Ukraine Brain Drain:

Exploring Factors and Methods

Allison Moore

STAT 692: Statistical Consulting

Texas A&M University

April 18, 2024

Table of Contents

Background

For the last 43 years, Client International Agency, Inc. (CIAI) has provided NATO leaders and decisions makers with high quality geopolitical analysis in key problem areas. With the recent increase in Russian aggression in Ukraine, CIAI has begun initial consideration of the Ukrainian brain drain problem. Originally coined in 1963 by the British Royal Society, *Brain Drain* is the emigration of scientists and other 'academic brains' (p. 363). Since 1963, social scientists have considered the implications of brain drain on the sending state (the emigrant's original country), with an overwhelming majority of case studies focusing on third world countries. However, little is known about what factors motivate Ukrainian emigrants to seek out specific areas within Europe.

This paper aims to provide CIAI personnel with a more wholistic understanding of the potential analytical methods and variables that should be used in their future research and analysis addressing Ukrainian emigration. Having a better understanding of the limitations and concerns associated with the Ukrainian brain drain use case will provide CIAI with the ability to assist decision makers more efficiently and with greater efficacy. Below, I will discuss the data and methodology used in this study, with the primary response variable being the number of Ukrainian emigrants, normalized by state population within Germany. Then I will provide key results and an expanded discussion of these results. Finally, I will conclude the paper with recommendations and future research opportunities that may be advantageous to CIAI.

The Data and Methodology

International Relations (IR) analysis typically takes place at any of three levels: the International System, the state (read as 'country'), or the individual level. To prevent confusion within this study, I will use 'country' to represent the state level of analysis and 'state' to refer to the individual states located inside of Germany. To gain the most insight about the aspects that impact and result from Ukrainian brain drain, I establish a concise, four-step methodology that spans across the system and country levels of analysis. Figure 1 summarizes the four steps used- which are discussed in detail below.

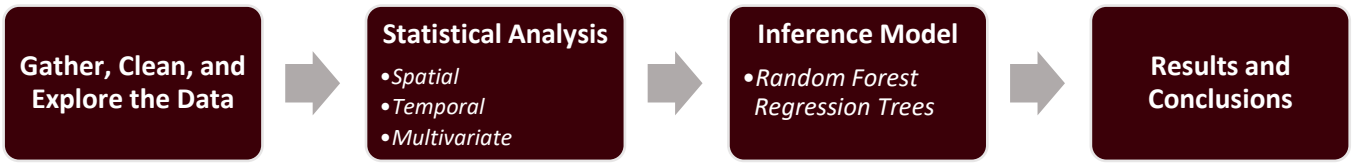


Figure 1. *The six-step methodology for this study. The first steps is the most important for ensuring that quality analysis can be conducted in steps 2 through 4. I implemented manual quality control steps throughout the study.*

Step 1: Gather, Clean, and Explore the Data. The first step revolves around identifying, acquiring, and understanding the data necessary for the analysis. Table 1 lists the variables used in the final analysis and acts as a quick reference tool as you review this report. Data at the country level is limited to 2005-2021 due to lack of data availability- or reliability- outside this range. After retrieving the data, I combined the variables into a machine-readable format (.csv) and checked the data for missing values. Appendix A contains the complete list of all variables that were used, and the methods used to modify values or fill in any missing values contained in the original data.

	Variable	Representation	Manipulation	Source
Europe	UkPop	Number of foreign-born Ukrainians in the specified country	Set Ukraine to 0	United Nations Statistics Division
Germany	State	The specific geographical states within Germany	Factorized	Federal Statistical Office Destatis
	Year	Year of measurement (2005-2021)	None	None
	CPI	Consumer Price Index	Per Capita normalization	Federal Statistical Office Destatis
	ICTPC	Information, Communication, and Technology industry value	Per Capita normalization	Federal Statistical Office Destatis
	UkPopPC	Number of foreign-born Ukrainians in the specified German state	Per Capita normalization	Federal and State Statistical Office
	PATPC	Number of patents for the specified German state and year	Per Capita normalization	Federal Statistical Office Destatis
	DeathT	Total sum of military and civilian deaths in Ukraine during conflict	Sum across all categories	Uppsala Universitet Department of Peace and Conflict
	EXAvg	Average number passing final degree exams	Average across all genders and nationalities	Federal Statistical Office Destatis
	LFEX	Average life expectancy of all German state citizens	Average across genders	Federal Statistical Office Destatis

Table 1. *The list of variables that were used in the final analysis. Variables will be referred to using their shorthand form in the 'Variable' column. Refer to Appendix A for the complete list of variables used throughout the study.*

Next, I conducted Exploratory Data Analysis (EDA), which is a process to understand the limitations of and the underlying relationships within the data collected. The EDA process also assists in identifying which independent variables should be included. I select Germany for country level analysis during this step. Appendix B details the EDA process, identified limitations, and describes how I mitigated these limitations in later analysis. Although most variables do not follow a normal distribution, I do not use any data transformations in the final analysis or inference model, as maintaining interpretability is important to CIAI. Python version 3.11.4 was used during the EDA process.

Step 2: Statistical Analysis. Following EDA, I conduct simple statistical analyses using R, version 4.3.1. This consists of spatial, temporal, and multivariate analyses. This step is extremely important because the presence of correlation within the data can lead to conclusions that are

inaccurate. Identifying and accounting for correlation is integral to the integrity of the results when precision is required.

Spatial analysis is done through the use of proximity matrices and Moran's I test for spatial autocorrelation, which are discussed in detail in Appendix C. Temporal analysis is done using autocorrelation functions (ACFs) for individual variables, and vector autoregression for multivariate temporal correlations. Temporal analysis is described in Appendix D. Lastly, multivariate analysis consists of Kruskal Wallis tests, backward step multivariate regression models, and backward step generalized least squares models which incorporate temporal autocorrelation. These are detailed in Appendix E.

Step 3: Inference Models. To provide CIAI with predictive capabilities, I utilized random forest regression trees. This type of inference model leverages numerous cross-validated iterations to reduce error and minimize variance in the final result. I selected random forest because tree-based inference models retain a high level of interpretability- which was noted to be of great importance to CIAI. Appendix F contains the full inference model results.

Step 4: Results and Conclusions. After all analysis and inference modeling has been summarized, I provide a discussion on my findings and recommendations. This report is intended to provide the most important "take aways." For detailed technical information regarding analysis and results, please refer to the appropriate appendices. Because this is an exploratory study, I used several statistical methods to analyze the data.

Compounding this aspect, within the IR domain, it is inherently difficult to identify causation. Therefore, many of the conclusions in this study are overarching and are geared towards identifying potential solutions for future analysis.

Results and Discussion

Throughout the analysis, I use a p-value of 0.05 as the primary demarcation to reject a null hypothesis. However, I highlight three levels of significance: 0.01, 0.05, and 0.1; this is to provide CIAI with the ability to assume more- or less- risk in the interpretations of this study.

Spatial Results. Spatial analysis was conducted for Europe and Germany using the two proximity matrices that were developed. This analysis used Moran's I test for spatial autocorrelation to test if a variable is representative of a spatial random process. Table 2 shows the most valuable information from this analysis. At the international level, Ukrainian population is highly significantly-positively correlated. This means that Ukrainian emigrants tend to cluster around the same areas within Europe. This clustering could potentially be the result of social network obligations (e.g., family, friends, religious groups, etc.), or cultural similarities present in the receiving states. The p-value for Germany's Ukrainian distribution is too large to reject the null hypothesis for Moran's I test; thus, I conclude that the distribution of Ukrainian emigrants across Germany is a random spatial process. Appendix C contains a detailed explanation of Moran's I test and an in-depth analysis of each variable used in the study.

	Region	Standard Deviate	p-value
UkPop	Europe	2.6400	0.0041 ***
UkPopPC	Germany	-0.0063	0.5025
PATPC	Germany	2.1348	0.0164 **
CPI	Germany	2.3263	0.0100***
ICTPC	Germany	0.9679	0.1665
EXAvg	Germany	1.3008	0.0967*
LFEX	Germany	2.1826	0.0145**

*Level of significance: *** 0.01 ** 0.05 *0.10*

Table 2. *Results for spatial autocorrelation of Ukrainian emigrants across Europe and Germany. The p-value for Europe is highly significant, which allows me to reject the null hypothesis and conclude that there is positive spatial correlation present. For Germany, the p-value is too large to reject the null, thus I conclude that the distribution of Ukrainians in Germany is a spatially random process. Of note, PATPC, CPI, and LFEX also indicate the presence of positive spatial correlation (clustering).*

Because there is spatial autocorrelation present in PATPC, CPI, and LFEX, I developed individual models at the state level. This removed the spatial component from the data all together and allowed me to provide more accurate models that are unique to the specific states within Germany. Next, I consider the temporal aspects of the data, as the data span the years 2005-2021.

Temporal Results. Data for the entirety of Europe was only used for spatial analysis and for guiding selection of an ideal country for the country level of analysis. Thus, temporal analysis was conducted only for the 16 states of Germany.

First, I plotted an autocorrelation function (ACF) for each state-variable combination. These are shown in Appendix D, Figure D-1. All state-variable combinations indicated either no temporal correlation, or an autoregressive (AR) type temporal correlation. To verify this visual inspection, each state-variable item was iteratively fit to AR models varying from order 0 to 5

(i.e., six different models were used from no temporal correlation to temporal dependence on the last 1-5 years). Table 3 shows the results from this iterative process.

	CPI	ICTPC	UkPopPC	PATPC	EXAvg	LFEX	DeathT
Baden Wurttemberg	**AR(2)**	**AR(2)**	**AR(4)**	*NONE*	*NONE*	*NONE*	*NONE*
Bayern	**AR(2)**	*NONE*	**AR(3)**	**AR(2)**	*NONE*	*NONE*	
Berlin	**AR(4)**	**AR(3)**	**AR(4)**	*NONE*	*NONE*	*NONE*	
Brandenburg	**AR(2)**	**AR(4)**	*NONE*	*NONE*	*NONE*	*NONE*	
Bremen	**AR(2)**	*NONE*	**AR(2)**	**AR(2)**	*NONE*	*NONE*	
Hamburg	**AR(4)**	*NONE*	*NONE*	*NONE*	*NONE*	*NONE*	
Hessen	**AR(3)**	*NONE*	*NONE*	**AR(2)**	*NONE*	*NONE*	
Mecklenburg Vorpommern	**AR(3)**	*NONE*	*NONE*	*NONE*	*NONE*	*NONE*	
Niedersachsen	**AR(3)**	**AR(2)**	**AR(3)**	*NONE*	*NONE*	*NONE*	
Nordrhein Westfalen	**AR(2)**	**AR(2)**	*NONE*	*NONE*	*NONE*	*NONE*	
Rheinland Pfalz	**AR(2)**	**AR(2)**	**AR(2)**	*NONE*	*NONE*	*NONE*	
Saarland	**AR(2)**	*NONE*	**AR(2)**	*NONE*	*NONE*	*NONE*	
Sachsen	**AR(2)**	**AR(2)**	*NONE*	*NONE*	*NONE*	*NONE*	
Sachsen Anhalt	**AR(2)**	**AR(3)**	**AR(2)**	*NONE*	*NONE*	*NONE*	
Schleswig Holstein	**AR(4)**	*NONE*	**AR(2)**	*NONE*	*NONE*	**AR(3)**	
Thüringen	**AR(2)**	*NONE*	**AR(2)**	*NONE*	*NONE*	*NONE*	

Table 3. *Consolidated autoregressive model fits. The best AR model was selected for each state-variable by minimum Akaike information criteria (AIC). Note that EXAvg and DeathT exhibit no temporal autocorrelation, while LFEX and PATPC have only a few states with temporal autocorrelation. CPI is the only variable with all states indicating temporal autocorrelation.*

Next, I considered bivariate temporal relationships between the response variable, UkPopPC, and each independent variable using models like equation 1 below, where t is the current year, const is a constant, and variable is the selected independent variable.

$$UkPopPC_t = const + UkPopPC_{t-1} + variable_{t-1} \quad (1)$$

All independent variables included at least one state in which the selected independent variable granger causes UkPopPC[1]. This indicates that there is temporal correlation present amongst the variables (not just within a variable). When accuracy is important, this correlation

[1] Granger Cause: When one time series can be used to predict another time series. This means that the current year value for UkPopPC is dependent upon the previous year's measurement for the selected independent variable.

must be accounted for. To further explore the intervariable effects, I explored multivariate methods.

Multivariate Results. Multivariate analysis encompasses uncovering how multiple variables interact with one another. The Kruskal Wallis (KW) test was used to compare how each individual German state influences the uniqueness of each of the other variables. This results in the declaration of which states are statistically the same across each variable. The null hypothesis for this test is that the two states share the same distribution for the specified variable. The most noteworthy results from this test were that EXAvg and CPI are statistically the same across all states. All remaining variables, including the response variable – UkPopPC, reject the null hypothesis in about 40% of the cases. This furthers the idea that constructing state level inference models will provide more accurate predictions than a single model for all of Germany. Appendix E provides the tabulated results for the KW test.

Following the KW test, I fit two distinct types of multivariate models: a simple linear model and a generalized least squares (GLS) model which accounts for a temporal AR correlation at a lag of one year[2]. I used backward step to find the minimum Akaike information criteria (AIC) for both model types[3]. Table 4 provides the results from these two models. The GLS model selected only a single variable of importance for all states, while the simple linear model identified numerous unique variable combinations across each state. The Shapiro Residual p-value column indicates whether the Shapiro- Wilk normality test indicated that the

[2] These models were developed using R's lm() and gls(correlation = ARMA(p=1)) functions.
[3] AIC: Akaike information criteria. A relative quality of measurement estimator commonly used for comparing statistical models. Backward step is an iterative process which involves removing variables from a model to find the model with the minimum AIC.

residuals were statistically considered to be white noise[4]. Those p-values marked as significant reject the null hypotheses of normality and are considered different from white noise (i.e., the model is a poor representation for the data).

	Simple Linear Model	Shapiro Residual p-val	GLS, ARMA(1,0)	Shapiro Residual p-val
Baden Wurttemberg	CPI + ICTPC + PATPC + LFEX	0.1905	CPI	0.9987
Bayern	CPI + ICTPC + PATPC + EXAvg + DeathT	0.3791	PATPC	0.0012***
Berlin	ICTPC + PATPC + LFEX	0.2839	PATPC	0.4003
Brandenburg	ICTPC + LFEX	0.0014***	PATPC	0.0001***
Bremen	PATPC + LFEX	0.0855*	PATPC	0.6708
Hamburg	EXAvg + DeathT + LFEX	0.6680	PATPC	0.3946
Hessen	CPI + ICTPC + PATPC + LFEX	0.7161	PATPC	0.8882
Mecklenburg Vorpommern	ICTPC + EXAvg	0.0701*	PATPC	0.0003***
Niedersachsen	ICTPC + PATPC + LFEX	0.3512	PATPC	0.6797
Nordrhein Westfalen	ICTPC + PATPC + LFEX	0.5237	PATPC	0.2445
Rheinland Pfalz	ICTPC + LFEX	0.0081***	PATPC	0.7701
Saarland	CPI + ICTPC + PATPC + EXAvg	0.6467	PATPC	0.3181
Sachsen	CPI + EXAvg	0.0147**	PATPC	0.0000***
Sachsen Anhalt	CPI + ICTPC + EXAvg + LFEX	0.3010	PATPC	0.6671
Schleswig Holstein	ICTPC + EXAvg + LFEX	0.8336	PATPC	0.3088
Thüringen	CPI + ICTPC + EXAvg + LFEX	0.6264	PATPC	0.7158

*Level of significance: *** 0.01 ** 0.05 *0.10*

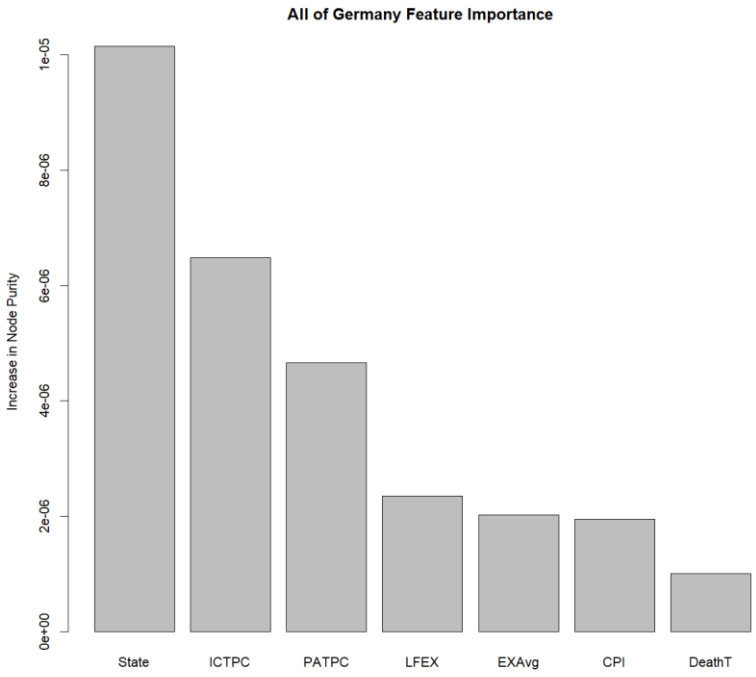
Table 4 *Comparison of the best linear model and the best generalized least squares (GLS), ARMA(1,0) model. The table shows the identified important variables along with the residuals Shapiro Wilk p-value. Most notably, the GLS model selects only a single variable in all cases. The simple linear model appears to provide more normal residuals; however, this is likely because the simple linear models do not account for the correlations present in the data.*

Although the simple linear models appear to be good fits at the 95% confidence level (only three states have non-normal residuals, versus four states for GLS), it is important to remember that these linear models do not account for the multivariate temporal correlation identified in the previous section (and Appendix D). Therefore, the GLS models are more

[4] Shapiro-Wilk test: A test of data normality. A model that is a 'good' fit for the data will have normally distributed residuals. Residuals are the differences between the model's prediction and the actual values. Residuals that are normally distributed are commonly referred to as 'white noise.'

accurate because they account for the granger causality that exists across most variables and the response variable.

Inference Models. To verify the theory that fitting individual models for each state is necessary, I first fit a cross validated random forest to all data across Germany. Figure 1 shows that State is the most important variable for increasing node purity. It also shows that there is substantial reduction in mean average error (MAE) with the transition to individual state models.



	MAE
Germany	1.501E-04
Berlin	1.184E-04
Mecklenburg Vorpommern	8.917E-05
Thuringen	6.679E-05
Sachsen	6.283E-05
Brandenburg	3.929E-05
Baden Wurttemberg	3.871E-05
Bayern	3.814E-05
Bremen	3.575E-05
Saarland	3.161E-05
Niedersachsen	3.114E-05
Sachsen Anhalt	2.920E-05
Schleswig Holstein	2.640E-05
Hamburg	2.448E-05
Rheinland Pfalz	1.804E-05
Hessen	1.560E-05
Nordrhein Westfalen	1.385E-05

Figure 1 *Feature importance for all of Germany random forest model (left) and all random forests' mean absolute errors, listed from largest error to smallest error (right). The model for the entirety of Germany indicates that State is the most important variable for increasing node purity and there is marked reduction in model error when fitting individual state models.*

Considering the individual state models, CPI, ICTPC, and PATPC rank among the most frequent important variables. The most important features for each state inference model and the corresponding model learning curves can be found in Appendix F.

Expanded Discussion. Each step of the analysis conducted further indicates that developing individual state-based inference models is critical to reducing the effects of correlation and minimizing inference error. Segregating the states completely removed the spatial correlation component and left only the temporal correlation, which was accounted for by using GLS models with temporal autocorrelation considerations. In this analysis, a generic AR(1) was used as the correlation component in the GLS models. For more accurate analysis in the future, CIAI personnel should consider fitting state-specific AR correlations within their GLS models.

There appears to be a clear relationship between Ukrainian populations across Germany and the number of patents issued in previous years. It is unknown whether the number of patents entices Ukrainian brains, or if the change in number of patents results from Ukrainian brains; however, this is a relationship that deserves additional study and may benefit from narrative or survey data.

The most unexpected finding from this research is that the number of military and civilian deaths in Ukraine during conflict between 2005-2021 does not appear to impact the number of Ukrainian emigrants arriving to most states of Germany. However, it is important to retain this variable in future analysis because, as data becomes available for 2021-2024, this finding may be invalidated.

Conclusion

This research applied diverse statistical methods to a variety of variables within the country of Germany with the goal of exploring methods that can be applied in future CIAI analysis across Europe. This study is limited in scope to the exploration of data related to

Ukrainian brain drain; and, due to the breadth of IR problem sets, it is often impossible to demonstrate causation. But, in this case, there is much to be gathered from correlation and inference. Between spatial, temporal, and multivariate analysis, several relationships were uncovered that warrant follow on analysis with more direct attention.

In this study, I found that Ukrainians tend to cluster in Northern/Northwestern Europe. I also found that the number of Ukrainian immigrants in German states maintains a strong relationship to number of patents when considering value changes across time. Information, Communication, and Technology industry value and Consumer Price Index also provided strong temporal relationships to Ukrainian immigrants but was identified as less important than number of patents for most German states. The number of deaths in Ukraine during periods of conflict did not impact number of Ukrainian immigrants in Germany in a statistically significant way.

The concept of developing individual models for each state is a key finding from this study that will likely provide CIAI with the most success moving forward. I recommend first developing a country analysis priority list, then conducting state/ regional level analysis within each country, leveraging GLS models when temporal correlation is present. All variables used in this study were important to specific states. Thus, I also recommend utilizing similar variables for other countries.

Future research is needed in two key areas. First, the factors leading to Ukrainian emigrant clustering within Europe. Potential additional variables to include at the system level are country democratic status, cultural similarity index (such as Dr. Roose's index), digital access index, and number of academic articles published. The second area revolves around the

qualitative analysis of Ukrainian emigration at all levels of analysis. I recommend pairing this qualitative analysis with quantitative methods- beginning with Germany is a natural starting point. Both areas will benefit from Ukrainian brain survey data. Narrative data may help CIAI to understand the causation behind the clustering and correlation uncovered in this work. And, understanding the factors that motivate Ukrainian brains to settle in specific regions around Europe will help CIAI personnel to provide more accurate advisement to NATO leaders.

Bibliography

Oldfield, R. C., Simmons, J. A., Jeffery, J. W., Cooper, W. M., Eden, R. J., Jones, G. O., Kondic, V., McMichael, J., Pike, E. R., Andrews, K. W., Bragg, W. L., Bagguley, D. M. S., Baker, J. M., Cooke, A. H., Elliott, R. J., Griffiths, J. H. E., ter Háar, D., Hatton, J., Hill, R. W., … Holliday, P. (1963). The Emigration of Scientists from the United Kingdom. *Minerva*, *1*(3), 358–380. http://www.jstor.org/stable/41821578

Federal and State Statistical Office. Specialist series / 1 / 2 <1982 - 2021>. Results of the Central Register of Foreigners. Wiesbaden. https://www.statistischebibliothek.de/mir/receive/DESerie_mods_00000018

Federal Statistical Office Destatis. Genesis- Online. *Database of the Federal Statistical Office of Germany.* https://www-genesis.destatis.de/genesis/online/data?operation=sprachwechsel&language=en

Roose, J. (2012). Cultural Similarity in Europe according to the Index on Cultural Similarity. https://userpage.fu-berlin.de/~jroose/index_en/main_indexvalues.htm#:~:text=The%20Index%20on%20Cultural%20Similarity%20shows%20how%20culutrally%20similar%20the,to%200%20for%20maximum%20dissimilarity.

UC Davis Geodata. (2022). Gadm41_ABW.gpkg. https://geodata.ucdavis.edu/gadm/gadm4.1/gpkg/

United Nations Statistics Division. (2024). Foreign-Born Population by Country/Area of Birth, Age and Sex. Updated 23 February 2024. https://data.un.org/Data.aspx?q=foreign+population&d=POP&f=tableCode%3a44

Uppsala Universitet Department of Peace and Conflict. (2024). Uppsala Conflict Data Program (UCDP): Ukraine. https://ucdp.uu.se/country/369

Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K. Wilke, C., Woo, K., Yuntani, H., Dunnington, D., and van den Brand, T. Create a Data Frame of Map Data. ggplot2. https://ggplot2.tidyverse.org/reference/map_data.html

Appendix A: List of Variables

This appendix provides a comprehensive list of all variables considered and any manipulation or imputation used.

	Variable	Representation	Missing Values	Manipulation	#Obs.	Source
Europe	UkPop	Number of foreign-born Ukrainians in the specified country	15 set to 0	Set Ukraine to 0	39	United Nations Statistics Division
	Latitude	Latitude of country borders	None	None	13909	Wickam, et al.
	Longitude	Longitude of country borders	None	None	13909	Wickam, et al.
Germany	State	The specific geographical states within Germany.	None	Factorized	272	Federal Statistical Office Destatis
	Latitude	Latitude of state borders	None	None	11302	UC Davis Geodata
	Longitude	Longitude of state borders	None	None	11302	US Davis Geodata
	Year	Year of measurement (2005-2021)	None	None	272	
	CPI	Consumer Price Index	20 filled by extrapolation	Per Capita normalization	272	Federal Statistical Office Destatis
	ICTPC	Information, Communication, and Technology industry value	None	Per Capita normalization	252	Federal Statistical Office Destatis
	UkPopPC	Number of foreign-born Ukrainians in the specified German state	None	Per Capita normalization	272	Federal and State Statistical Office
	PATPC	Number of patents for the specified German state and year	None	Per Capita normalization	272	Federal Statistical Office Destatis
	DeathA	Number of receiving state deaths (Ukrainian Military)	None	Summarized to mitigate correlation	2990	Uppsala Universitet Department of Peace and Conflict
	DeathB	Number of attacking state deaths	None	Summarized to mitigate correlation	2990	Uppsala Universitet Department of Peace and Conflict
	DeathC	Number of civilian Ukrainian deaths	None	Summarized to mitigate correlation	2990	Uppsala Universitet Department of Peace and Conflict
	DeathT	Total sum of DeathA, DeathB, and DeathC	None	Sum across all categories	2990	Uppsala Universitet Department of Peace and Conflict
	EXAvg	Average number passing final degree exams	None	Average across all genders and nationalities	272	Federal Statistical Office Destatis
	LFEXM	Life expectancy of men	None	Summarized to mitigate correlation	272	Federal Statistical Office Destatis
	LFEXW	Life expectancy of women	None	Summarized to mitigate correlation	272	Federal Statistical Office Destatis
	LFEX	Average life expectancy of all German state citizens	None	Average across genders	272	Federal Statistical Office Destatis
	Rent	Cost of rent	None	Dropped due to correlation with and similarity to CPI	272	Federal Statistical Office Destatis
	GDP	Gross Domestic Product	None	Dropped due to correlation with and similarity to ICTPC	272	Federal Statistical Office Destatis

Variable dropped from analysis. Variable created by summarization.

Appendix B: Exploratory Data Analysis

Appendix B is maintained in a separate file (Appendix_B.ipynb) and is available upon request.

Appendix C: Spatial Analysis

This appendix consolidates the spatial analysis performed for all three levels of analysis (International System, State, and Individual). For the International system, data was collected on 31 countries. At the country level, data was collected for all 16 states in Germany.

Developing the Proximity Matrices. A proximity matrix (also known as a neighbor matrix) was created for both Europe and Germany. Figure C-1 shows the Germany proximity matrix, a 16x16 binary matrix. '1' indicates that the state listed in the corresponding row shares a border with the state in the corresponding column. '0' indicates that the two states do not share a border. The Europe proximity matrix, a 39x39 matrix, is not shown due to its size.

	Baden Wurttemberg	Bayern	Berlin	Brandenburg	Bremen	Hamburg	Hessen	Mecklenburg Vorpommern	Niedersachsen	Nordrhein Westfalen	Rheinland Pfalz	Saarland	Sachsen	Sachsen Anhalt	Schleswig Holstein	Thuringen
Baden Wurttemberg	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0
Bayern	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1
Berlin	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Brandenburg	0	0	1	0	0	0	0	1	1	0	0	0	1	1	0	0
Bremen	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Hamburg	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0
Hessen	1	1	0	0	0	0	0	0	1	1	1	0	0	0	0	1
Mecklenburg	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0
Niedersachsen	0	0	0	1	1	1	1	1	0	1	0	0	0	1	1	1
Nordrhein Westfalen	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0
Rheinland Pfalz	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0
Saarland	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Sachsen	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1
Sachsen Anhalt	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1
Schleswig Holstein	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0
Thuringen	0	1	0	0	0	0	1	0	1	0	0	0	1	1	0	0

Figure C-1. *The German state proximity matrix. This matrix is used in the test for spatial autocorrelation (Moran's I test). Note that each state will contain a '0' when comparing to itself (i.e., a state is considered to NOT share a border with itself).*

Spatial Analysis of the International System. After the proximity matrices are developed, I conduct a geospatial distribution analysis of Ukrainian emigrants across Europe. Figure C-2 shows a clear concentration of Ukrainians in Northern/Western Europe and a lack of data for several states. Because we want to see the pure distribution of Ukrainian emigrants, I do not

normalize the data in the per capita method used for the state analysis. This allows us to view the raw spatial distribution of our variable.

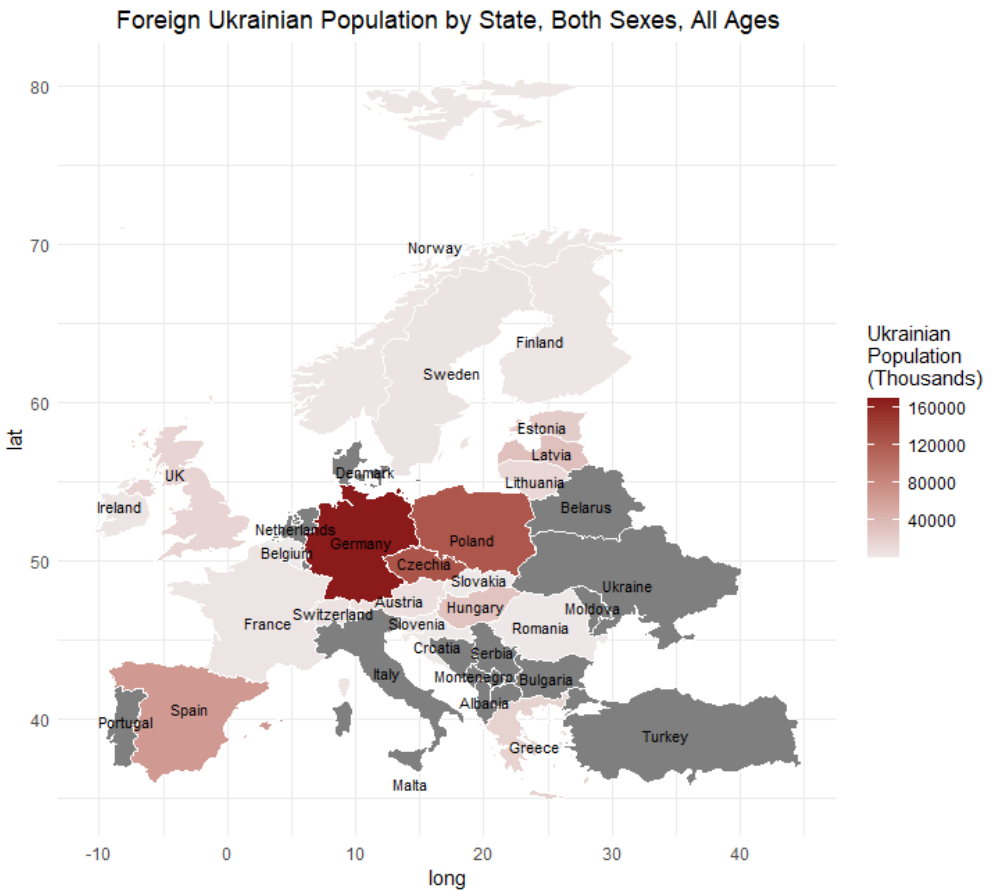


Figure C-2. *Spatial distribution of Ukrainians in Europe. The data used is the most recent measurement available for each state, which is inconsistent and ranges from 2001-2021. Several states did not have any data available. These states are shown in grey. I also exclude Ukraine from this analysis; thus, it is also shown in grey.*

Figure C-3 is a surface plot of this data. The surface plot is a visualization of the data after a thin plate spline regression (TPSR) has been applied to smooth the state boundaries. The TPSR surface plot suggests the distribution of Ukrainians is highly correlated, with a large concentration in North-Central Europe.

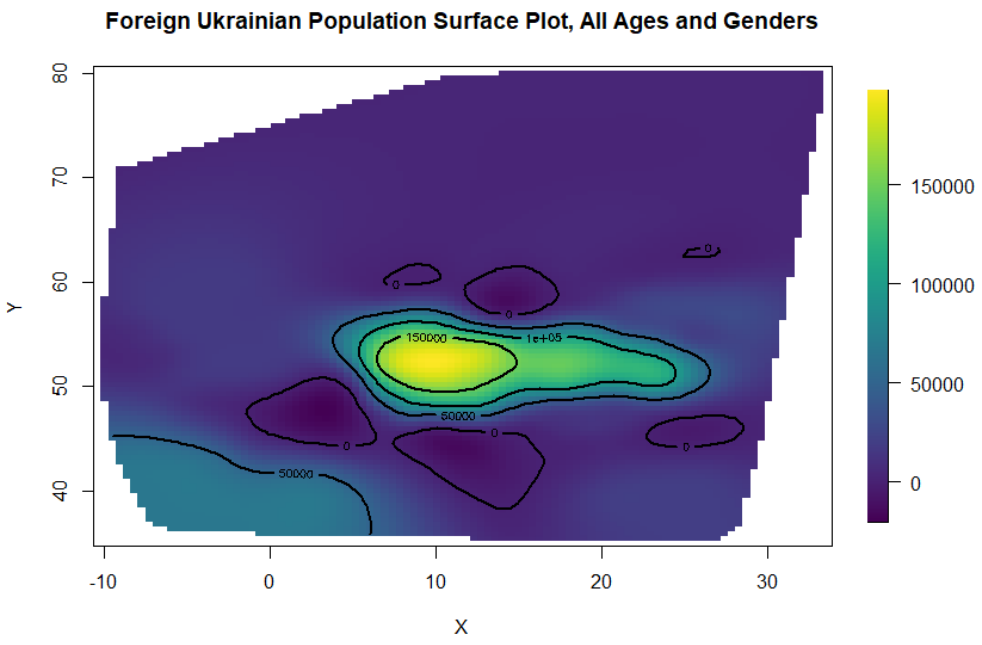


Figure C-3. *Thin plate spline surface of Ukrainian population in European states. This plot suggests the distribution of Ukrainians across Europe is not random and is consolidated heavily around Germany. This is evident through the clustering of colors. Missing values were removed for this analysis.*

To confirm my visual interpretation of the surface plot, I conduct Moran's I test for spatial autocorrelation. This test is a spatial statistical test that checks data for the presence of spatial autocorrelation; and it requires the use of the proximity matrix. The null and alternative hypotheses for this test are displayed below.

$$H_0: The\ variable\ is\ randomly\ distributed\ across\ space$$

$$H_A: The\ variable\ is\ not\ a\ random\ spatial\ processes$$

The result of the test is the p-value 0.004145 and the standard deviate 2.64. This p-value allows me to reject the null hypothesis and conclude there is a presence of spatial correlation, being 99.6% confident. The positive standard deviate shows that this correlation is

positive spatial correlation (i.e., we will tend to have high values surrounding high values and low values surrounding low values). This confirms my visual surface plot check.

Because I use analysis of the International System to identify an ideal state for deeper analysis, I only consider the response variable's spatial distribution. As there is significant clustering around Germany, I opt to use Germany as the country for further analysis.

Spatial Analysis of Country Level. Like the international level of analysis, I begin by considering the geospatial distribution of Ukrainian emigrants across German states. Figure C-4 shows high concentrations of Ukrainians in the darker shaded regions. Now that I am considering the country level, I normalize the Ukrainian population values by each state's total population, to get a per capita or percent of total population value. Visually, the Ukrainian population in each state appears to be random.

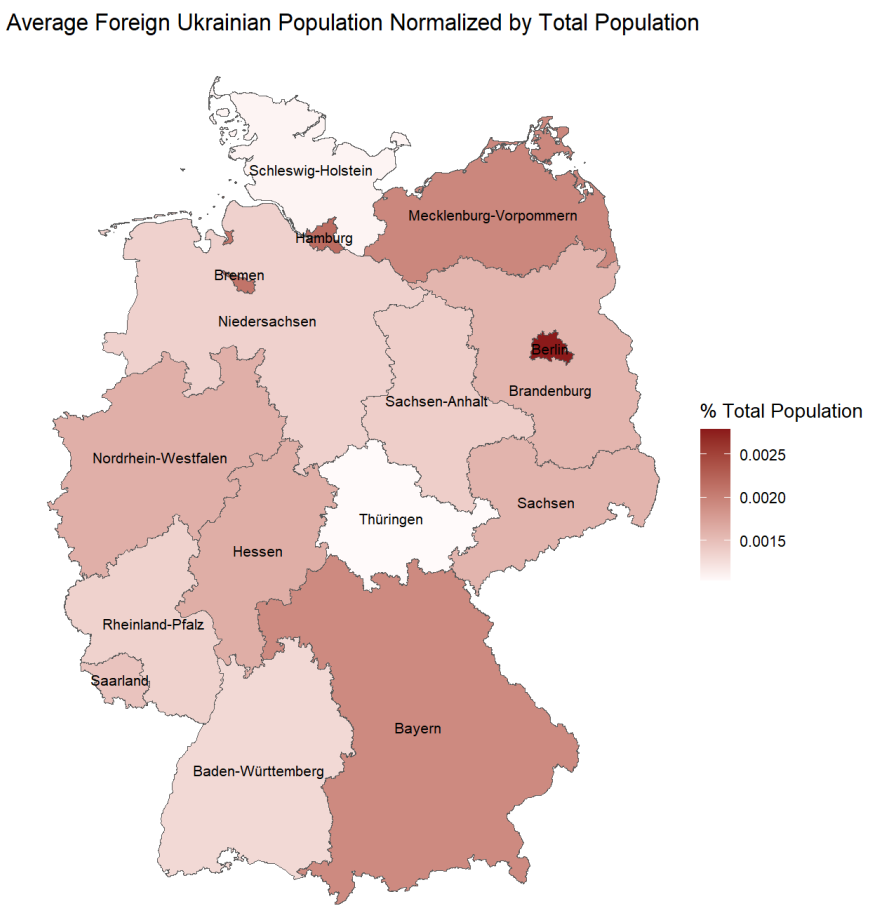


Figure C-4. *Spatial distribution of Ukrainian population across German states. Ukrainian population was normalized by total state population; representation is shown as percent of state population. Ukrainian population appears to be fairly randomly distributed.*

To confirm this visual inspection, I conduct Moran's I test with the same null and alternative hypotheses listed in Equations 1 and 2 in the previous section. Table C-1 lists the Moran's I test for spatial autocorrelation values. Because the data is also temporal, I consider each year, as well as the overall temporal average. From these values, I cannot reject the null hypothesis and conclude that the variable is a random spatial process.

YEAR	STANDARD DEVIATE	P-VALUE
2005	0.10209	0.4593
2006	0.10052	0.4600
2007	0.05748	0.4771
2008	0.0989	0.4606
2009	0.076094	0.4697
2010	0.10692	0.4574
2011	0.14702	0.4416
2012	0.16784	0.4334
2013	0.14605	0.4419
2014	-0.13413	0.5533
2015	-0.49515	0.6898
2016	-0.40685	0.6579
2017	-0.1705	0.5677
2018	-0.03712	0.5148
2019	0.067637	0.4730
2020	0.25026	0.4012
2021	0.58198	0.2803
AVERAGED ACROSS ALL YEARS	-0.0063	0.5025

Table C-1. *Moran's I Test for spatial autocorrelation. All results indicate that Ukrainian population is likely representative of a random spatial process and does not exhibit autocorrelative behaviors. An interesting feature here is the shift from positive to negative, back to positive standard deviate values.*

Table C-2 contains the Moran's I test results for the independent variables that were selected during EDA for use in the final analysis. Spatial correlation is evident in PATPC (number of patents per capita), CPI (Consumer Price Index), and LFEX (average life expectancy). ICTPC (Information, Communication, and Technology value per capita) and EXAvg (average university final exam pass rate) generally do not demonstrate spatial correlation.

YEAR	PATPC		CPI		ICTPC		EXAvg		LFEX	
	Std. Dev.	p-value	Std. Dev.	p-value	Std. Dev.	p-value	Std. Dev.	p-value	Std. Dev.	p-value
2005	3.4103	0.0003	2.6764	0.0037	0.94551	0.1722	1.0587	0.1449	1.7974	0.0361
2006	2.7488	0.0030	2.7232	0.0032	0.93585	0.1747	0.63358	0.2632	1.9115	0.0280
2007	2.5674	0.0051	2.6646	0.0039	0.94268	0.1729	1.3568	0.0874	1.9063	0.0283
2008	2.1946	0.0141	2.5271	0.0058	1.0155	0.1549	0.72743	0.2335	1.9267	0.0270
2009	2.3495	0.0094	2.4076	0.0080	1.0342	0.1505	0.96757	0.1666	2.1326	0.0165
2010	2.3495	0.0094	2.2508	0.0122	1.0588	0.1448	0.53098	0.2977	2.175	0.0148
2011	2.1658	0.0152	2.1637	0.0152	1.0262	0.1524	-1.0907	0.8623	2.3706	0.0089
2012	2.2969	0.0108	2.1721	0.0149	1.0107	0.1561	0.14735	0.4414	2.4101	0.0080
2013	2.1494	0.0158	1.8933	0.0292	0.95226	0.1705	-0.10786	0.5429	2.3451	0.0095
2014	1.9828	0.0237	1.6643	0.0480	0.98944	0.1612	0.18279	0.4275	2.3018	0.0107
2015	1.7306	0.0418	-0.33833	0.6324	0.94625	0.1720	-0.00943	0.5038	2.2319	0.0128
2016	1.7231	0.0424	0.34095	0.3666	0.91047	0.1813	1.028	0.1520	2.2553	0.0121
2017	1.7436	0.0406	0.65529	0.2561	0.94292	0.1729	1.5471	0.0609	2.1309	0.0166
2018	1.5198	0.0643	0.20477	0.4189	0.94699	0.1718	1.526	0.0635	2.2375	0.0126
2019	1.4153	0.0785	-0.33832	0.6324	0.92156	0.1784	0.61488	0.2693	2.25	0.0122
2020	1.6462	0.0499	no variation		0.94835	0.1715	0.97967	0.1636	2.1173	0.0171
2021	1.7896	0.0368	0.55106	0.2908	0.92786	0.1767	1.8315	0.0335	2.2827	0.0112
Average	2.1348	0.0164	2.3263	0.0100	0.9679	0.1665	1.3008	0.0967	2.1826	0.0145

Significance at 0.01 0.05 0.1

Table C-2. *Moran's I Test for spatial autocorrelation in all independent variables. Note the legend in the bottom of the table indicates the level of significance.*

Appendix D: Temporal Analysis

This appendix consolidates all temporal analysis performed. Because I only used spatial analysis at the International System level of analysis, this appendix only contains results at the Country level.

Temporal analysis considers how a single variable is related to itself over time. To uncover this relationship, autocorrelation functions (ACFs) are used. All ACF plots are provided at the end of this appendix in Figure D-1. ACFs can be plotted to visually reveal a variable's underlying relationship with its past self and can aid in identifying if the variable tends to adhere to an autoregressive (AR) model or a moving average (MA) model. There are added complexities when a variable demonstrates behaviors of both types of models; however, we are fortunate that all of our variables exhibit classic autoregressive behaviors, or do not appear to have a temporal relationship at all.

	CPI	ICTPC	UkPopPC	PATPC	EXAvg	LFEX	DeathT
Baden Wurttemberg	**AR(2)**	**AR(2)**	**AR(4)**	*NONE*	*NONE*	*NONE*	*NONE*
Bayern	**AR(2)**	*NONE*	**AR(3)**	**AR(2)**	*NONE*	*NONE*	
Berlin	**AR(4)**	**AR(3)**	**AR(4)**	*NONE*	*NONE*	*NONE*	
Brandenburg	**AR(2)**	**AR(4)**	*NONE*	*NONE*	*NONE*	*NONE*	
Bremen	**AR(2)**	*NONE*	**AR(2)**	**AR(2)**	*NONE*	*NONE*	
Hamburg	**AR(4)**	*NONE*	*NONE*	*NONE*	*NONE*	*NONE*	
Hessen	**AR(3)**	*NONE*	*NONE*	**AR(2)**	*NONE*	*NONE*	
Mecklenburg Vorpommern	**AR(3)**	*NONE*	*NONE*	*NONE*	*NONE*	*NONE*	
Niedersachsen	**AR(3)**	**AR(2)**	**AR(3)**	*NONE*	*NONE*	*NONE*	
Nordrhein Westfalen	**AR(2)**	**AR(2)**	*NONE*	*NONE*	*NONE*	*NONE*	
Rheinland Pfalz	**AR(2)**	**AR(2)**	**AR(2)**	*NONE*	*NONE*	*NONE*	
Saarland	**AR(2)**	*NONE*	**AR(2)**	*NONE*	*NONE*	*NONE*	
Sachsen	**AR(2)**	**AR(2)**	*NONE*	*NONE*	*NONE*	*NONE*	
Sachsen Anhalt	**AR(2)**	**AR(3)**	**AR(2)**	*NONE*	*NONE*	*NONE*	
Schleswig Holstein	**AR(4)**	*NONE*	**AR(2)**	*NONE*	*NONE*	**AR(3)**	
Thuringen	**AR(2)**	*NONE*	**AR(2)**	*NONE*	*NONE*	*NONE*	

Table D-1. *Fitted time series models for each state across all variables. Of note, DeathT does not vary by state as it is the total number of conflict deaths in Ukraine.*
**AR(X) = Autoregressive model of order X*

Knowing this, I iterate over multiple orders of AR models. The results of this iterative process are listed in Table D-1. An order of an AR model indicates how many time units in the past the current value is related to. For example, in Berlin, the Consumer Price Index (CPI) for the current year is related to the past 4 years' values of CPI. The minimum Akaike information criteria (AIC) was used for selection.

The models discussed in Table D-1 account for the autocorrelation of a single variable over time; however, the response variable may also have interactions with independent variables throughout time. To check for this, we fit models that use "lagging." Lagging a variable shifts the position of each measurement over time. For example, to consider how the response variable, UkPopPC, relates to ICTPC over time, we use vector autoregression (VAR) models D1 and D2 to account for both autocorrelation and multivariate lagged relationships.

$$UkPopPC_t = const + UkPopPC_{t-1} + ICTPC_{t-1} \quad \text{(D1)}$$

$$UkPopPC_t = const + UkPopPC_{t-1} + UkPopPC_{t-2} + ICTPC_{t-1}\, ICTPC_{t-2} \quad \text{(D2)}$$

Here, const represents a constant and t represents the year of measurement. Model D1 captures the relationships between UkPopPC and ICTPC lagged up to one year, while Model D2 captures the relationship lagged over two years. When one time series can be used to predict another time series, it is said to 'Granger' cause the other time series. Table D-2 contains the results from fitting the VAR models D1 and D2, which includes both Granger causality and instantaneous p-values.

LAG 1

STATE	ICTPC		CPI		PATPC		EXAvg		DeathT		LFEX	
	Granger p	Instant p	Granger p	Instant p	Granger p	Instant p	Granger p	Instant p	Granger p	Instant p	Granger p	Instant p
Baden Wurttemberg	0.92478	0.14958	0.59929	0.06527	0.50867	0.25743	0.70598	0.92726	0.81968	0.22373	0.30554	0.09513
Bayern	0.04273	0.13364	0.01929	0.86616	0.69150	0.85580	0.55204	0.95862	0.00713	0.01718	0.09052	0.17738
Berlin	0.15855	0.04803	0.07933	0.83835	0.71168	0.01762	0.47341	0.78213	0.21518	0.08379	0.15841	0.20529
Brandenburg	0.00053	0.02450	0.00005	0.03812	0.03299	0.02914	0.36704	0.20008	0.42373	0.45894	0.00080	0.00599
Bremen	0.15503	0.03286	0.00809	0.09052	0.00725	0.10631	0.85234	0.58036	0.80722	0.39527	0.37995	0.01820
Hamburg	0.55938	0.21903	0.53851	0.36951	0.24328	0.87078	0.57884	0.19482	0.13931	0.44618	0.95042	0.42324
Hessen	0.00059	0.28144	0.00009	0.70593	0.00009	0.40620	0.74889	0.57583	0.91907	0.47315	0.00106	0.08638
Mecklenburg Vorpommern	0.31403	0.06313	0.04658	0.78460	0.88179	0.13492	0.13434	0.82770	0.00132	0.02052	0.07733	0.14602
Niedersachsen	0.00004	0.02764	0.00001	0.20107	0.00415	0.32844	0.91361	0.52086	0.82234	0.22192	0.00005	0.03382
Nordrhein Westfalen	0.00647	0.02781	0.01134	0.05115	0.11090	0.06574	0.67382	0.49592	0.22863	0.79673	0.05881	0.03539
Rheinland Pfalz	0.00005	0.44479	0.00006	0.94837	0.04350	0.04819	0.69259	0.76599	0.78167	0.40440	0.00921	0.16558
Saarland	0.10300	0.74428	0.35314	0.44474	0.99997	0.24131	0.94032	0.06178	0.25558	0.30003	0.00760	0.69364
Sachsen	0.41850	0.06434	0.36176	0.02323	0.40297	0.79398	0.68070	0.07364	0.54498	0.85137	0.02550	0.01155
Sachsen Anhalt	0.00001	0.64325	0.00000	0.57405	0.00009	0.03088	0.37643	0.59865	0.46791	0.63837	0.00371	0.16166
Schleswig Holstein	0.00646	0.26877	0.00338	0.62319	0.05093	0.13368	0.21605	0.60430	0.64907	0.26695	0.00669	0.62586
Thuringen	0.09900	0.10673	0.17569	0.14937	0.25996	0.05382	0.00709	0.83859	0.82427	0.29986	0.53561	0.08244

LAG 2

STATE	ICTPC		CPI		PATPC		EXAvg		DeathT		LFEX	
	Granger p	Instant p	Granger p	Instant p	Granger p	Instant p	Granger p	Instant p	Granger p	Instant p	Granger p	Instant p
Baden Wurttemberg	0.91783	0.16133	0.54312	0.05780	0.19925	0.15124	0.90822	0.90506	0.95287	0.54824	0.09234	0.08263
Bayern	0.00847	0.25174	0.03533	0.31157	0.93052	0.75642	0.32605	0.13678	0.03717	0.01887	0.04317	0.29609
Berlin	0.32019	0.06049	0.07250	0.98293	0.34238	0.02403	0.72555	0.62335	0.26774	0.11673	0.06483	0.57320
Brandenburg	0.02030	0.06220	0.00484	0.03219	0.42699	0.23488	0.56833	0.32420	0.70005	0.11412	0.10882	0.00774
Bremen	0.19306	0.02122	0.03335	0.15122	0.02585	0.29380	0.82096	0.59001	0.61583	0.66694	0.00342	0.23480
Hamburg	0.36837	0.56819	0.26898	0.17598	0.25104	0.31560	0.96884	0.22510	0.54602	0.59788	0.01165	0.24593
Hessen	0.00078	0.52016	0.02425	0.55794	0.01023	0.23722	0.70232	0.50945	0.76402	0.20604	0.00588	0.50609
Mecklenburg Vorpommern	0.23641	0.02030	0.07318	0.30264	0.23274	0.18622	0.50868	0.25550	0.01228	0.02503	0.07893	0.20840
Niedersachsen	0.00443	0.06265	0.00413	0.29215	0.04937	0.20976	0.61807	0.30371	0.48666	0.11574	0.01317	0.02950
Nordrhein Westfalen	0.01116	0.05964	0.04244	0.07085	0.24478	0.08230	0.37585	0.34358	0.24080	0.18572	0.00418	0.40279
Rheinland Pfalz	0.00453	0.11495	0.03852	0.77984	0.00818	0.85806	0.20077	0.32630	0.98534	0.64765	0.10553	0.55684
Saarland	0.09671	0.85765	0.10560	0.30984	0.89610	0.43653	0.78561	0.34121	0.40892	0.17124	0.12383	0.51432
Sachsen	0.09648	0.33072	0.18220	0.05450	0.35949	0.16619	0.35226	0.02215	0.74315	0.27962	0.00021	0.10059
Sachsen Anhalt	0.01957	0.24715	0.00992	0.98867	0.20530	0.05124	0.46236	0.52966	0.97704	0.42854	0.00109	0.67782
Schleswig Holstein	0.00674	0.27877	0.01445	0.41736	0.20280	0.23295	0.68331	0.07271	0.25752	0.23610	0.00259	0.99052
Thuringen	0.05943	0.23402	0.61890	0.18146	0.16919	0.04632	0.05212	0.36792	0.58350	0.24258	0.93039	0.11427

Significance at 0.01 0.05 0.1

Table D-2. *Vector autoregression p-values for lag 1 and lag 2 models. Each variable is fit to the response variable, UkPopPC, for each German state. Note the legend at the bottom indicates the level of significance for the tests. It is interesting to note that variables Granger cause UkPopPC across more states than instantaneously cause UkPopPC.*

Autocorrelation Functions

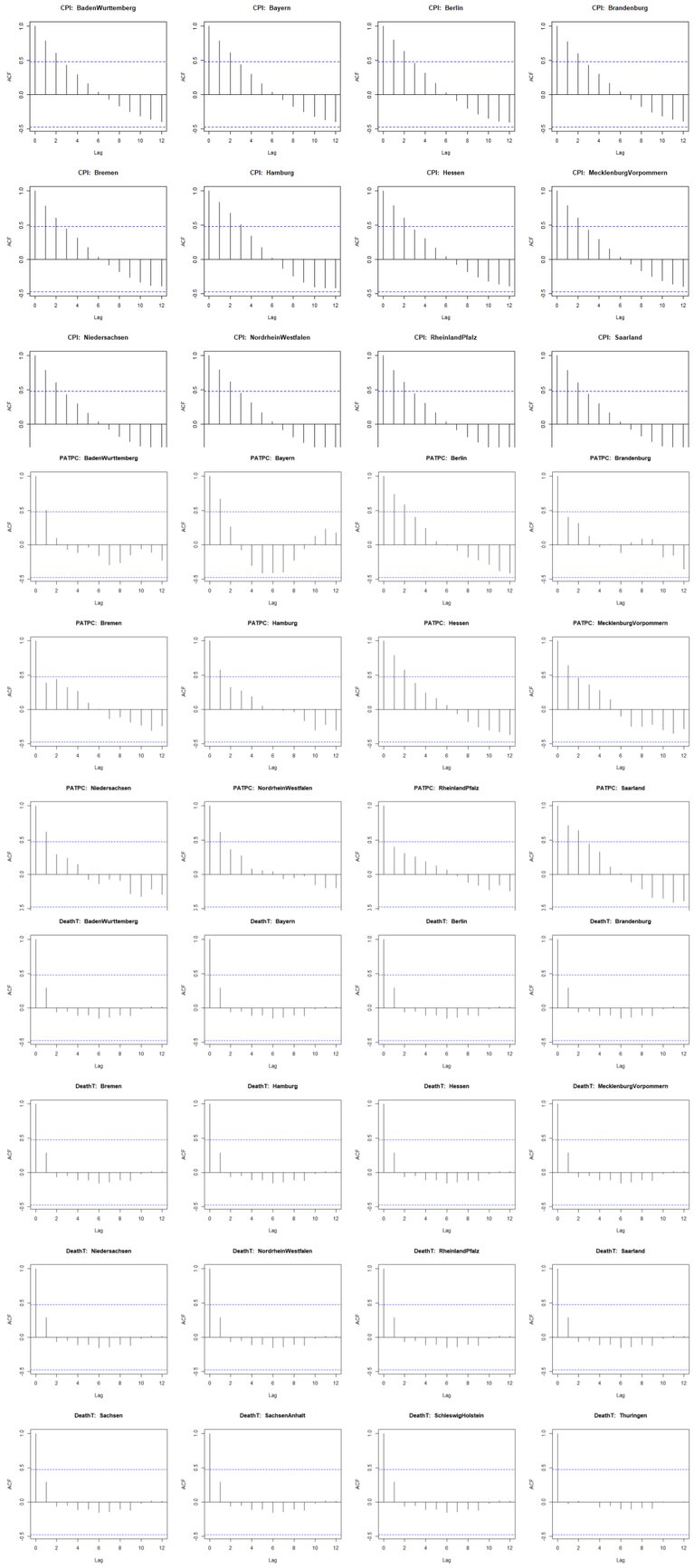


Figure D-1. *Autocorrelation functions (ACFs) for each state across all variables. The blue dotted line indicates statistical significance (e.g., a lag bar crossing above the top line, or below the bottom line, indicates the lag is statistically significantly different from 0). Statistical significance is expected at Lag 0 as this is the current point in time. Many of these ACFs adhere to the autoregressive (AR) model format. For AR models, the ACF demonstrates statistical significance up to a lag, and then does not cross back over the blue line.*

Appendix E: Multivariate Analysis

This appendix provides details on the analysis conducted which entailed multiple variables (non-spatial and non-temporal analysis).

Because most variables were identified as non-normal during exploratory data analysis (EDA), I opted to use the Kruskal Wallis (KW) test to compare all state pairs. The KW test is used in statistics to compare two or more groups. It is a non-parametric test, which means that it does not require normal data; however, it does mandate that all groups' data are from the same distribution family. Below are the null and alternative hypotheses for the KW test.

$$H_0\colon\ \tilde{y}_A = \tilde{y}_B$$

$$H_A\colon \tilde{y}_A \neq \tilde{y}_B$$

Where $\tilde{y}_x$ is the variable's population median for state x. Note that I am running this test for pairs and not larger sets of states. This results in 120 pairs. Table E-1 provides a quick summary of the results, while Table E-2 provides the detailed state to state results. Interestingly, EXAvg and CPI are the only two variables in which we cannot reject the null hypothesis in all cases (i.e., we assume that all states belong to the same group for these variables). In more than 40% of the pairs, we can reject the null hypothesis for the Ukrainian population per capita values. This suggests that several states are unique in their Ukrainian born population and encourages the use of individual statistical and inference models for each state.

	EXAVG	ICTPC	UkPopPC	PATPC	CPI	LFEX	% Unique
BADEN WURTTEMBERG	0	5	7	10	0	10	36%
BAYERN	0	5	7	10	0	7	32%
BERLIN	0	5	9	4	0	2	22%
BRANDENBURG	0	7	3	8	0	2	22%
BREMEN	0	9	7	5	0	6	30%
HAMBURG	0	10	7	8	0	4	32%
HESSEN	0	8	5	4	0	5	24%
MECKLENBURG VORPOMMERN	0	7	6	10	0	6	32%
NIEDERSACHSEN	0	4	7	6	0	2	21%
NORDRHEIN WESTFALEN	0	5	5	6	0	2	20%
RHEINLAND PFALZ	0	3	7	5	0	4	21%
SAARLAND	0	4	5	4	0	6	21%
SACHSEN	0	6	3	5	0	4	20%
SACHSEN ANHALT	0	8	5	8	0	9	33%
SCHLESWIG HOLSTEIN	0	4	9	6	0	2	23%
THURINGEN	0	10	10	5	0	3	31%
OVERALL	0.00%	41.67%	42.50%	43.33%	0.00%	30.83%	

Table E-1. *Percentage of state pairs that are statistically different from one another by state. Each state has 15 other states to compare to. The numbers listed in the variable columns are the total number of pairs that the indicated state is statistically different from. The percentages at the bottom and right represent the total percent of pairs that are statistically different for each variable (bottom) and each state across all variables (right).*

Following Kruskal Wallis, I fit two types of multivariate models. First is a simple linear-multivariate model. Although I have identified the presence of spatial correlation within and amongst our variables (see Appendix D), it is pertinent to see which variables' raw values lead to the "best fit" for each state. I use backward step to find the model with the minimum Akaike information criteria (AIC). To validate these models, I view plots of the model residuals and use the Shapiro-Wilk normality test to determine if the residuals are equivalent to white noise. Table E-2 contains each state's identified "best" linear model where UkPopPC is the response variable. Interestingly, most models identify ICTPC and LFEX as necessary for the best fit, while only two states include DeathT.

Baden Wurttemberg	Variable	**UkPopPC**	**CPI**	**ICTPC**	**PATPC**	**LFEX**	
	Estimate	-0.00614	9.60E-06	3.00E-05	-0.18162	8.07E-05	
	p-value	0.026	0.017	0.094	0.002	0.030	
Bayern	Variable	**UkPopPC**	**CPI**	**ICTPC**	**PATPC**	**EXAvg**	**DeathT**
	Estimate	0.00405	-5.13E-05	0.00043	-0.55757	0.00095	5.5E-08
	p-value	0.000	0.002	0.000	0.081	0.178	0.091
Berlin	Variable	**UkPopPC**	**ICTPC**	**PATPC**	**LFEX**		
	Estimate	0.01429	2.41E-04	-4.47335	-0.00015		
	p-value	0.021	0.000	0.004	0.051		
Brandenburg	Variable	**UkPopPC**	**ICTPC**	**LFEX**	**PATPC**		
	Estimate	0.01992	0.00021	-2.35E-04	-3.17535		
	p-value	0.010	0.023	0.016	0.199		
Bremen	Variable	**UkPopPC**	**PATPC**	**LFEX**			
	Estimate	0.01706	-1.40129	-0.00018			
	p-value	0.008	0.137	0.016			
Hamburg	Variable	**UkPopPC**	**EXAvg**	**DeathT**	**LFEX**		
	Estimate	-0.00505	-6.71E-04	-3.08E-08	9E-05		
	p-value	0.017	0.021	0.084	0.00176		
Hessen	Variable	**UkPopPC**	**CPI**	**ICTPC**	**PATPC**	**LFEX**	
	Estimate	0.01293	2.23E-05	5.8E-05	0.98793	-0.00018	
	p-value	0.042	0.058	0.060	0.079	0.029	
Mecklenburg Vorpommern	Variable	**UkPopPC**	**ICTPC**	**EXAvg**			
	Estimate	-0.00106	0.00069	0.00334			
	p-value	0.011	0.000	0.011			
Niedersachsen	Variable	**UkPopPC**	**ICTPC**	**PATPC**	**LFEX**		
	Estimate	0.02213	0.00025	-0.68945	-0.00027		
	p-value	0.001	0.000	0.037	0.001		
Nordrhein Westfalen	Variable	**UkPopPC**	**ICTPC**	**PATPC**	**LFEX**		
	Estimate	0.01263	4.79E-05	-6.38E-01	-0.00014		
	p-value	0.000	0.001	0.028	0.000		
Rheinland Pfalz	Variable	**UkPopPC**	**ICTPC**	**LFEX**			
	Estimate	0.01435	7.68E-05	-1.66E-04			
	p-value	0.000	0.007	0.000			
Saarland	Variable	**UkPopPC**	**CPI**	**ICTPC**	**PATPC**	**EXAvg**	
	Estimate	0.00212	-1.67E-05	0.00012	0.90839	5.66E-04	
	p-value	0.002	0.023	0.019	0.090	0.028	
Sachsen	Variable	**UkPopPC**	**CPI**	**EXAvg**			
	Estimate	-0.00083	2.5E-05	0.00222			
	p-value	0.104	0.000	0.043			
Sachsen Anhalt	Variable	**UkPopPC**	**CPI**	**ICTPC**	**EXAvg**	**LFEX**	
	Estimate	0.00875	-8.08E-06	0.00017	0.00029	-9.21E-05	
	p-value	0.012	0.241	0.008	0.111	0.050	
Schleswig Holstein	Variable	**UkPopPC**	**ICTPC**	**EXAvg**	**LFEX**		
	Estimate	0.02303	1.62E-04	2.97E-04	-0.00029		
	p-value	0.000	0.000	0.071	0.000		
Thuringen	Variable	**UkPopPC**	**CPI**	**ICTPC**	**EXAvg**	**LFEX**	
	Estimate	0.00676	2.13E-05	0.00012	-0.0002	-0.0001	
	p-value	0.021	0.022	0.199	0.219	0.016	

Table E-2. *Best linear model fits by state. Each state's identified best fit is drastically different from the others, with various combinations of identified important variables and estimated coefficients changing signs throughout. This supports the concept that it is integral to consider each state as a unique entity.*

The second type of multivariate model I use is a generalized least squares (GLS) model with an ARMA(1,0) correlation structure[5]. This type of model takes the temporal correlation uncovered in Appendix D into consideration and is anticipated to provide more accurate results because it accounts for the correlation in our variables. Table E-3 compiles the "best" GLS models and compares the resulting Shapiro Wilk p-value to the linear model results. Notably, the GLS model selects only a single variable as important for each state, with most states identifying PATPC as important. Although the simple linear model appears to provide more normal residuals, the results from the GLS are likely more accurate because the linear model does not account for correlation within the data.

	Simple Linear Model	Shapiro Residual p-val	GLS, ARMA(1,0)	Shapiro Residual p-val
Baden Wurttemberg	CPI + ICTPC + PATPC + LFEX	0.1905	CPI	0.9987
Bayern	CPI + ICTPC + PATPC + EXAvg + DeathT	0.3791	PATPC	0.0012
Berlin	ICTPC + PATPC + LFEX	0.2839	PATPC	0.4003
Brandenburg	ICTPC + LFEX	0.0014	PATPC	0.0001
Bremen	PATPC + LFEX	0.0855	PATPC	0.6708
Hamburg	EXAvg + DeathT + LFEX	0.6680	PATPC	0.3946
Hessen	CPI + ICTPC + PATPC + LFEX	0.7161	PATPC	0.8882
Mecklenburg Vorpommern	ICTPC + EXAvg	0.0701	PATPC	0.0003
Niedersachsen	ICTPC + PATPC + LFEX	0.3512	PATPC	0.6797
Nordrhein Westfalen	ICTPC + PATPC + LFEX	0.5237	PATPC	0.2445
Rheinland Pfalz	ICTPC + LFEX	0.0081	PATPC	0.7701
Saarland	CPI + ICTPC + PATPC + EXAvg	0.6467	PATPC	0.3181
Sachsen	CPI + EXAvg	0.0147	PATPC	0.0000
Sachsen Anhalt	CPI + ICTPC + EXAvg + LFEX	0.3010	PATPC	0.6671
Schleswig Holstein	ICTPC + EXAvg + LFEX	0.8336	PATPC	0.3088
Thuringen	CPI + ICTPC + EXAvg + LFEX	0.6264	PATPC	0.7158

Significance at 0.01 0.05 0.1

Table E-3. *Comparison of the best linear model and the best generalized least squares (GLS), ARMA(1,0) model. The figure shows the identified important variables along with the Shapiro Wilk p-value. Most notably, the GLS model selects only a single variable in all cases. The simple linear model provides more normal residuals; however, this is likely because the models do not account for the correlations present in the data.*

[5] ARMA= Autoregressive Moving Average. I use an ARMA model of order 1, 0. This equates to an AR(1). Please refer to Appendix D for details.

State Pair	Critical Diff	Exam Average		ICT Value Per Capita		UkPop Per Capita		Patents Per Capita		CPI		Life Expectancy	
		Observed Diff	Statistically Significant	Observed Diff	Statistically Significant	Observed Diff	Statistically Significant	Observed Diff	Statistically Significant	Observed Diff	Statistically Significant	Observed Diff	Statistically Significant
BadenWurttemberg-Bayern	95.2256	1.5882	FALSE	13.8824	FALSE	149.1176	TRUE	16.5294	FALSE	1.5000	FALSE	35.0294	FALSE
BadenWurttemberg-Berlin	95.2256	4.8824	FALSE	15.1765	FALSE	199.0000	TRUE	145.2941	TRUE	4.4412	FALSE	103.8529	TRUE
BadenWurttemberg-Brandenburg	95.2256	21.7941	FALSE	118.7647	TRUE	82.7059	FALSE	225.4706	TRUE	9.4706	FALSE	134.5294	TRUE
BadenWurttemberg-Bremen	95.2256	18.8824	FALSE	66.3529	FALSE	157.2353	TRUE	160.7647	TRUE	2.6471	FALSE	188.9118	TRUE
BadenWurttemberg-Hamburg	95.2256	14.2941	FALSE	83.4706	FALSE	170.9412	TRUE	47.2941	FALSE	24.8529	FALSE	81.8529	FALSE
BadenWurttemberg-Hessen	95.2256	25.0000	FALSE	39.0588	FALSE	107.7059	TRUE	83.2941	FALSE	18.2647	FALSE	44.7353	FALSE
BadenWurttemberg-MecklenburgVorpommern	95.2256	16.8235	FALSE	126.2941	TRUE	124.5294	TRUE	247.5294	TRUE	4.3529	FALSE	188.3529	TRUE
BadenWurttemberg-Niedersachsen	95.2256	13.8824	FALSE	60.5882	FALSE	10.1176	FALSE	64.4706	FALSE	13.9412	FALSE	118.8824	TRUE
BadenWurttemberg-NordrheinWestfalen	95.2256	13.6471	FALSE	0.5882	FALSE	105.8824	TRUE	61.7647	FALSE	2.6176	FALSE	131.0882	TRUE
BadenWurttemberg-RheinlandPfalz	95.2256	21.7059	FALSE	73.1176	FALSE	9.9412	FALSE	122.1765	TRUE	9.3824	FALSE	90.2941	FALSE
BadenWurttemberg-Saarland	95.2256	5.7647	FALSE	60.0000	FALSE	51.1765	FALSE	143.4706	TRUE	10.4412	FALSE	191.7941	TRUE
BadenWurttemberg-Sachsen	95.2256	2.7941	FALSE	110.1176	TRUE	82.1176	FALSE	164.2353	TRUE	3.1765	FALSE	90.0000	FALSE
BadenWurttemberg-SachsenAnhalt	95.2256	1.7059	FALSE	146.2941	TRUE	22.3529	FALSE	238.3529	TRUE	5.9118	FALSE	222.7353	TRUE
BadenWurttemberg-SchleswigHolstein	95.2256	9.0000	FALSE	40.7059	FALSE	41.2353	FALSE	192.4118	TRUE	29.1176	FALSE	110.2353	TRUE
BadenWurttemberg-Thuringen	95.2256	0.9412	FALSE	156.7647	TRUE	44.2941	FALSE	123.1765	TRUE	2.8235	FALSE	158.0588	TRUE
Bayern-Berlin	95.2256	6.4706	FALSE	29.0588	FALSE	49.8824	FALSE	128.7647	TRUE	5.9412	FALSE	68.8235	FALSE
Bayern-Brandenburg	95.2256	23.3824	FALSE	132.6471	TRUE	66.4118	FALSE	208.9412	TRUE	10.9706	FALSE	99.5000	TRUE
Bayern-Bremen	95.2256	20.4706	FALSE	52.4706	FALSE	8.1176	FALSE	144.2353	TRUE	4.1471	FALSE	153.8824	TRUE
Bayern-Hamburg	95.2256	15.8824	FALSE	69.5882	FALSE	21.8235	FALSE	30.7647	FALSE	23.3529	FALSE	46.8235	FALSE
Bayern-Hessen	95.2256	26.5882	FALSE	25.1765	FALSE	41.4118	FALSE	66.7647	FALSE	19.7647	FALSE	9.7059	FALSE
Bayern-MecklenburgVorpommern	95.2256	18.4118	FALSE	140.1765	TRUE	24.5882	FALSE	231.0000	TRUE	2.8529	FALSE	153.3235	TRUE
Bayern-Niedersachsen	95.2256	15.4706	FALSE	74.4706	FALSE	139.0000	TRUE	47.9412	FALSE	15.4412	FALSE	83.8529	FALSE
Bayern-NordrheinWestfalen	95.2256	15.2353	FALSE	13.2941	FALSE	43.2353	FALSE	45.2353	FALSE	4.1176	FALSE	96.0588	TRUE
Bayern-RheinlandPfalz	95.2256	23.2941	FALSE	87.0000	FALSE	139.1765	TRUE	105.6471	TRUE	10.8824	FALSE	55.2647	FALSE
Bayern-Saarland	95.2256	7.3529	FALSE	73.8824	FALSE	97.9412	TRUE	126.9412	TRUE	11.9412	FALSE	156.7647	TRUE
Bayern-Sachsen	95.2256	4.3824	FALSE	124.0000	TRUE	67.0000	FALSE	147.7059	TRUE	1.6765	FALSE	54.9706	FALSE
Bayern-SachsenAnhalt	95.2256	0.1176	FALSE	160.1765	TRUE	126.7647	TRUE	221.8235	TRUE	7.4118	FALSE	187.7059	TRUE
Bayern-SchleswigHolstein	95.2256	10.5882	FALSE	54.5882	FALSE	190.3529	TRUE	175.8824	TRUE	27.6176	FALSE	75.2059	FALSE
Bayern-Thuringen	95.2256	2.5294	FALSE	170.6471	TRUE	193.4118	TRUE	106.6471	TRUE	1.3235	FALSE	123.0294	TRUE
Berlin-Brandenburg	95.2256	16.9118	FALSE	103.5882	TRUE	116.2941	TRUE	80.1765	FALSE	5.0294	FALSE	30.6765	FALSE
Berlin-Bremen	95.2256	14.0000	FALSE	81.5294	FALSE	41.7647	FALSE	15.4706	FALSE	1.7941	FALSE	85.0588	FALSE
Berlin-Hamburg	95.2256	9.4118	FALSE	98.6471	TRUE	28.0588	FALSE	98.0000	TRUE	29.2941	FALSE	22.0000	FALSE
Berlin-Hessen	95.2256	20.1176	FALSE	54.2353	FALSE	91.2941	FALSE	62.0000	FALSE	13.8235	FALSE	59.1176	FALSE
Berlin-MecklenburgVorpommern	95.2256	11.9412	FALSE	111.1176	TRUE	74.4706	FALSE	102.2353	TRUE	8.7941	FALSE	84.5000	FALSE
Berlin-Niedersachsen	95.2256	9.0000	FALSE	45.4118	FALSE	188.8824	TRUE	80.8235	FALSE	9.5000	FALSE	15.0294	FALSE
Berlin-NordrheinWestfalen	95.2256	8.7647	FALSE	15.7647	FALSE	93.1176	FALSE	83.5294	FALSE	1.8235	FALSE	27.2353	FALSE
Berlin-RheinlandPfalz	95.2256	16.8235	FALSE	57.9412	FALSE	189.0588	TRUE	23.1176	FALSE	4.9412	FALSE	13.5588	FALSE
Berlin-Saarland	95.2256	0.8824	FALSE	44.8235	FALSE	147.8235	TRUE	1.8235	FALSE	6.0000	FALSE	87.9412	FALSE
Berlin-Sachsen	95.2256	2.0882	FALSE	94.9412	FALSE	116.8824	TRUE	18.9412	FALSE	7.6176	FALSE	13.8529	FALSE
Berlin-SachsenAnhalt	95.2256	6.5882	FALSE	131.1176	TRUE	176.6471	TRUE	93.0588	FALSE	1.4706	FALSE	118.8824	TRUE
Berlin-SchleswigHolstein	95.2256	4.1176	FALSE	25.5294	FALSE	240.2353	TRUE	47.1176	FALSE	33.5588	FALSE	6.3824	FALSE
Berlin-Thuringen	95.2256	3.9412	FALSE	141.5882	TRUE	243.2941	TRUE	22.1176	FALSE	7.2647	FALSE	54.2059	FALSE
Brandenburg-Bremen	95.2256	2.9118	FALSE	185.1176	TRUE	74.5294	FALSE	64.7059	FALSE	6.8235	FALSE	54.3824	FALSE
Brandenburg-Hamburg	95.2256	7.5000	FALSE	202.2353	TRUE	88.2353	FALSE	178.1765	TRUE	34.3235	FALSE	52.6765	FALSE
Brandenburg-Hessen	95.2256	3.2059	FALSE	157.8235	TRUE	25.0000	FALSE	142.1765	TRUE	8.7941	FALSE	89.7941	FALSE
Brandenburg-MecklenburgVorpommern	95.2256	4.9706	FALSE	7.5294	FALSE	41.8235	FALSE	22.0588	FALSE	13.8235	FALSE	53.8235	FALSE
Brandenburg-Niedersachsen	95.2256	7.9118	FALSE	58.1765	FALSE	72.5882	FALSE	161.0000	TRUE	4.4706	FALSE	15.6471	FALSE
Brandenburg-NordrheinWestfalen	95.2256	8.1471	FALSE	119.3529	TRUE	23.1765	FALSE	163.7059	TRUE	6.8529	FALSE	3.4412	FALSE
Brandenburg-RheinlandPfalz	95.2256	0.0882	FALSE	45.6471	FALSE	72.7647	FALSE	103.2941	TRUE	0.0882	FALSE	44.2353	FALSE
Brandenburg-Saarland	95.2256	16.0294	FALSE	58.7647	FALSE	31.5294	FALSE	82.0000	FALSE	0.9706	FALSE	57.2647	FALSE
Brandenburg-Sachsen	95.2256	19.0000	FALSE	8.6471	FALSE	0.5882	FALSE	61.2353	FALSE	12.6471	FALSE	44.5294	FALSE
Brandenburg-SachsenAnhalt	95.2256	23.5000	FALSE	27.5294	FALSE	60.3529	FALSE	12.8824	FALSE	3.5588	FALSE	88.2059	FALSE
Brandenburg-SchleswigHolstein	95.2256	12.7941	FALSE	78.0588	FALSE	123.9412	TRUE	33.0588	FALSE	38.5882	FALSE	24.2941	FALSE
Brandenburg-Thuringen	95.2256	20.8529	FALSE	38.0000	FALSE	127.0000	TRUE	102.2941	TRUE	12.2941	FALSE	23.5294	FALSE
Bremen-Hamburg	95.2256	4.5882	FALSE	17.1176	FALSE	13.7059	FALSE	113.4706	TRUE	27.5000	FALSE	107.0588	TRUE
Bremen-Hessen	95.2256	6.1176	FALSE	27.2941	FALSE	49.5294	FALSE	77.4706	FALSE	15.6176	FALSE	144.1765	TRUE
Bremen-MecklenburgVorpommern	95.2256	2.0588	FALSE	192.6471	TRUE	32.7059	FALSE	86.7647	FALSE	7.0000	FALSE	0.5588	FALSE
Bremen-Niedersachsen	95.2256	5.0000	FALSE	126.9412	TRUE	147.1176	TRUE	96.2941	TRUE	11.2941	FALSE	70.0294	FALSE
Bremen-NordrheinWestfalen	95.2256	5.2353	FALSE	65.7647	FALSE	51.3529	FALSE	99.0000	TRUE	0.0294	FALSE	57.8235	FALSE
Bremen-RheinlandPfalz	95.2256	2.8235	FALSE	139.4706	TRUE	147.2941	TRUE	38.5882	FALSE	6.7353	FALSE	98.6176	TRUE
Bremen-Saarland	95.2256	13.1176	FALSE	126.3529	TRUE	106.0588	TRUE	17.2941	FALSE	7.7941	FALSE	2.8824	FALSE
Bremen-Sachsen	95.2256	16.0882	FALSE	176.4706	TRUE	75.1176	FALSE	3.4706	FALSE	5.8235	FALSE	98.9118	TRUE
Bremen-SachsenAnhalt	95.2256	20.5882	FALSE	212.6471	TRUE	134.8824	TRUE	77.5882	FALSE	3.2647	FALSE	33.8235	FALSE
Bremen-SchleswigHolstein	95.2256	9.8824	FALSE	107.0588	TRUE	198.4706	TRUE	31.6471	FALSE	31.7647	FALSE	78.6765	FALSE
Bremen-Thuringen	95.2256	17.9412	FALSE	223.1176	TRUE	201.5294	TRUE	37.5882	FALSE	5.4706	FALSE	30.8529	FALSE
Hamburg-Hessen	95.2256	10.7059	FALSE	44.4118	FALSE	63.2353	FALSE	36.0000	FALSE	43.1176	FALSE	37.1176	FALSE
Hamburg-MecklenburgVorpommern	95.2256	2.5294	FALSE	209.7647	TRUE	46.4118	FALSE	200.2353	TRUE	20.5000	FALSE	106.5000	TRUE
Hamburg-Niedersachsen	95.2256	0.4118	FALSE	144.0588	TRUE	160.8235	TRUE	17.1765	FALSE	38.7941	FALSE	37.0294	FALSE
Hamburg-NordrheinWestfalen	95.2256	0.6471	FALSE	82.8824	FALSE	65.0588	FALSE	14.4706	FALSE	27.4706	FALSE	49.2353	FALSE
Hamburg-RheinlandPfalz	95.2256	7.4118	FALSE	156.5882	TRUE	161.0000	TRUE	74.8824	FALSE	34.2353	FALSE	8.4412	FALSE
Hamburg-Saarland	95.2256	8.5294	FALSE	143.4706	TRUE	119.7647	TRUE	96.1765	TRUE	35.2941	FALSE	109.9412	TRUE
Hamburg-Sachsen	95.2256	11.5000	FALSE	193.5882	TRUE	88.8235	FALSE	116.9412	TRUE	21.6765	FALSE	8.1471	FALSE
Hamburg-SachsenAnhalt	95.2256	16.0000	FALSE	229.7647	TRUE	148.5882	TRUE	191.0588	TRUE	30.7647	FALSE	140.8824	TRUE
Hamburg-SchleswigHolstein	95.2256	5.2941	FALSE	124.1765	TRUE	212.1765	TRUE	145.1176	TRUE	4.2647	FALSE	28.3824	FALSE
Hamburg-Thuringen	95.2256	13.3529	FALSE	240.2353	TRUE	215.2353	TRUE	75.8824	FALSE	22.0294	FALSE	76.2059	FALSE
Hessen-MecklenburgVorpommern	95.2256	8.1765	FALSE	165.3529	TRUE	16.8235	FALSE	164.2353	TRUE	22.6176	FALSE	143.6176	TRUE
Hessen-Niedersachsen	95.2256	11.1176	FALSE	99.6471	TRUE	97.5882	TRUE	18.8235	FALSE	4.3235	FALSE	74.1471	FALSE
Hessen-NordrheinWestfalen	95.2256	11.3529	FALSE	38.4706	FALSE	1.8235	FALSE	21.5294	FALSE	15.6471	FALSE	86.3529	FALSE
Hessen-RheinlandPfalz	95.2256	3.2941	FALSE	112.1765	TRUE	97.7647	TRUE	38.8824	FALSE	8.8824	FALSE	45.5588	FALSE
Hessen-Saarland	95.2256	19.2353	FALSE	99.0588	TRUE	56.5294	FALSE	60.1765	FALSE	7.8235	FALSE	147.0588	TRUE
Hessen-Sachsen	95.2256	22.2059	FALSE	149.1765	TRUE	25.5882	FALSE	80.9412	FALSE	21.4412	FALSE	45.2647	FALSE
Hessen-SachsenAnhalt	95.2256	26.7059	FALSE	185.3529	TRUE	85.3529	FALSE	155.0588	TRUE	12.3529	FALSE	178.0000	TRUE
Hessen-SchleswigHolstein	95.2256	16.0000	FALSE	79.7647	FALSE	148.9412	TRUE	109.1176	TRUE	47.3824	FALSE	65.5000	FALSE
Hessen-Thuringen	95.2256	24.0588	FALSE	195.8235	TRUE	152.0000	TRUE	39.8824	FALSE	21.0882	FALSE	113.3235	TRUE
MecklenburgVorpommern-Niedersachsen	95.2256	2.9412	FALSE	65.7059	FALSE	114.4118	TRUE	183.0588	TRUE	18.2941	FALSE	69.4706	FALSE
MecklenburgVorpommern-NordrheinWestfalen	95.2256	3.1765	FALSE	126.8824	TRUE	18.6471	FALSE	185.7647	TRUE	6.9706	FALSE	57.2647	FALSE
MecklenburgVorpommern-RheinlandPfalz	95.2256	4.8824	FALSE	53.1765	FALSE	114.5882	TRUE	125.3529	TRUE	13.7353	FALSE	98.0588	TRUE
MecklenburgVorpommern-Saarland	95.2256	11.0588	FALSE	66.2941	FALSE	73.3529	FALSE	104.0588	TRUE	14.7941	FALSE	3.4412	FALSE
MecklenburgVorpommern-Sachsen	95.2256	14.0294	FALSE	16.1765	FALSE	42.4118	FALSE	83.2941	FALSE	1.1765	FALSE	98.3529	TRUE
MecklenburgVorpommern-SachsenAnhalt	95.2256	18.5294	FALSE	20.0000	FALSE	102.1765	TRUE	9.1765	FALSE	10.2647	FALSE	34.3824	FALSE
MecklenburgVorpommern-SchleswigHolstein	95.2256	7.8235	FALSE	85.5882	FALSE	165.7647	TRUE	55.1176	FALSE	24.7647	FALSE	78.1176	FALSE
MecklenburgVorpommern-Thuringen	95.2256	15.8824	FALSE	30.4706	FALSE	168.8235	TRUE	124.3529	TRUE	1.5294	FALSE	30.2941	FALSE
Niedersachsen-NordrheinWestfalen	95.2256	0.2353	FALSE	61.1765	FALSE	95.7647	TRUE	2.7059	FALSE	11.3235	FALSE	12.2059	FALSE
Niedersachsen-RheinlandPfalz	95.2256	7.8235	FALSE	12.5294	FALSE	0.1765	FALSE	57.7059	FALSE	4.5588	FALSE	28.5882	FALSE
Niedersachsen-Saarland	95.2256	8.1176	FALSE	0.5882	FALSE	41.0588	FALSE	79.0000	FALSE	3.5000	FALSE	72.9118	FALSE
Niedersachsen-Sachsen	95.2256	11.0882	FALSE	49.5294	FALSE	72.0000	FALSE	99.7647	TRUE	17.1176	FALSE	28.8824	FALSE
Niedersachsen-SachsenAnhalt	95.2256	15.5882	FALSE	85.7059	FALSE	12.2353	FALSE	173.8824	TRUE	8.0294	FALSE	103.8529	TRUE
Niedersachsen-SchleswigHolstein	95.2256	4.8824	FALSE	19.8824	FALSE	51.3529	FALSE	127.9412	TRUE	43.0588	FALSE	8.6471	FALSE
Niedersachsen-Thuringen	95.2256	12.9412	FALSE	96.1765	TRUE	54.4118	FALSE	58.7059	FALSE	16.7647	FALSE	39.1765	FALSE
NordrheinWestfalen-RheinlandPfalz	95.2256	8.0588	FALSE	73.7059	FALSE	95.9412	TRUE	60.4118	FALSE	6.7647	FALSE	40.7941	FALSE
NordrheinWestfalen-Saarland	95.2256	7.8824	FALSE	60.5882	FALSE	54.7059	FALSE	81.7059	FALSE	7.8235	FALSE	60.7059	FALSE
NordrheinWestfalen-Sachsen	95.2256	10.8529	FALSE	110.7059	TRUE	23.7647	FALSE	102.4706	TRUE	5.7941	FALSE	41.0882	FALSE
NordrheinWestfalen-SachsenAnhalt	95.2256	15.3529	FALSE	146.8824	TRUE	83.5294	FALSE	176.5882	TRUE	3.2941	FALSE	91.6471	FALSE
NordrheinWestfalen-SchleswigHolstein	95.2256	4.6471	FALSE	41.2941	FALSE	147.1176	TRUE	130.6471	TRUE	31.7353	FALSE	20.8529	FALSE
NordrheinWestfalen-Thuringen	95.2256	12.7059	FALSE	157.3529	TRUE	150.1765	TRUE	61.4118	FALSE	5.4412	FALSE	26.9706	FALSE
RheinlandPfalz-Saarland	95.2256	15.9412	FALSE	13.1176	FALSE	41.2353	FALSE	21.2941	FALSE	1.0588	FALSE	101.5000	TRUE
RheinlandPfalz-Sachsen	95.2256	18.9118	FALSE	37.0000	FALSE	72.1765	FALSE	42.0588	FALSE	12.5588	FALSE	0.2941	FALSE
RheinlandPfalz-SachsenAnhalt	95.2256	23.4118	FALSE	73.1765	FALSE	12.4118	FALSE	116.1765	TRUE	3.4706	FALSE	132.4412	TRUE
RheinlandPfalz-SchleswigHolstein	95.2256	12.7059	FALSE	32.4118	FALSE	51.1765	FALSE	70.2353	FALSE	38.5000	FALSE	19.9412	FALSE
RheinlandPfalz-Thuringen	95.2256	20.7647	FALSE	83.6471	FALSE	54.2353	FALSE	1.0000	FALSE	12.2059	FALSE	67.7647	FALSE
Saarland-Sachsen	95.2256	2.9706	FALSE	50.1176	FALSE	30.9412	FALSE	20.7647	FALSE	13.6176	FALSE	101.7941	TRUE
Saarland-SachsenAnhalt	95.2256	7.4706	FALSE	86.2941	FALSE	28.8235	FALSE	94.8824	FALSE	4.5294	FALSE	30.9412	FALSE
Saarland-SchleswigHolstein	95.2256	3.2353	FALSE	19.2941	FALSE	92.4118	FALSE	48.9412	FALSE	39.5588	FALSE	81.5588	FALSE
Saarland-Thuringen	95.2256	4.8235	FALSE	96.7647	TRUE	95.4706	TRUE	20.2941	FALSE	13.2647	FALSE	33.7353	FALSE
Sachsen-SachsenAnhalt	95.2256	4.5000	FALSE	36.1765	FALSE	59.7647	FALSE	74.1176	FALSE	9.0882	FALSE	132.7353	TRUE
Sachsen-SchleswigHolstein	95.2256	6.2059	FALSE	69.4118	FALSE	123.3529	TRUE	28.1765	FALSE	25.9412	FALSE	20.2353	FALSE
Sachsen-Thuringen	95.2256	1.8529	FALSE	46.6471	FALSE	126.4118	TRUE	41.0588	FALSE	0.3529	FALSE	68.0588	FALSE
SachsenAnhalt-SchleswigHolstein	95.2256	10.7059	FALSE	105.5882	TRUE	63.5882	FALSE	45.9412	FALSE	35.0294	FALSE	112.5000	TRUE
SachsenAnhalt-Thuringen	95.2256	2.6471	FALSE	10.4706	FALSE	66.6471	FALSE	115.1765	TRUE	8.7353	FALSE	64.6765	FALSE
SchleswigHolstein-Thuringen	95.2256	8.0588	FALSE	116.0588	TRUE	3.0588	FALSE	69.2353	FALSE	26.2941	FALSE	47.8235	FALSE

Table E-4. *Complete breakdown of the Kruskal Wallis test results. Those pairs that are statistically different for a specified variable are labeled 'TRUE' and highlighted green.*

Appendix F: Inference Models

This appendix provides additional details on the inference models developed using Random Forest Regression Trees, with k-fold cross validation (k=6)[6]. Figure F-1 provides the learning curve (how the model performs across varying train-test splits) and the feature importance plot for all data (all of Germany). Notably, State is the most important feature. This encourages follow on- individual state models.

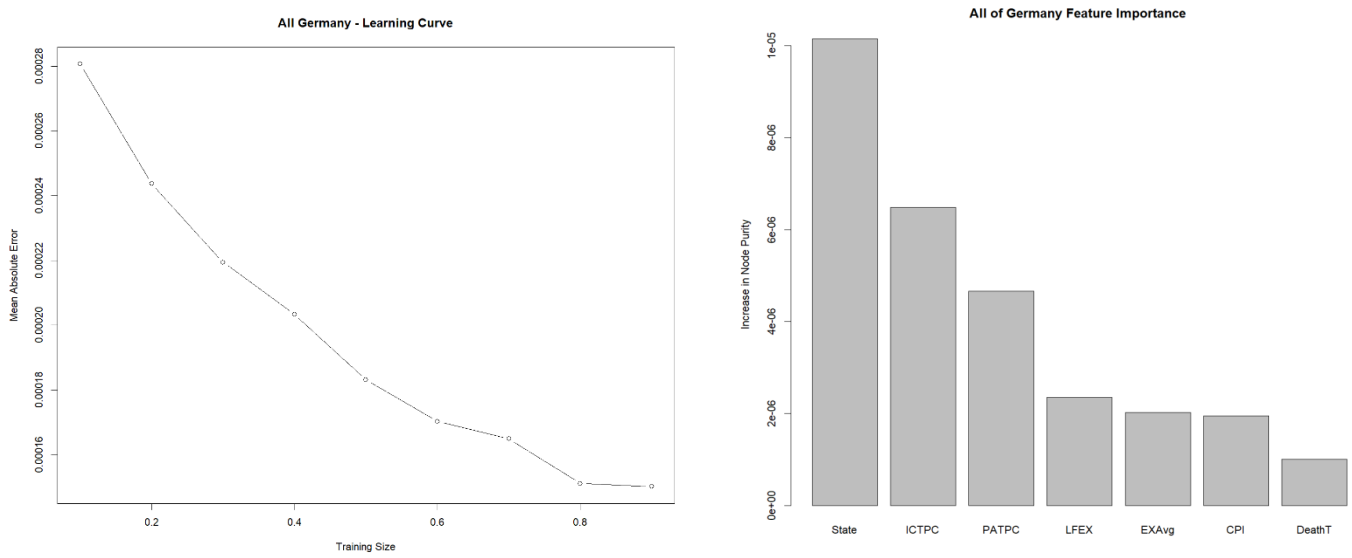


Figure F-1. *All of Germany random forest results. Train-test split versus mean absolute error (left) shows that the model performance improves with more training data. To avoid overfitting the model, I select a conservative train-test split ratio of 0.6 for mean absolute error measurements. The feature importance plot (right) shows that State, ICTPC, and PATPC are the top three most important features when considering all of the data across Germany.*

Figures F-2 and F-3 provide the state-specific learning curves and feature importance plots. Similar to the All of Germany Feature Importance, most states contain DeathT in the bottom two.

[6] The R library caret was used to generate the random forest models.

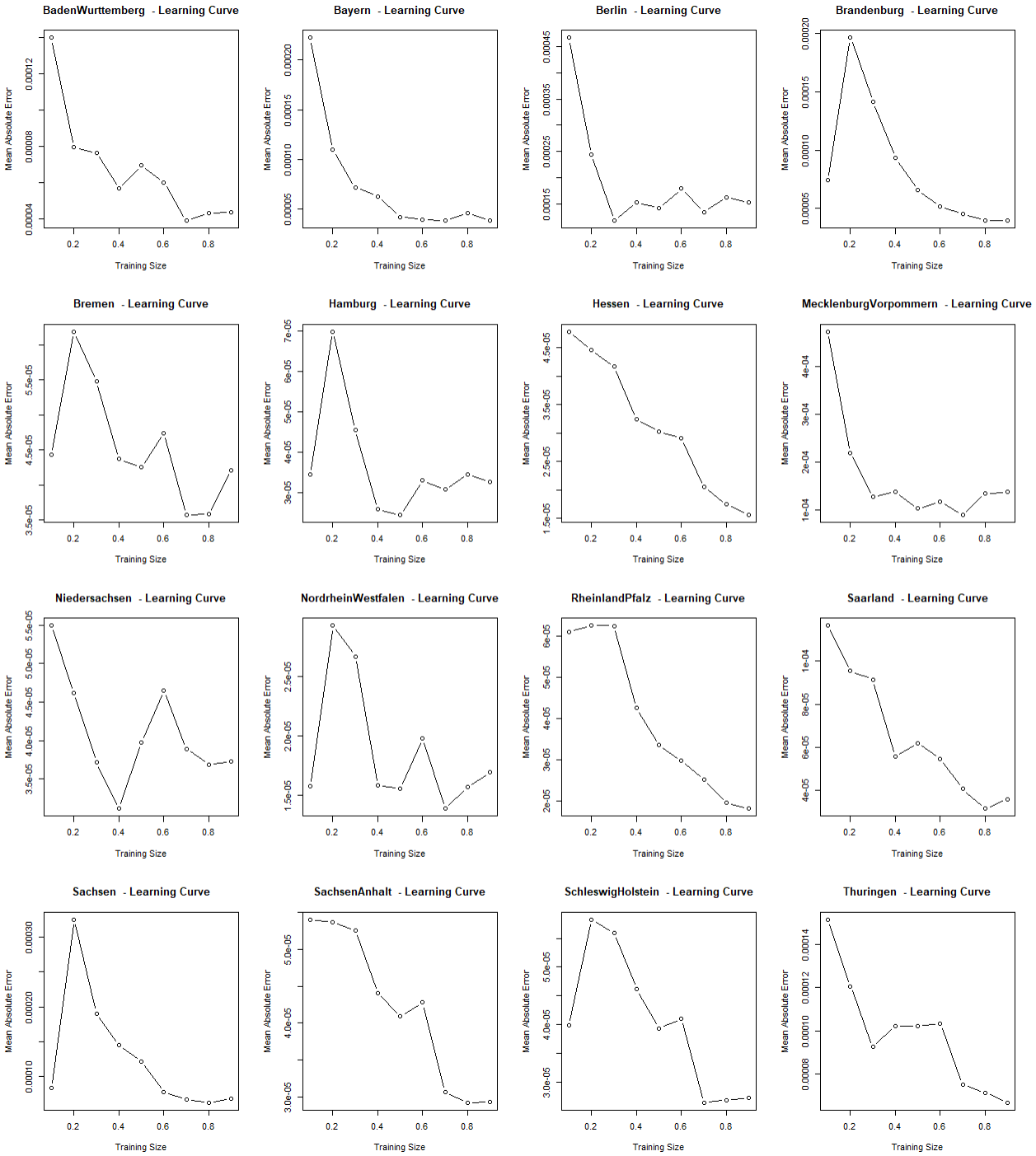


Figure F-2. *Individual state learning curves. Several states exhibit slightly irregular performance in early train-test splits; however, most models improve substantially between 0.6-0.8.*

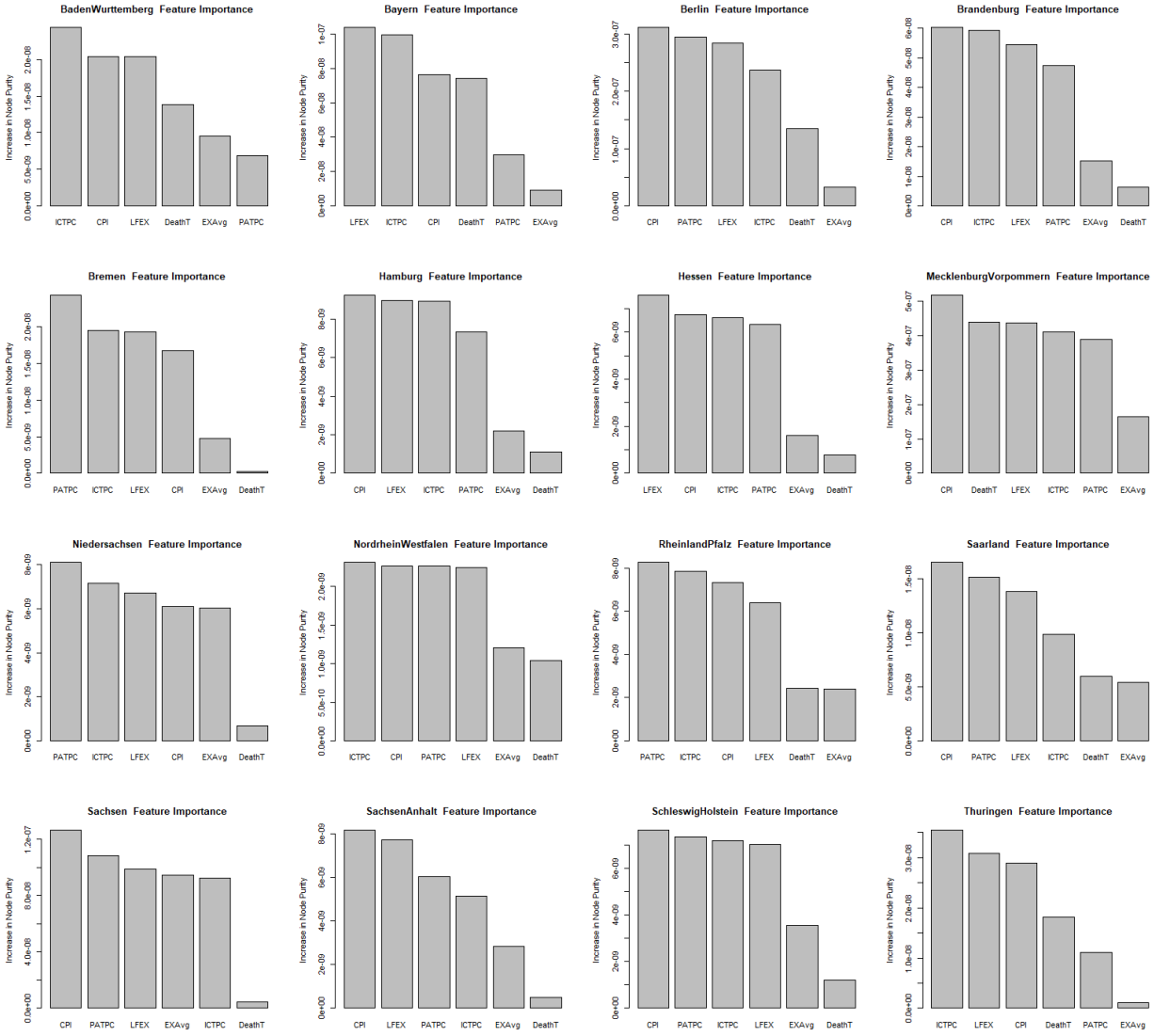


Figure F-3. *Individual state feature importance plots. CPI, ICTPC, and PATPC occur in the top two most important features the most often. Notably, DeathT appears in the bottom two in most states.*