

Data Appendix

Purushottam Mohanty¹

1 Bartik Indicator

The Bartik indicator essentially captures the differences in wages across men and women in each year, country, region and age-group. In order to estimate it, first a measure of weighted average wage by gender, year, country, region, age-group and industry needs to be computed:

$$\bar{w}_{gycaj} = \sum_j \gamma_{gcra_j} w_{-rgycaj}$$

where, g indexes gender, y year, c country, r region, a age-group and j industry.

γ_{gcra_j} is the proportion of female (or male) workers of country c , region r , age-group a working in industry j in 2002 (or 2005 for those countries which uses LFS data). $w_{-rgycaj}$ is the annual gender specific wage in EUR, in age-group a , industry j , country c and year y , excluding the region r .

1.1 Data Input

- Structure of Earnings Survey (SES) - 2002, 2006, 2010 and 2014
- Labour Force Survey (LFS) - 2010, 2014
- EU-SILC (ilc_di01) - Distribution of income by quantiles - EU-SILC and ECHP surveys.
https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ilc_di01&lang=en
- Exchange Rates - European Central Bank: Statistical Data Warehouse
<http://sdw.ecb.europa.eu/browse.do?node=9691296>

1.2 Data

The Bartik indicator was calculated using SES and LFS data. The SES data was preferred because it provides point estimates of annual wage (in EUR) and also because it provides an indicator for public and private ownership. For the SES data, only those observations where the individual was employed in a privately-owned firm was used for the estimation of the Bartik indicator. However, due to lack of any such distinction in LFS, the entire set of observations was used for those countries whose data was calculated using LFS. The list of countries and the data used for each country in each year is listed in table 1.

The data was used in the following manner due to serious constraints regarding the availability of data. Essentially, data was required for the years - 2004, 2008, 2010 and 2016. However, the SES follows a 4 year cycle implying that data was available only for the years 2002, 2006, 2010 and 2014. On the other hand, the LFS data is available for every year but the data on the wage variable is only available post 2009. Hence, it was decided to use a 2 year lagged measure of the Bartik indicator. Hence, SES 2002 was to be used

¹Paris School of Economics (purushottam.mohanty[at]gmail.com)

for 2004, SES 2006 for 2008, SES 2010 for 2010 (since there is no data for 2008 in SES or LFS) and SES 2014 was to be used for 2016. However, another issue is that SES is only available for a subset of the 23 countries (mentioned in table 1) that the data was needed for. Hence, for those countries whose data was not available in the SES dataset, the LFS data was used. Also, for a few countries, SES did not provide any regional variation, i.e. the data was at the country level, so LFS was also used for those countries.

1.2.1 LFS

The EU Labour Force Survey was used only for the years 2010 and 2014. It was used for the following countries not included in the EU SES in 2010 and 2014.

- 2010: Austria (AT), Switzerland (CH), Denmark (DK), Ireland (IE), Iceland (IS) and Slovenia (SI)
- 2014: Austria (AT), Switzerland (CH), Denmark (DK), Ireland (IE), Iceland (IS), Greece (GR) and Slovenia (SI)

The LFS was also used for the following countries for 2010 and 2014, due to lack of regional variation in the SES data.

- 2010: Czech Republic (CZ), Denmark (DK), Portugal (PT) and Slovakia (SK)
- 2014: Czech Republic (CZ), Denmark (DK), Portugal (PT) and Slovakia (SK)

1.3 Methodology

As mentioned earlier, for countries whose indicator was constructed using SES data, only those individuals who worked in a privately-owned firm, were used for the estimation of the indicator. Also, for all countries only employed individuals and those with data on industry of occupation (NACE classification) were used for estimation. Further, for the construction of the Bartik indicator, the age groups used are 20-39 years and 50-69 years.

The gamma variable which varies at the country, region, age, sex and industry level was constructed using the SES 2002 and LFS 2005 for respective countries as mentioned in table 1.

1.3.1 Income Deciles

The SES data provides annual wages while LFS only contains wages in deciles (1-10) which differs across countries and years. However, the major cause of concern was the lack of any documentation on the absolute cut-off values for each decile for each year and country. Hence, a reasonable assumption was made that the deciles are the same as those mentioned in the EU SILC data which is publicly available in the EUROSTAT website and can be obtained using the link mentioned in the data input section. Using the data, mid-points for each decile was computed. Additionally, for the mid-point of the 10th decile, the upper cut-off of the 9th decile was used. The methodology used does introduce some degree of measurement error but given the constraints regarding the data, it was perhaps the best that could have been done.

Country	ISO	Gamma	2002	2006	2010	2014
Austria	AT	LFS 2005	NA	NA	LFS 2010	LFS 2014
Belgium	BE	SES 2002	SES 2002	SES 2006	SES 2010	SES 2014
Switzerland	CH	LFS 2005	NA	NA	LFS 2010	LFS 2014
Czech Republic	CZ	LFS 2005	NA	NA	LFS 2010	LFS 2014
Germany	DE	SES 2006	NA	SES 2006	SES 2010	SES 2014
Denmark	DK	LFS 2005	NA	NA	LFS 2010	LFS 2014
Estonia	EE	SES 2002	SES 2002	SES 2006	SES 2010	SES 2014
Spain	ES	SES 2002	SES 2002	SES 2006	SES 2010	SES 2014
Finland	FI	LFS 2005	NA	NA	LFS 2010	LFS 2014
France	FR	SES 2002	SES 2002	SES 2006	SES 2010	SES 2014
Greece	GR	SES 2002	SES 2002	SES 2006	SES 2010	LFS 2014
Hungary	HU	SES 2002	SES 2002	SES 2006	SES 2010	SES 2014
Ireland	IE	LFS 2005	NA	NA	LFS 2010	LFS 2014
Iceland	IS	LFS 2005	NA	NA	LFS 2010	LFS 2014
Italy	IT	SES 2002	SES 2002	SES 2006	SES 2010	SES 2014
Netherlands	NL	SES 2002	SES 2002	SES 2006	SES 2010	SES 2014
Norway	NO	SES 2002	SES 2002	SES 2006	SES 2010	SES 2014
Poland	PL	SES 2002	SES 2002	SES 2006	SES 2010	SES 2014
Portugal	PT	LFS 2005	NA	NA	LFS 2010	LFS 2014
Sweden	SE	SES 2002	SES 2002	SES 2006	SES 2010	SES 2014
Slovenia	SI	LFS 2005	NA	NA	LFS 2010	LFS 2014
Slovakia	SK	LFS 2005	NA	NA	LFS 2010	LFS 2014
United Kingdom	UK	SES 2002	SES 2002	SES 2006	SES 2010	SES 2014

Table 1: Country-wise data input

Country	ISO	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
Austria	AT	5456	12505.5	15316	17599	19861.5	22254.5	24918.5	28292	33978.5	37759
Switzerland	CH	8005	18146.5	21995.5	25369.5	28848.5	32593	36716	42405.5	50964	56029
Czech Republic	CZ	2179.5	4825.5	5593	6174.5	6756.5	7406.5	8232.5	9326.5	11057	12171
Denmark	DK	6983.5	15793	18953	21639	24329.5	26990.5	29784	33108.5	38019.5	41077
Finland	FI	6000	13299	15832	18107.5	20249	22492.5	25029.5	28089	32711.5	35668
Greece	GR	2750	6324	8000.5	9593	11148	12935.5	14865	17212	20967.5	23333
Ireland	IE	5587	12284.5	14434	16594	19113.5	21949.5	24981	29283	36028.5	40066
Iceland	IS	5561	12282	14236	15799	17444	19270	21314.5	24020	28660.5	31690
Portugal	PT	2090.5	4828	5955.5	6968	8089	9360.5	10833.5	12749.5	16425	18975
Slovenia	SI	3271.5	7403	8876	10022	11145.5	12347.5	13685.5	15354.5	17874	19451
Slovakia	SK	1716.5	3901.5	4665	5246	5824.5	6468.5	7251.5	8236	9753	10717

Table 2: Income Deciles (2010)

Country	ISO	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
Austria	AT	6196.5	14097.5	17097.5	17599	19605.5	22014.5	27431.5	31180	37481	41572
Switzerland	CH	10259	23076.5	27831	31944.5	36012	40463.5	45837	52832.5	64144	71532
Czech Republic	CZ	2304	5101.5	5958.5	6625	7275	7980.5	8800	10002	11992.5	13242
Denmark	DK	7841.5	17508.5	20762	23602	26443	29421	32751	36903	43272.5	47271
Finland	FI	6684	14755.5	17377	19822.5	22368	25010.5	27870.5	31420.5	36860	40301
Greece	GR	1500	3678.5	4923.5	5959	7054	8236.5	9470	11159	13841.5	15512
Ireland	IE	5211	11741.5	14236.5	16505	18883.5	21650.5	24941.5	29330.5	35930.5	39951
Iceland	IS	7123.5	15465	17575	19504.5	21514.5	23578.5	25833	28763	33804	37078
Portugal	PT	1747.5	4240	5526	6596.5	7677.5	8817.5	10210.5	12141	15478.5	17690
Slovenia	SI	3078.5	7119.5	8808.5	10120	11307	12545.5	13918	15668.5	18391.5	20100
Slovakia	SK	1842.5	4242.5	5189	5853	6468.5	7178	7977	8953	10533	11567

Table 3: Income Deciles (2014)

1.3.2 Exchange Rates

For some countries, the wage variable consisted of wages in their own currencies which needed to be converted to euros (EUR) for consistency. Although, the wage ratios are robust to currencies used so conversion to euros might not be required in the first place. The exchange rates which are yearly averages, were obtained for the European Central Bank database. The link in the data input section can be used for further reference.

1.4 Regions dis-aggregation

The regional dis-aggregation for the countries is at the level of NUTS 1. However, for some countries data was at country level for certain years while at the NUTS 1 level for the other years. These involve Hungary, Sweden and Poland. Hence, for the those years where data was at country level, data was disaggregated into the NUTS 1 level where each region represented the same values as those at the country level. The benefit of doing so is to avoid losing out on regional variation for the years where data exists at the NUTS 1 level which otherwise would have been the case had the data been aggregated to the country-level instead.

So in the 2002 data, the regions were disaggregated for *Hungary (HU)* and *Sweden (SE)*. HU was disaggregated into HU1, HU2 and HU3 whereas SE was disaggregated into SE1, SE2 and SE3. Similarly, regions for *Poland (PL)* was disaggregated in both 2002 and 2006 into PL1, PL2, PL3, PL4, PL5 and PL6.

1.4.1 NACE Harmonization

To avoid NACE rev 1.1 and NACE rev 2 confusions, a common harmonization index was used which largely followed the NACE rev 1.1 classification. Table 4 shows the NACE index used and its correspondence to NACE 1.1 ² and NACE 2 ³.

The first column mentions the actual ID used as part of the construction of the Bartik indicator with it's corresponding labels in the fourth column.

²EU Commission documentation on NACE Rev 1.1: http://www.instat.gov.al/media/2956/nace_rev11.pdf

³Eurostat documentation for NACE Rev 2: <https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF>

NACE ID	NACE 1.1	NACE 2	Labels
1	A, B	A	Agriculture, hunting, forestry and fishing
2	C	B	Mining and quarrying
3	D	C	Manufacturing
4	E	D, E	Electricity, gas and water supply
5	F	F	Construction
6	G	G	Wholesale and retail trade; repair of motor vehicles motorcycles and personal and household goods
7	H	I	Hotels and restaurants
8	I	H, J	Transportation and storage, Information and communication
9	J	K	Financial and insurance activities
10	K	L, M, N	Real estate, renting and business activities
11	L	O	Public administration and defence; compulsory social security
12	M	P	Education
13	N	Q	Human health and social work activities
14	O	R, S	Other community, social and personal service activities
15	P	T	Activities of households
16	Q	U	Extra-territorial organizations and bodies

Table 4: NACE Harmonization Correspondence

1.5 Output Data Quality

The Bartik indicator was computed for all countries, except *Estonia, Iceland and Norway* because data was only available at the country level for these countries both in the LFS and SES datasets. So their wages exogenous of their own region couldn't be computed and correspondingly the Bartik indicator couldn't be computed. On the other hand for *Slovenia*, regions in LFS 2005 were unable to be matched to the regions in LFS 2010 and 2014.

1.6 Summary Statistics

	(1)	(2)	(3)	(4)	(5)
	2002	2006	2010	2014	Avg.
AT	0.709	0.715	0.721	0.733	0.719
BE	0.822	0.795	0.844	0.878	0.835
CZ	0.745	0.735	0.844	0.874	0.800
DE	0.735	0.748	0.634	0.703	0.705
DK	0.792	0.772	0.875	0.899	0.835
EE	0.731	0.693			0.712
ES	0.677	0.696	0.703	0.696	0.693
FI	0.818	0.767	0.802	0.810	0.799
FR	0.669	0.729	0.770	0.773	0.737
GR	0.701	0.746	0.878		0.775
HU	0.901	0.863	0.908	0.881	0.888
IE	0.833	0.761	0.828	0.834	0.814
IS	0.717				0.717
IT	0.797	0.822	0.674	0.714	0.762
NL	0.549	0.523	0.539	0.537	0.537
NO	0.787	0.816			0.802
PL	0.867	0.828	0.861	0.857	0.853
PT	0.788	0.872	0.810	0.860	0.832
SE	0.736	0.757	0.721	0.799	0.753
SI	0.938	0.927			0.933
SK	0.722	0.726	0.851	0.811	0.778
UK	0.582	0.625	0.602	0.613	0.606
Total	0.733	0.744	0.752	0.764	0.748
CH			0.716	0.709	0.712

Table 5: Female-to-male wage ratios for each country in each year (modified)

2 European Value Survey (EVS)

The main results of the paper have been computed using the region of residence at the time of the survey. The norm (NTN) variable is essentially the average norm of the "potential" grandmothers in the region of residence. However, if the respondent had migrated to a region different from its region at birth then the norm of the respondent suffers from measurement error. However, the error does not cause bias long as migration is random across the level of the norm. This is indeed testable but unfortunately the European Social Survey (ESS) dataset doesn't provide any information of the region of the respondent at birth.

Hence, the European Value Survey (EVS) for the year 2008 was used for the robustness check. The EVS includes a similar question to the one included in the ESS, i.e. Job scarce: men should have more right to a job than women (agree or disagree)? However, the variable only has 3 levels (agree/disagree/neither) rather than the 5 levels (strongly agree/agree/neither/disagree/strongly disagree) in the ESS survey so the variation is indeed lower. It also includes all the other individual level controls that was used in the main regressions involving the FE strategy. However, the main advantage of using the EVS is that it provides the respondents region when they were of age 14 along with the region of the their residence (i.e. region where the interview was undertaken). Although, it's not the region at birth, it's the best that can be done with the available data.

2.1 Regressions

The regression involving the FE strategy of the paper was replicated using the EVS data. All the necessary control variables were constructed except the variable "ever unemployment \geq 5 years" since no information was available on the same in the EVS data.

Essentially two sets of regressions were computed, one using region at the age of 14 and the other using the region at the time of interview. With each table, regressions were computed for women with first child aged 0-5 years and those without any children. Unfortunately due to less observations and low variation of the norm (due to 3 levels than the 5 in the ESS), the point estimates of the norm are insignificant. However, the estimates of women with first child aged 0-5 years old are higher than women with no children and sign of the coefficients are as expected.

2.2 Migration Checks

All computations and figures in this section was calculated only using respondents who had migrated. This refers to all those respondents whose region at the age of 14 was different from the region of residence at the NUTS 2 level. (Note: Regions refers to regions at the NUTS 2 level).

2.2.1 Norm

Due to low variation in the NTN variable in the EVS data, the following analysis was done using the NTN computed from the ESS data. The computed NTN from the ESS was merged with the EVS data for both the region of residence and the region at age 14 for each individual. Thereafter, the mean of the norm of the region of residence was computed for each region at the age of 14. Finally, this mean value of the norm of region of residence was compared with the norm of the region at the age of 14 for each region at the age of 14. This was computed for all young women aged 20 to 40 years. Three separate figures for computed involving all young women, young women with first child aged 0-5 years, and young women with no children.

2.2.2 Employment Rate

The employment rate for computed for each region of residence and each region at age 14 separately. Respondents who are employed full-time, part-time or self-employed are considered employed. The employment

rate for each region is the proportion of employed respondents in each region. Thereafter, the mean of the employment rate was computed for each region of residence was computed for each region at age 14. Finally, the employment rates were compared for region of residence and region at age 14 for each region at age 14; and three figures were plotted involving all young women, young women with first child aged 0-5 years, and young women with no children.

2.2.3 Bartik Indicator

The Bartik wage ratio from the ESS data was merged with the EVS data for each region of residence and each region at age 14 separately. Thereafter, mean of the wage ratio was taken for each region at age 14. Finally, wage ratios were compared for region of residence and region at age 14 for each region at age 14; and three figures were plotted involving all young women, young women with first child aged 0-5 years, and young women with no children.

3 Heterogeneity Analysis

3.1 Motherhood Gap Estimator: Heterogenous Effects

Female-Male Differentials in Labour Market

1. Female to Male Gap in Tertiary Education (*gap_tert*) - Eurostat ([here](#))

It has been calculated by the author from the Eurostat data and refers to the ratio of females aged 24-64 with tertiary education to that of males aged 24-64 with tertiary education at the NUTS-2 region level in the year 2000.

2. Female to Male Gap in Unemployment Rate (*gap_urate*) - Eurostat ([here](#))

It refers to the ratio of unemployment rate of females to that of males for each NUTS-2 region in the year 2000.

3. Bartik Wage Ratio (*wage_ratio*) - Self calculated by author using Structure of Earnings Survey (SES) and Labour Force Survey (LFS). (Refer to section on Bartik Indicator)

Demographic and Economic Factors

1. Share of Immigrants in Region (*migr_share*) - Eurostat ([here](#))

It refers to the share of immigrants (i.e. non-citizens) in the total population of the region at the NUTS-2 level in the year 2000.

2. Median Population Age (*medage_T*) - Eurostat ([here](#))

It refers to the median age for all individuals in a particular region at the NUTS-2 level in the year 2000.

3. GDP Per-Capita (*gdp_pc*) - Calculated by author using GDP ([here](#)) and population ([here](#)) data at the NUTS-2 region level.

It refers to the GDP Per-Capita at the current market prices for each NUTS-2 region.

Historical and Ethnic Factors

1. Plough Use in Agriculture (*plowtype*) - Using [Alesina et al. \(2013\)](#) and Ethnographic Atlas.
It refers to the use of plough by ethnic groups which have been mapped to the NUTS-2 region level. (Refer to section on Ethnographic Atlas for more detail.)
2. Female Participation in Agriculture (*femagr*) - Using [Alesina et al. \(2013\)](#) and Ethnographic Atlas.
It refers to the degree of participation of female in agriculture by ethnic groups which have been correspondingly mapped to the NUTS-2 region level. (Refer to section on Ethnographic Atlas for more detail.)
3. Family Type Indicator (*famtype*) - Using [Alesina et al. \(2013\)](#) and [Gray \(1999\)](#)
It refers to the to the average family type in the region historically which ranges from nuclear families to small extended families.

3.2 Motherhood Gap Estimator: Heterogenous Effects by Cross-Country Institutional Characteristics

Parental Leave

1. Weeks of Maternity Leave - OECDStat ([here](#))
It refers the number of weeks of maternity leave available to mothers. The indicator is at the country level and is used for the year 2000.
2. Job Protected Parental Leave - OECDStat ([here](#))
It refers to the weeks of job protected parental leave available to the mothers, regardless of the income support at the country level in the year 2000.
3. Parental Leave Payment - OECDStat ([here](#))
It refers to the weeks of payment associated with parental leave available to mothers at the country level in the year 2000.

Other Parental Leave Regulations

1. Parental Leave Payment Rates - OECDStat ([here](#))
It refers to the proportion of gross earnings replaced by parental leave benefits across weeks of paid parental and home care leave available to mothers by level of earnings. The indicator is at the country level and available for the year 2014.
2. Public Expenditure on Paternity and Maternity Leaves - OECDStat ([here](#))
It refers to the public expenditure on paternity and maternity leave per live birth at USD 2010 (PPP 2015) at the country level.
3. Net Equalised Income over the period of parental and home care leave - OECDStat ([here](#))
It refers to the net equalised household income one month after the birth of the child as a percent

of income one year before the birth, by level of earnings. The indicator is at the country level and available for the year 2014.

Labour Market Institutions

1. Strictness of Employee Protection - OECDStat ([here](#))

It is synthetic indicator of the strictness of regulation on dismissals and the use of temporary contracts at the country level for the year 2000.

2. Marginal Effective Tax Rate - OECDStat ([here](#))

It refers to the fraction of any additional earnings that is lost to either higher taxes or lower benefits when an employed person increases marginally their working hours at the country level for the year 2001.

3. Net Replacement Rate (NRR) in Unemployment - OECDStat ([here](#))

It refers to the proportion of income that is maintained after 2 months of unemployment including social assistance and housing benefits for each country for the year 2001.

3.3 Family Types and Intimate Partner Violence [Tur-Prats \(2019\)](#)

- *World Values Survey*: 6 waves (publicly available)
- *Survey (Instituto de la Mujer, 1999, 2002, 2006)* Intimate Partner Violence (IPV)
- Instrument: self-constructed

3.4 On the origins of gender roles: Women and the plough. [Alesina et al. \(2013\)](#)

- *Ethnographic Atlas*: traditional female labour force participation in agriculture using information on gender-based division of labour in agriculture.(and, plough use) (and, controls: presence of large domesticated animals, measure of economic development and political complexity.)
- *FAO Global Agro-Ecological Zones (GAEZ) - 2002: (Fisher, et. al 2002)* Reports crop suitability measures for 5 arc-min by 5 arc-min (56km x 56km) grids globally.
- *Standard Cross-Cultural Sample (SCCS): Murdock and White (1969)* For specific tasks within agriculture.
- *15th ed. Ethnologue Languages of the World (Gordon, 2005)* Current geographic distribution of 7,612 different languages.
- *Landscan 2000 database* Estimates of world population in 2000 for a 30 arc-sec by 30 arc-sec (1km x 1km) grid cells globally.
- *Geo-Referencing of Ethnic Groups (GREG) database (Weidmann et. al 2010)* divides world's land into polygons with each polygon indicating the location of specific ethnicity.

4 List of Regional Controls

4.1 Controls

1. Median Age in Region (*medage_T*) - Eurostat ([here](#))

It refers to the median age for all individuals in a particular region at the NUTS-2 level in a given year.

2. Population (*pop*) - Eurostat ([here](#))

It refers to the total population in a particular region at the NUTS-2 level in a given year.

3. Fertility Rate (*fertilityrate*) - Eurostat ([here](#))

It refers to the average number of children that would be born to a woman over the course of her lifetime assuming the current age specific fertility rate to be constant throughout her lifetime for each NUTS-2 region in each year.

4. Number of Births (*births*) - Eurostat ([here](#))

It refers to the total number of live births for each NUTS-2 region in each year.

5. Share of Population with Tertiary Education (*sh_tertiary_T*) - Eurostat ([here](#))

It refers to the ratio of the population with tertiary education aged 24-64 to the total population of those aged 24-64 at the NUTS-2 region level for each year.

6. Female to Male Gap in Tertiary Education (*gap_tert*) - Eurostat ([here](#))

It has been calculated by the author from the Eurostat data and refers to the ratio of females aged 24-64 with tertiary education to that of males aged 24-64 with tertiary education at the NUTS-2 region level for each year.

7. Unemployment Rate (*urate_T*) - Eurostat ([here](#))

It refers to the unemployment rate for each NUTS-2 region in each year.

8. Female to Male Gap in Unemployment Rate (*gap_urate*) - Eurostat ([here](#))

It refers to the ratio of unemployment rate of females to that of males for each NUTS-2 region in each year.

4.2 Ethnographic Controls

The data for the ethnographic indicators (plough type and female participation in agriculture) were obtained from [Alesina et al. \(2013\)](#). Thereafter, it was merged with the ethnographic atlas from [Gray \(1999\)](#) using ethnicity in order to obtain the family type indicator variable in addition to the plough type and female participation indicators. Finally, this was merged with the ESS data from 2010, 2012 and 2014 using the ancestry variable. The remaining observations for which ethnicity wasn't available was merged using the language family variable. The following ESS waves were specifically chosen due to the availability of the ancestry variable in these particular waves. While, language family was available in other ESS waves we

chose not to include them as most European languages were categorized into the "Indo-European" language family, the inclusion of which would have taken away lot of variation in the data. The final indicator for the three ethnographic controls (plough type, female participation in agriculture and family type) is an average of these variables at the NUTS-2 region level.

The variable, plough type refers to historical use of plough in agriculture at the NUTS-2 region level. It ranges from 0 to 1 with higher values signifying higher use of plough where 1 implies all ethnic groups in the region used plough and 0 implying no ethnic group used plough. Female participation in agriculture refers to the historical participation of females in agriculture among ethnic groups in the region with higher values implying higher participation. Family type refers to the average family type in the region historically and ranges between 1 and 7 where 1 implies nuclear families and 7 implies small extended families.

References

- Alesina, A., Giuliano, P., and Nunn, N. (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics*, 128(2):469–530.
- Gray, J. P. (1999). A corrected ethnographic atlas. *World Cultures*, 10(1):24–85.
- Tur-Prats, A. (2019). Family types and intimate partner violence: A historical perspective. *The Review of Economics and Statistics*, 101(5):878–891.