EUROSÜSTEEM

THE GENDER WEALTH GAP IN EUROPE: APPLICATION OF MACHINE LEARNING TO PREDICT INDIVIDUAL-LEVEL WEALTH

Merike Kukk[a,b], Jaanika Meriküll[a,c], Tairi Rõõm[a,b]

(a) Bank of Estonia, (b) Tallinn University of Technology, (c) University of Tartu

Pathways to Gender Equality in Central Banks, 18-19 May 2022

The authors acknowledge financial support from the project InWeGe (Income, Wealth and Gender), financed by European Union’s Rights, Equality and Citizenship Programme (2014-2020), No 820778. Merike Kukk acknowledges financial support from the European Union’s Horizon 2020 Research and Innovation programme under Grant Agreement No 952574. The European Commission and Bank of Estonia do not accept any responsibility for use that may be made of the information the presentation contains.

MOTIVATION

- Little is known about the **gender wealth gap**
  - Abundant evidence on gender wage gap (e.g. Bertrand 2011)
  - **Gender** differences in **wealth function** ↔ link btw gender wage and wealth gap is not one-to-one
  - Lack of individual-level wealth data, few studies on **single country** (Sierminska et al, 2010, 2018; D’Alessio 2018, Bonnet et al 2013, Frémeaux and Leturcq 2020, Meriküll et al. 2021)
  - Comparative evidence only on **single-member households** (Schneebaum et al. 2018, Ravazzini and Chesters 2018)
- The wealth **inequality** has increased in many countries (e.g. Piketty 2013)
  - The inequality of wealth is higher for **individuals** than for **households** (D’Alessio 2018, Frémeaux and Leturcq 2020)
  - The independence of family members and **individualisation of family wealth** has increased in recent decades (Sonnenberg 2008, Borgoyne et al 2007, Frémeaux and Leturcq 2020)
  - → the contribution of within-household inequality to overall inequality has been and will likely continue to be increasing

RELATED LITERATURE – GENDER WEALTH GAP

- International evidence on **raw gender wealth gap at mean** [(men-women)/men]
  - 33% in Germany in 2012 (Sierminska et al. 2018)
  - 31% in Estonia in 2013 (Meriküll et al. 2021)
  - 20% in Italy in 2016 (D'Alessio 2018)
  - 16% in France in 2015 (Frémeaux and Leturcq 2020)
  - around 300% higher in India, 100% higher in Ghana and slightly in favour of women in Ecuador (Doss et al. (2014))
- Regularities behind the gap (Sierminska et al. 2010, 2018, Bonnet et al. 2013, D'Alessio 2018, Meriküll et al. 2021)
  - The gap **enlarges at the top**
  - The raw gap is the largest in **financial and business assets** and among **partner-headed** households

AIM AND CONTRIBUTION

- **The aim** of this study is to estimate the gender wealth gap in 22 European countries
  - We are the first to derive **whole population** based and **comparative** gender wealth gap estimates for a large set of countries
  - Use Household Finance and Consumption Survey (HFCS) data from 2017
  - Apply **machine learning** and **Bayesian model averaging** techniques to predict individual-level wealth of multi-member households from a wealth function of single-member households
- **Findings**
  - The wealth gap is much **larger** in the whole population than in single-member households
  - Men have 24% more wealth than women on average (wage gap is 14%)
  - The wealth gap across countries is related to **overall earnings gap** and participation in **stocks** while higher **home-ownership** rate is related to lower wealth gap
  - **Individual-level wealth inequality** is on average 3.2 pp **higher** than the household-level wealth inequality in multi-member households

WEALTH ACCUMULATION FUNCTION

$$W_t = \sum_{a=1}^{n} (1 + r_a) w_{a,t-1} + S_t + H_t$$

where $w_{\alpha,t-1}$ shows different asset types *a* at the end of period *t-1*, $(1 + r_\alpha)$ returns of asset *a*, S_t denotes savings (Y_t - C_t) and H_t inheritances from period *t*.

- Wealth accumulation function of men and women
  - Men have higher wages and income (e.g. Bertrand 2011)
  - Men have more riskier assets as stocks (e.g. Bajtelsmit and Bernasek (1996), Hinz et al (1997), Embrey and Fox (1997)), make riskier occupational choices and are more likely self-employed (Alesina et. al. (2013))
  - Inconclusive evidence about differences in savings rate (higher for men by Sunden and Surette 1998 lower for men by Agnew 2005)
  - No differences in inheritances (e.g. Edlund and Kopczuk (2009))

DATA

- HFCS data from 2017 and for 22 countries
- Detailed information on household-level assets and liabilities
- Apply survey weights and use data of all 5 implicates
- Use individuals from single-member hhs to predict individual-level wealth of multi-member hhs
- **For each country estimate gendered wealth functions** for **11 quantiles** in single-member households (q5, q10, q20, …q90, q95)
- Apply **recentered influence function** or unconditional quantile regression by Firpo et al (2009)

$$RIF^{g,q} = f(X) + \epsilon \quad if\ i\ \in single - member\ hhs$$

- Estimate the wealth functions with **3 methods**: supervised machine learning methods **elastic net** (lasso + ridge) and **random forest**, **Bayesian model averaging**

	Number of individuals (adults only)	Number of men (adults only)	Number of women (adults only)	Number of households
Total, observations	170163	81501	88658	91235
Total, fraction single	0.206	0.170	0.239	0.367
Total, fraction couples	0.564	0.581	0.548	0.506
Total, fraction other	0.230	0.249	0.213	0.127

EUROSÜSTEEM

METHODOLOGY: ROADMAP

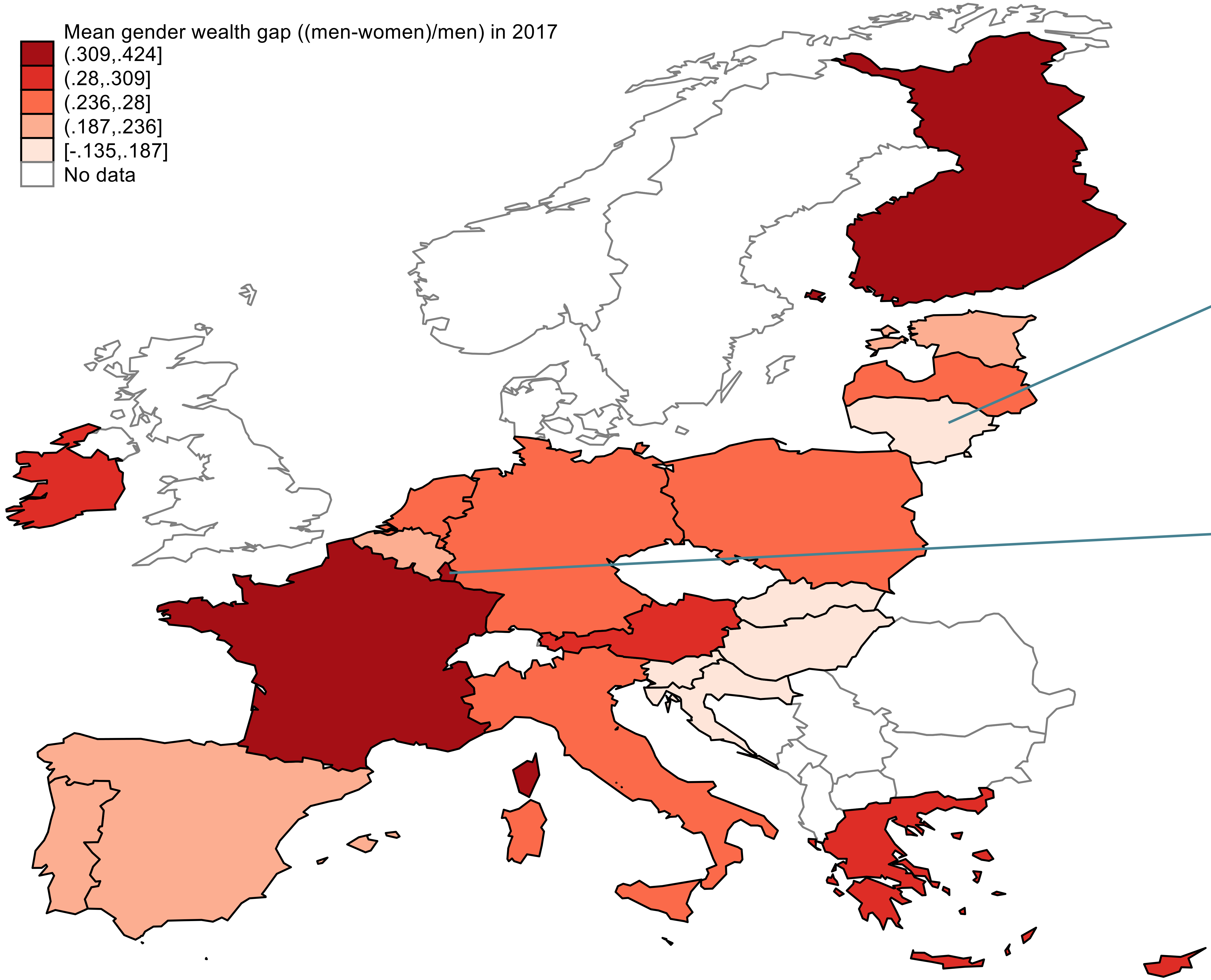
Training sample = single-member hhs

- Obtain wealth function estimates $\widehat{RIF^{g,q}} = \hat{f}(X)$ for each **quantile** and **gender**
- Estimations by **country**

Test sample = multi-member hhs

- Predict wealth around each quantile of **individualised wealth**
- Assess **prediction performance** by RMSE btw survey-collected hhs wealth and hhs sum of prediction
- The best prediction from **random forest**

EESTI PANK
EUROSÜSTEEM



The lowest:
LT insig.
HR 13%
HU 13%

The highest:
LU 42%
MT 38%
FR 35%

EUROSÜSTEEM

GENDER WEALTH GAP: VALIDITY AND DISTRIBUTION

	Mean gap	Mean gap, other studies	Median gap	p90 gap	p95 gap
Austria	0.282**		0.409***	0.309***	0.335***
Belgium	0.188***		0.000	0.181**	0.229***
Croatia	0.125**		0.166*	0.032	0.205*
Cyprus	0.286**		0.230	0.376	0.355*
Estonia	0.236***	0.31	0.254***	0.153*	0.196
Finland	0.348***		0.328***	0.305***	0.389***
France	0.350***	0.16	0.352***	0.333***	0.377***
Germany	0.239***	0.33	0.277***	0.118	0.250***
Greece	0.302***		0.119	0.237**	0.424**
Hungary	0.127***		0.014	0.115**	0.211**
Ireland	0.309***		0.189**	0.264***	0.288*
Italy	0.237***	0.20	0.110*	0.170***	0.285***
Latvia	0.280***		0.141	0.235**	0.304
Lithuania	-0.135		-0.004	0.046	-0.147
Luxembourg	0.424***		0.324***	0.308***	0.519***
Malta	0.377***		0.051	0.341***	0.346***
Netherlands	0.253***		0.390***	0.085	0.107
Poland	0.280***		0.117**	0.111**	0.198***
Portugal	0.220***		0.219***	0.223***	0.337**
Slovakia	0.187**		0.059	0.070	0.189
Slovenia	0.146***		0.226***	0.039	0.060
Spain	0.189***		0.150*	0.150	0.252***
Cross-country average	0.239		0.187	0.191	0.260

EUROSÜSTEEM

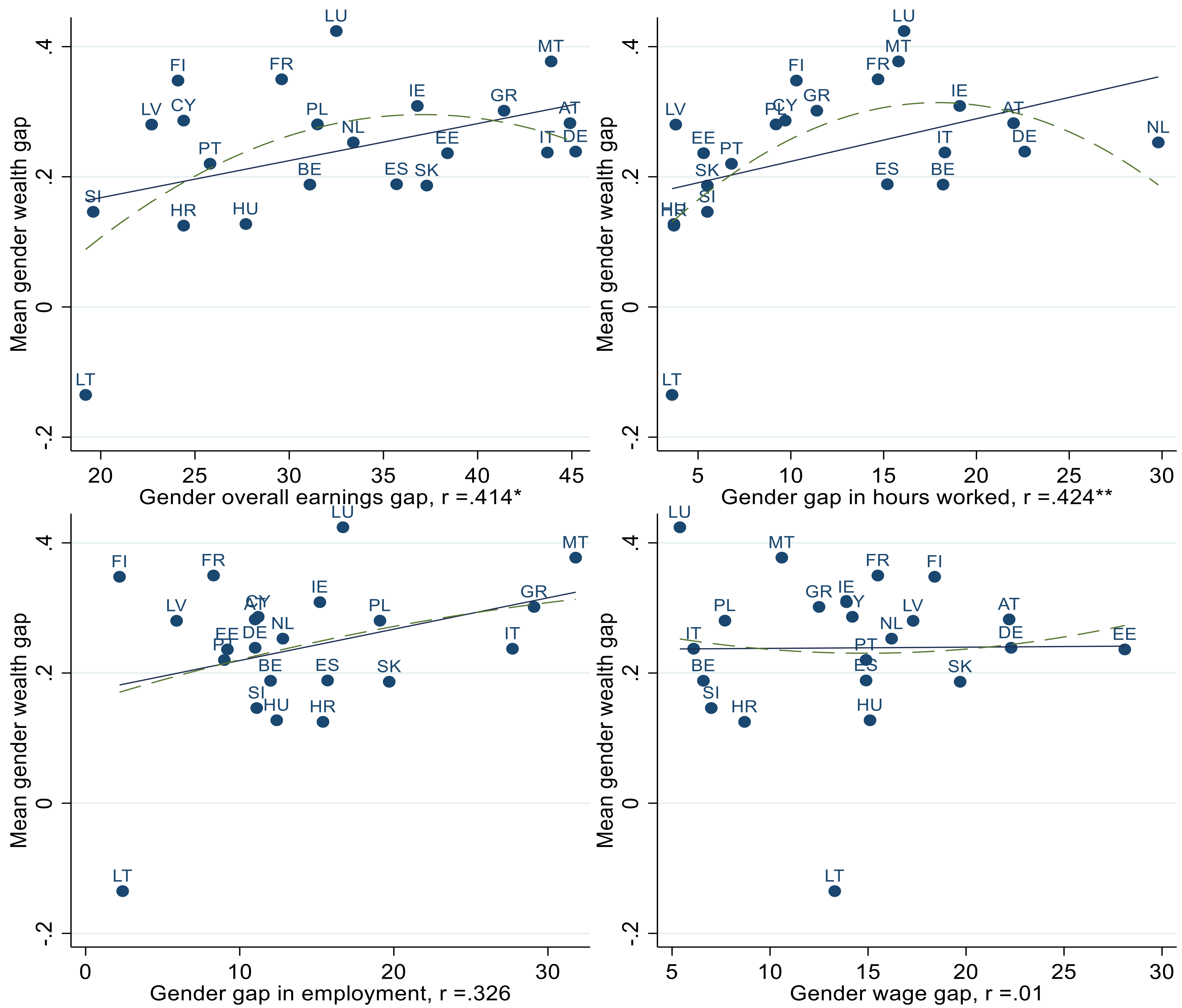
RESULTS: HETEROGENEITY

- Gender wealth gap is **in favour of men**, men have 24% more wealth than women on average
- The gap is **enlarging at the top** (as found in majority of studies)
- The gap is **much higher for multi-member hhs** than for single-member hhs
  - Couples have higher gender wealth gap than singles in 19 out of 22 countries

Wealth gap	Mean	Median	p90	p95
Cross-country average, singles	0.088	-0.159	0.076	0.092
Cross-country average, couples	0.298	0.252	0.237	0.299
Cross-country average, other	0.245	0.151	0.197	0.226
Cross-country average, all	0.239	0.187	0.191	0.260

EUROSÜSTEEM

CROSS-COUNTRY VIEW



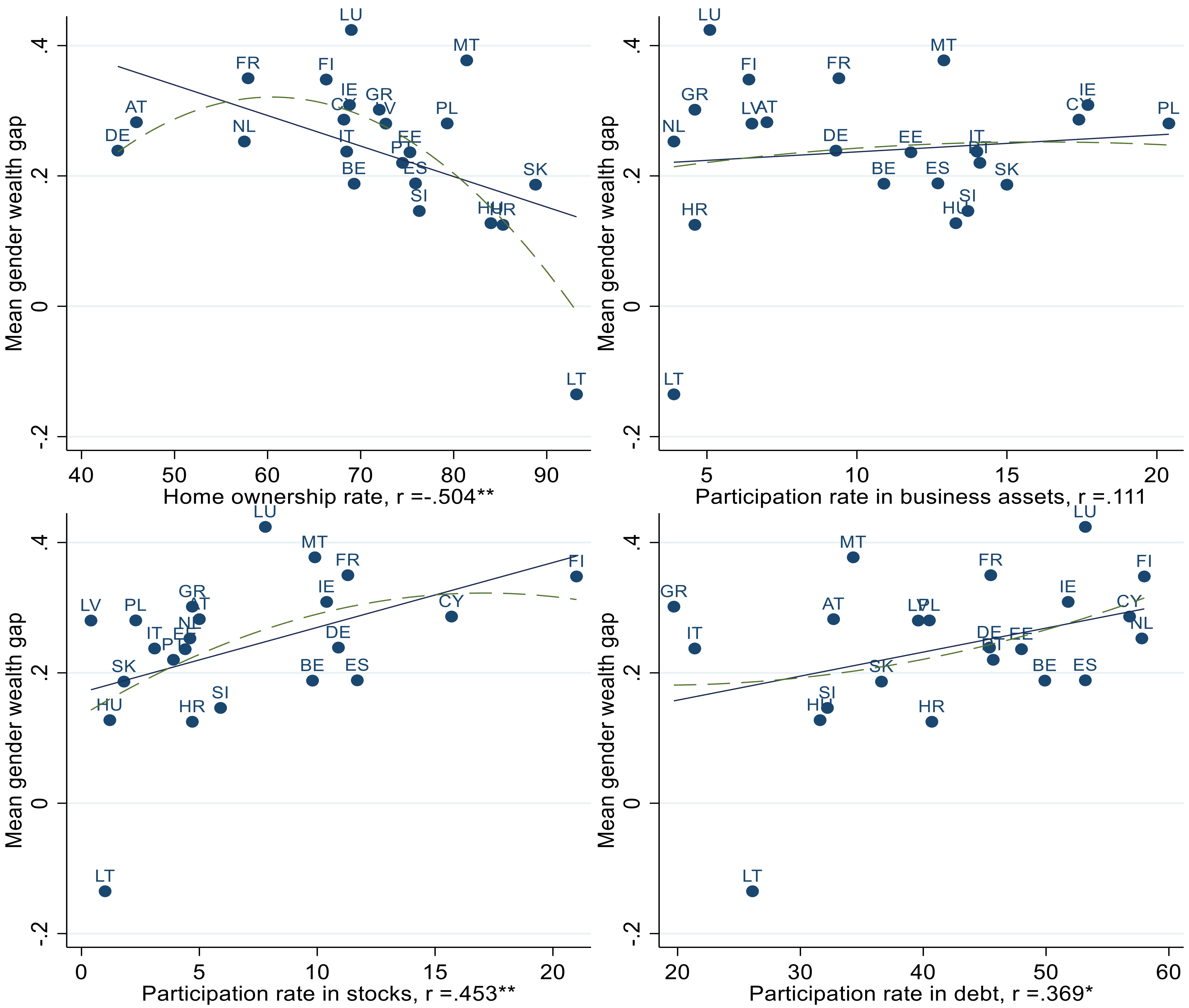
The gender wealth gap is

- **higher** in countries where the gender **overall earnings gap** is larger

Overall earnings gap is composed from gender gaps in three aspects: (1) differences in employment; (2) differences in hours worked; (3) differences in wages

- The **strongest** correlation of wealth gap is with gender gap in **hours worked**

CROSS-COUNTRY VIEW



The gender wealth gap is

- **lower** in countries with high **home-ownership** (the gap is usually the lowest in real estate ☺)
- **higher** in countries with more participation in **riskier financial assets** ☺
- **higher** in countries where more households **hold debt**

EUROSÜSTEEM

WEALTH INEQUALITY: HOUSEHOLDS VS INDIVIDUALS, MULTI-MEMBER HOUSEHOLDS ONLY

	(1) Net wealth Gini, households, Survey-collected	(2) Net wealth Gini, households, sum of prediction	(3) Net wealth Gini, individuals, prediction	(4) Difference: column (3) - column (2)
Austria	0.656	0.644	0.676	0.032
Belgium	0.590	0.614	0.646	0.032
Croatia	0.594	0.546	0.578	0.032
Cyprus	0.730	0.628	0.668	0.039
Estonia	0.677	0.602	0.636	0.034
Finland	0.602	0.603	0.626	0.024
France	0.616	0.631	0.641	0.010
Germany	0.684	0.660	0.684	0.025
Greece	0.566	0.724	0.777	0.053
Hungary	0.641	0.658	0.678	0.021
Ireland	0.648	0.679	0.713	0.034
Italy	0.576	0.618	0.662	0.044
Latvia	0.652	0.658	0.665	0.006
Lithuania	0.560	0.561	0.559	-0.002
Luxembourg	0.632	0.709	0.755	0.046
Malta	0.562	0.672	0.732	0.061
Netherlands	0.723	0.640	0.667	0.027
Poland	0.537	0.611	0.638	0.027
Portugal	0.670	0.659	0.692	0.033
Slovakia	0.522	0.591	0.638	0.047
Slovenia	0.566	0.576	0.629	0.053
Spain	0.661	0.750	0.780	0.030
Cross-country average	0.621	0.638	0.670	0.032

Smallest diff:
LT -0.02pp
LV 0.06pp
FR 1.0pp

Largest diff:
MT 6.1pp
GR 5.3pp
SI 5.3pp

EESTI PANK
EUROSÜSTEEM

SUMMARY

- **Single-member households** are not representative for gender wealth gap estimates
  - the wealth gap is much larger in the whole population than in single-member households
- Gender wealth gap is **large**
  - men have 24% more wealth than women on average (wage gap is 14%)
  - The wealth gap across countries is related to **overall earnings gap** and participation in **stocks** while higher **home-ownership** rate is related to lower wealth gap
- **Individual-level wealth inequality** is higher than household-level wealth inequality
  - 3.2 pp higher on average in multi-member households

EESTI PANK
EUROSÜSTEEM

EESTI PANK
EUROSÜSTEEM

THANK YOU! QUESTIONS, COMMENTS?

MERIKE.KUKK@EESTIPANK.EE

ELASTIC NET

- **Linear prediction** similar to OLS, but in addition to residual sum of squares term, the minimisation function contains also a **penalty term**
- The aim of the penalty term is to shrink the set of variables and to address **overfitting** problem
- **Elastic net** has penalty term as weighted average of two penalty terms:
  - **Lasso** $\min_{\beta}\left\{\sum_{i=1}^{N}\left(Y_i-\beta_0-\sum_{j=1}^{p}X_{ij}\beta_j\right)^2+\lambda\sum_{j=1}^{p}\left|\beta_j\right|\right\}$
  - **Ridge** $\min_{\beta}\left\{\sum_{i=1}^{N}\left(Y_i-\beta_0-\sum_{j=1}^{p}X_{ij}\beta_j\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2\right\}$
- Lasso sets some of the coefficients to zero, Ridge shrinks all the coefficients somewhat
- The penalty term λ and the weight of either of the penalty terms is found by cross-validation (sub-sampling in training sample)
  - Variables need to be standardised first (see the penalty term)

EUROSÜSTEEM

RANDOM FOREST

- **Decision tree** based approach
- For each **predictor Xj** and **cut-point s** seek the covariate X_{j1}^{*} and cut-point s_1^{*} that minimises
$$\sum_{i:X_i \in I_1^{(j,s)}} (Y_i - \bar{Y}_1)^2 + \sum_{i:X_i \in I_2^{(j,s)}} (Y_i - \bar{Y}_2)^2$$
- In the next step seek the covariate X_{j2}^{*} and cut-point s_2^{*} that minimises
$$\sum_{i:X_i \in I_1^{(j,s)}} (Y_i - \bar{Y}_1)^2 + \sum_{i:X_i \in I_2^{(j,s)}} (Y_i - \bar{Y}_2)^2 + \sum_{i:X_i \in I_3^{(j,s)}} (Y_i - \bar{Y}_3)^2$$
- Continue till some stopping rule is reached (leave size 5 obs)
- **Recursive algorithm**, at each step add the split that improves the prediction the most
- **Overfitting** is addressed by estimating **many trees** (100) in **random subsamples** and **averaging the predictions**, many random trees -> random forest
  - For the larger variation in individual trees, each tree selects variables not from the full set of explanatory variables p but from a random subset m, $m \approx \sqrt{p}$.

BAYESIAN MODEL AVERAGING

- Model **selection** and **estimation** in one step
- Allows more **structural intervention** than machine learning tools
  - **Focus regressors** always in the model (income and its squared term, age and its squared term and education groups)
  - **Auxiliary regressors** enter in random combination, 7 variables -> 128 possible models (status groups, employment and its squared term, immigrant, children)
- Estimate coefficients as **weighted average** from all the estimated coefficients
  - Weights depend on the relative performance of the model
- Use these coefficients for prediction

METHODOLOGY: STEP 2

- **Individualise** the survey-collected wealth in multi-member households
    - There are only 2 wealth items for which some information is available at the **individual-level**
        - **Defined contribution pension assets** are collected at the level of individual
        - **Self-employed business assets** are assigned to those household members that are employed in these firms
    - The rest of the wealth items are split equally between adult hhs members
- Assign all members of multi-member hhs to **11 individualised wealth groups** by gender. The wealth groups are defined by quantiles of single-member households
    - As distribution of wealth in single- and multi-member hhs differs, we do not assume that the wealth function applies to the same **quantiles**, but to the same groups defined by **monetary value**

EESTI PANK
EUROSÜSTEEM

METHODOLOGY: STEP 3

- **Predict individual-level wealth of each member in multi-member hhs** from wealth functions of single-member hhs

$$\widehat{RIF^{g,q}} = \hat{f}(X) \quad if\ i\ \in multi - member\ hhs\ \&\ q \in \widetilde{W}^q$$

- Wealth function of q5 is used to predict wealth below q5, wealth function of q10 to predict wealth btw q5 and q15 and so on
- It is assumed that the **wealth functions**, for the same value of wealth, are **the same for single-member hhs** and for individuals from **multi-member hhs**
    - The plausibility of this assumption is tested on Estonian registry data and there are hardly any statistically significant differences in these wealth functions
- The **prediction performance** for multi-member hhs is estimated with root-mean-square error (**RMSE**) between the survey-collected wealth at household-level and the sum of predicted individual wealth in a household

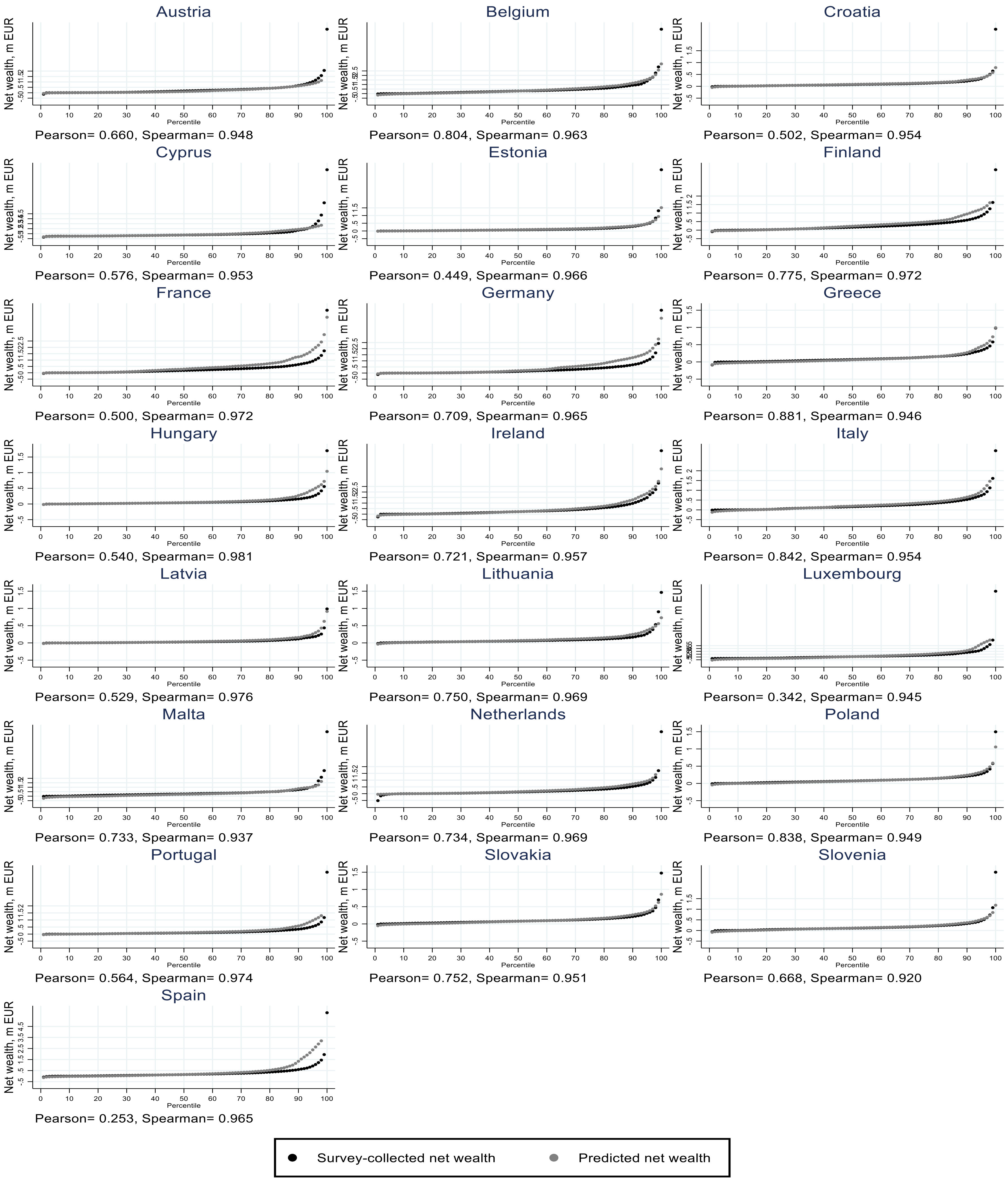
$$RMSE = \sqrt{\frac{\sum_{h=1}^{H}(\hat{y}_h - y_h)^2}{H}} \quad if\ h\ \in multi - member\ hhs$$

EUROSÜSTEEM

RESULTS: OUT-OF-SAMPLE RMSE -> RANDOM FOREST!

	Elastic net	Random forest	Bayesian model averaging
Austria	808.4	747.9	882.5
Belgium	549.7	490.9	625.1
Croatia	294.5	287.0	586.1
Cyprus	1416.3	1346.1	1347.3
Estonia	459.9	430.8	520.1
Finland	672.6	339.7	789.6
France	996.6	964.2	1267.7
Germany	694.9	564.2	884.5
Greece	217.9	82.2	137.9
Hungary	228.3	211.7	255.5
Ireland	3214.8	601.3	2381.8
Italy	673.3	233.9	420.2
Latvia	155.8	144.4	155.7
Lithuania	120.1	118.8	125.1
Luxembourg	5021.5	4850.3	4971.2
Malta	3094.7	568.0	3750.6
Netherlands	947.4	418.6	1301.6
Poland	299.9	163.4	205.9
Portugal	591.7	530.1	584.8
Slovakia	135.2	123.1	143.2
Slovenia	300.3	276.5	401.9
Spain	1749.8	1767.9	1956.2

EUROSÜSTEEM


Austria
Pearson= 0.660, Spearman= 0.948
Belgium
Pearson= 0.804, Spearman= 0.963
Croatia
Pearson= 0.502, Spearman= 0.954
Cyprus
Pearson= 0.576, Spearman= 0.953
Estonia
Pearson= 0.449, Spearman= 0.966
Finland
Pearson= 0.775, Spearman= 0.972
France
Pearson= 0.500, Spearman= 0.972
Germany
Pearson= 0.709, Spearman= 0.965
Greece
Pearson= 0.881, Spearman= 0.946
Hungary
Pearson= 0.540, Spearman= 0.981
Ireland
Pearson= 0.721, Spearman= 0.957
Italy
Pearson= 0.842, Spearman= 0.954
Latvia
Pearson= 0.529, Spearman= 0.976
Lithuania
Pearson= 0.750, Spearman= 0.969
Luxembourg
Pearson= 0.342, Spearman= 0.945
Malta
Pearson= 0.733, Spearman= 0.937
Netherlands
Pearson= 0.734, Spearman= 0.969
Poland
Pearson= 0.838, Spearman= 0.949
Portugal
Pearson= 0.564, Spearman= 0.974
Slovakia
Pearson= 0.752, Spearman= 0.951
Slovenia
Pearson= 0.668, Spearman= 0.920
Spain
Pearson= 0.253, Spearman= 0.965
Net wealth, m EUR
Percentile
Survey-collected net wealth
Predicted net wealth

GENDER WEALTH GAP BEFORE AND AFTER PREDICTION