dependent variable



The **number of new divorces** across countries for each year is normalized by the country population.

We can't infer any trend or pattern from these (and the previous) plots, but it is evident that we have some missing data and our panel dataset could be considered ***unbalanced***!

We decide to focus on the number of new marriages.

Exploratory Data Analysis

Independent variables (1)

We collected independent variables such as *economic* indicators and *welfare* indicators, sometimes distinguishable by gender, referring to a particular country in a particular year:

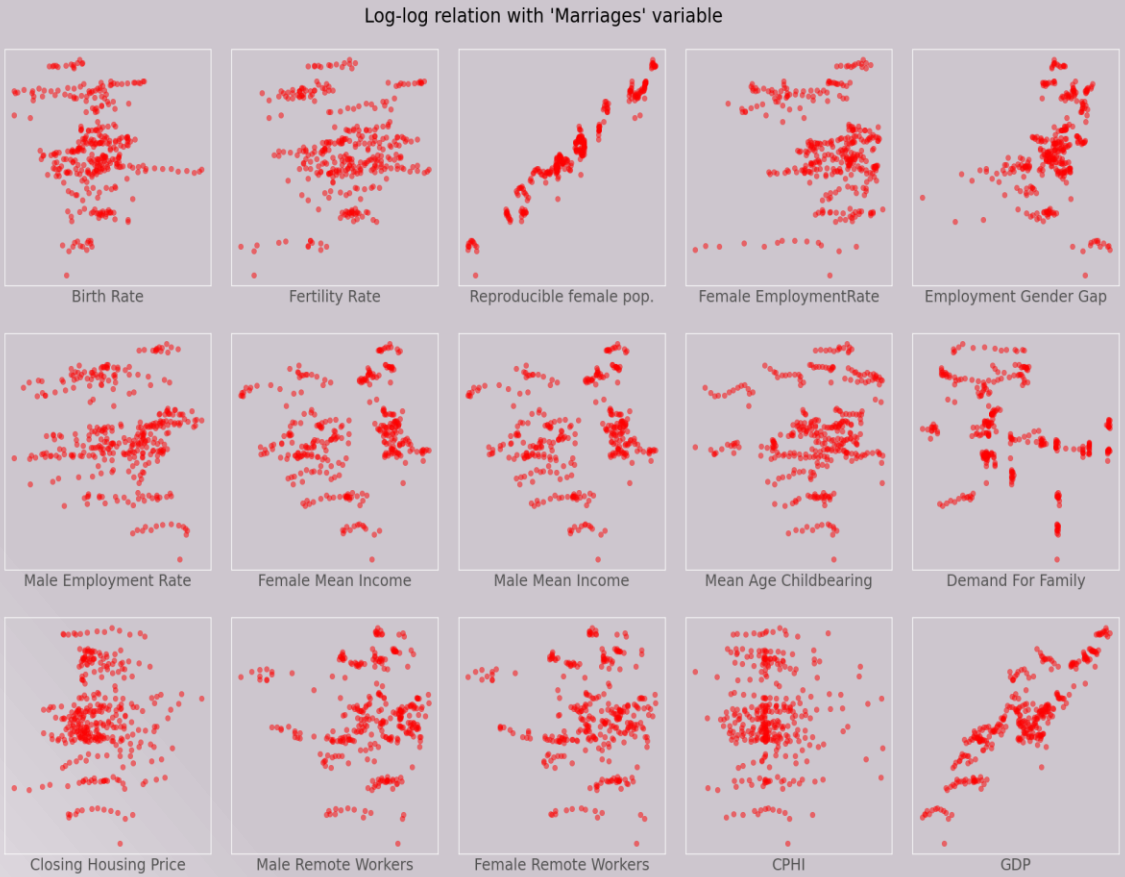
Variable	Unit	Description
Female/male median and mean incomes	Euro	Average or median net income by gender
Female/male employment rates	% of tot. population	Proportion of new employees hired by gender
GDP	Mln of euro	Gross domestic product
Closing housing price	NSA in (2015=100)	Average house price in the last quarter of the year
Female/male tot. population	Thousands of people	Ratio of inhabitants by gender
Mean age of childbearing	Floating number	Mean age of women at birth of first child is the weighted average of the different childbearing ages, using as weights the age-specific fertility rates of first-order births.
Fertility rate	Floating number	Ratio of the number of live births to mothers of age x to the average female population of age x
Number of female people in reproducible age	Counting number	Female people in reproducible age (15-49 y.o.).
Female people working from home	% of female population	Ratio of female people working from home over total female population

Exploratory Data Analysis

Independent variables (2)

Variable	Unit	Description
Female in Parliament	% of female population	Percentage of female population that are in Parliament
Percentage male work home	% of male population	Percentages of males that work at home
CPHI	Floating number	indicator of inflation and price stability for the European Central Bank (ECB)
Female Mean Age Marriage	Floating number	Mean age of the female at the first marriage
Woman in Senior Positions	% of female population	Percentage of woman that are in senior positions
Tertiary Education Female	% of female population	Percentage of woman that have a tertiary education level
Tertiary Education Male	% of male population	Percentage of man that have a tertiary education level
Birth rate	Floating number	Average (per month) number of children born alive

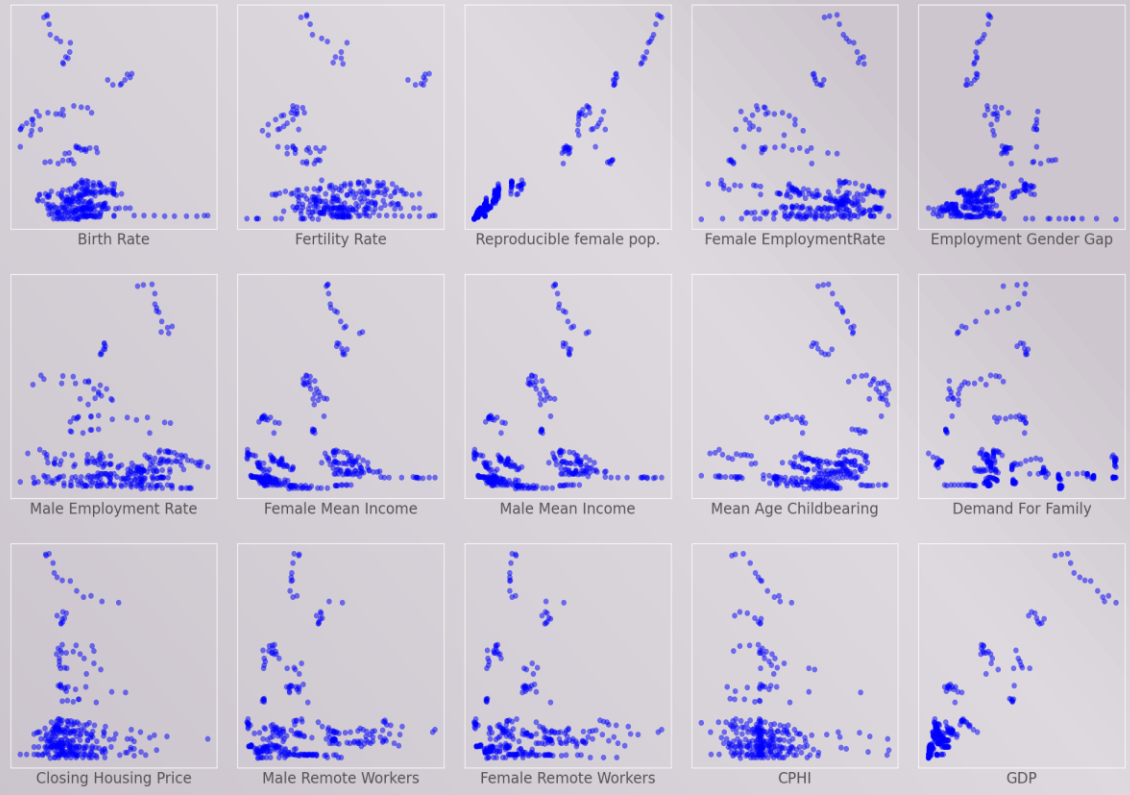
Exploratory Data Analysis
Inspecting relations (1)



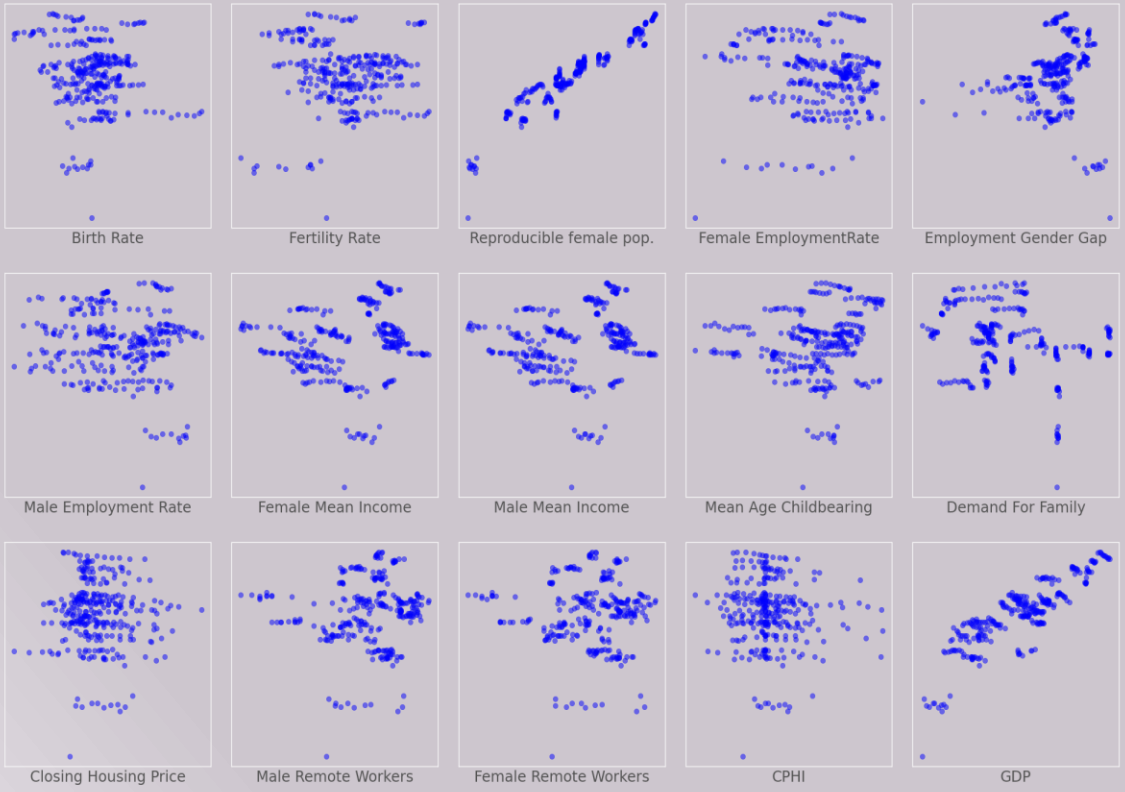
Exploratory Data Analysis

Inspecting relations (2)

Level-level relation with 'Divorces' variable



Log-log relation with 'Divorces' variable



Exploratory Data Analysis

Inspecting relations (3)

Exploratory Data Analysis
Inspecting relations (4)

The dataset (Eurostat and UN database) is full of variables that can explain the number of marriages. We choose the independent variables based on our intuition in terms of causal relationships.

Recall that our goal is not to build a reliable predictive model, but to analyze causal ratios… thus we include variables that are not necessarily informative in terms of correlation.

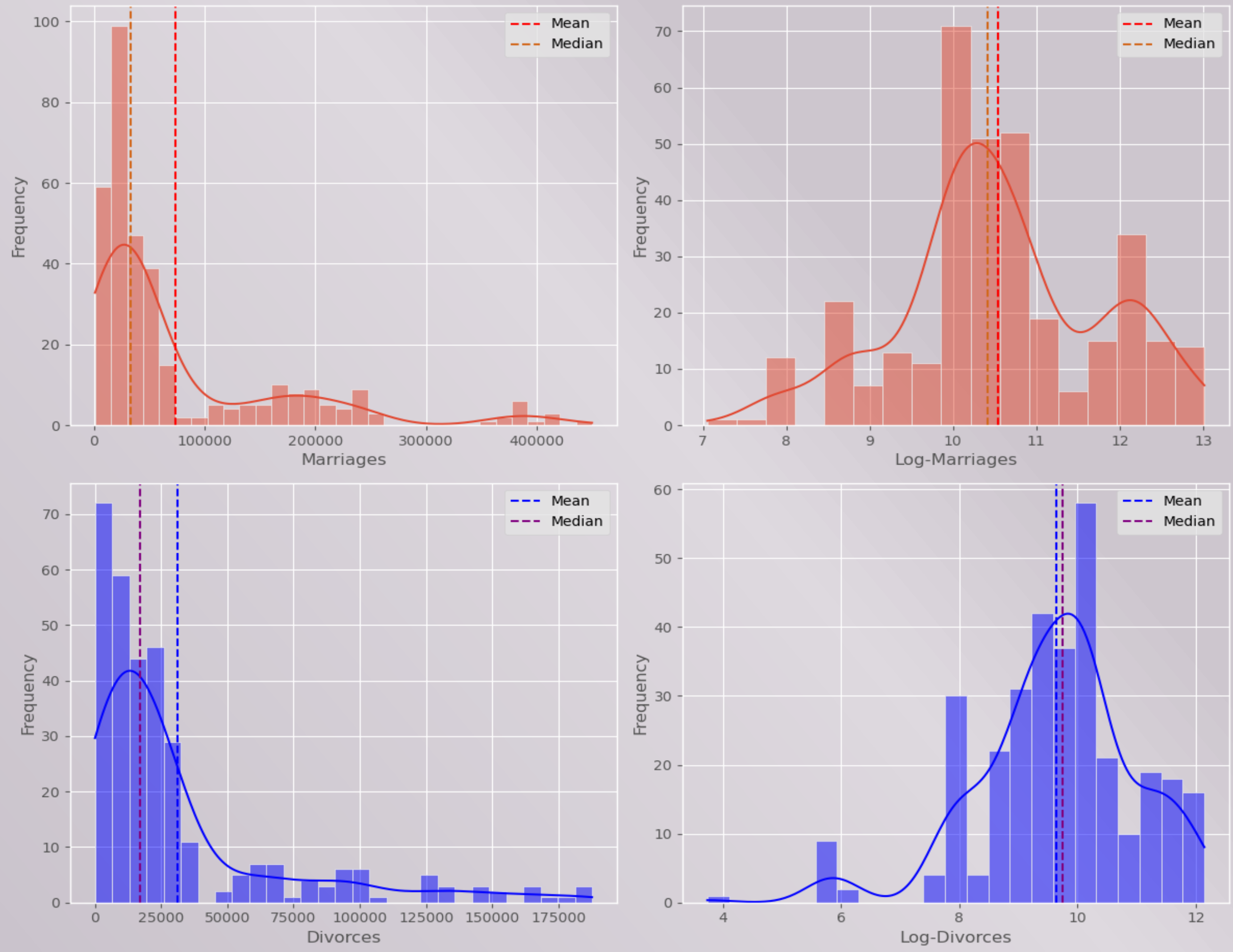
Model	Dependent Variable	Independent Variable	Interpretation of β_1
Level-level	y	x	$\Delta y = \beta_1 \Delta x$
Level-log	y	$\log(x)$	$\Delta y = (\beta_1/100)\%\Delta x$
Log-level	$\log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$
Log-log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

Note how the choice of scale grants a different interpretation of the parameters!

SAPIENZA
UNIVERSITÀ DI ROMA

Exploratory Data Analysis

Looking for a distribution



We would like to start our analysis by considering our dataset as a **pooled cross-sectional** one.
Thus, we are interested in understanding the empirical distribution of the *y* variable.

As can be deduced from the side plots, it is interesting to apply a monotonic transformation (log) to the data empirical distribution, gaining nice statistical properties.

By using the log(y) as dependent variable we are able to control for the skewness of data (simply look at the median-mean gap).

stata

SAPIENZA
UNIVERSITÀ DI ROMA

Exploratory Data Analysis

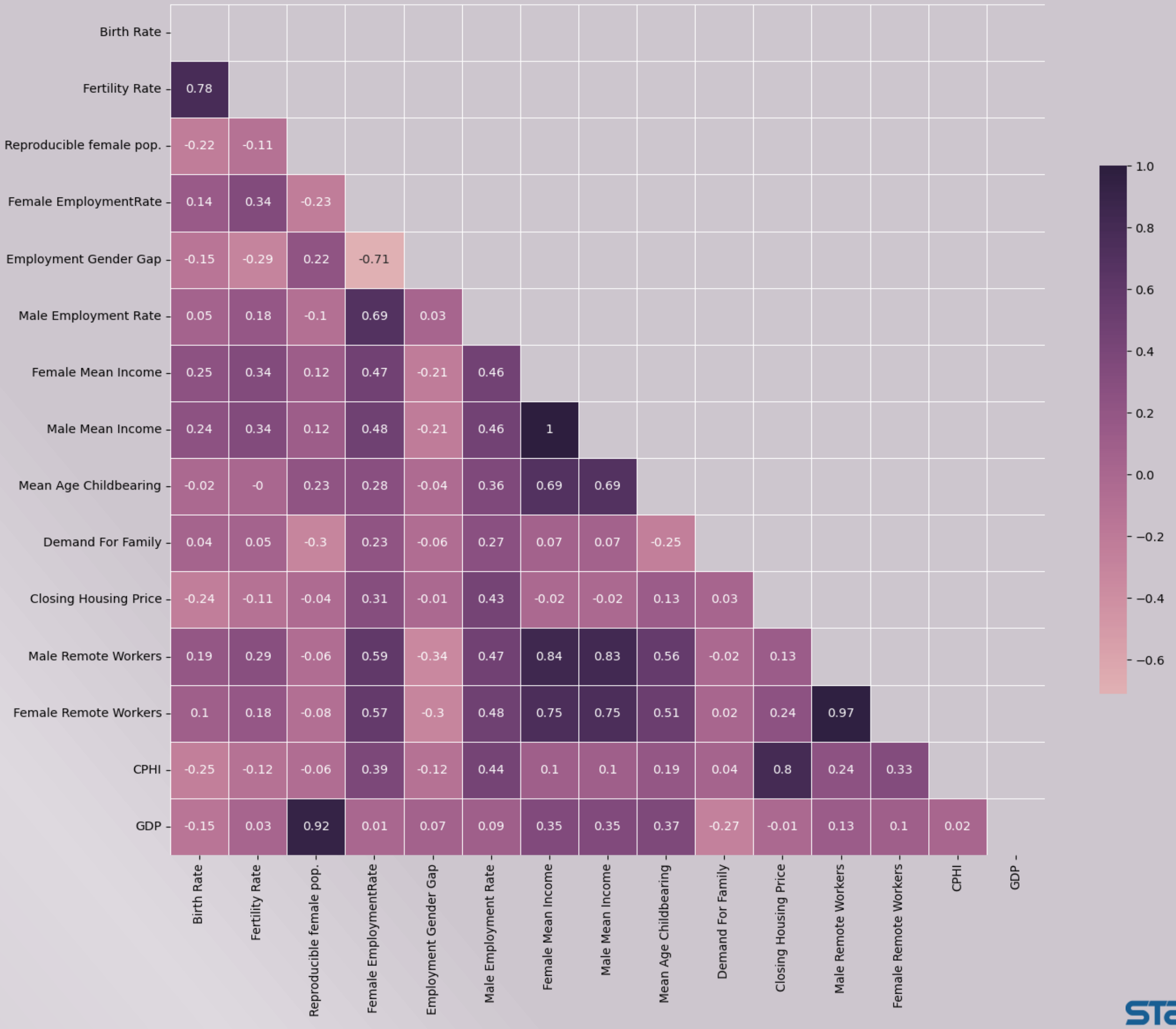
Correlation Matrix

Given the high number of independent variables, it is useful to analyze any strong correlations among them.

Some collinearity arises from the fact that:

- certain variables are proportional to each other (e.g. fertility and birth rate)
- one variable defines a subset (or the complementary set) of another

It will be necessary to **exclude high collinearity variables** when we test our models



stata

Pooled OLS

Using the information form the previous slides, we try to fit our fist model: the Pooled OLS Regression Model.

Two important choices:

1. Use the log(y) as dependent variable
2. Pay attention to highly correlated features

Notice that the Pooled OLS regression exploits the cross-sectional nature of data by stacking samples for different periods of time and introduces the relative period **dummy variables:**

$$y_{it} = \delta_1 + \delta_2 d_2 + \cdots + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + a_i + u_{it}$$

Source	SS	df	MS			
				Number of obs	=	258
				F(29, 228)	=	100.10
Model	356.124861	29	12.2801676	Prob > F	=	0.0000
Residual	27.9710828	228	.122680188	R-squared	=	0.9272
				Adj R-squared	=	0.9179
Total	384.095944	257	1.49453675	Root MSE	=	.35026

log_marriages	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
birth_rate	.0288517	.0384195	0.75	0.453	-.046851	.1045545
fertility_rate	.7656916	.2496227	3.07	0.002	.2738293	1.257554
n_repro_female	3.85e-07	1.91e-08	20.14	0.000	3.47e-07	4.22e-07
f_employment_rate	.0215976	.0067541	3.20	0.002	.0082891	.0349061
m_employment_rate	-.0226198	.0076386	-2.96	0.003	-.0376712	-.0075685
f_mean_income	-.0003147	.0001013	-3.11	0.002	-.0005142	-.0001152
m_mean_income	.0002488	.0000956	2.60	0.010	.0000603	.0004372
gdp	-1.38e-06	1.45e-07	-9.54	0.000	-1.67e-06	-1.10e-06
closing_housing_price	.0053451	.0018345	2.91	0.004	.0017304	.0089598
mean_age_childbearing	-.0038524	.0404036	-0.10	0.924	-.0834645	.0757597
demand_for_family	-.0151782	.0081042	-1.87	0.062	-.0311469	.0007905
percentage_female_work_home	-.1312566	.0143422	-9.15	0.000	-.1595168	-.1029964
percentage_male_work_home	.1545174	.0165167	9.36	0.000	.1219724	.1870624
consumer_price_harmonized_index	.0029252	.0092137	0.32	0.751	-.0152298	.0210802
senior_position_woman	.0277907	.0041806	6.65	0.000	.0195531	.0360284
tertiary_ed_f	-.0760931	.0069182	-11.00	0.000	-.0897249	-.0624613
tertiary_ed_m	.0637968	.0080484	7.93	0.000	.0479381	.0796555
parliament_female	.0122769	.0044056	2.79	0.006	.0035959	.0209578
year_dummy2	.0823204	.1150796	0.72	0.475	-.1444351	.3090759
year_dummy3	-.0104061	.1262893	-0.08	0.934	-.2592494	.2384373
year_dummy4	.1117287	.1324158	0.84	0.400	-.1491865	.3726439
year_dummy5	.072999	.1449343	0.50	0.615	-.2125829	.3585809
year_dummy6	0	(omitted)				
year_dummy7	.2233304	.1473116	1.52	0.131	-.0669357	.5135966
year_dummy8	.2532186	.1496656	1.69	0.092	-.041686	.5481233
year_dummy9	.23464	.1647134	1.42	0.156	-.0899151	.559195
year_dummy10	.2484557	.1771379	1.40	0.162	-.1005809	.5974923
year_dummy11	.2394433	.1936126	1.24	0.217	-.1420555	.620942
year_dummy12	.0467192	.2091927	0.22	0.823	-.365479	.4589173
year_dummy13	0	(omitted)				
year_dummy14	.3039781	.3280467	0.93	0.355	-.3424127	.950369
_cons	9.752684	1.68025	5.80	0.000	6.44188	13.06349

VIF correction for Multicollinearity

Variable	VIF	1/VIF
f_mean_inc~e	1636.94	0.000611
m_mean_inc~e	1589.43	0.000629
percentage..	62.02	0.016125
p~female_w~e	40.58	0.024644
gdp	25.42	0.039338
n_repro_fe~e	17.90	0.055870
year_dummy14	16.18	0.061790
consumer_p~x	12.21	0.081906
tertiary_e~f	10.61	0.094286
tertiary_e~m	8.95	0.111789
year_dummy12	7.18	0.139305
year_dummy11	6.40	0.156218
birth_rate	6.36	0.157268
f_employme~e	5.96	0.167661
year_dummy10	5.15	0.194284
fertility_~e	5.13	0.194967
closing_ho~e	4.83	0.207128
mean_age_c~g	4.44	0.225267
year_dummy9	4.27	0.234406
parliament~e	4.12	0.242492
m_employme~e	4.09	0.244274
year_dummy7	3.85	0.259713
year_dummy8	3.83	0.261429
senior_pos~n	3.57	0.279760
year_dummy5	3.45	0.290213
year_dummy4	3.11	0.321431
year_dummy3	2.18	0.459385
year_dummy2	1.90	0.526315
demand_for~y	1.78	0.563129
Mean VIF	120.75	

Multicollinearity

High (but not perfect) correlation between two or more independent variables is called multicollinearity.

VIF

Variance Inflation Factor statistic is intended to determine the severity of multicollinearity. Thresholding the values of VIF at 10 to choice which variables potentially affect our model

$$VIF_j = \frac{1}{(1-R_j)} \quad \rightarrow \quad Var(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{SST_j} VIF_j$$

Testing the dummy variables

```
( 1)  year_dummy2 = 0
( 2)  year_dummy3 = 0
( 3)  year_dummy4 = 0
( 4)  year_dummy5 = 0
( 5)  year_dummy7 = 0
( 6)  year_dummy8 = 0
( 7)  year_dummy9 = 0
( 8)  year_dummy10 = 0
( 9)  year_dummy11 = 0
(10)  year_dummy12 = 0
(11)  year_dummy14 = 0

      F( 11,   228) =    1.08
           Prob > F =    0.3754
```

We want to test that, once we controlled for all the other variables, the year dummy variables are not significant for our model.

Thus, we test 12 **exclusion restrictions** using the **F test statistic**

$$F \equiv \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)}$$

Where:

- SSR_r is the sum of squared residuals from the *restricted* model
- SSR_{ur} is the sum of squared residuals from the *unrestricted* model

We cannot reject the null hypothesis → we can exclude the dummy variables from our model

New Pooled OLS

New VIF and exclusion of temporal information

We adjust our results for **multicollinearity** and remove the dummy variables that are **jointly non-significant**

We can verify that we have no more issues with VIF, but some variable are still statistically non-significant

Source	SS	df	MS			
				Number of obs	=	258
				F(15, 242)	=	38.06
Model	269.761063	15	17.9840709	Prob > F	=	0.0000
Residual	114.334881	242	.472458188	R-squared	=	0.7023
				Adj R-squared	=	0.6839
Total	384.095944	257	1.49453675	Root MSE	=	.68736

log_marriages	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
birth_rate	.1535324	.0670106	2.29	0.023	.0215339	.2855308
fertility_rate	.3482304	.4527306	0.77	0.443	-.5435652	1.240026
f_employment_rate	-.0189045	.0102025	-1.85	0.065	-.0390015	.0011925
m_employment_rate	-.0128915	.013462	-0.96	0.339	-.0394091	.0136261
f_mean_income	-.0000682	.0000145	-4.69	0.000	-.0000969	-.0000396
closing_housing_price	.0076761	.0030154	2.55	0.012	.0017363	.013616
mean_age_childbearing	-.0499527	.0750999	-0.67	0.507	-.1978856	.0979803
demand_for_family	-.0130612	.0145774	-0.90	0.371	-.041776	.0156536
gdp	1.06e-06	8.63e-08	12.25	0.000	8.87e-07	1.23e-06
percentage_female_work_home	.0025006	.0083945	0.30	0.766	-.0140351	.0190363
consumer_price_harmonized_index	-.0015768	.0099297	-0.16	0.874	-.0211365	.0179829
senior_position_woman	.0099645	.0069318	1.44	0.152	-.0036898	.0236189
tertiary_ed_f	-.0674843	.0126977	-5.31	0.000	-.0924964	-.0424723
tertiary_ed_m	.0622379	.0150133	4.15	0.000	.0326644	.0918114
parliament_female	.0576356	.0076335	7.55	0.000	.042599	.0726722
_cons	12.28692	2.889505	4.25	0.000	6.59513	17.97871

Variable	VIF	1/VIF
tertiary_e~f	9.28	0.107789
f_mean_inc~e	8.76	0.114098
tertiary_e~m	8.08	0.123722
birth_rate	5.02	0.199089
fertility_~e	4.38	0.228265
mean_age_c~g	3.98	0.251100
consumer_p~x	3.68	0.271584
p~female_w~e	3.61	0.277032
f_employme~e	3.53	0.282975
closing_ho~e	3.39	0.295229
m_employme~e	3.30	0.302887
parliament~e	3.21	0.311067
senior_pos~n	2.55	0.391896
gdp	2.34	0.427980
demand_for~y	1.49	0.670281
Mean VIF	4.44	

New Pooled OLS

Final feature selection

Finally, we both remove statistically irrelevant and conceptually wrong variables: we don't want to include in our model independent variables that are mostly an **effect** of our dependent variable rather than a **cause**.

Endogeneity issue

Looking at the table at slide 11, we can interpret the parameter of the *'female employment rate'* variable in the following way:
A marginal increment in the female employment rate causes a decrement of about 2.6% in the number of marriages

Source	SS	df	MS
Model	256.833927	8	32.1042408
Residual	129.967905	261	.497961322
Total	386.801832	269	1.43792503

Number of obs	=	270
F(8, 261)	=	64.47
Prob > F	=	0.0000
R-squared	=	0.6640
Adj R-squared	=	0.6537
Root MSE	=	.70566

log_marriages	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
f_employment_rate	-.0258076	.0076224	-3.39	0.001	-.0408169	-.0107984
f_mean_income	-.0000292	8.20e-06	-3.56	0.000	-.0000453	-.000013
closing_housing_price	.0046755	.0018671	2.50	0.013	.000999	.0083521
mean_age_childbearing	-.2265932	.0584719	-3.88	0.000	-.3417298	-.1114566
gdp	1.08e-06	7.42e-08	14.63	0.000	9.39e-07	1.23e-06
tertiary_ed_f	-.0437079	.010731	-4.07	0.000	-.0648383	-.0225775
tertiary_ed_m	.0544838	.0134108	4.06	0.000	.0280767	.0808909
parliament_female	.0479613	.00632	7.59	0.000	.0355166	.060406
_cons	17.38958	1.724129	10.09	0.000	13.99461	20.78456

A unitary increment in the 'mean age childbearing' causes a decrement of about 22.6% in the number of marriages

Heteroskedasticity

Testing time

Homoskedasticity definition:
The OLS regression model exhibits Homoskedasticity if the error u has the same variance given any values of the explanatory variables, i.e.:

$$Var(u \mid x_1, \dots, x_k) = \sigma^2$$

Otherwise, the model shows **Heteroskedasticity**

We use the **Breusch-Pagan test** and **White's test** to check for homoskedasticity

```
White's test
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity

   chi2(44) = 214.39
Prob > chi2 = 0.0000

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	214.39	44	0.0000
Skewness	92.05	8	0.0000
Kurtosis	5.90	1	0.0152
Total	312.34	53	0.0000

```
Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of log_marriages

H0: Constant variance

    chi2(1) =    4.51
Prob > chi2 = 0.0338
```

The In both tests we reject the null hypothesis, detecting the presence of heteroskedasticity!

Heteroskedasticity

Correction methods

We now want to select the method to control heteroskedasticity that best fits our case. We want to use the "*hetreg*" command in Stata and we have two options:

1. Maximum Likelihood (ML) method → more efficient if the mean and variance function are correctly and the errors are normally distributed

2. Feasible Generalized Least-Squares (FGLS) → more robust if the variance function is incorrect or the errors are nonnormal.

GLS method accounts for the heteroskedasticity structure by applying a suitable weighting matrix to the regression equations.

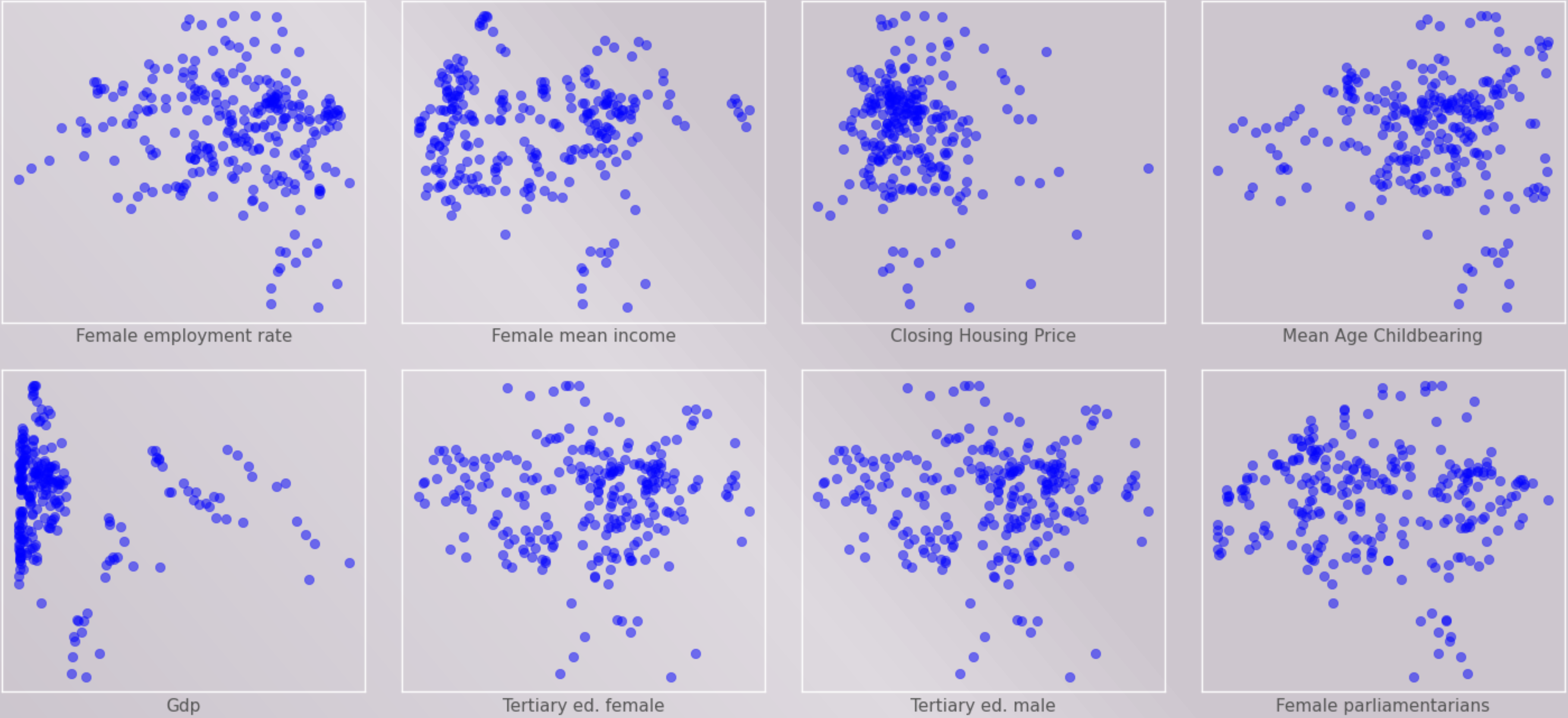
From GLS to FGLS →

FGLS method, the heteroskedasticity-consistent estimated variance-covariance matrix is used to calculate the GLS weights.

Heteroskedasticity

Inspecting the variables we want to use to model error variance in FGLS

But first all we inspect the variables on which we want to model the variance, we select the variables that seem to show some relation with the residuals



Heteroskedasticity

Feasible GLS

FGLS is also called Two-step GLS since assuming that:

$$Var(u \mid x) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k)$$

it can be divided in two steps:

1. We define the error as a function of the variance and the independent variables

$$u_2 = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k)$$

$$\log(u_2) = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k + error$$

2. From this regression the fitted values $\hat{g}_i$ are used for the estimates $\hat{h}_i$

$$\hat{h}_i = \exp(\hat{g}_i)$$

We thus correct the homoskedasticity defining the model variance for different combinations of independent variables

Heteroskedasticity

Heteroskedastic robust OLS regression

After heteroskedasticity correction we found that the *'mean age childbearing'* is no more statically relevant.

Heteroskedastic linear regression
ML estimation

Number of obs = 270
Wald chi2(8) = 1021.55
Log likelihood = **-220.1782**
Prob > chi2 = 0.0000

log_marriages	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
log_marriages						
f_employment_rate	-.0185399	.0052051	-3.56	0.000	-.0287416	-.0083382
f_mean_income	-.0000169	3.53e-06	-4.79	0.000	-.0000239	-.00001
closing_housing_price	-.0016667	.0013651	-1.22	0.222	-.0043423	.0010089
mean_age_childbearing	-.0355575	.0470504	-0.76	0.450	-.1277746	.0566596
gdp	8.85e-07	4.38e-08	20.19	0.000	7.99e-07	9.71e-07
tertiary_ed_f	-.0352465	.0069394	-5.08	0.000	-.0488475	-.0216455
tertiary_ed_m	.045256	.0087837	5.15	0.000	.0280402	.0624717
parliament_female	.0248366	.0052013	4.78	0.000	.0146422	.0350311
_cons	12.32983	1.349487	9.14	0.000	9.684881	14.97477
lnsigma2						
f_mean_income	-.0000721	.000018	-4.00	0.000	-.0001073	-.0000368
mean_age_childbearing	.5260506	.1526125	3.45	0.001	.2269356	.8251655
parliament_female	-.106242	.0130167	-8.16	0.000	-.1317543	-.0807298
_cons	-12.90114	4.358397	-2.96	0.003	-21.44344	-4.358842

LR test of lnsigma2=0: chi2(3) = **128.46** Prob > chi2 = **0.0000**

Heteroskedastic linear regression
ML estimation

Number of obs = 270
Wald chi2(7) = 1038.89
Log likelihood = **-220.4656**
Prob > chi2 = 0.0000

log_marriages	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
log_marriages						
f_employment_rate	-.0178707	.0050277	-3.55	0.000	-.0277249	-.0080166
f_mean_income	-.0000179	3.28e-06	-5.45	0.000	-.0000243	-.0000114
closing_housing_price	-.002124	.0012152	-1.75	0.080	-.0045057	.0002577
gdp	8.77e-07	4.25e-08	20.62	0.000	7.94e-07	9.61e-07
tertiary_ed_f	-.0348958	.0069206	-5.04	0.000	-.0484599	-.0213317
tertiary_ed_m	.0450726	.0087567	5.15	0.000	.0279097	.0622355
parliament_female	.0228137	.0044921	5.08	0.000	.0140092	.0316181
_cons	11.33793	.3080196	36.81	0.000	10.73423	11.94164
lnsigma2						
f_mean_income	-.0000707	.000018	-3.94	0.000	-.0001059	-.0000355
mean_age_childbearing	.5189918	.1550547	3.35	0.001	.2150902	.8228934
parliament_female	-.1098532	.0121838	-9.02	0.000	-.1337331	-.0859734
_cons	-12.60656	4.422061	-2.85	0.004	-21.27364	-3.939476

LR test of lnsigma2=0: chi2(3) = **142.99** Prob > chi2 = **0.0000**

The female income and the other variables seems to be relevant for modeling the variance

The **LR test** statistic supports model's heteroskedastic nature.

Random Effect Model

We introduce Panel Data methods, trying to solve the omitted variables problem and accounting for the **unobserved effect** a_i. We start from the definition of the **unobserved effect model**

$$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \cdots + \beta_k x_{itk} + a_i + u_{it}, \qquad t = 1,2,\ldots,T$$

RE assumption a_i is uncorrelated with each explanatory variable:

$$Cov(x_{itj}, \alpha_i) = 0,\ \ t = 1,2,\ldots,T, j = 1,2,\ldots,k$$

The unobserved effect model with composite error term $v_{it} = a_i + u_{it}$:

$$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \cdots + \beta_k x_{itk} + v_{it},\ \ t = 1,2,\ldots,T$$

under the random affect assumption, we have that v_{it} are correlated across time:

$$Corr(v_{it}, v_{is}) = \frac{\sigma_a^2}{(\sigma_u^2 + \sigma_a^2)}, \qquad t \neq s$$

The **RE estimator** is given by:

$$y_{it} - \theta\bar{y}_i = \beta_0(1-\theta) + \beta_1(x_{it1} - \theta\bar{x}_{i1}) + \cdots + \beta_k(x_{itk} - \theta\hat{x}_{ik}) + (v_{it} - \theta\bar{v}_i), \qquad \theta = 1 - \left[\frac{\sigma_u^2}{(\sigma_u^2 + T\sigma_u^2)}\right]^{\frac{1}{2}}$$

RE vs OLS


```
Breusch and Pagan Lagrangian multiplier test for random effects

        log_marriages[numeric_Nation,t] = Xb + u[numeric_Nation] + e[numeric_Nation,t]

        Estimated results:
                         |       Var     SD = sqrt(Var)
                ---------+-----------------------------
                log_mar~s |   1.437925       1.199135
                        e |   .0191729       .1384663
                        u |   .6109999       .7816648

        Test: Var(u) = 0
                             chibar2(01) =   972.75
                          Prob > chibar2 =   0.0000
```

From the **Breush and Pagan Lagrangian Multiplier test** we deduce that we would prefer to consider the Random Effect model to the Pooled OLS

An increment of 1 thousands of euro in the *female mean income* causes a decrease of the 2.5% in the number of marriages

Here we report the new results for the RE model.

Notice the change in the coefficients and in the significance levels of the selected variables.

After the utilization of the RE model we notice that the *level of education* for females is no more relevant

log_marriages	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
f_employment_rate	.0083368	.0041272	2.02	0.043	.0002475	.016426
f_mean_income	-.0000247	7.60e-06	-3.26	0.001	-.0000396	-9.84e-06
closing_housing_price	.0016775	.0006345	2.64	0.008	.0004339	.0029211
gdp	3.20e-07	9.55e-08	3.35	0.001	1.33e-07	5.07e-07
tertiary_ed_f	.0049353	.004821	1.02	0.306	-.0045136	.0143842
tertiary_ed_m	-.013178	.0066952	-1.97	0.049	-.0263004	-.0000555
parliament_female	-.0071125	.0028989	-2.45	0.014	-.0127943	-.0014307
_cons	10.44116	.2531555	41.24	0.000	9.944985	10.93734
sigma_u	.78166481					
sigma_e	.13846629					
rho	.96957515	(fraction of variance due to u_i)				

stata

Fixed Effect Model

The fixed effect model is built using the **Between estimator** (average equation over time)

$$\bar{y}_{it} = \beta_1\bar{x}_{it1} + \beta_2\bar{x}_{it2} + \cdots + \beta_k\bar{x}_{itk} + a_i + \bar{u}_{it}, \qquad t = 1,2,\ldots,T$$

Then we deduce the real fixed effect model i.e., the **Within estimator**

$$y_{it} - \bar{y}_{it} = \beta_1(x_{it1} - \bar{x}_{it1}) + \beta_2(x_{it2} - \bar{x}_{it2}) + \cdots + \beta_k(x_{itk} - \bar{x}_{itk}) + u_{it} - \bar{u}_i$$

$$\ddot{y}_{it} = \beta_1\ddot{x}_{it1} + \beta_2\ddot{x}_{it2} + \cdots + \beta_k\ddot{x}_{itk} + \ddot{u}_{it}, \qquad t = 1,2,\ldots,T$$

Obs. 1: FE model is equivalent to a RE model with θ=1 (when θ=0 RE is equivalent to a Pooled OLS model)

Obs. 2: fixed effects allows arbitrary correlation between a_i and x_{itk}, while random effects does not
→ FE is widely thought to be a more convincing tool for estimating ceteris paribus effects.

FE vs RE

SAPIENZA
UNIVERSITÀ DI ROMA

The fixed effect results are quite similar to the ones obtained with RE

The coefficients of the variable of interest (regarding the female empowerment) don't change that much and we can give a similar interpretation.

log_marriages	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
f_employment_rate	.0090992	.0039349	2.31	0.022	.0013475	.0168509
f_mean_income	-.0000288	7.79e-06	-3.70	0.000	-.0000441	-.0000134
closing_housing_price	.0020063	.0006216	3.23	0.001	.0007818	.0032307
gdp	1.31e-07	9.83e-08	1.33	0.185	-6.30e-08	3.24e-07
tertiary_ed_f	.0062642	.0045818	1.37	0.173	-.0027618	.0152903
tertiary_ed_m	-.0129645	.0063843	-2.03	0.043	-.0255415	-.0003876
parliament_female	-.0077849	.0027556	-2.83	0.005	-.0132134	-.0023564
_cons	10.43113	.1836708	56.79	0.000	10.0693	10.79296
sigma_u	1.2623653					
sigma_e	.13846629					
rho	.98811158	(fraction of variance due to u_i)				

F test that all u_i=0: F(24, 238) = 288.78 Prob > F = 0.0000

	Coefficients (b) fixed	(B) random	(b-B) Difference	sqrt(diag(V_b-V_B)) Std. err.
f_employme~e	.0090992	.0083368	.0007624	.0005856
f_mean_inc~e	-.0000288	-.0000247	-4.05e-06	3.21e-06
closing_ho~e	.0020063	.0016775	.0003288	.0001761
gdp	1.31e-07	3.20e-07	-1.89e-07	4.16e-08
tertiary_e~f	.0062642	.0049353	.0013289	.000564
tertiary_e~m	-.0129645	-.013178	.0002134	.0009576
parliament~e	-.0077849	-.0071125	-.0006724	.0003439

b = Consistent under H0 and Ha; obtained from **xtreg**.
B = Inconsistent under Ha, efficient under H0; obtained from **xtreg**.

Test of H0: Difference in coefficients not systematic

chi2(6) = (b-B)'[(V_b-V_B)^(-1)](b-B)
= 33.81
Prob > chi2 = 0.0000

We use the **Hausman test** to select the FE model against the RE model.

We have systematic differences in the coefficients thus we opt for the FE model.

stata

Improving the RE and FE model

Surprisingly, there is space for improvements for the both RE and FE model by reintroducing some independent variables that we previously excluded during the analysis of the Pooled OLS results.

We get this new final model that allows for better **Within and Between Goodness-of-fit performances**.

```
R-squared:                                 Obs per group:
     Within  = 0.5151                                    min =          1
     Between = 0.6273                                    avg =        9.2
     Overall = 0.6946                                    max =         12

                                           F(11, 187)          =      18.06
corr(u_i, Xb) = 0.0487                     Prob > F            =     0.0000
```

marriages_log	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
birth_rate	.0626054	.0157382	3.98	0.000	.0315581	.0936526
n_repro_female	1.08e-07	5.20e-08	2.08	0.039	5.47e-09	2.11e-07
gdp	6.60e-07	1.62e-07	4.07	0.000	3.40e-07	9.81e-07
m_employment_rate	.0251695	.0073866	3.41	0.001	.0105978	.0397412
f_employment_rate	-.0075409	.0088639	-0.85	0.396	-.025027	.0099452
percentage_female_work_home	-.008366	.0017519	-4.78	0.000	-.0118221	-.00491
consumer_price_harmonized_index	.0074592	.0015986	4.67	0.000	.0043057	.0106128
f_mean_age_marriage	-.0733563	.020978	-3.50	0.001	-.1147402	-.0319724
year_dummy7	.0615827	.0213065	2.89	0.004	.0195507	.1036147
year_dummy8	.0739005	.0212724	3.47	0.001	.0319358	.1158653
year_dummy9	.0508279	.0215025	2.36	0.019	.0084092	.0932466
_cons	9.38697	.633315	14.82	0.000	8.13761	10.63633
sigma_u	.74737436					
sigma_e	.08884017					
rho	.98606687	(fraction of variance due to u_i)				

F test that all u_i=0: F(23, 187) = 290.10 Prob > F = 0.0000

```
R-squared:                                 Obs per group:
     Within  = 0.5129                                    min =          1
     Between = 0.6480                                    avg =        9.2
     Overall = 0.6968                                    max =         12

                                           Wald chi2(11)       =     248.20
corr(u_i, X) = 0 (assumed)                 Prob > chi2         =     0.0000
```

marriages_log	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
birth_rate	.0609624	.0155766	3.91	0.000	.0304328	.091492
n_repro_female	1.24e-07	2.24e-08	5.55	0.000	8.04e-08	1.68e-07
gdp	6.20e-07	1.12e-07	5.53	0.000	4.00e-07	8.39e-07
m_employment_rate	.0200875	.0070075	2.87	0.004	.006353	.033822
f_employment_rate	-.0017123	.0081792	-0.21	0.834	-.0177432	.0143185
percentage_female_work_home	-.008341	.0017199	-4.85	0.000	-.011712	-.0049699
consumer_price_harmonized_index	.0071475	.0016111	4.44	0.000	.0039897	.0103052
f_mean_age_marriage	-.071439	.0198922	-3.59	0.000	-.1104271	-.032451
year_dummy7	.0626679	.0220219	2.85	0.004	.0195058	.10583
year_dummy8	.0735575	.0219603	3.35	0.001	.030516	.116599
year_dummy9	.0518281	.0221712	2.34	0.019	.0083734	.0952828
_cons	9.256084	.5603753	16.52	0.000	8.157769	10.3544
sigma_u	.57196305					
sigma_e	.08884017					
rho	.97644248	(fraction of variance due to u_i)				

A chance for the First Differencing

The First Differencing model also accounts for the fixed effect term, but this we do the difference between samples at adjacent periods (instead of the difference with the average):

$$\Delta y_{it} = \alpha_0 + \alpha_3 d3_t + \alpha_4 d4_t + \ \ldots\ \alpha_T dT_t + \beta_1 \Delta \mathrm{x}_{it1} + \ \ldots\ + \beta_k \Delta x_{itk} + \Delta u_{it}, \qquad t = 2,3,\ldots,T$$

Source	SS	df	MS		
				Number of obs =	184
				F(11, 172) =	17.19
Model	1.57842167	11	.14349288	Prob > F =	0.0000
Residual	1.43593394	172	.008348453	R-squared =	0.5236
				Adj R-squared =	0.4932
Total	3.01435562	183	.016471889	Root MSE =	.09137

d_log_marriages	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
d_birth_rate	.0610467	.0293095	2.08	0.039	.0031941	.1188993
d_n_repro_female	4.14e-07	1.47e-07	2.81	0.006	1.23e-07	7.04e-07
d_gdp	1.90e-06	2.49e-07	7.62	0.000	1.40e-06	2.39e-06
d_m_employment_rate	.0119063	.0098635	1.21	0.229	-.0075627	.0313753
d_f_employment_rate	.0178085	.0117805	1.51	0.132	-.0054445	.0410615
d_percentage_female_work_home	-.0048029	.0026304	-1.83	0.070	-.0099948	.0003891
d_cpriceharmonizedindex	.0037982	.0020778	1.83	0.069	-.0003031	.0078995
d_f_mean_age_marriage	-.0906559	.029343	-3.09	0.002	-.1485747	-.0327371
d_year_dummy7	.0325812	.0175578	1.86	0.065	-.0020753	.0672377
d_year_dummy8	.0367377	.0212109	1.73	0.085	-.0051295	.0786049
d_year_dummy9	.0084505	.0180997	0.47	0.641	-.0272757	.0441767
_cons	-.0112042	.0138322	-0.81	0.419	-.0385068	.0160984

When $T > 2$, the FE and FD estimators are not the same but:

- both are unbiased under Assumptions FE.1 through FE.4 → we cannot use unbiasedness as a criterion
- both are consistent under FE.1 through FE.4.

For large N and small T, the choice between FE and FD hinges on the serial correlation in the idiosyncratic errors, u_{it} :

u_{it} serially uncorrelated → FE is better

Comments

SAPIENZA UNIVERSITÀ DI ROMA

- The dummy for the **year 2020** has great significance and magnitude, representing the abrupt decrease in marriages provoked by the **COVID-19 pandemic**. A great correlation with the error forced us to not include it in our models; our variables aren't able to totally capture the complex effects the pandemic had on the society and the economy.

- Our model seem to capture that higher economic security (‘*m_employment_rate*’, ‘*consumer_price_harmonized_index*’) has a positive effect on the number of marriages.

- Another observed trend is the negative correlation with factors commonly attributed to an **increased involvement of women in education and workforce** (‘*f_employment_rate*’, ‘*tertiary_ed_f*,’ ‘*f_mean_age_marriage*’).

- We consider particularly meaningful the last one: as more people decide to delay marriage to pursue education and careers relationships evolve differently, with alternative arrangements like cohabitation becoming commoner and formalizations as marriage becoming of secondary importance.

stata

LDV with Poisson regression

Choosing a distribution for *counting data*:

$$Y \in (0,1,2,\dots)$$

$$Y \sim Poiss(\lambda)$$

Poisson assumption:

$$E(Y) = Var(Y) = \lambda$$

$$E(y|x_1, x_2, \dots, x_k) = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$$

FE

```
Iteration 0:   Log likelihood = -149617.02
Iteration 1:   Log likelihood = -118310.15
Iteration 2:   Log likelihood = -118305.76
Iteration 3:   Log likelihood = -118305.76

Conditional fixed-effects Poisson regression        Number of obs     =        270
Group variable: numeric_Nation                      Number of groups  =         25

                                                    Obs per group:
                                                                  min =          9
                                                                  avg =       10.8
                                                                  max =         12

                                                    Wald chi2(8)      =   62682.16
Log likelihood = -118305.76                         Prob > chi2       =     0.0000
```

marriages	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
f_employment_rate	.0108164	.000178	60.78	0.000	.0104676	.0111652
f_mean_income	-.0000346	3.00e-07	-115.56	0.000	-.0000352	-.000034
closing_housing_price	.0015083	.0000228	66.24	0.000	.0014636	.0015529
mean_age_childbearing	-.029148	.0016077	-18.13	0.000	-.032299	-.025997
gdp	1.82e-07	1.74e-09	104.58	0.000	1.79e-07	1.85e-07
tertiary_ed_f	.0054198	.0001777	30.49	0.000	.0050714	.0057682
tertiary_ed_m	-.0088019	.0002338	-37.65	0.000	-.0092601	-.0083437
parliament_female	-.0101233	.0000729	-138.92	0.000	-.0102661	-.0099805

RE

```
Random-effects Poisson regression                   Number of obs     =        270
Group variable: numeric_Nation                      Number of groups  =         25

Random effects u_i ~ Gamma                          Obs per group:
                                                                  min =          9
                                                                  avg =       10.8
                                                                  max =         12

                                                    Wald chi2(8)      =   62682.30
Log likelihood = -118670.01                         Prob > chi2       =     0.0000
```

marriages	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
f_employment_rate	.0108154	.000178	60.77	0.000	.0104666	.0111642
f_mean_income	-.0000346	3.00e-07	-115.56	0.000	-.0000352	-.000034
closing_housing_price	.0015081	.0000228	66.24	0.000	.0014635	.0015527
mean_age_childbearing	-.0291421	.0016076	-18.13	0.000	-.0322929	-.0259913
gdp	1.82e-07	1.74e-09	104.60	0.000	1.79e-07	1.85e-07
tertiary_ed_f	.0054188	.0001777	30.49	0.000	.0050704	.0057672
tertiary_ed_m	-.0088024	.0002338	-37.65	0.000	-.0092606	-.0083442
parliament_female	-.0101231	.0000729	-138.92	0.000	-.0102659	-.0099802
_cons	11.99252	.2147884	55.83	0.000	11.57155	12.4135
/lnalpha	.1081761	.2466399			-.3752292	.5915813
alpha	1.114244	.274817			.6871318	1.806843

```
LR test of alpha=0: chibar2(01) = 4.2e+06                  Prob >= chibar2 = 0.000
```

Checking Overdispersion

From the Poisson to the Negative Binomial model

Poisson main drawback:
The Poisson assumption may not hold

Check for overdispersion
(*Variance* $\gg$ *Mean*)

Negative Binomial Model

Marriages				
	Percentiles	Smallest		
1%	2546	1157		
5%	5499	2276		
10%	6671	2353	Obs	344
25%	20670	2546	Sum of wgt.	344
50%	33453		Mean	73990.53
		Largest	Std. dev.	92560.83
75%	72074.5	407466		
90%	203850	410426	Variance	8.57e+09
95%	241831	416324	Skewness	2.008607
99%	407466	449466	Kurtosis	6.665913

$$Y \in (0,1,2,\ldots)$$

$$Y \sim NB(r,p)$$

$$\Pr(Y=k) = \binom{k+r-1}{k}(1-p)^k p^r$$

stata

LDV with Negative Binomial

SAPIENZA UNIVERSITÀ DI ROMA

Here we find that the RE model seems to have a better log-likelihood than FE (and it is not correct to compare the log-likelihood for models based on different distributions, but we have a difference of 3 orders of magnitude with the Poisson model).

FE

```
Conditional FE negative binomial regression         Number of obs     =    270
Group variable: numeric_Nation                      Number of groups  =     25

                                                    Obs per group:
                                                                  min =      9
                                                                  avg =   10.8
                                                                  max =     12

                                                    Wald chi2(8)      =  47.63
Log likelihood = -2392.0806                         Prob > chi2       = 0.0000
```

marriages	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
f_employment_rate	.0149411	.0038531	3.88	0.000	.0073892	.022493
f_mean_income	-.0000136	8.04e-06	-1.69	0.090	-.0000294	2.14e-06
closing_housing_price	.0009912	.0005603	1.77	0.077	-.0001069	.0020893
mean_age_childbearing	-.0572244	.0367293	-1.56	0.119	-.1292125	.0147637
gdp	1.41e-07	7.16e-08	1.97	0.049	7.42e-10	2.82e-07
tertiary_ed_f	.0060548	.0041166	1.47	0.141	-.0020135	.0141231
tertiary_ed_m	-.0139253	.0057404	-2.43	0.015	-.0251763	-.0026742
parliament_female	-.0050666	.0024748	-2.05	0.041	-.0099172	-.0002161
_cons	5.169604	.9536638	5.42	0.000	3.300458	7.038751

RE

```
Random-effects negative binomial regression         Number of obs     =    270
Group variable: numeric_Nation                      Number of groups  =     25

Random effects u_i ~ Beta                           Obs per group:
                                                                  min =      9
                                                                  avg =   10.8
                                                                  max =     12

                                                    Wald chi2(7)      =  41.10
Log likelihood = -2759.9183                         Prob > chi2       = 0.0000
```

marriages	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
f_employment_rate	.0144171	.0037982	3.80	0.000	.0069729	.0218614
f_mean_income	-8.43e-06	7.44e-06	-1.13	0.258	-.000023	6.16e-06
closing_housing_price	.000887	.0005549	1.60	0.110	-.0002006	.0019746
mean_age_childbearing	-.0614486	.036586	-1.68	0.093	-.133156	.0102587
tertiary_ed_f	.0056297	.0040845	1.38	0.168	-.0023757	.0136352
tertiary_ed_m	-.0130803	.0056911	-2.30	0.022	-.0242346	-.001926
parliament_female	-.0037788	.0025049	-1.51	0.131	-.0086883	.0011306
_cons	5.270684	.9520635	5.54	0.000	3.404674	7.136694
/ln_r	-.2195921	.2458157			-.7013819	.2621978
/ln_s	5.521454	.3468834			4.841575	6.201333
r	.8028462	.1973522			.4958995	1.299784
s	249.9982	86.72021			126.6686	493.4061

```
LR test vs. pooled: chibar2(01) = 947.39              Prob >= chibar2 = 0.000
```

stata

LDV with Negative Binomial

For the interpretability of the results, we prefer to report the coefficients in terms of **Incidence Rate Ratio:**

$$\boxed{IRR_i = \exp(\beta_i)}$$

FE

marriages	IRR	Std. err.	z	P>|z|	[95% conf.	interval]
f_employment_rate	1.015053	.0039111	3.88	0.000	1.007417	1.022748
f_mean_income	.9999864	8.04e-06	-1.69	0.090	.9999706	1.000002
closing_housing_price	1.000992	.0005608	1.77	0.077	.9998931	1.002092
mean_age_childbearing	.9443821	.0346865	-1.56	0.119	.8787872	1.014873
gdp	1	7.16e-08	1.97	0.049	1	1
tertiary_ed_f	1.006073	.0041416	1.47	0.141	.9979885	1.014223
tertiary_ed_m	.9861712	.0056611	-2.43	0.015	.975138	.9973294
parliament_female	.9949462	.0024623	-2.05	0.041	.9901318	.9997839
_cons	175.8452	167.6972	5.42	0.000	27.12505	1139.963

RE

marriages	IRR	Std. err.	z	P>|z|	[95% conf.	interval]
f_employment_rate	1.014522	.0038533	3.80	0.000	1.006997	1.022102
f_mean_income	.9999916	7.44e-06	-1.13	0.258	.999977	1.000006
closing_housing_price	1.000887	.0005554	1.60	0.110	.9997994	1.001977
mean_age_childbearing	.9404012	.0344056	-1.68	0.093	.8753286	1.010311
tertiary_ed_f	1.005646	.0041075	1.38	0.168	.9976271	1.013729
tertiary_ed_m	.9870049	.0056171	-2.30	0.022	.9760567	.9980759
parliament_female	.9962283	.0024954	-1.51	0.131	.9913493	1.001131
_cons	194.5489	185.2229	5.54	0.000	30.10447	1257.265

Interpretation:

- If IRR_i<1 → negative effect on the y variable
- Otherwise → positive effect on the y variable

Refactoring: changing the target variable

SAPIENZA
UNIVERSITÀ DI ROMA

Let us now introduce a deeper insight: we try to capture a more interesting signal for the analysis of the causal relationship between the number of marriages and the independent variables.

Old setup

The number of marriages (log scale) previously used contained only the information about the new number of marriages for each year.

If not used in conjunction with the number of divorces, however, it did not capture the information of the actual number of married people w.r.t. the population of each country.

New setup

We propose as a new dependent variable the **number of women of reproductive age currently married** (for each year).

In this way we are able to take into account at the same time:

- new marriage relationships being formed
- old marital relationships that break down

stata

FE and RE with the new variable

Here we report our best model for this new setup

FE

R-squared:
Within = 0.8703
Between = 0.9914
Overall = 0.9908

Obs per group:
min = 2
avg = 7.6
max = 10

F(6, 152) = 169.91
corr(u_i, Xb) = -0.2715   Prob > F = 0.0000

log_tot_f_married_in_~e	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
log_income_gender_gap	-.0362771	.0072576	-5.00	0.000	-.0506159	-.0219383
log_demand_for_family	1.179177	.2316849	5.09	0.000	.7214382	1.636915
log_f_mean_age_marriage	-.9253754	.1070278	-8.65	0.000	-1.13683	-.7139212
log_tertiary_ed_f	-.1318934	.020641	-6.39	0.000	-.1726737	-.0911131
log_avg_population_male	1.061082	.0862905	12.30	0.000	.8905984	1.231566
log_birth_rate	.2062438	.031935	6.46	0.000	.1431501	.2693376
_cons	-4.083164	1.715056	-2.38	0.019	-7.471589	-.6947385
sigma_u	.1153685					
sigma_e	.01922563					
rho	.97297969	(fraction of variance due to u_i)				

F test that all u_i=0: F(23, 152) = 96.50   Prob > F = 0.0000

RE

R-squared:
Within = 0.8680
Between = 0.9927
Overall = 0.9922

Obs per group:
min = 2
avg = 7.6
max = 10

Wald chi2(6) = 4984.12
corr(u_i, X) = 0 (assumed)   Prob > chi2 = 0.0000

log_tot_f_married_in_~e	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
log_income_gender_gap	-.0352153	.0071931	-4.90	0.000	-.0493136	-.0211171
log_demand_for_family	.8190265	.2041732	4.01	0.000	.4188544	1.219199
log_f_mean_age_marriage	-.8669252	.1069877	-8.10	0.000	-1.076617	-.6572331
log_tertiary_ed_f	-.1333931	.0203754	-6.55	0.000	-.1733282	-.093458
log_avg_population_male	1.029935	.0160126	64.32	0.000	.9985512	1.061319
log_birth_rate	.2178384	.0315343	6.91	0.000	.1560324	.2796445
_cons	-2.250691	1.01903	-2.21	0.027	-4.247953	-.2534301
sigma_u	.08429791					
sigma_e	.01922563					
rho	.95055692	(fraction of variance due to u_i)				

In this case we use the log-log relationship: notice that the coefficients in this case are directly interpreted as percentage values

An increment in the difference of median income between genders of 100 euros, causes a decrement of 3.5% in the married females.

New FE vs New RE

We newly pursue the Hausman test → the FE model is preferred

```
. hausman FE_ RE_

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      FE_          RE_         Difference       Std. err.
-------------+----------------------------------------------------------------
log_income~p |   -.0362771    -.0352153       -.0010617        .0009653
log_demand~y |    1.179177     .8190265        .3601501        .1095043
log_f_mea~ge |   -.9253754    -.8669252       -.0584503        .0029298
log_tertia~f |   -.1318934    -.1333931        .0014996        .0033004
log_avg_po~e |    1.061082     1.029935        .0311468        .0847918
log_birth_~e |    .2062438     .2178384       -.0115946        .0050431
------------------------------------------------------------------------------
                          b = Consistent under H0 and Ha; obtained from xtreg.
           B = Inconsistent under Ha, efficient under H0; obtained from xtreg.

Test of H0: Difference in coefficients not systematic

    chi2(6) = (b-B)'[(V_b-V_B)^(-1)](b-B)
            =  26.24
Prob > chi2 = 0.0002
(V_b-V_B is not positive definite)
```

Notice that the RE assumption of zero-correlation doesn't really hold (the FE shows some correlation).

Conclusions

- Increase in the median/mean female income and female employment rate cause a decrease in the number of marriages,
- while the male counterpart causes an increment.

- Obviously, the effect of age on marriage is negative, so it is plausible to think that countries that support young people poorly see fewer marriages.

- It is not generally true how access to senior female positions in business and government actually affect the number of marriages

- The level of tertiary education also seems to have almost no statistical significance.

- The gender gap in employment rate has no significant relevance in most of the models reported, while as the gender gap in income increases, there is a reduction in the number of marriages.

→ **Further improvements:** pursue a joint analysis with the number of new divorces per year.

References

1) Puiu, Andreea-Ionela. (2019). THE ECONOMETRIC ANALYSIS BETWEEN DIVORCE PHENOMENON AND ECONOMIC-SOCIAL VARIABLES IN ROMANIA. 48. 104-119.
2) Prox L, Farine D. A framework for conceptualizing dimensions of social organization in mammals. Ecol Evol. 2020; 10: 791–807. https://doi.org/10.1002/ece3.5936.
3) T. Clutton-Brock, Science373, eabc9699(2021). DOI: 10.1126/science.abc9699
4) Penis size: an evolutionary prespective- Carole Jahme - The Guardian
5) NADLER, R., & BARTLETT, E. (1997). Penile Erection: A Reflection of Sexual Arousal and Arousability in Male Chimpanzees. In Physiology & Behavior (Vol. 61, Issue 3, pp. 425–432). Elsevier BV.
6) Review Lecture: Mammalian mating systems. (1989). In Proceedings of the Royal Society of London. B. Biological Sciences (Vol. 236, Issue 1285, pp. 339–372). The Royal Society.
7) Brouwer, L., & Griffith, S. C. (2019). Extra-pair paternity in birds. In Molecular Ecology (Vol. 28, Issue 22, pp. 4864–4882). Wiley.
8) Díaz-Muñoz, S. L., & Bales, K. L. (2015). “Monogamy” in Primates: Variability, Trends, and Synthesis: Introduction to special issue on Primate Monogamy. In American Journal of Primatology (Vol. 78, Issue 3, pp. 283–287). Wiley.
9) Petrie, M., & Kempenaers, B. (1998). Extra-pair paternity in birds: explaining variation between species and populations. In Trends in Ecology & Evolution (Vol. 13, Issue 2, pp. 52–58). Elsevier BV
10) Petersdorf, M., & Higham, J. P. (2018). Mating systems (primates). In The International Encyclopedia of Biological Anthropology (pp. 1–5). Wiley.
11) Hohmann, G. (2015). Bonobos. In The International Encyclopedia of Human Sexuality (pp. 1–5). Wiley.
12) Harcourt, A. H., Purvis, A., & Liles, L. (1995). Sperm Competition: Mating System, Not Breeding Season, Affects Testes Size of Primates. In Functional Ecology (Vol. 9, Issue 3, p. 468). JSTOR.
13) why did humans evolve big penises but small testicles? - The Conversation
14) Penis size and Morphology - CARTA
15) Brandon, M. (2016). Monogamy and Nonmonogamy: Evolutionary Considerations and Treatment Challenges. In Sexual Medicine Reviews (Vol. 4, Issue 4, pp. 343–352). Oxford University Press (OUP).
16) Stoller, M. (2002). In International Journal of Primatology (Vol. 23, Issue 3, pp. 697–699). Springer Science and Business Media LLC

Thank you

Nicola Grieco	2081607
Enrico Grimaldi	1884443
Mario Edoardo Pandolfo	1835189
Lorenzo Pannacci	1948926