

Supplementary Material for Can Regional Gender Ideologies Account for Variation of Gender Pay Gaps? The Case of Germany

Small Area Estimation

Let \mathbf{D}_p be an $N \times R$ matrix indicating the membership of population unit i belonging to region r and let \mathbf{f}_p denote a vector of population values for gender ideologies of N observations. The population mean of factor scores of gender ideologies \bar{f}_d is predicted for each region based on knowledge about \mathbf{D}_p , on \mathbf{X}_p as an $N \times k$ matrix for k auxiliary variables at the individual level and about \mathbf{Z} as a vector of the region-level predictors. Therefore, each matrix is partitioned into observed values and non-sampled values, the latter marked by the subscript j :

$$\mathbf{f}_p = \begin{bmatrix} \mathbf{f} \\ \mathbf{f}_j \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X}_j \end{bmatrix} \boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \begin{bmatrix} \mathbf{D} \\ \mathbf{D}_j \end{bmatrix} \mathbf{v} + \begin{bmatrix} \mathbf{e} \\ \mathbf{e}_j \end{bmatrix},$$

where the error matrix \mathbf{e} and random intercepts for region-averages \mathbf{v} are independent, with $\mathbf{e} \sim N(0, \sigma^2)$ and $\mathbf{v} \sim N(0, \sigma^2 \boldsymbol{\lambda})$. $\boldsymbol{\lambda}$ represents the ratio of within-region variance and total variance. Thereby, synthetic gender ideologies \mathbf{f}_j for non-sampled units will be estimated by the auxiliary information and their associations $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ as well as by the region's random intercept v_d (based on the ratio of global and its region-level variance λ_r). The lower the number of sampled units within a region n_r in comparison to its actual population size N_d , the more weight is given to the synthetic gender ideologies, imputing also reliable values for potentially non-sampled regions. Hierarchical Bayes estimates for the gender ideology index were obtained using uninformative priors on $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, σ^2 and $\boldsymbol{\lambda}$ (coefficients are shown in Table S1). Finally, a vector of region-level adjusted means of gender ideologies is predictors to reflect the region-specific gender ideologies.

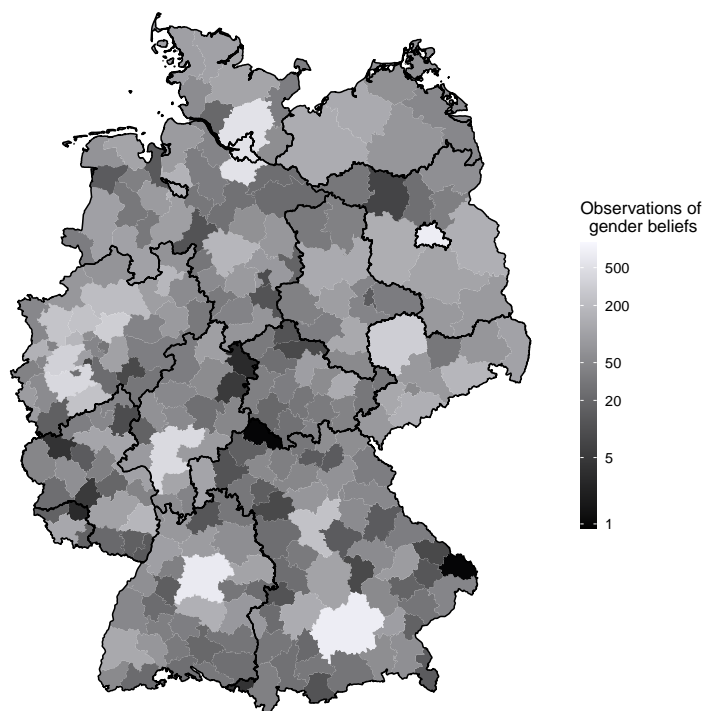
Table S1: Small Area Estimation Coefficients and Modelfit

	Coefficient
Gender (1 = female)	−0.267** (0.027)
Age (Ref.: 35-44):	
18-24	−0.055 (0.056)
25-34	−0.027 (0.039)
45-54	0.131** (0.032)
55-64	0.220** (0.039)
65 and above	0.263** (0.040)
Urbanity (Ref.: Densd labor market region):	
Rural labor market region	0.043 (0.053)
Metropolitan labor market region	0.041 (0.047)
College Degree? (1 = yes)	−0.494** (0.023)
Female Employment? (1 = yes)	−0.202** (0.031)
Migrationbackground? (1 = yes)	0.425** (0.026)
Religious denomination (Ref.: None or others)	
Catholic	0.147** (0.027)
Ev. Protestant	0.072** (0.026)
Subpart of Germany (1 = East)	−0.939** (0.048)
Intercept	3.030** (0.056)
Within variance ($\hat{\sigma}_e^2$)	1.797** (0.019)
Variance ratio (λ)	0.025
R^2 (individual level)	0.176
AIC	62315
Nr. of individuals	18244
Nr. of regions	257

Source: SOEP v35 and auxiliary information by Destatis, own estimation

Note: Levels of significance: * $p < 0.05$, ** $p < 0.01$. Standard errors shown in parentheses.

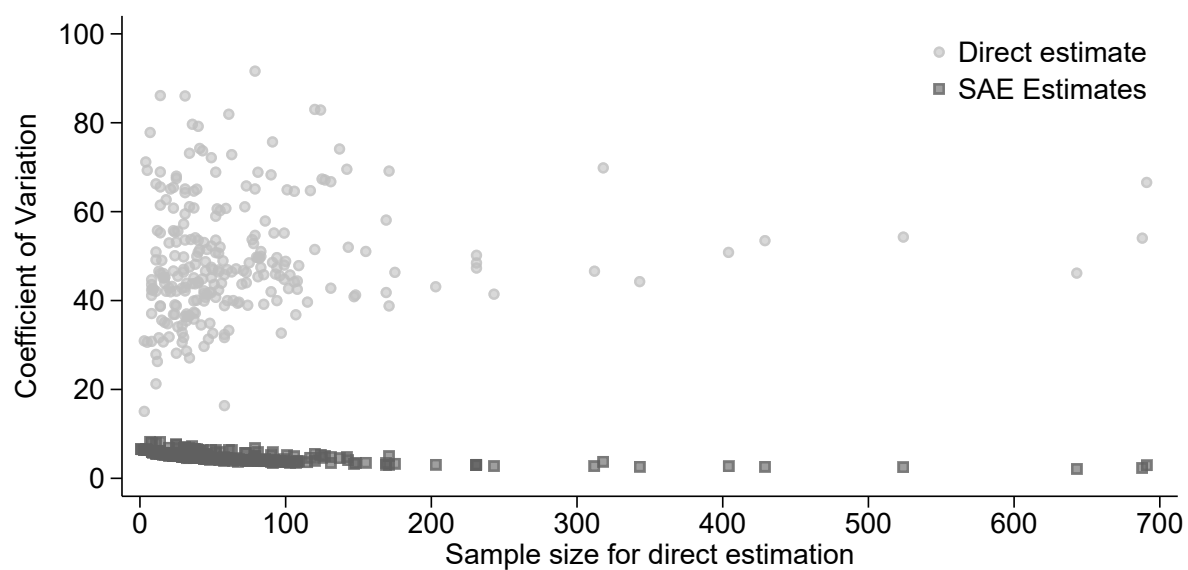
Figure S1: Number of Observations of Individual-Level Gender Beliefs per Labor Market Region



Source: SOEPv35; own estimation

Note: Number of observations lies between 1 and more than 500 in agglomerated regions (Stuttgart in the south-west, Munich in south-east, Berlin, and Hamburg)

Figure S2: Coefficient of Variation of Aggregated Gender Ideology Index and Small Area Estimates by Region's Sample Size



SOEP v35, own estimation

Table S2: Multi-level Models of Log Hourly Gross Wages by Gender, Gender Ideologies and two Mediating Variables

	(1) Individual level index	(2) Regional level childcare
Gender (1 = female) (β)	-0.102** (0.010)	-0.114 (0.073)
Gender ideology	0.005 (0.010)	-0.006 (0.011)
Gender \times gender ideology	-0.014* (0.007)	-0.023* (0.010)
Individual level gender beliefs index	-0.001 (0.005)	
Gender \times individual level index	-0.032** (0.007)	
Share of children < 6 in childcare		-0.245+ (0.139)
Gender \times share in childcare		0.028 (0.119)
sd(Gender) (σ_γ)	0.000 (0.000)	0.048** (0.006)
Socio-demogr. controls ^c	Yes	Yes
Human capital controls ^c	Yes	Yes
Segregation ^c	Yes	Yes
Sectorial controls ^c	Yes	Yes
AIC	5378.31	39793.46
Counties	255	256
Individuals	9013	17515
Pearson-Years	9013	51365

Source: SOEP v35, own estimation

Note: Levels of significance: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$. Standard errors shown in parentheses. Random intercepts not shown.

^a Controls are equivalent to Model 8 in Table 2.

Table S3: Robustness Checks of Multi-Level Models of Log Hourly Gross Wages by Gender and Gender Ideologies

					Within subparts	
	(1) Index w/o SAE ^a	(2) $n_c \geq 30$	(3) Mundlak approach	(4) Spatial dep. ^b	(5) West	(6) East
Gender (1 = female) (β)	-0.096** (0.007)	-0.098** (0.007)	-0.076** (0.009)	-0.096** (0.008)	-0.093** (0.013)	-0.078 (0.049)
Gender ideology	0.006 (0.007)	-0.001 (0.010)	0.000 (0.011)	-0.006 (0.010)	0.001 (0.013)	-0.014 (0.019)
Gender \times gender ideology	-0.025** (0.006)	-0.025** (0.006)	-0.021** (0.007)	-0.018* (0.008)	-0.028* (0.012)	0.008 (0.025)
Spatial lag of gender ideology				0.028* (0.013)		
Gender \times spatial lag of gender ideology				-0.014 (0.012)		
sd(Gender) (σ_γ)	0.049** (0.006)	0.047** (0.006)	0.045** (0.007)	0.049** (0.006)	0.046** (0.008)	0.053** (0.012)
Socio-demogr. controls ^c	Yes	Yes	Yes	Yes	Yes	Yes
Human capital controls ^c	Yes	Yes	Yes	Yes	Yes	Yes
Segregation ^c	Yes	Yes	Yes	Yes	Yes	Yes
Sectorial controls ^c	Yes	Yes	Yes	Yes	Yes	Yes
Individual level means of covariates			Yes			
Cross-level interaction with urbanity					Yes	Yes
AIC	39795.26	39649.98	19643.75	39791.43	32475.00	6945.74
Counties	256	242	256	256	203	53
Person-Years	51365	51121	51365	51365	41820	9545

Source: SOEP v35, own estimation

Note: Levels of significance: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$. Standard errors shown in parentheses. Random intercepts not shown.

^a Individual-level means of indicators on gender attitudes and aggregated within regions without SAE adjustment.

^b Builds upon inverse distance matrix of centroids of all regions.

^c Controls are equivalent to Model 8 in Table 2.