Another One Rides the Bus ? Identifying Spatial Influences on the Job Matching Process of VET Students

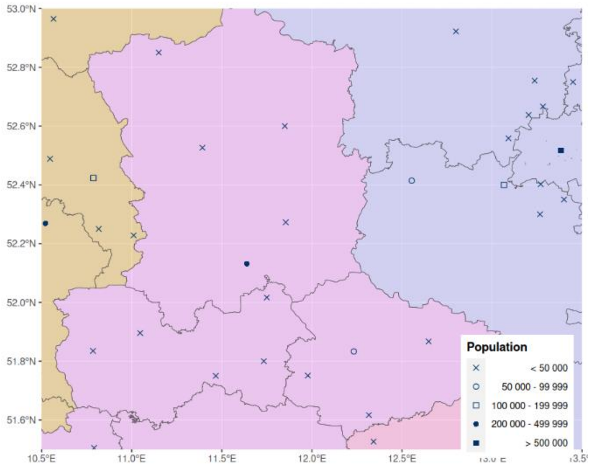
Dennis Oliver Kubitza

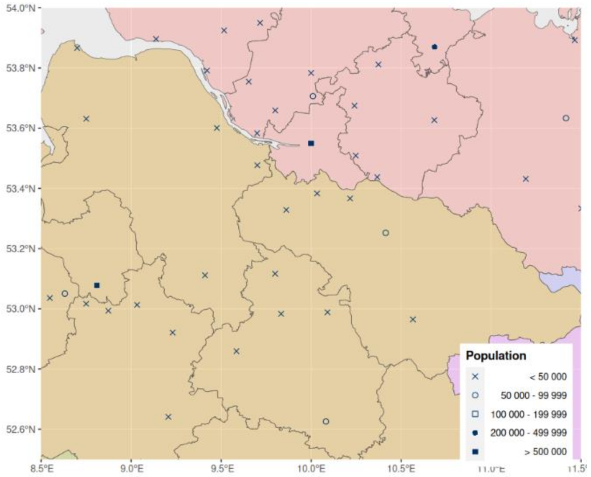
Federal Institute for Vocational Education and Training (bibb)
Bonn, Germany

Research Centre for Education and the Labour Market
Maastricht University, The Netherlands

Joint Spring Meeting BBS, BTU & SOEP RegioHub– The Power of Where

Problems with Regional Analysis of Labor Markets:





- ➢ (vacant) Jobs and (unmatched) Applicants that are surveyed in the same area might be unreachable of each other
- ➢ (vagcant) Jobs and (unmatched) Applicants might not be surveyed in the same area units

Sources: Own Visualization, [MAP-Data] Gebietsstruktur der Bundesagentur für Arbeit - Statistik der Bundesagentur für Arbeit (arbeitsagentur.de)

Solution Strategies from existing Research

Use Regional Controls: **City vs Rural Districts**
Density or other AD-Hoc Indicators

Modelling of Spillovers : **Include Neighbors**
Spatial Buffers
Commuting Patterns

Functional Regions: **Define Areas such that spillovers are minimized**

Topic: What is the best concept for identifying Spatial Influences on the Job Matching Process for the German Vocational Education and Training (VET) Market

Working Assumption:

- **Labor markets that are better (or worse) connected to their surroundings and internally should have a different matching efficiency of jobs and available applicants.**

Approach:

- **Regression Analysis on areal units with regional indicators**
- **Dependent Variable: Share of Job Vacancies**
- Compare different strategies for measuring spatial influences in their modelling power
  - Use Additional Information on the internal structure of Regions and connections between regions.
  - By translating this information to indexes on a [0,1] scale, such that 1 resembles better connectedness and 0 less connectedness.

Regression Analysis on Regional Units

As we only have aggregated job market information on areal units (**Labour Agency Districts)** the analysis will be divided in two parts:

a) Compare concepts measuring spatial Influences **within a LAD**
b) Compare concepts measuring spatial Influences **within a LAD and** spatial spillovers over **LAD borders**

In both parts we will use **Adjusted R2** to compare the goodness of fit.

Regression Analysis on Regional Units

INTRA Regional Analysis (linear Regression)

Share of Job Vacancies = Share of Free Applicants + Educational Labour Market Characteristics + Interactions +
Index of Regional Spatial Mismatch +
Index of Regional Spatial Mismatch * Relative Applicant Supply

INTER Regional Analysis (Spatial Durbin Error Model)

Share of Job Vacancies = Share of Free Applicants + Educational Labour Market Characteristics + Interactions +
Index of Regional Spatial Mismatch +
Index of Regional Spatial Mismatch * Relative Applicant Supply
Spatial Lag * Relative Applicant Supply Neighbours

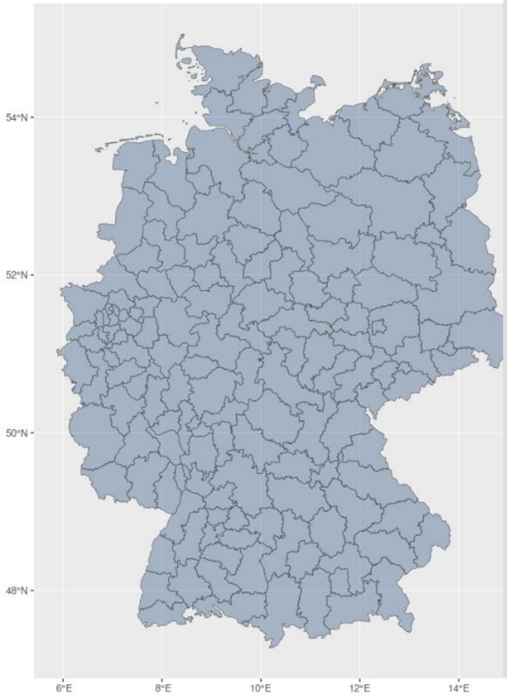
Concepts for Comparison I: Intra- Regional

East-West	EW	Indicator for a LAD being in East Germany (NOT INTPRETABLE {0,1}
Single City Region	SC	Indicator for being a Single City LAD
Frequency of Cities	FC	1/ Number of Cities within a LAD
Urban Permeation	UP	It measures the intensity and spread of buildings in the landscape. Max normalised to [0,1].
Share of Traffic Infrastructure	TI	Share of Traffic Infrastructure in Living Areas.
Population Density	PD	Population Density, rescaled from [0, max] to [0,1]
Functional Regions after Kosfeld	FK FK_S	1/ Number of functional regions that intersect the LAD Share of biggest Functional Region
Labour Market Regions	LM_I LM_S	1/ Number of Labour Market Regions that are present in that LAD. Share of biggest Labour Market Region
Distance	DIST_NUM DIST_POP	Share of cities/ population that can be reached by an average city/ citizen within x km distance.
Nearest Neighbour	NN_NUM NN_POP	Share of cities/ population that can be reached by an average city/ citizen within the x Nearest Neighbours
Car Transport Time	CAR_NUM CAR_POP	Share of cities/ population that can be reached by an average city/ citizen within x minutes of car driving,
Public Transport Time	PUB_NUM PUB_POP	Share of cities/ population that can be reached by an average city/ citizen within x minutes of taking public transport.

Concepts for Comparison II: Inter - Regional

Neighbourhoods	WN1 -WN3	Neighbourhoods of degree 1,2 and 3
Higher Level Aggregation	WHLA	Indicator, whether two LAD belong to the same higher-level aggregation. In this case we use 10 administrative regions defined by the Labour Agency.
Functional Regions after Kosfeld	WFK_B WFK_C WFK_P	Whether two LAD belong to the same functional region, with either binary encoding or city/ population weighted weights.
Labour Market Regions	WLK_B WLK_C WLK_P	Whether two LAD belong to the same labour market region, with either binary encoding or city/ population weighted weights.
Centroid Distances	WCENT	Weighting by distances of LAD Centroids
Commuters	WCOM_I WCOM_O	Inward and Outward Commuters between LAD
Distance Index	WDIST_B WDIST_C WDIST_P	Indicator or Share of cities/ population that can be reached by an average city/ citizen within x km distance.
Nearest Neighbour	WKNN_B WKNN_C WKNN_P	Indicator or Share of cities/ population that can be reached by an average city/ citizen within the x Nearest Neighbours,
Car Transport Time	WCAR_B WCAR_C WCAR_P	Indicator or Share of cities/ population that can be reached by an average city/ citizen within x minutes of car driving,
Public Transport Time	WPUB_B WPUB_C WPUB_P	Indicator or Share of cities/ population that can be reached by an average city/ citizen within x minutes of taking public transport.

Data Sources I

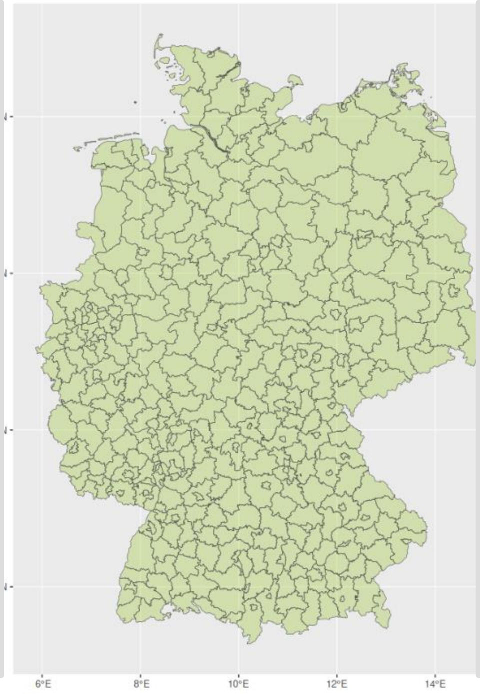


BIBB Survey: Newly Concluded Training Contracts [A]

Job Vacancies 2023
Applicants 2023
New Contracts 2023

-

7 profession groups (KLDB 3-digits) for 154 Employment Agency Districts



INKAR Regional Database [B]

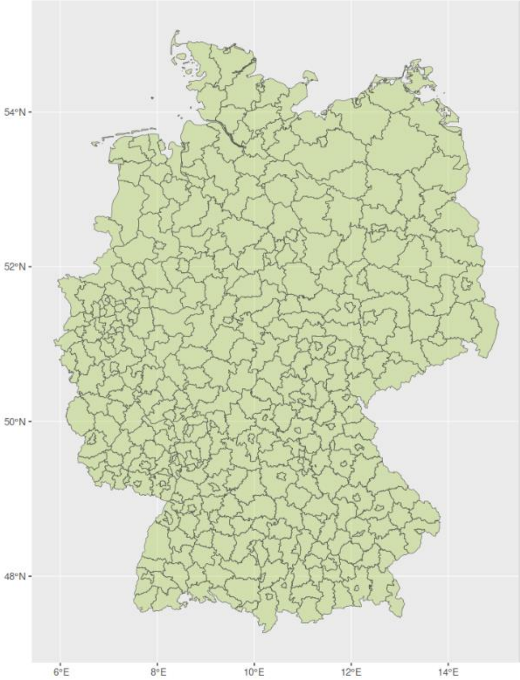
Median Income

Shares of
Unemployment <25
Big Companies
Medium Companies
Cars
UNI Students
LSE Students
MSE Students

For 401 federal districts (NUTS3)

Sources: [A] https://www.bibb.de/naa309, BIBB (Bundesinstitut für Berufsbildung)
[B] BBSR. "INKAR Indikatoren und Karten zur Raum-und Stadtentwicklung." (2021).
[MAPS] https://gdz.bkg.bund.de/index.php/default/open-data/verwaltungsgebiete-1-1-000-000-stand-31-12-vg1000-31-12.html) License: dl-de/by-2-0, www.govdata.de/dl-de/by-2-0)

Data Sources II



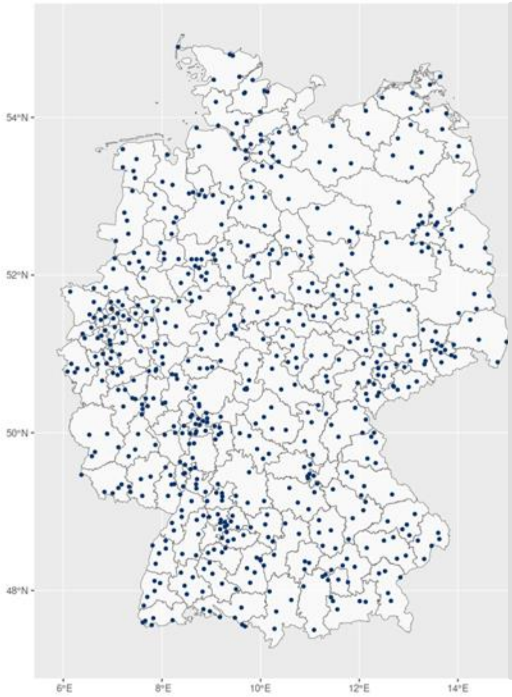
IÖR-Monitor[C]

Urban Permeation
Share of Traffic Infrastructure
Population Density

For 401 federal districts (NUTS3)

Sources [C] Monitor der Siedlugs-und Freiraumentwicklung (IOER-Monitor), Leibniz-Institut für ökologische Raumentwicklung. (2021).

Data Sources III

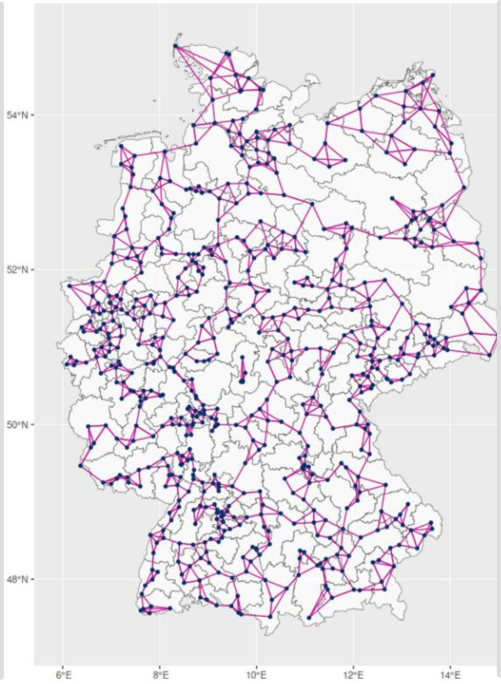


Wikidata Sparql Endpoint [D]

All Cities and Villages in Germany (>10000)

Geolocation
Popultion

Sample such that 50% Population per region is covered



Google Distance Matrix API [E]

For all sample cities in dataset with less than 50km distance or within neighbouring regions

Public Transport Times
Driving Times

Sources: [D] query.wikidata.org , Wikimedia Foundation
[E] Google Distance Matrix API (Google Maps Platform Documentation | Routes API | Google Developers - Fahrtzeiten ÖPNV)

Prelimnary Results – Intra District (Linear Regression)

Best Distances: 13.06/13.06
Best Car Transport Times: 25.57/25.57
Best Public Transport Times 43.45/43.45

	Base	EW	SC	FC	UP	DEN	TI	FK	DIST_NUM	CAR_NUM	PUB_NUM	KNN_NUM	DIST_POP	CAR_POP	PUB_POP	KNN_POP
(Intercept)	8.168	3.748	17.078 *	21.699 **	23.513 **	23.882 **	22.006 **	17.059 *	22.686 **	21.234 **	25.479 **	23.614 **	24.298 **	24.471 **	25.616 **	23.943 **
	(5.789)	(6.834)	(6.654)	(7.319)	(7.266)	(7.400)	(7.301)	(7.351)	(7.880)	(7.439)	(7.658)	(7.687)	(8.039)	(7.597)	(7.645)	(7.823)
RelativeSupply	-0.001	0.170	-1.720 ^	-2.447 *	-2.862 *	-2.526 ^	-2.443 ^	-0.500	-2.487 ^	-1.967 ^	-2.709 *	-2.690 *	-2.542 ^	-2.333 ^	-2.651 *	-2.633 *
	(0.782)	(0.865)	(1.004)	(1.163)	(1.226)	(1.275)	(1.331)	(1.334)	(1.359)	(1.149)	(1.254)	(1.219)	(1.415)	(1.200)	(1.289)	(1.264)
NewContracts	-0.017 *	-0.020 **	-0.017 *	-0.015 *	-0.015 *	-0.015 *	-0.014 ^	-0.015 *	-0.014 *	-0.015 *	-0.012 ^	-0.015 *	-0.013 ^	-0.013 ^	-0.011	-0.014 ^
	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
medSE	0.115	0.073	-0.015	-0.048	-0.091	-0.053	-0.042	0.138	-0.042	-0.011	-0.043	-0.033	-0.043	-0.031	-0.030	-0.023
	(0.105)	(0.101)	(0.114)	(0.116)	(0.119)	(0.114)	(0.116)	(0.105)	(0.119)	(0.113)	(0.112)	(0.113)	(0.118)	(0.113)	(0.111)	(0.113)
lowSE	0.143	0.490 ^	0.048	-0.018	0.032	-0.000	0.080	-0.099	-0.031	0.012	-0.097	-0.048	-0.057	-0.028	-0.088	-0.049
	(0.188)	(0.247)	(0.191)	(0.194)	(0.185)	(0.186)	(0.183)	(0.208)	(0.195)	(0.190)	(0.194)	(0.196)	(0.195)	(0.188)	(0.192)	(0.195)
RelativeSupply:medSE	-0.020	-0.014	0.005	0.009	0.016	0.005	0.006	-0.024	0.006	0.000	0.005	0.005	0.005	0.002	0.002	0.003
	(0.014)	(0.013)	(0.017)	(0.017)	(0.018)	(0.017)	(0.018)	(0.016)	(0.018)	(0.017)	(0.016)	(0.016)	(0.018)	(0.017)	(0.016)	(0.016)
RelativeSupply:lowSE	0.013	-0.008	0.030	0.040	0.040	0.047	0.033	0.033	0.042	0.027	0.048	0.047	0.045	0.034	0.048	0.047
	(0.039)	(0.044)	(0.039)	(0.039)	(0.038)	(0.040)	(0.039)	(0.040)	(0.040)	(0.038)	(0.039)	(0.040)	(0.040)	(0.038)	(0.039)	(0.040)
innerSpatial		4.933 ^	-5.577 *	-9.017 *	-0.460 **	-19.151 **	-17.409 **	-8.355 *	-9.940 *	-10.363 **	-12.301 **	-11.146 **	-11.583 **	-13.040 **	-13.412 ***	-11.912 **
		(2.805)	(2.674)	(3.457)	(0.154)	(5.888)	(5.746)	(3.624)	(3.812)	(3.866)	(3.735)	(4.111)	(3.989)	(4.020)	(3.922)	(4.296)
RelativeSupply:innerSpatial		0.438	0.984 **	1.437 **	0.068 **	2.399 **	2.283 *	0.474	1.439 *	1.363 *	1.628 **	1.691 **	1.492 *	1.563 *	1.636 **	1.671 **
		(0.519)	(0.371)	(0.504)	(0.022)	(0.867)	(0.887)	(0.692)	(0.618)	(0.577)	(0.587)	(0.586)	(0.662)	(0.608)	(0.622)	(0.620)
adj.r.squared	0.276	0.342	0.311	0.321	0.334	0.335	0.325	0.317	0.311	0.313	0.336	0.322	0.320	0.333	0.340	0.320
r.squared	0.317	0.392	0.363	0.372	0.385	0.386	0.377	0.369	0.364	0.365	0.386	0.374	0.372	0.384	0.391	0.371
N. obs.	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; ^ $p < 0.1$.

Regression Results: Energy Technicians

Preliminary Results – Intra District (Linear Regression)

Table:: Modelled Profession and ranking of model performances **without controls without spillovers**

	B)	C)	D)	E)	F)	G)	#Significant	Avg (adj. R2)
PUB_NUM	2	3	2	2	2	4	6	0,258
PUB_POP		2	1	4		5	4	0,250
DIST_NUM	1	12			1		3	0,250
EW				1		1	2	0,249
CAR_NUM	3	10	3		6		4	0,245
CAR_POP		3	4		5		3	0,240
DIST_POP		8			3		2	0,238
KNN_NUM	4	6			4	2	4	0,234
FK		10	5	3			3	0,230
KNN_POP		8			7	3	3	0,225
FC	5	5			8		3	0,225
UP		1	7		10		3	0,225
SC		7			9		2	0,220

Preliminary Results – Intra District (Linear Regression)

Table: Modelled Profession and ranking of model performances **with controls without spillovers**

	A)	B)	C)	D)	F)	#Significant	Avg (adj. R2)
PUB_NUM	2		8	2		3	0,338
PUB_POP	1	1	11			3	0,336
DIST_NUM			9	4		2	0,335
CAR_NUM	4	3	5	3		4	0,332
CAR_POP	3	2	7			3	0,332
DIST_POP		4	10			2	0,327
KNN_NUM			2		1	2	0,327
SC			3			1	0,323
KNN_POP			6		2	1	0,323
EW			12	1		2	0,321
FC			4			1	0,321
UP			1			1	0,319
FK			12		3	2	0,317

Preliminary Results – Inter District

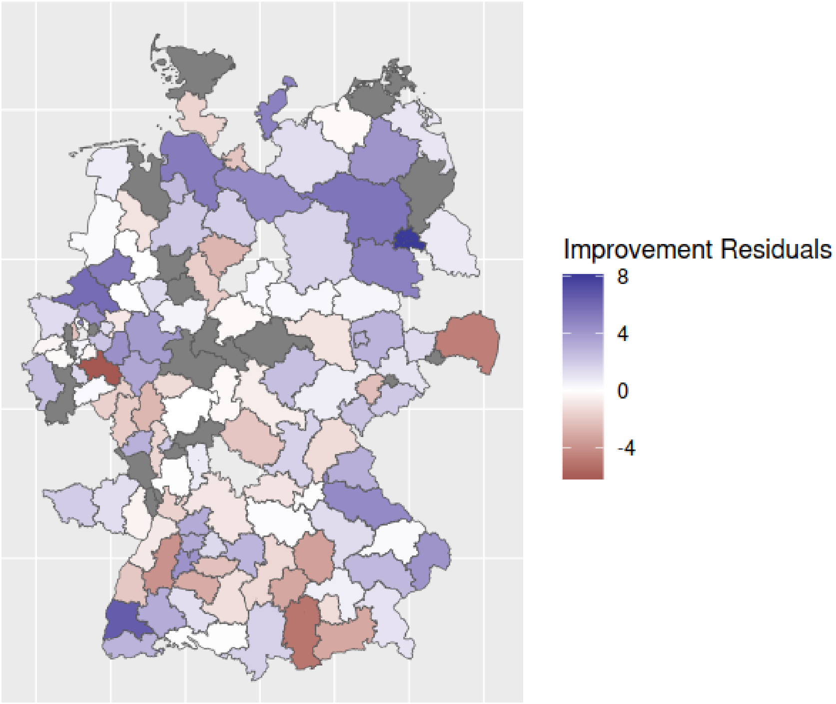
Table: Modelled Profession and ranking of model performances **without controls with spillovers**

	B)	C)	D)	E)	F)	G)	#Significant	Avg (R2)
N2	1	1	1	1	5	1	6	0.317
WCAR_B	9	5	2		1	5	5	0.301
WDIST_B	4	4	3		2	9	5	0.296
WCAR_C	8	11	4		4	11	5	0.296
WDIST_P	4	8	11		5	5	5	0.295
WDIST_C	2	8	8		7	8	5	0.295
WCAR_P	3	6	5		8	5	5	0.293
N3		2	8		16	2	4	0.292
WPUB_P	4	8	6		8	3	5	0.292
WPUB_B	10	3	6		3	10	5	0.291
WPUB_C	7	12	8		10	11	5	0.289
N1	11	14	12		15	13	5	0.282
FK_C	13	6	14		12	15	5	0.281
FK_P	14	12	13		12	16	5	0.281
FK_B		15	15		11	14	4	0.280
HLA	11	16	16		14	3	5	0.276

Results – Intermediate Findings

- Significant Controls & Best Spatial Measures vary between Professions.
- Buffer-Based Methods outperform other measures for intra-regional Analysis.
- For inter-regional analysis
  - N2 outperforms all measures
  - Buffer Based Methods outperform all other measures

Results – Visualised



Future Tasks:

- Implement further measures:
    - Labour Market Regions (IAB)
    - Functional Regions (Kropp, Schwengler)

    […]

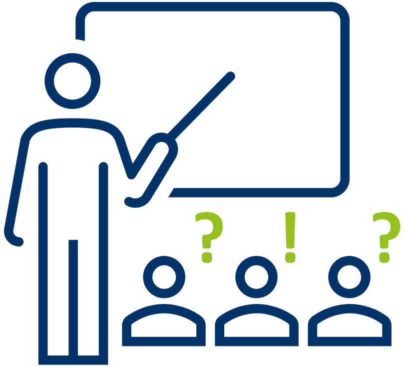
- Assure robustness of results by Simulation (Random Merges of Neighbours)

- Introduce a priori and a-posterior measures of Moran’S I of Error Terms for the analysis.

Contact Information

Dennis.Kubitza@bibb.de

Questions & Comments ?



	unemp25	bigCo	cohort	income	medCo	density	cars	shareVET	shareUNI	medSE	lowSE	urban_permeation	share_traffic	inhabitants
unemp25	1.00	0.28	0.38	-0.22	0.26	0.21	-0.55	-0.42	0.15	-0.29	-0.39	0.41	0.30	0.06
bigCo	0.28	1.00	-0.02	0.46	0.71	0.49	-0.52	-0.37	0.57	-0.28	-0.14	0.60	0.63	[illegible]
cohort	0.38	-0.02	1.00	-0.40	0.02	0.03	-0.43	-0.36	0.11	-0.03	-0.53	0.01	-0.08	0.05
income	-0.22	0.46	-0.40	1.00	0.03	0.65	-0.16	-0.32	0.38	-0.26	0.19	0.49	0.61	0.27
medCo	0.26	0.71	0.02	0.03	1.00	0.07	-0.33	-0.07	0.29	0.03	-0.17	0.26	0.24	-0.20
density	0.21	0.49	0.03	0.65	0.07	1.00	-0.66	-0.63	0.52	-0.46	-0.21	0.87	0.91	0.44
cars	-0.55	-0.52	-0.43	-0.16	-0.33	-0.66	1.00	0.75	-0.58	0.46	0.43	-0.69	-0.60	-0.37
shareVET	-0.42	-0.37	-0.36	-0.32	-0.07	-0.63	0.75	1.00	-0.67	0.45	0.44	-0.58	-0.56	-0.38
shareUNI	0.15	0.57	0.11	0.38	0.29	0.52	-0.58	-0.67	1.00	-0.39	-0.25	0.50	0.52	0.20
medSE	-0.29	-0.28	-0.03	-0.26	0.03	-0.46	0.46	0.45	-0.39	1.00	0.16	-0.49	-0.46	-0.19
lowSE	-0.39	-0.14	-0.53	0.19	-0.17	-0.21	0.43	0.44	-0.25	0.16	1.00	-0.16	-0.13	-0.08
urban_permeation	0.41	0.60	0.01	0.49	0.26	0.87	-0.69	-0.58	0.50	-0.49	-0.16	1.00	0.96	0.32
share_traffic	0.30	0.63	-0.08	0.61	0.24	0.91	-0.60	-0.56	0.52	-0.46	-0.13	0.96	1.00	0.28
inhabitants	0.06	[illegible]	0.05	0.27	-0.20	0.44	-0.37	-0.38	0.20	-0.19	-0.08	0.32	0.28	1.00

$$y = X\gamma + G_y y\beta + G_x X\theta + G_v v\lambda + \varepsilon.$$

$$Ind^x_{num}(A) = \sum_{a,b \in A} \left(\frac{1}{|A| \cdot |A|} \right) \cdot \mathbb{1}x(a,b) \tag{1}$$

$$Ind^x_{pop}(A) = \sum_{a,b \in A} \left(\frac{P(a) \cdot P(b)}{P(A) \cdot P(A)} \right) \cdot \mathbb{1}x(a,b) \tag{2}$$

$$W^x_{num}(A,B) = \sum_{a \in A} \sum_{b \in B} \left(\frac{1}{|A| \cdot |B|} \right) \cdot \mathbb{1}^x(a,b) \tag{3}$$

$$W^x_{pop}(A,B) = \sum_{a \in A} \sum_{b \in B} \left(\frac{P(a) \cdot P(b)}{P(A) \cdot P(A)} \right) \cdot \mathbb{1}^x(a,b) \tag{4}$$

$$W^x_{bin}(A,B) = \max_{a \in A, b \in B} \left(\mathbb{1}^x(a,b) \right) \tag{5}$$