# Explore the multifaceted determinants of regional employment rates in Germany

*How do economic performance, labor market dynamics, educational attainment, and industry composition affect regional employment rates in Germany?*

Stephen Elvis Ampah elvis.ampah@tu-dortmund.de

Technische Universität Dortmund - Ruhr Alliance, Germany

July 30, 2024

# Table of Contents

**Importance of Studying Regional Employment Rates**

- **Economic Health Indicator:** Employment rates are critical indicators of economic health and labour market performance
- **Policy Relevance:** Understanding regional variations in employment can guide targeted policy interventions
- **Social Implications:** High employment rates are associated with better social outcomes, including reduced poverty and improved quality of life (Möller 2010).

# Research Question

**Research Question**

- How do economic performance, labour market dynamics, educational attainment, and industry composition affect regional employment rates in Germany?

**Literature**

- How elastic is labor demand? A meta-analysis for the German labor market. *Journal for Labour Market Research*, 57(1), 14. (Popp 2023).
- Another economic miracle? The German labor market and the Great Recession. *IZA Journal of Labor Policy*, 1, 1-21. (Rinne and Zimmermann 2012).

# Data Sources

- **INKAR.de**: A comprehensive database provided by the Federal Institute for Research on Building, Urban Affairs, and Spatial Development (BBSR) in Germany.
- **Variables**:
  - **Employment Rate (ER)**: Proportion of the working-age population that is employed.
  - **Higher Education Enrollment Quota (HEEQ)**: Percentage of the population enrolled in higher education.
  - **Employees in Knowledge-Intensive Industries (EKII)**: Share of employment in sectors requiring advanced knowledge and skills.
  - **Household Income (HI)**: Average income of households.
  - **Unemployment Rate (UR)**: Percentage of the labour force that is unemployed and seeking work.
  - **Gross Domestic Product (GDP)**: Total economic output of a region.

# Baseline Model: Ordinary Least Squares (OLS) Regression

- **Purpose**: Establish initial relationships between the dependent variable $ER$ and independent variables $HEEQ, EKII, HI, UR, GDP$.
- **Procedure**: Fit an OLS regression model to the data to determine the direct effects of the independent variables on the employment rate.
- **Formula**:

$$ER = \beta_0 + \beta_1 HEEQ + \beta_2 EKII + \beta_3 HI + \beta_4 UR + \beta_5 GDP + \epsilon \quad (1)$$

where $\beta_i$ are parameters to be estimated and $\epsilon$ is the error term.

*Note: The model formulation follows the guidelines set out in Wooldridge (2012).*

# Moran's I Test (Moran 1950)

- **Purpose**: Detect spatial patterns and autocorrelation in the residuals of the OLS model.
- **Formula**:

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \tag{2}$$

  where:
  - $N$: Number of observations
  - $W$: Sum of all spatial weights $w_{ij}$
  - $x_i$: Value of the variable at location $i$
  - $x_j$: Value of the variable at location $j$
  - $\bar{x}$: Mean of the variable
  - $w_{ij}$: Spatial weight between location $i$ and $j$
- **Result**: Indicated patterns of spatial autocorrelation, suggesting that the employment rates in one district are influenced by those in neighbouring regions.

# Spatial Econometric Models

- **Spatial Lag Model (SLM) (Anselin 1988)**:
  - **Purpose**: Account for the dependence of employment rates in neighbouring district.
  - **Method**: Incorporates a spatially lagged dependent variable to capture the influence of neighbouring district' employment rates on the local employment rate.
  - **Equation**:

$$ER = \rho WER + X\beta + \epsilon \tag{3}$$

    where:

    - $ER$: Employment rate in region
    - $W$: The spatial weights matrix
    - $WER$: Spatially lagged employment rate
    - $X$: Matrix of independent variables
    - $\beta$: The vector of coefficients for the independent variables
    - $\epsilon$: The vector of error terms
    - $\rho$: Spatial autoregressive coefficient

# Spatial Econometric Models (Robustness Test)

- **Spatial Error Model (SEM)**:
  - **Purpose**: Account for spatially correlated error terms.
  - **Method**: Adjusts for spatial dependence in the error term, capturing unobserved district factors affecting employment rates.
  - **Equation**:

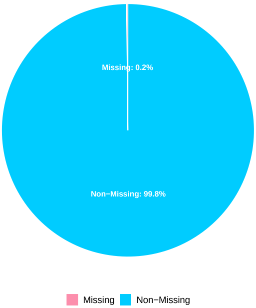
$$ER = X\beta + u \tag{4a}$$

$$u = \lambda W u + \epsilon \tag{4b}$$

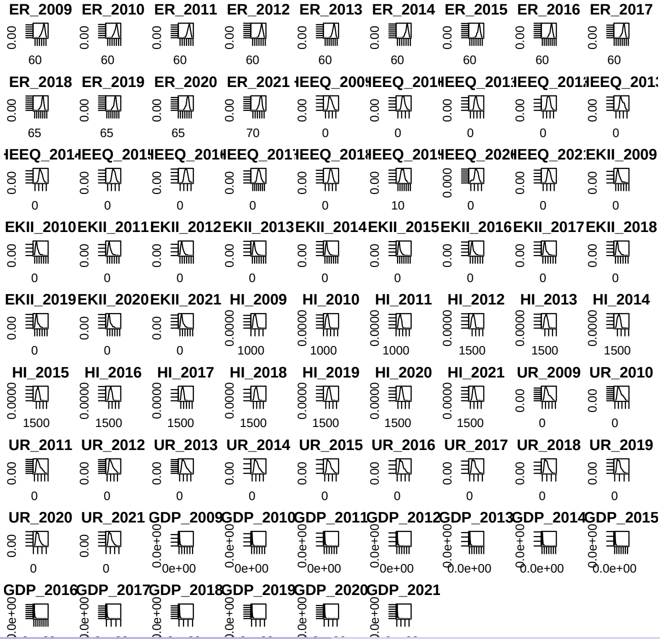
  where:
    - $ER$ is the dependent variable vector
    - $X$ is the matrix of independent variables
    - $\beta$ is the vector of coefficients for the independent variables
    - $u$ is the vector of error terms that incorporates spatial autocorrelation
    - $\lambda$ is the coefficient of spatial autocorrelation for the error terms
    - $W$ is the spatial weights matrix
    - $\epsilon$ is the vector of independently and identically distributed (iid) error terms

- *The robustness of the SEM in comparison to SLM is discussed extensively in Spatial Econometrics (Anselin 1988; Elhorst 2014).*

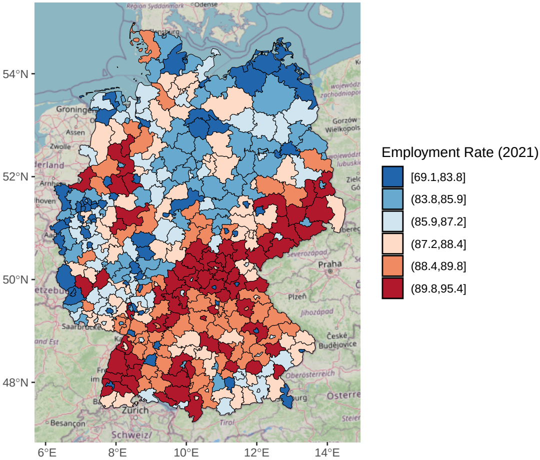
# Percentage of missing vs non-missing data points



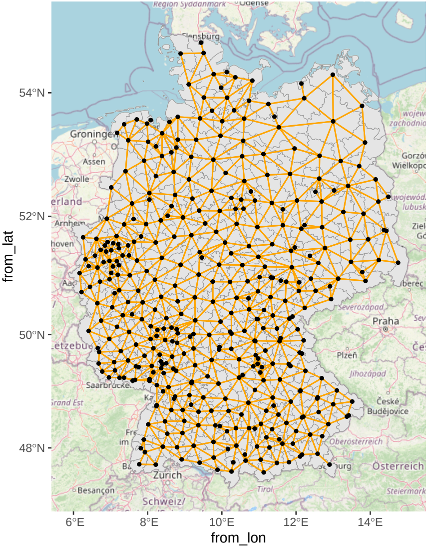
# Density Plots



# Employmemt rate in Germany - 2021



# Neighborhood Structure (Spatial dependence)



# Correlation Matrix of Numeric Variables

# Descriptive Statistics for Various Variables

| Variable | Mean | Median | SD | Min | Max | Q25 | Q75 | Skewness |
|---|---|---|---|---|---|---|---|---|
| ER | 4.41 | 4.42 | 0.05 | 4.09 | 4.57 | 4.38 | 4.45 | -0.77 |
| HEEQ | 3.41 | 3.44 | 0.38 | -2.81 | 4.25 | 3.23 | 3.64 | -4.35 |
| EKII | 2.12 | 2.15 | 0.74 | -1.47 | 4.04 | 1.71 | 2.61 | -0.63 |
| HI | 7.47 | 7.47 | 0.14 | 7.09 | 8.14 | 7.37 | 7.56 | 0.12 |
| UR | 1.68 | 1.69 | 0.47 | 0.22 | 2.88 | 1.33 | 2.02 | 0.00 |
| GDP | 15.45 | 15.37 | 0.77 | 13.73 | 18.92 | 14.90 | 15.88 | 0.90 |

# OLS Regression Results

| Variable | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 3.32 | 0.04 | 75.07 | $< 2.00e^{-16}$ *** |
| HEEQ | -0.03 | 0.002 | -16.72 | $< 2.00e^{-16}$ *** |
| EKII | 0.00 | 0.001 | 4.43 | $9.51e^{-6}$ *** |
| HI | 0.19 | 0.006 | 31.28 | $< 2.00e^{-16}$ *** |
| UR | 0.00 | 0.002 | 2.20 | $2.76e^{-2}$ * |
| GDP | -0.02 | 0.001 | -19.77 | $< 2.00e^{-16}$ *** |

*Residual standard error*: 0.042 on 5194 degrees of freedom
*Multiple R-squared*: 0.333, *Adjusted R-squared*: 0.332
*F-statistic*: 518.1 on 5 and 5194 DF, *p-value*: $< 2.2e^{-16}$

# VIF values by year and variable

| Year | HEEQ | EKII | HI | UR | GDP |
|---|---|---|---|---|---|
| 2021 | 1.45 | 1.22 | 2.00 | 2.42 | 1.24 |
| 2020 | 1.08 | 1.19 | 2.02 | 2.12 | 1.20 |
| 2019 | 1.51 | 1.23 | 2.50 | 3.09 | 1.23 |
| 2018 | 1.51 | 1.28 | 2.50 | 3.06 | 1.27 |
| 2017 | 1.47 | 1.26 | 2.19 | 2.75 | 1.23 |
| 2016 | 1.52 | 1.27 | 2.49 | 3.07 | 1.25 |
| 2015 | 1.52 | 1.27 | 2.59 | 3.16 | 1.26 |
| 2014 | 1.49 | 1.28 | 2.60 | 3.06 | 1.29 |
| 2013 | 1.57 | 1.26 | 2.51 | 3.01 | 1.29 |
| 2012 | 1.41 | 1.26 | 2.49 | 2.79 | 1.28 |
| 2011 | 1.09 | 1.26 | 2.33 | 2.38 | 1.19 |
| 2010 | 1.47 | 1.24 | 2.30 | 2.79 | 1.19 |
| 2009 | 1.69 | 1.22 | 2.16 | 3.02 | 1.15 |

# Moran's I Statistics and P-Values by Year

| Year | Moran's I Statistic | P-Value |
|---|---|---|
| 2009 | 0.29 | 5.61e-19 |
| 2010 | 0.32 | 5.07e-22 |
| 2011 | 0.34 | 4.46e-26 |
| 2012 | 0.32 | 1.61e-22 |
| 2013 | 0.28 | 3.35e-18 |
| 2014 | 0.28 | 5.60e-18 |
| 2015 | 0.28 | 1.12e-17 |
| 2016 | 0.28 | 2.01e-17 |
| 2017 | 0.25 | 1.17e-14 |
| 2018 | 0.24 | 8.62e-14 |
| 2019 | 0.22 | 6.35e-12 |
| 2020 | 0.19 | 1.66e-09 |
| 2021 | 0.19 | 1.51e-09 |

# Robust LM Diagnostics for Spatial Dependence in OLS Model (Anselin 1988)

**Formulas and Hypotheses**
*RLM Lag Test:*

$$LM_{\text{lag}} = n \cdot (\hat{\rho}^2)$$

Where $\hat{\rho}$ is the estimated spatial autoregressive coefficient.
**Hypotheses:**
$H_0 : \rho = 0$ (no spatial lag dependence),
$H_1 : \rho \neq 0$ (spatial lag dependence).

*RLM Error Test:*

$$LM_{\text{error}} = n \cdot (\hat{\lambda}^2)$$

Where $\hat{\lambda}$ is the estimated coefficient of the spatial error model.
**Hypotheses:**
$H_0 : \lambda = 0$ (no spatial error autocorrelation),
$H_1 : \lambda \neq 0$ (spatial error autocorrelation).

**Results**

| Test | Statistic | df | p-value |
|---|---|---|---|
| RLM Lag | 3.585 | 1 | 0.05829 |
| RLM Error | 24.873 | 1 | 6.12e-07 |

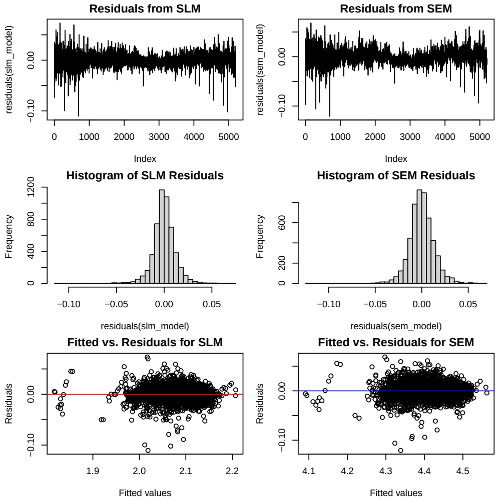
# Spatial Lag Model (SLM) and Spatial Error Model (SEM) with individual fixed effects

| | SLM | | SEM | |
|---|---|---|---|---|
| **Term** | **Estimate** | **Pr(>\|t\|)** | **Estimate** | **Pr(>\|t\|)** |
| **Spatial Coefficient** | | | | |
| lambda/rho | 0.526 | $<$ 2.2e-16 *** | 0.535 | $<$ 2.2e-16 *** |
| **Coefficients** | | | | |
| HEEQ | -0.002 | 0.013 * | -0.002 | 0.010 * |
| EKII | 0.003 | 0.022 * | 0.001 | 0.296 |
| HI | 0.112 | $<$ 2.2e-16 *** | 0.139 | $<$ 2.2e-16 *** |
| UR | 0.008 | 0.012 * | 0.016 | 3.50e-08 *** |
| GDP | -0.003 | 0.532 | 0.008 | 0.061 . |
| wHEEQ | 0.002 | 0.070 . | 0.001 | 0.413 |
| wEKII | -0.003 | 0.151 | -0.008 | 0.003 ** |
| wHI | 0.057 | 1.068e-06 *** | 0.163 | $<$ 2.2e-16 *** |
| wUR | 0.028 | 6.017e-15 *** | 0.046 | $<$ 2.2e-16 *** |
| wGDP | 0.028 | 5.259e-06 *** | 0.070 | $<$ 2.2e-16 *** |

# Impact Measures and Simulated p-values for the Spatial Lag Model (SLM)

| Variable | Direct | Indirect | Total | p-value (Total) |
|---|---|---|---|---|
| HEEQ | -0.0026 | -0.0025 | -0.0051 | 0.0174 |
| EKII | 0.0029 | 0.0028 | 0.0057 | 0.0369 |
| HI | 0.1194 | 0.1163 | 0.2357 | $< 2.2e\text{-}16$ |
| UR | 0.0083 | 0.0081 | 0.0164 | 0.0083 |
| GDP | -0.0029 | -0.0028 | -0.0056 | 0.4777 |
| wHEEQ | 0.0024 | 0.0023 | 0.0047 | 0.0640 |
| wEKII | -0.0033 | -0.0032 | -0.0064 | 0.1154 |
| wHI | 0.0606 | 0.0590 | 0.1196 | 9.37e-07 |
| wUR | 0.0295 | 0.0288 | 0.0583 | 1.13e-14 |
| wGDP | 0.0297 | 0.0289 | 0.0586 | 1.14e-05 |

# Diagnostics for Spatial Models



# Conclusion

**Robustness of Findings:**

- **Consistent Variables:** HEEQ, HI, and UR consistently show significant effects across all models.
- **Spatial Effects:** Both SLM and SEM highlight significant spatial dependencies, underscoring the importance of incorporating spatial econometric techniques.
- **Household Income (HI):** Emerges as a critical factor, with both direct and spatial spillover effects significantly influencing employment rates.

# Conclusion

**Policy Implications:**

- **Enhancing Education:** Addressing the negative impact of higher education enrollment on employment may involve aligning educational outcomes with market needs.
- **Supporting Knowledge-Intensive Industries:** Encouraging growth in these sectors can positively influence employment.
- **Boosting Household Income:** Policies aimed at increasing household income can have substantial direct and spillover benefits for regional employment.
- **Addressing Unemployment:** Targeted interventions are needed to mitigate structural unemployment and its regional impacts.

# References


- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers.
- Elhorst, J. P. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer-Verlag.
- Moran, P. A. P. (1950). "Notes on Continuous Stochastic Phenomena". In: *Biometrika* 37.1/2, pp. 17–23.
- Möller, J. (2010). "The German labor market response in the world recession – De-mystifying a miracle". In: *Zeitschrift für ArbeitsmarktForschung* 42.4, pp. 325–336.
- Popp, M. (2023). "How elastic is labor demand? A meta-analysis for the German labor market". In: *Journal for Labour Market Research* 57.1, p. 14.
- Rinne, U. and K. F. Zimmermann (2012). "Another economic miracle? The German labor market and the Great Recession". In: *IZA Journal of Labor Policy* 1, pp. 1–21.