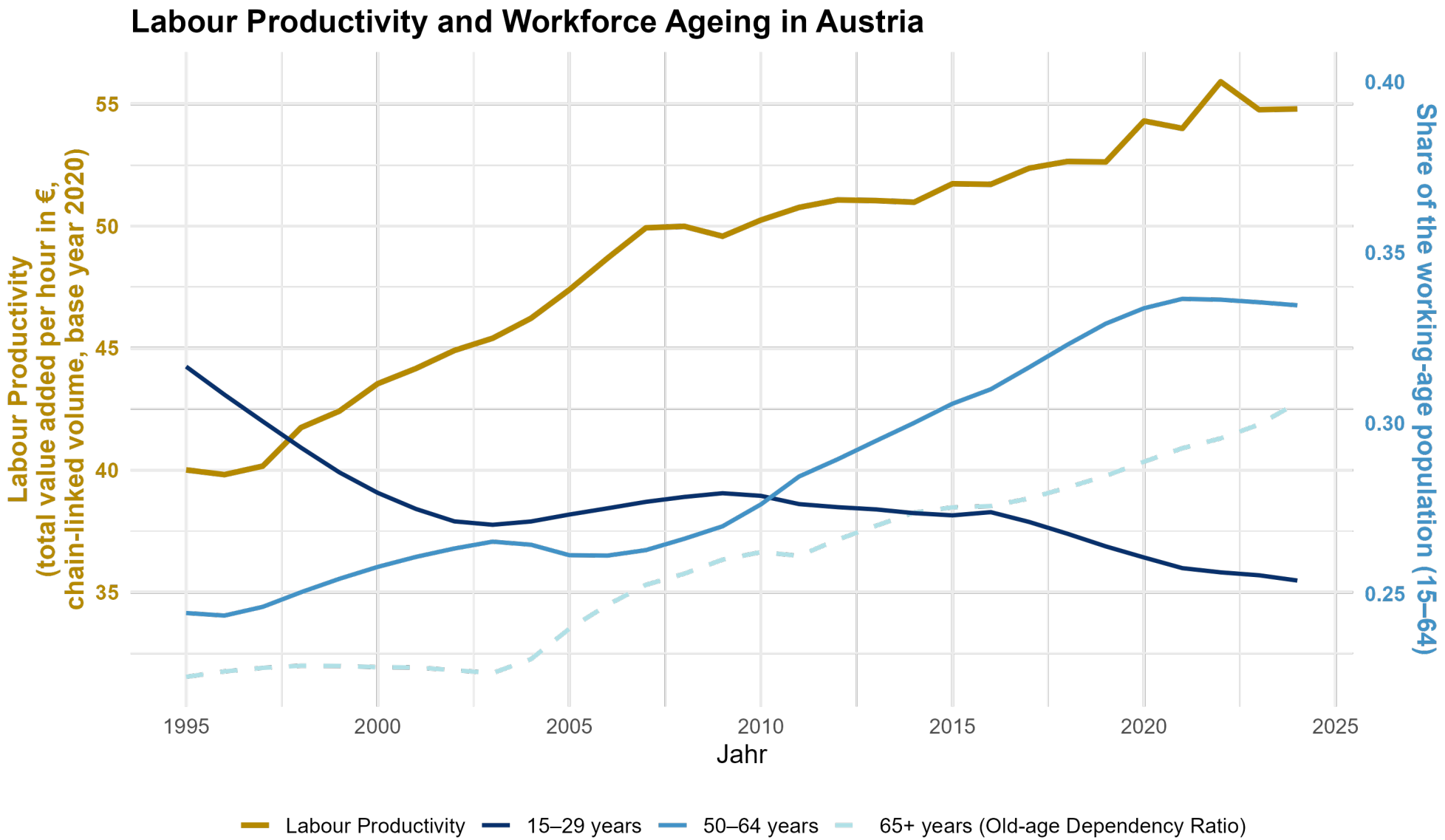
Machine Learning for Economic Demography: Bayesian Approach to Analysing Ageing and Productivity

Isabel Gerstner (1, 2)

1 Vienna Institute of Demography, Austrian Academy of Sciences, 1010 Vienna, Austria
2 Institute of Statistics and Mathematical Methods in Economics, Research Unit Economics, University of Technology, 1040 Vienna, Austria

Make it (Net)Work! - MLA2S Networking Seminar #10
Vienna, 09.06.2026



Source: OECD Data Explorer.

Data – Linked Employer-Employee Data Analysis (LEDA)

Firms

Structural Business Statistics Survey (LSE - RE)

firm size, investments, total value added, personnel costs, legal status, region, OENACE, year of foundation, some inventory, ...

Individuals

Register-based Labour Market Statistics (AEST REGZ)

age, gender, education level and field, nationality, part-time/full-time, occupation status (self-employed/employed/parental leave/etc.), number of children, ...

(1) Individuals are **linked** to firms & (2) staff information is **aggregated to firm level**

Result: firm-level data (2013-2022)
~250.000 firms with ~2 million actively working individuals per year

Full list of firm characteristics, workforce characteristics and NACE sectors

Research Questions

- What is the **relationship between productivity and the age** composition of the workforce at the firm level?

- How does this relationship **vary with firm heterogeneity**, such as firm size or sector?

- What role does the **educational attainment of the workforce** play?

Model & Methodology I

production, where the most important input factors are capital and labour:

$$Y_f = A_f K_f^{\alpha} L_f^{1-\alpha}$$

here we assume that labour is heterogeneous across age groups:

$$L_f = \theta_y L_{fy} + \theta_p L_{fp} + \theta_o L_{fo}$$

By dividing by L_f and assuming $\ln(1+x) \approx x$, we obtain:

$$\ln\left(\frac{Y_f}{L_f}\right) = \alpha \ln\left(\frac{K_f}{L_f}\right) + \beta_y s_{fy} + \beta_p s_{fp} + \beta_o s_{fo} + \ln A_f + \varepsilon_f$$

When further decomposing A_f into observed (firm- & workforce-level) controls and time shocks and drop the share of prime-aged workers as the reference group, we are left with the following for the regression:

$$\ln\left(\frac{Y_f}{L_f}\right) = \alpha \ln\left(\frac{K_f}{L_f}\right) + \boldsymbol{\beta_y} s_{fy} + \boldsymbol{\beta_o} s_{fo} + Z_f^F \gamma + Z_f^W \delta + \omega_t + \varepsilon_f$$

ÖAW VID

Model & Methodology II

$$\ln\left(\frac{Y_f}{L_f}\right) = \alpha \ln\left(\frac{K_f}{L_f}\right) + \sum_{i=1}^{N-1} \frac{L_{fi}}{L_f}\beta_i + Z_f^F\gamma + Z_f^W\delta + \omega_t + \varepsilon_f$$

Kuhn, M., Prskawetz, A., Wrzaczek, S., & Feichtinger, G. (2007). *Health, Survival and Consumption over the Life Cycle: An Optimal Control Approach*. Working paper.

$\frac{Y_f}{L_f}$ *Labour Productivity*

def. as *total value added per worker* (in an extension *per fte*)

$\frac{L_{fi}}{L_f}$ *Age Groups*

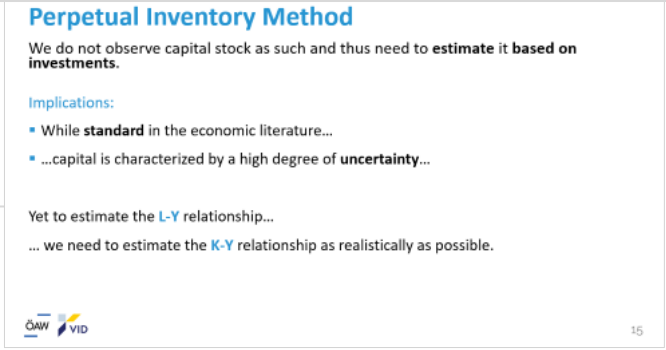
(1) Broad age groups (15-29, 30-49, 50+)
(2) 5-year age groups
(3) Joint shares of age & education

$\frac{K_f}{L_f}$ *Capital per worker*

estimated based on investments and the *perpetual inventory method*

Bayesian time-FE model for each NACE sector to understand between-firm variation in labour productivity

Annex: priors



Shortly about Bayesian Econometrics

$$\ln\left(\frac{Y_f}{L_f}\right) = \alpha \ln\left(\frac{K_f}{L_f}\right) + \sum_{i=1}^{N-1} \frac{L_{fi}}{L_f}\beta_i + Z_f^F\gamma + Z_f^W\delta + \omega_t + \varepsilon_f$$

Central idea of Bayesian updating:

$$p(\beta|y) \propto p(y|\beta)p(\beta)$$

Posterior distribution — Likelihood — Prior distribution

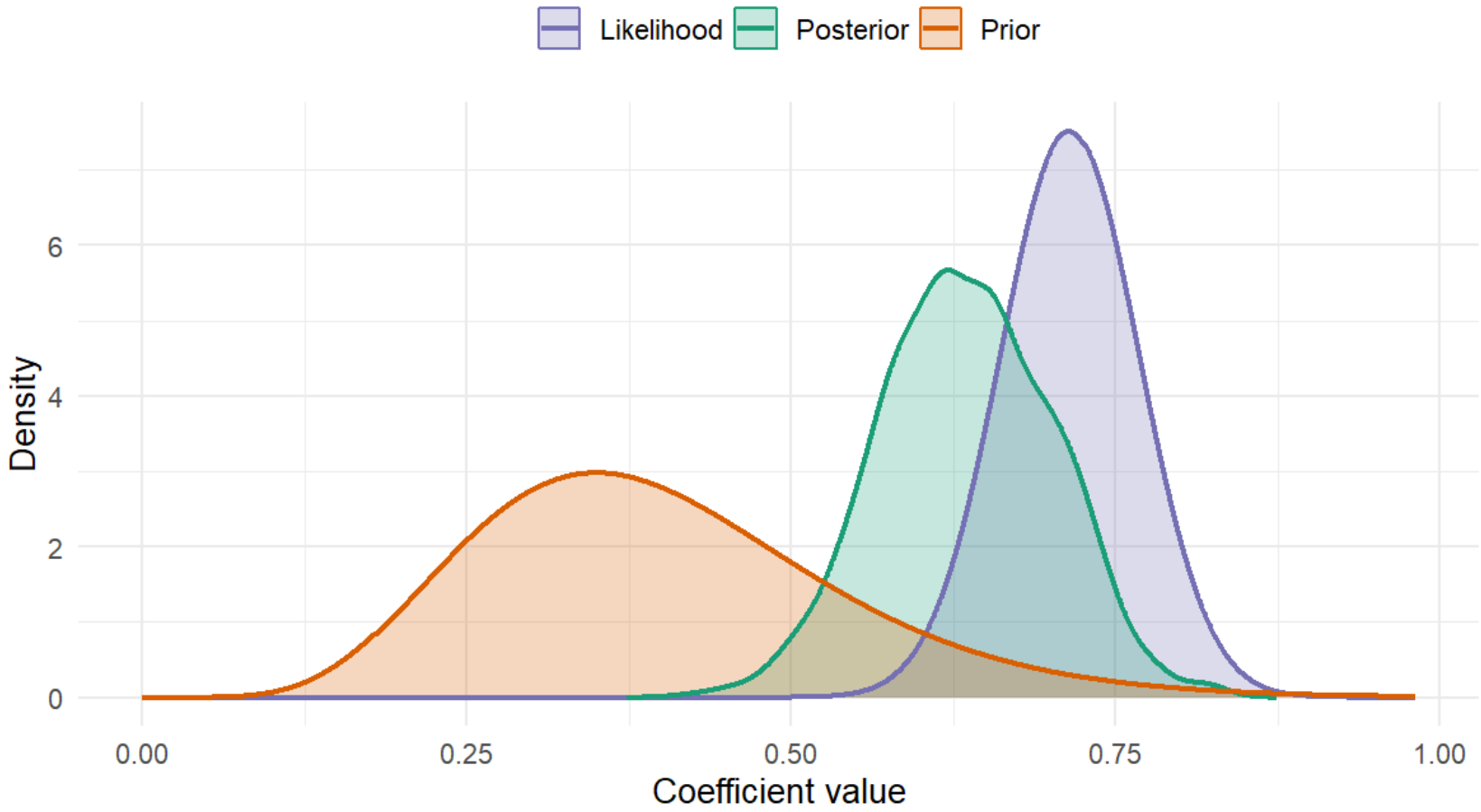
Selection of priors

- Weakly informative priors for the variables of interest (i.e., age)
- Priors for firm characteristics based on economic theory & similar studies

ÖAW VID

Prior on Capital

$$\ln\left(\frac{Y_f}{L_f}\right) = \boldsymbol{\alpha}\, \boldsymbol{ln}\left(\frac{\boldsymbol{K_f}}{\boldsymbol{L_f}}\right) + \sum_{i=1}^{N-1} \frac{L_{fi}}{L_f}\beta_i + Z_f^F\gamma + Z_f^W\delta + \omega_t + \varepsilon_f$$



Results

$$\ln\left(\frac{Y_f}{L_f}\right) = \alpha \ln\left(\frac{K_f}{L_f}\right) + \sum_{i=1}^{N-1} \frac{L_{fi}}{L_f}\beta_i + Z_f^F\gamma + Z_f^W\delta + \omega_t + \varepsilon_f$$

- **hump-shaped** relationship between **age shares** and **productivity**…
- …partly driven by **extremes of the age** distributions…
- …where **low education may strengthen** this negative relationship.

→ These results differ by economic activity

Getting back to the bigger picture:

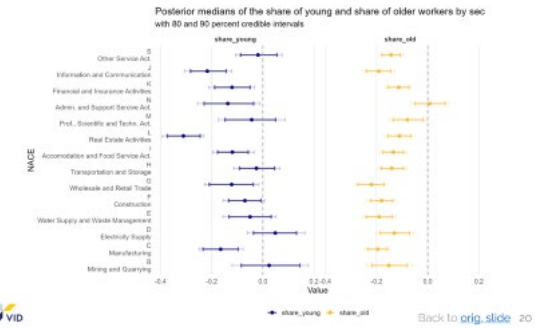
$$\ln\left(\frac{Y_f}{\boldsymbol{L_f}}\right) = \alpha \ln\left(\frac{\boldsymbol{K_f}}{\boldsymbol{L_f}}\right) + \sum_{i=1}^{N-1} \frac{L_{fi}}{\boldsymbol{L_f}}\beta_i + \boldsymbol{Z_f^F}\gamma + Z_f^W\delta + \omega_t + \varepsilon_f$$

- Firm characteristics – strongly related to productivity
- Labour intensity – results differ when using *per head* or *per fte* measure

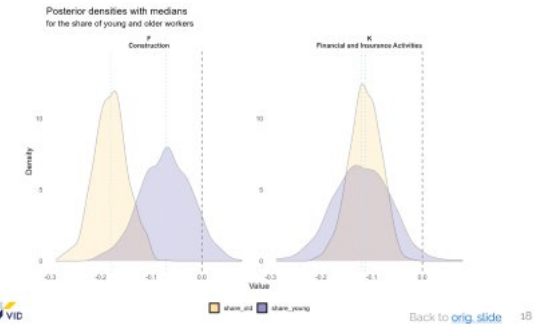
Results

$$\ln\left(\frac{Y_f}{L_f}\right) = \alpha \ln\left(\frac{K_f}{L_f}\right) + \sum_{i=1}^{N-1} \frac{\boldsymbol{L_{fi}}}{\boldsymbol{L_f}} \beta_i + Z_f^F \gamma + Z_f^W \delta + \omega_t + \varepsilon_f$$

Spec. 1 – Broad age groups

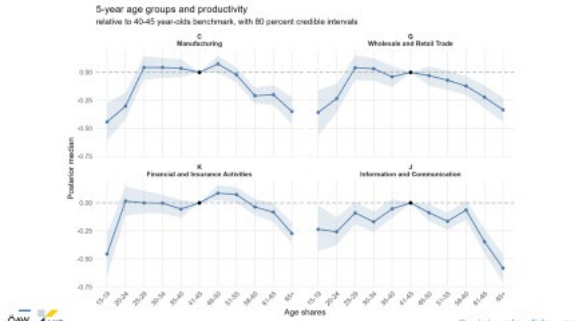


Back to orig. slide 20

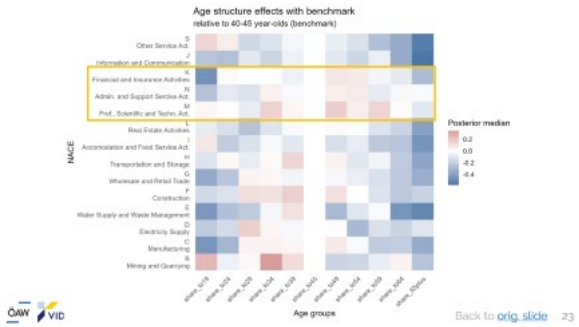


Back to orig. slide 18

Spec. 2 – Narrow age groups

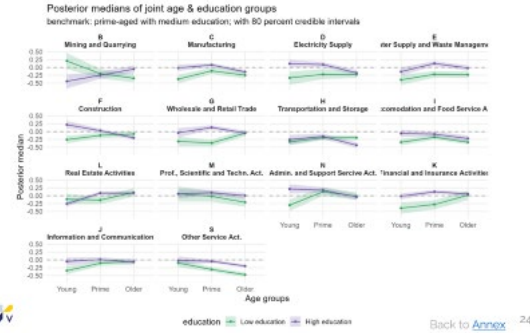


Back to orig. slide 21

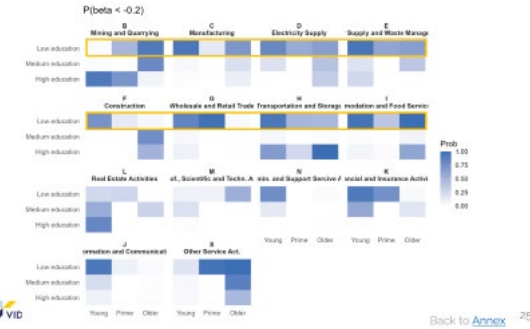


Back to orig. slide 23

Spec. 3 – Age & education



Back to Annex 24



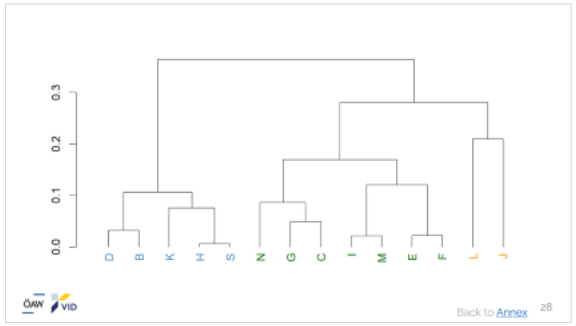
Back to Annex 25

Further possible ML applications

$$\ln\left(\frac{Y_f}{L_f}\right) = \alpha\, ln\left(\frac{K_f}{L_f}\right) + \sum_{i=1}^{N-1} \frac{L_{fi}}{L_f}\beta_i + Z_f^F\gamma + Z_f^W\delta + \omega_t + \varepsilon_f$$

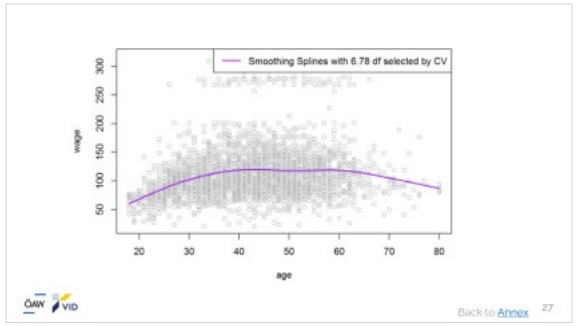
Clustering

- based on selected characteristics such as age
- supports interpretation regarding NACE-sector similarities



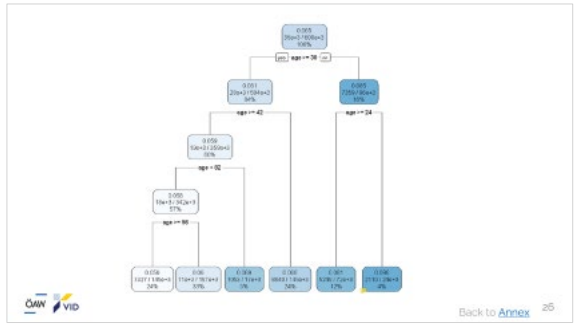
Smoothing splines

- Depart from linearity & one coefficient for each variable
- Relates to idea of benefits of age diversity ↔ disadvantage of age concentration



Random Forests

- Partitioning of feature space to obtain homogeneous subsets for regression
- Coefficient on age shares may depend on other firm characteristics



Conclusion on the Ageing-Productivity Nexus

- **Inverse U-relationship** between Ageing and Productivity
- More decisive: **capital, technology, firm size, age of the firm**

On a more general note

- **Socio-demographic composition** of age groups (such as education or part-time rates) relevant
- Definition of **labour intensity** matters (*per fte* instead of *per head*)
- **Register data:** allows for detailed country-specific analysis and shows **strong heterogeneity in firms** (computational limits, however)

Thank you!

LEDA project

Period: June 2024 – November 2026

Funding Body: Data:Research:Austria funding programme of the Austrian Academy of Sciences (Call 2023).

Contacts: alexia.fuernkranz-prskawetz@oeaw.ac.at & isabel.gerstner@oeaw.ac.at

Website: https://www.oeaw.ac.at/vid/research/research-projects/leda
or scan the QR code:

Annex

Perpetual Inventory Method

We do not observe capital stock as such and thus need to **estimate** it **based on investments**.

Implications:

- While **standard** in the economic literature…
- …capital is characterized by a high degree of **uncertainty**…

Yet to estimate the **L-Y** relationship…

… we need to estimate the **K-Y** relationship as realistically as possible.

Model & Methodology II

$$\ln\left(\frac{Y_f}{L_f}\right) = \alpha \ln\left(\frac{K_f}{L_f}\right) + \sum_{i=1}^{N-1} \frac{L_{fi}}{L_f}\beta_i + Z_f^F\gamma + Z_f^W\delta + \omega_t + \varepsilon_f$$

$\frac{Y_f}{L_f}$ *Labour Productivity*

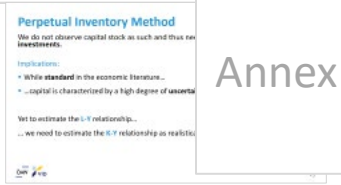
def. as *total value added per worker* (in an extension *per fte*)

$\frac{L_{fi}}{L_f}$ *Age Groups*

(1) Broad age groups (15-29, 30-49, 50+)
(2) 5-year age groups
(3) Joint shares of age & education

$\frac{K_f}{L_f}$ *Capital per worker*

estimated based on investments and the *perpetual inventory method*



Bayesian time-FE model for each NACE sector to understand between-firm variation in labour productivity

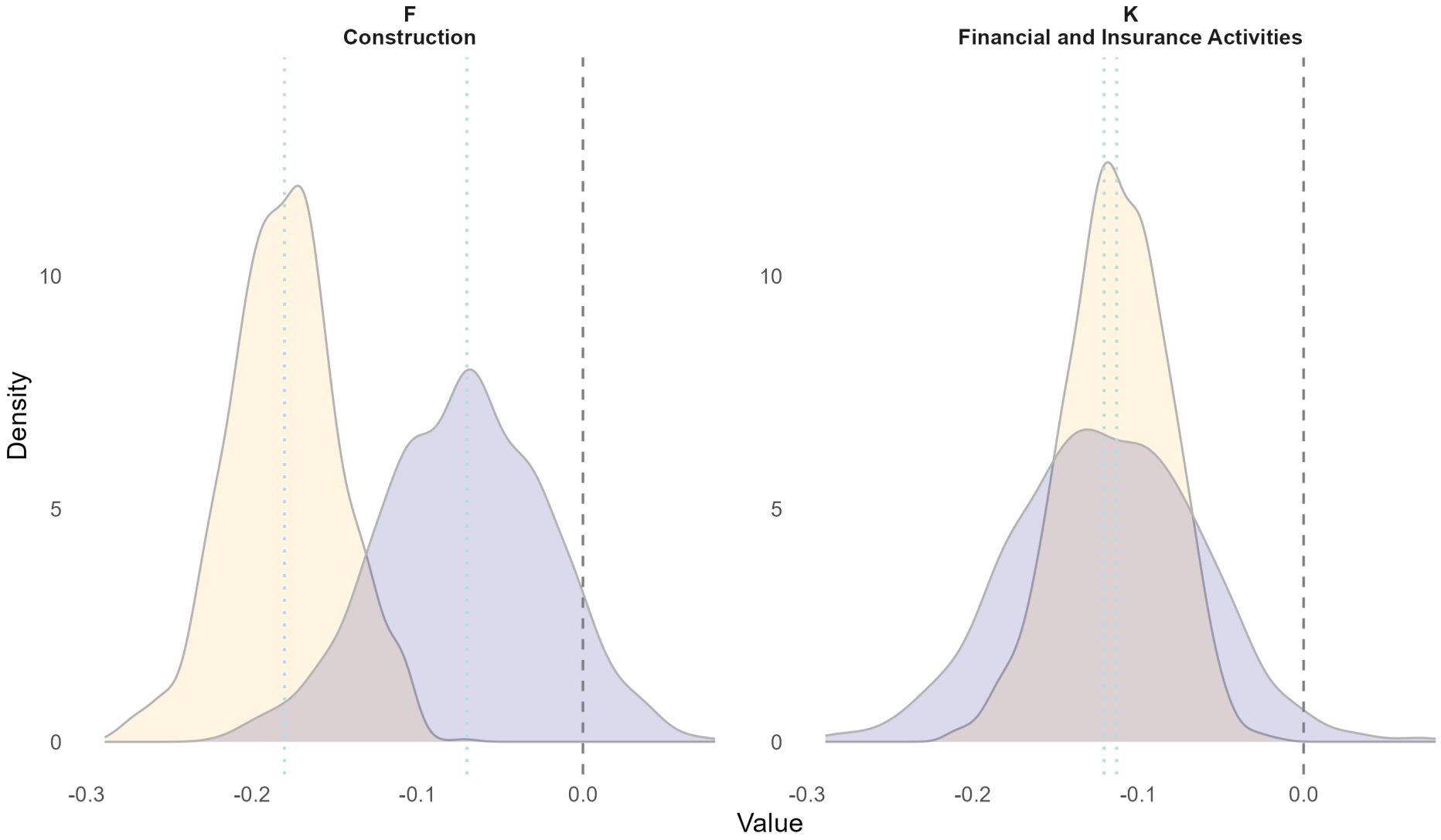
Annex: priors

Kuhn, M., Prskawetz, A., Wrzaczek, S., & Feichtinger, G. (2007). *Health, Survival and Consumption over the Life Cycle: An Optimal Control Approach*. Working paper.

Back to orig. slide

Posterior densities with medians

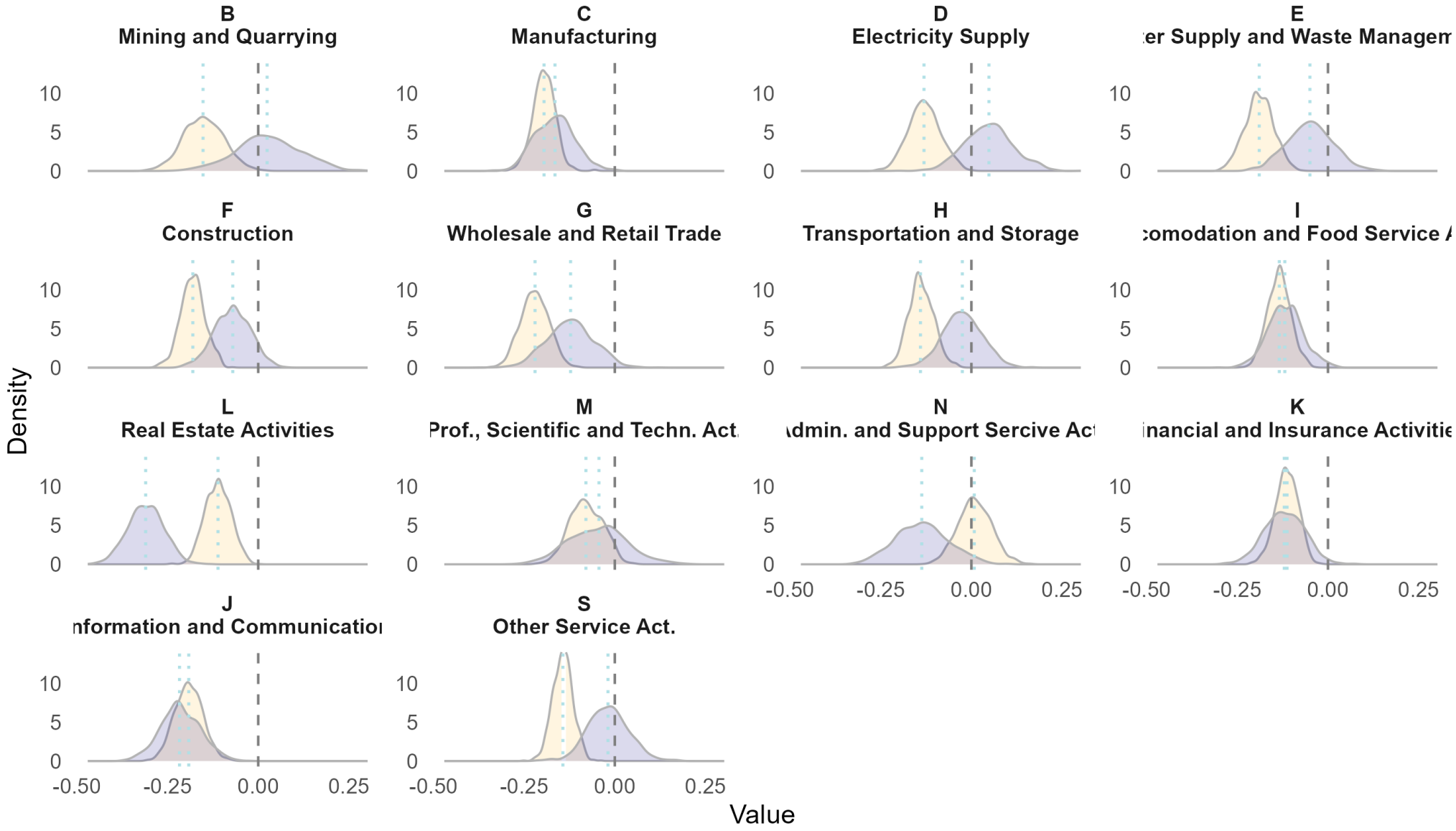
for the share of young and older workers



Back to orig. slide

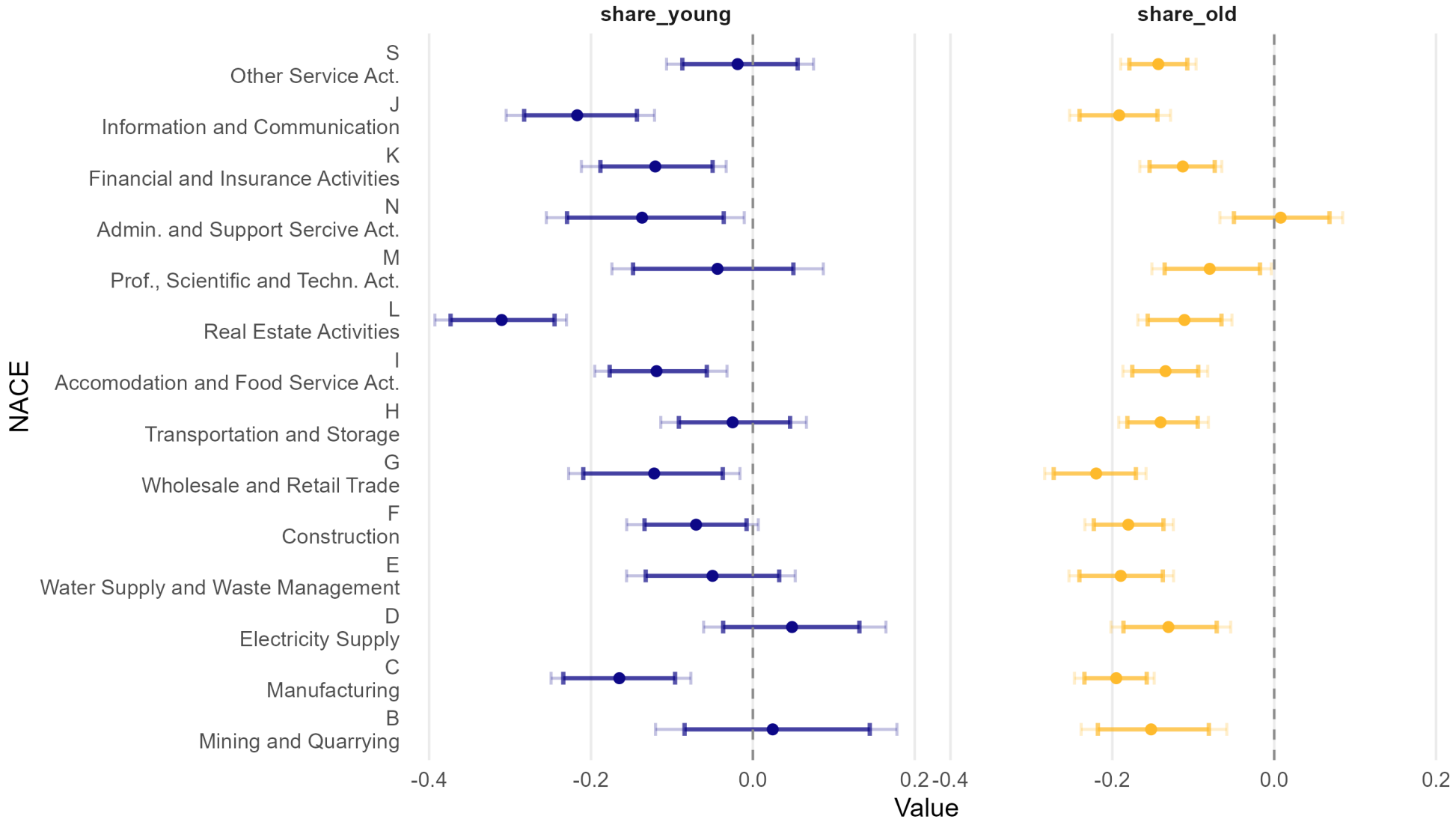
Posterior densities with medians

for the share of young and older workers



Posterior medians of the share of young and share of older workers by sec

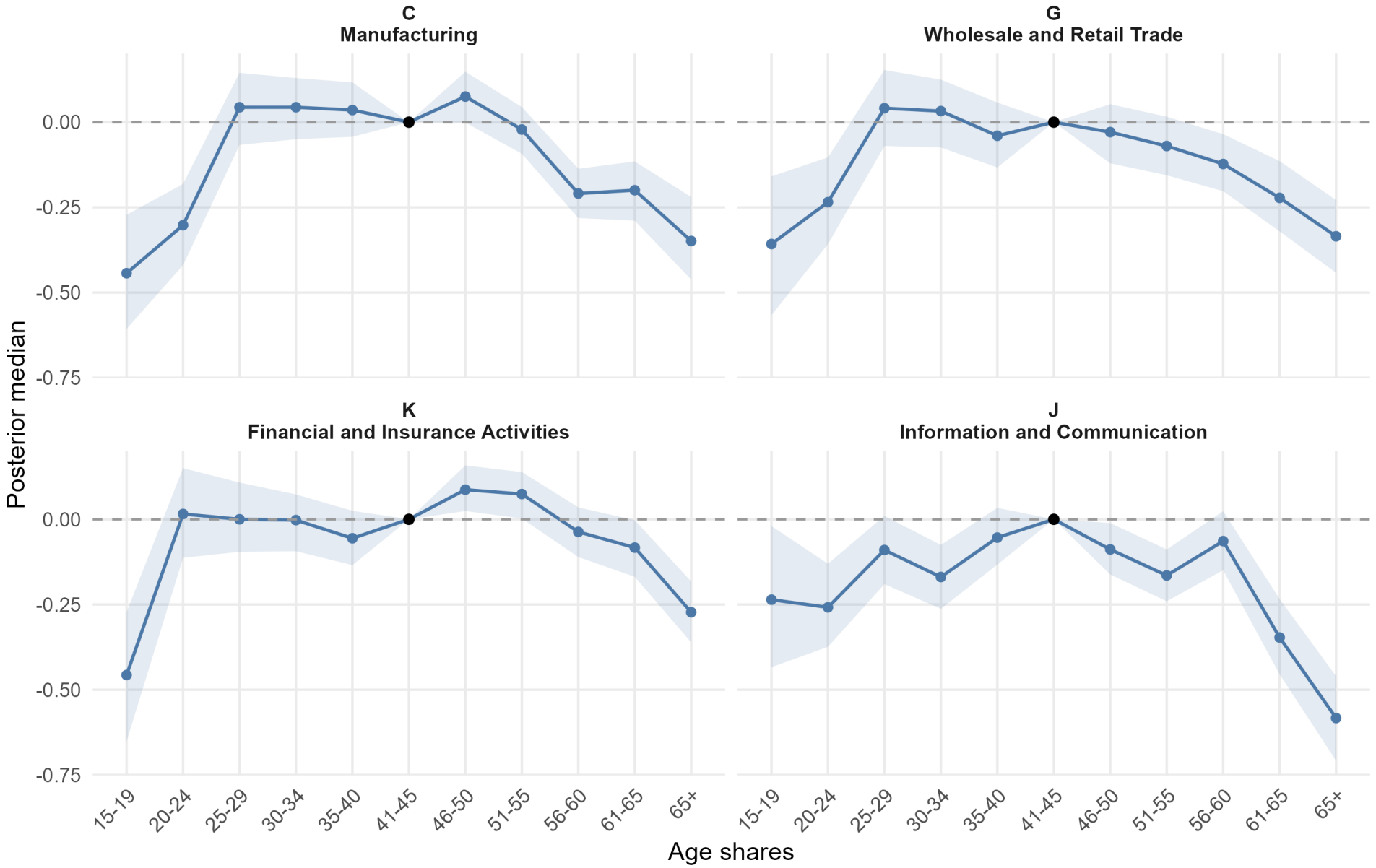
with 80 and 90 percent credible intervals



Back to orig. slide

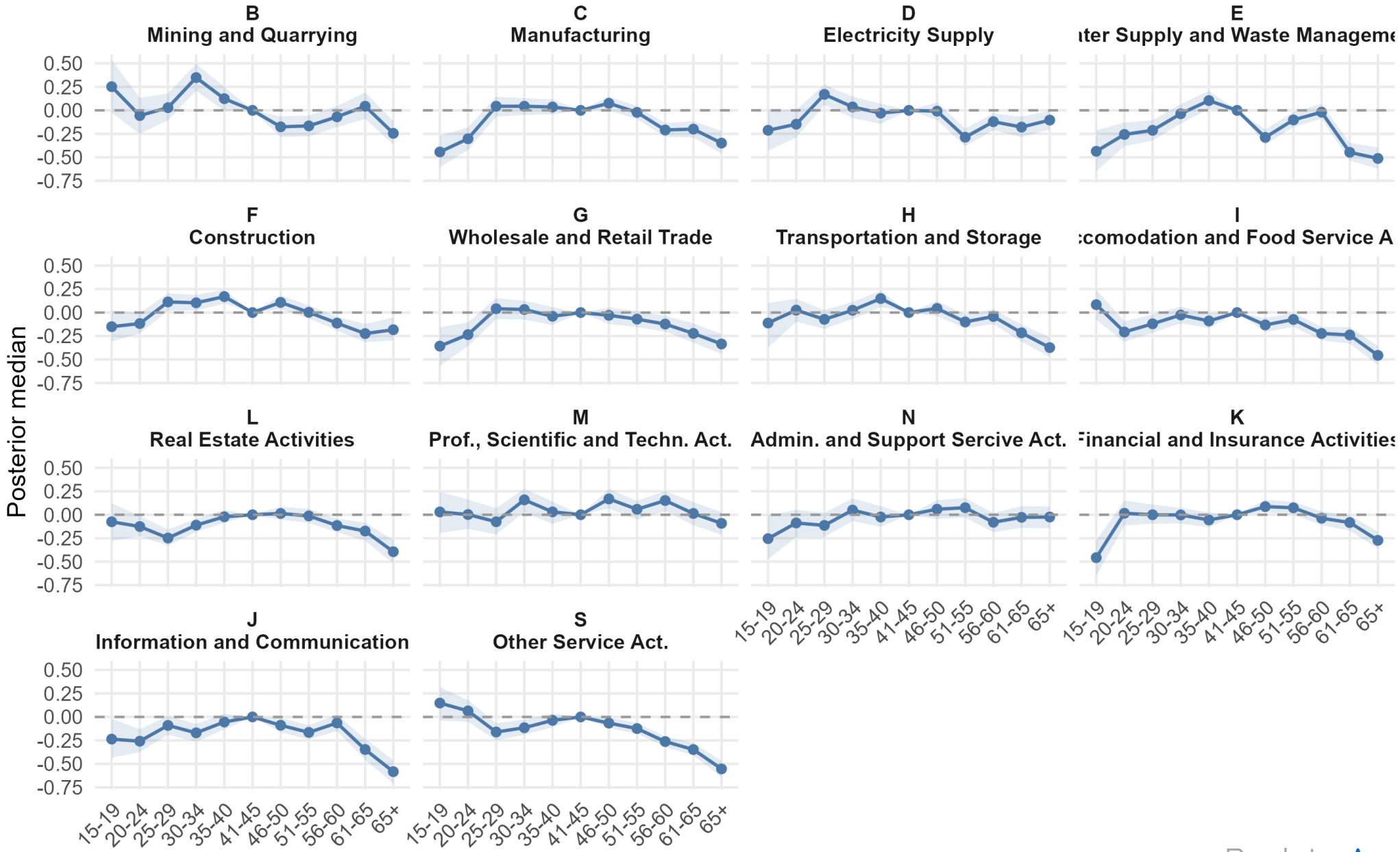
5-year age groups and productivity

relative to 40-45 year-olds benchmark, with 80 percent credible intervals



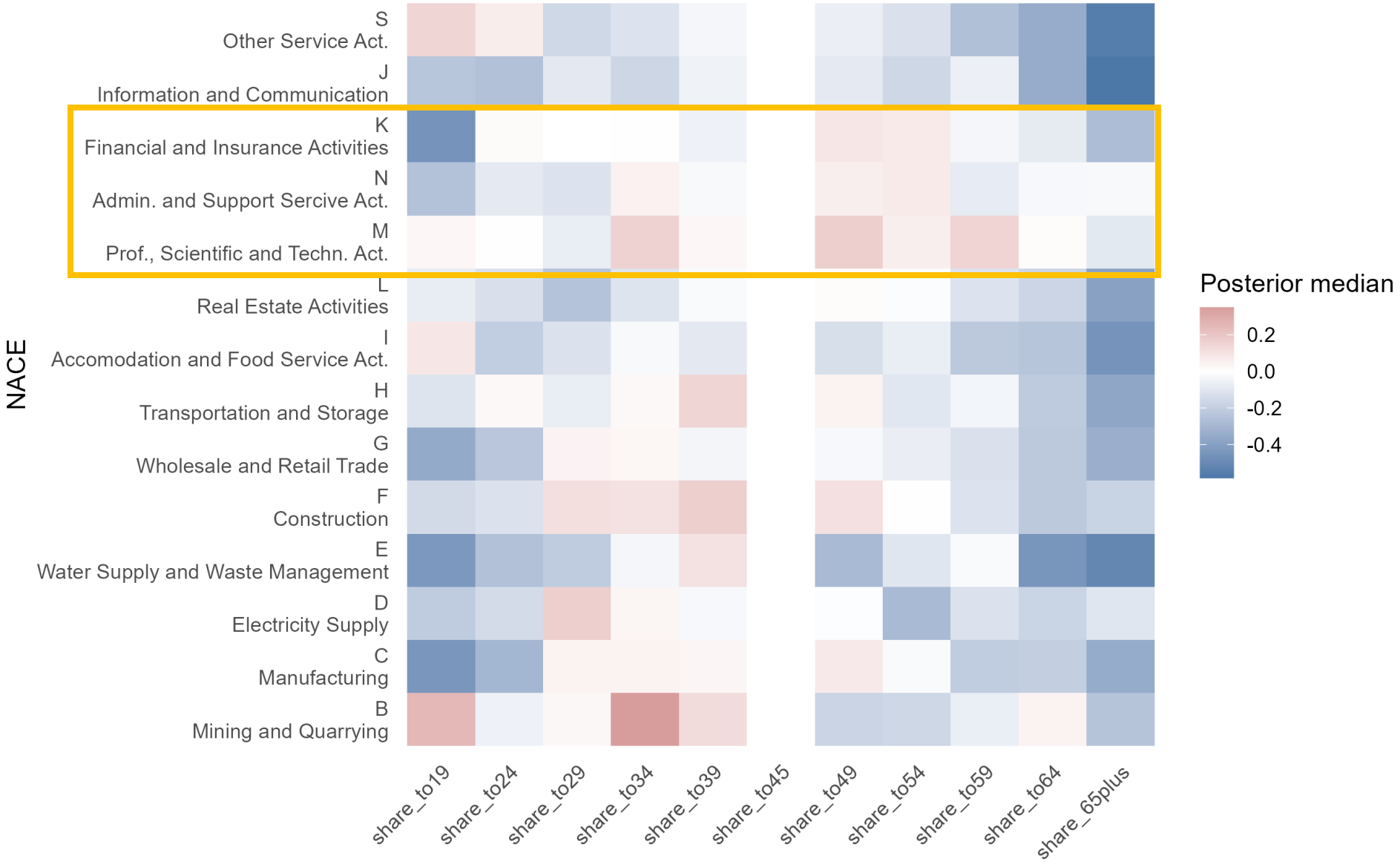
5-year age groups and productivity

relative to 40-45 year-olds benchmark, with 80 percent credible intervals



Age structure effects with benchmark

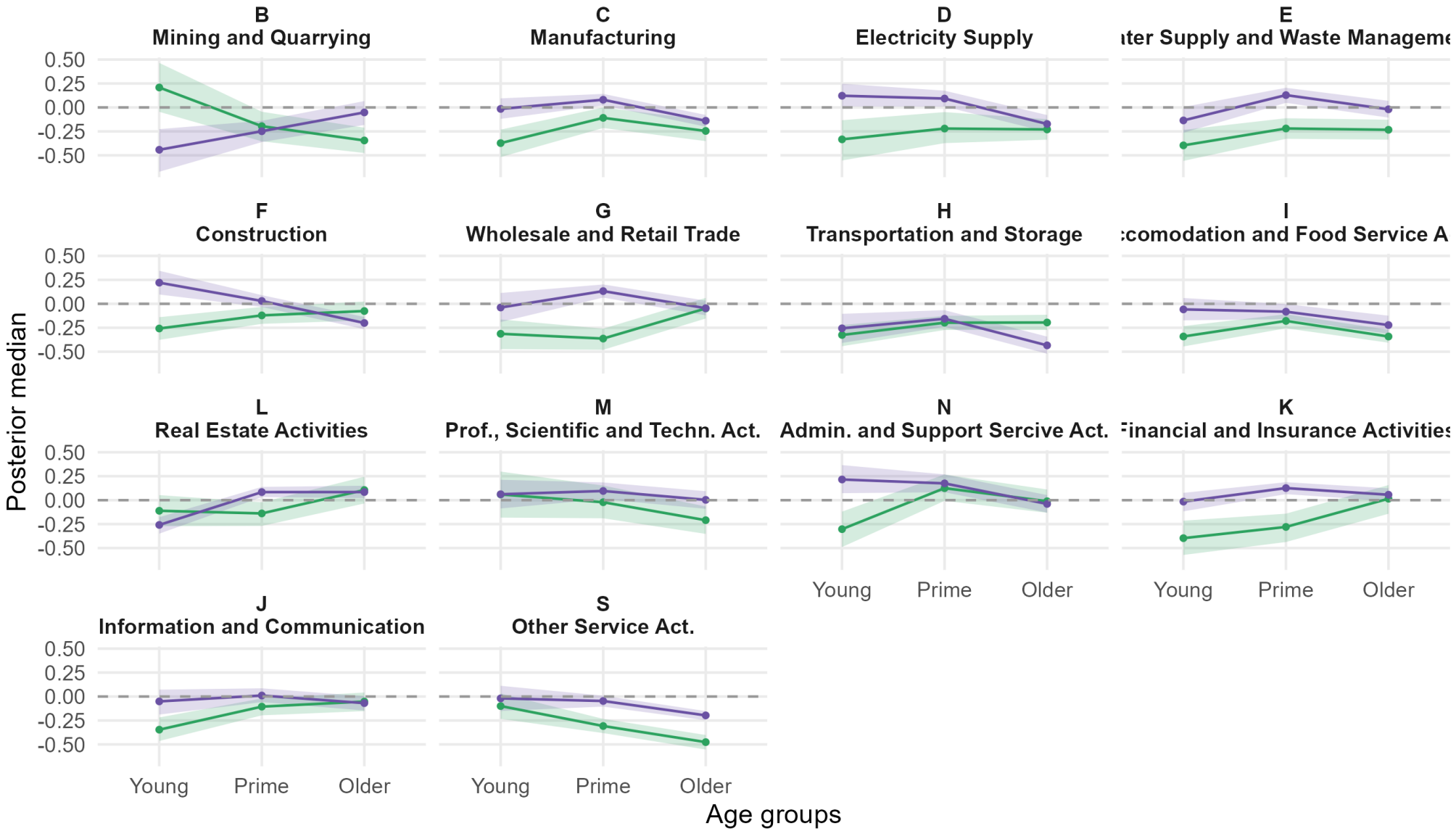
relative to 40-45 year-olds (benchmark)



Back to orig. slide

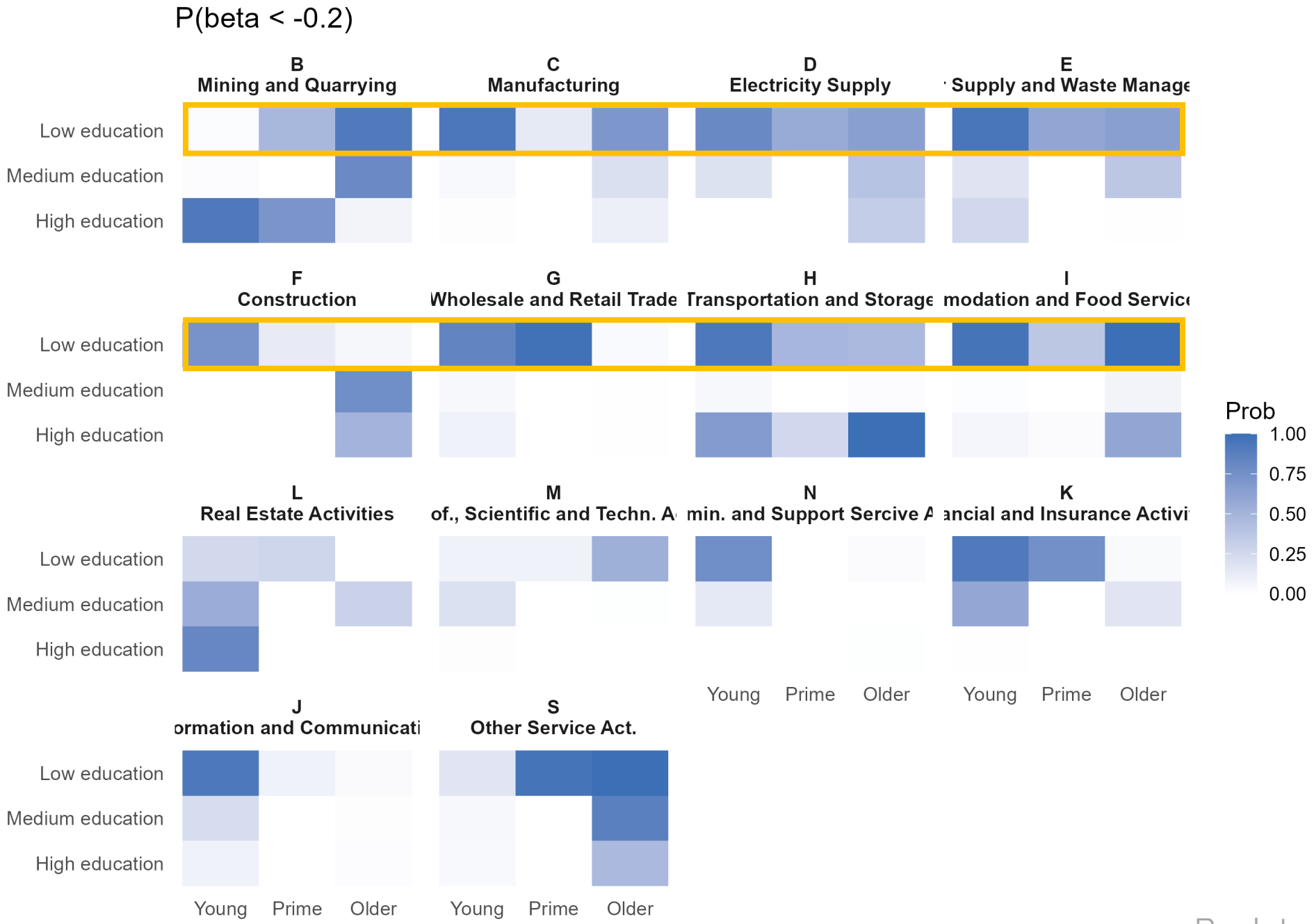
Posterior medians of joint age & education groups

benchmark: prime-aged with medium education; with 80 percent credible intervals

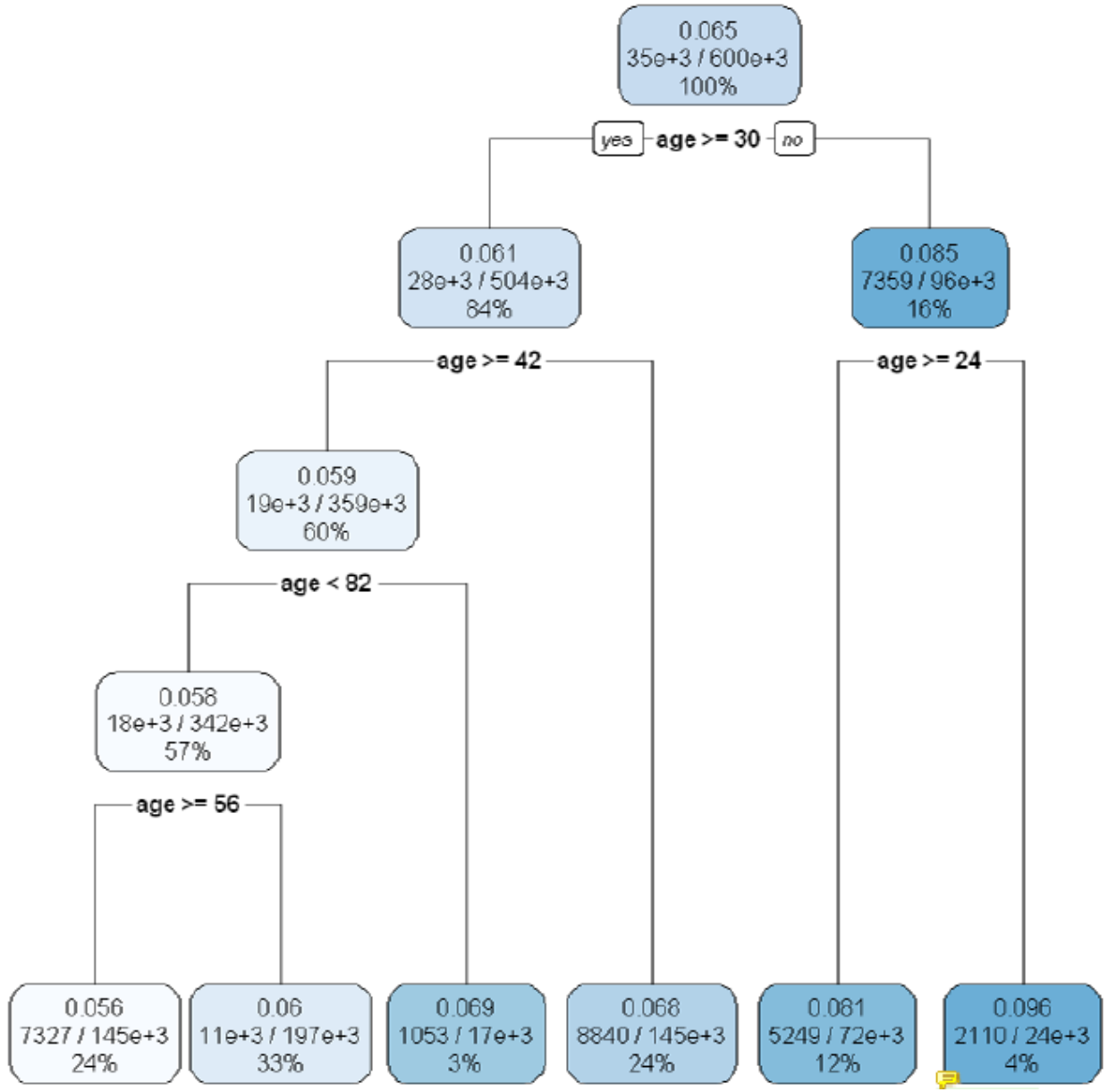


education Low education High education

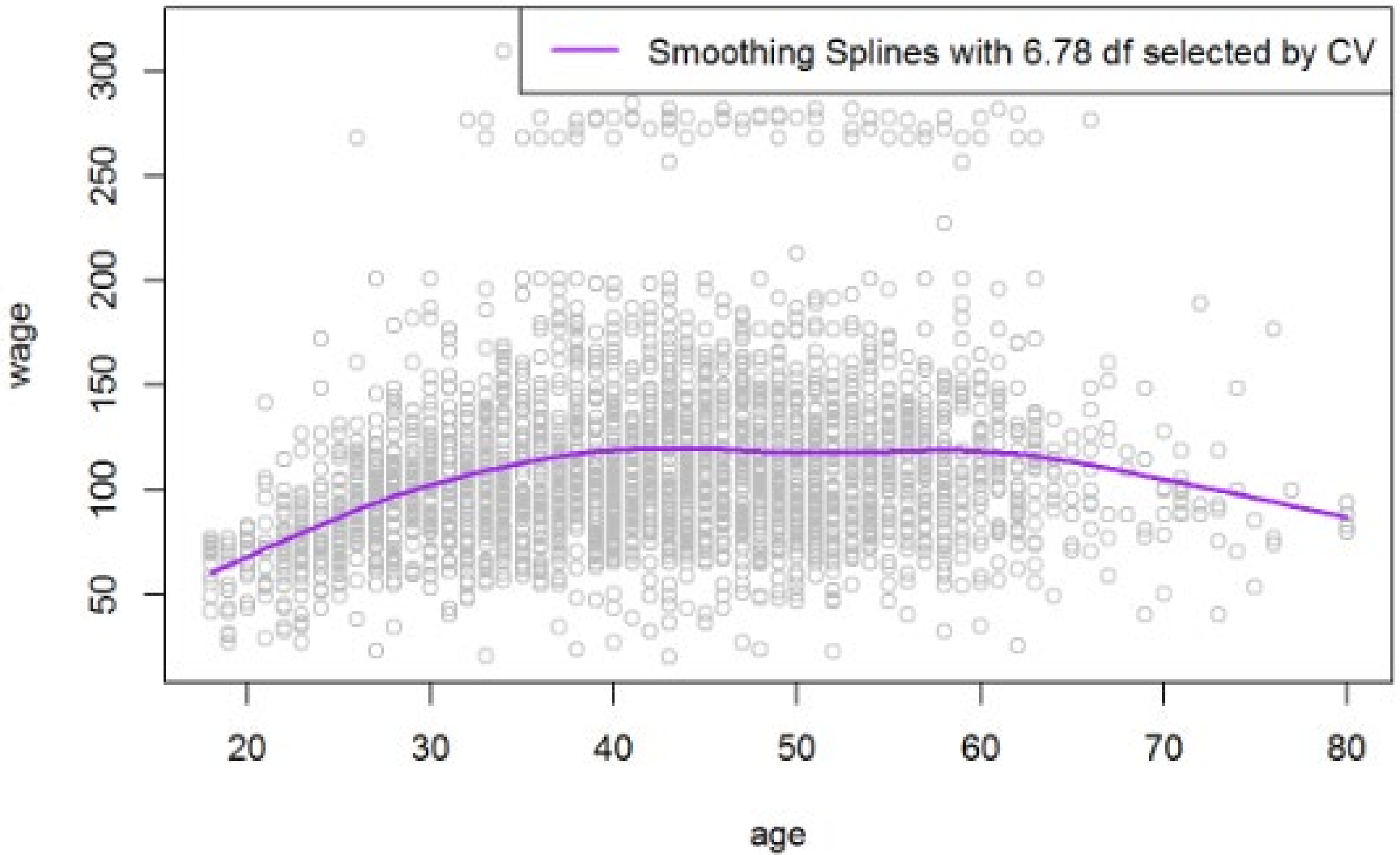
Back to Annex

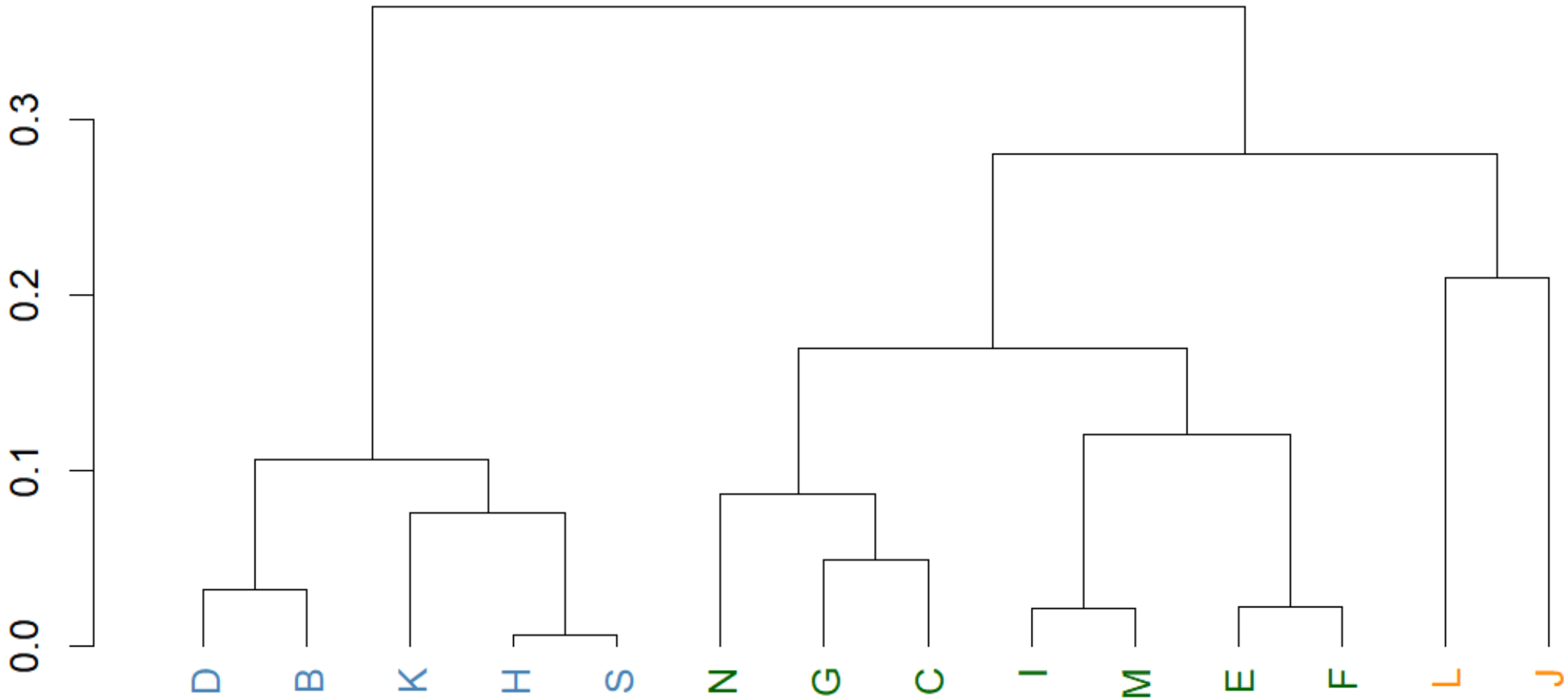


Back to Annex

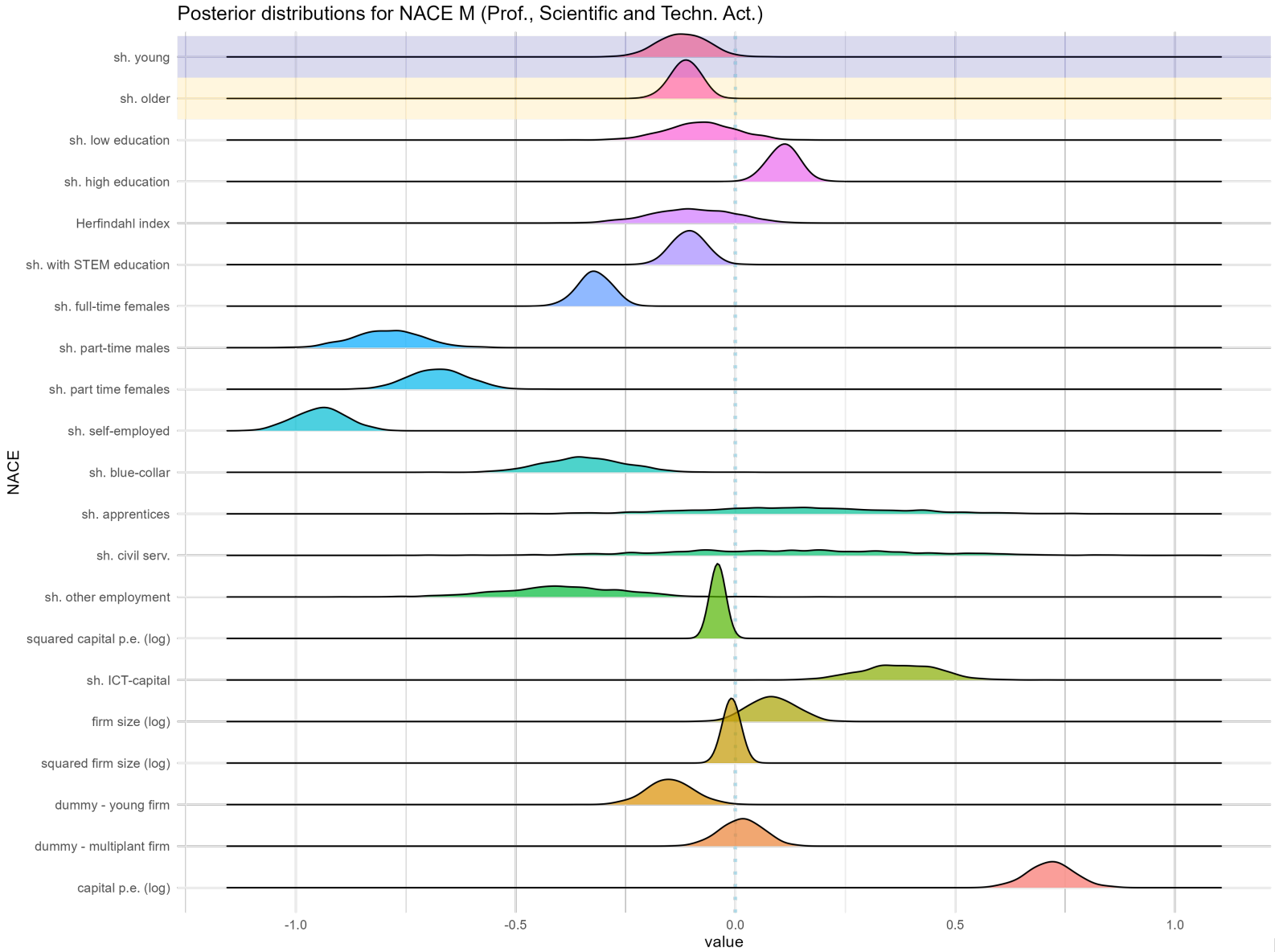


Back to Annex


Smoothing Splines with 6.78 df selected by CV
wage
50
100
150
200
250
300
20
30
40
50
60
70
80
age


0.0
0.1
0.2
0.3
D
B
K
H
S
N
G
C
I
M
E
F
L
J

Results for NACE M



Back to Annex

Variables – Employee Characteristics

Category	Variable Name	Description / Notes
Dependent Variable	ln_val_add_e_d_AEST	Log of value added per employee
	L.ln_val_add_e_d_AEST	Lagged value added
Age Structure	share_young	Share of young (15–29 years)
	share_middle	Share of middle aged
	share_old	Share of old (50+ years)
	herf_age	Herfindahl index of age concentration
Birth Cohorts	sharecohort_generation_z	Share of Gen Z cohort (1997–2010)
	sharecohort_generation_y	Share of Gen Y cohort (1981–1996)
	sharecohort_generation_x	Share of Gen X cohort (1965–1980)
	sharecohort_baby_boomer	Share of Baby Boomer cohort (≤1964)
	share_kid15	Share of employees with children under 15
Education	share_loweducated	Share of low-educated employees (primary/lower secondary)
	share_mediumeducated	Share of medium-educated (upper secondary)
	share_higheducated	Share of highly educated employees (mainly tertiary)
Employment Composition	female_share_AEST	Female employee share
	part_time_share_AEST	Part-time employee share
	part_female_share_AEST	Part-time female employee share
	blue_collar_share_AEST	Blue-collar employee share
	white_collar_share_AEST	White-collar employee share
	apprentices_share_AEST	Apprentice share
	other_employed_share_AEST	Other employed share

Variables – Firm Characteristics

Category	Variable Name	Description / Notes
Firm Characteristics	ln_size_firm_AEST	Log of firm size
	ln_capital_total_e_d_AEST	Log of total capital per employee
	multiplant_dummy	Dummy for multi-plant firms
	young_firm_dummy	Dummy for young firms
	ICT_capital_share	Share of ICT capital
Legal Form Dummies	lf_AG_dummy	Corporation dummy
	lf_GEN_dummy	General partnership dummy
	lf_GES_dummy	Limited partnership dummy
	lf_partnership_dummy	Other partnership dummy
Industry Dummies (NACE)	nace_B_dummy	Mining and quarrying
	nace_C_dummy	Manufacturing
	nace_D_dummy	Electricity, gas, steam
	nace_E_dummy	Water supply, sewerage
	nace_F_dummy	Construction
	nace_H_dummy	Transportation and storage
	nace_I_dummy	Accommodation and food service
	nace_J_dummy	Information and communication
	nace_K_dummy	Financial and insurance activities
	nace_L_dummy	Real estate activities
	nace_M_dummy	Professional, scientific and technical services
	nace_N_dummy	Administrative and support services
	nace_S_dummy	Other service activities
Region Dummies	9 federal state dummies	Burgenland to Vorarlberg (NUTS-2)

Priors – Firm characteristics

Priors

- Log(capital/head) ~Gamma(8,20) (right skewed)
- Log(capital/head)^2 ~normal(-0.05, 0.1) (decr. returns to scale)
- ICT_capital_share ~normal(0.1, 0.1)
- Log(firmsize) ~normal(0.1, 0.2)
- Log(firmsize)^2 ~normal(-0.1, 0.2) (decr. returns to scale)
- Young_firm_dummy ~normal(-0.1, 0.2)
- Multiplant_dummy ~normal(0.05, 0.1)

Priors – Employee characteristics

Priors

- Share_young, share_old ~normal(0, 0.3)
- Herf_age ~normal(0, 0.3)
- Share_loweducated, share_higheducated ~normal(0, 0.3)
- Share_STEM_education ~normal(0, 0.3)
- Share_fullt_female ~normal(0, 0.2)
- Share_partt_male, share_partt_female ~normal(-0.1, 0.3)
- Self_employed_share, blue_collar_share, apprentices_share, other_empl_share ~normal(0, 0.3)

NACE sectors

- ÖNACE - Klassifikation der Wirtschaftstätigkeiten – WKO

nace	sector	nace_name	nace_name_english
B	production	Bergbau	Mining and Quarrying
C	production	Herstellung von Waren	Manufacturing
D	production	Energieversorgung	Electricity Supply
E	production	Wasserversorgung und Abfallentsorgung	Water Supply and Waste Management
F	production	Bautaetigkeiten	Construction
G	service	Gross- und Einzelhandel	Wholesale and Retail Trade
H	service	Verkehr und Lagerei	Transportation and Storage
I	service	Beherbergung und Gastronomie	Accomodation and Food Service Act.
J	ICT	Information und Kommunikation	Information and Communication
K	financial	Finanz- und Verischerungsdienstl.	Financial and Insurance Activities
L	service	Immobilien	Real Estate Activities
M	service	Freiberufliche/techn. Dienstl.	Prof., Scientific and Techn. Act.
N	service	Sonst. wirtschaftl. Dienstl.	Admin. and Support Sercive Act.
S	ICT	Sonst. Dienstl.	Other Service Act.

Results of Study by Firm Sizes

- inverse u-relationship between age and productivity…
  - … but only in micro (1-9 ind.) and small (10-49) firms
  - … and not in medium (50-249) or large (250+) firms

Limitations

- On the individual level: no information about
  - Experience
  - Skills
  - Tasks
  - Position within the firm
- On the firm level: no information about
  - capital stock observed
  - managerial practices
  - Technology usage