Essays on Labor Economics and Applied Econometrics

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von

Kristina Maria Zapp

im Frühjahrs-/Sommersemester 2022

Abteilungssprecher	Prof. Volker Nocke, Ph.D.
Referent	Prof. Dr. Sebastian Siegloch
Korreferent	Prof. Christina Gathmann, Ph.D.

Tag der Verteidigung	25.05.2022

Acknowledgments

I am grateful to my first supervisor Sebastian Siegloch and my second supervisor and co-author Christina Gathmann for their constant guidance. This thesis has benefited greatly from their invaluable feedback and suggestions. I thank Christina Gathmann especially for the possibility to start with a joint project. During the fruitful collaboration I have learned a lot about research. I want to thank Sebastian Siegloch especially for his feedback and advice in the process of writing my solo-authored chapter and for providing guidance during the last steps of this thesis.

I want to thank my further co-authors Enno Mammen, Ralf Wilke and Terry Gregory for the fruitful collaboration. I thank Ralf Wilke for his advice at different stages of my PhD, for the constructive cooperation and for hosting me at Copenhagen Business School which was an inspiring experience. I thank Enno Mammen for broadening my perspective on chapter 1.

I have written this thesis during my time as a doctoral researcher at the ZEW – Leibniz Centre for European Economic Research in Mannheim. I thank for the collaborative working environment at the department Labour Markets and Human Resources and the inspiring seminars and discussions. I thank for helpful feedback, especially from Eduard Brüll, Sarra Ben Yahmed and Boris Ivanov. I am grateful for the support of my research stay at Copenhagen Business School and for the possibility to join international workshops and conferences. I want to thank my fellow PhD students at the University of Mannheim for a helpful and friendly atmosphere also during the coursework. I thank the Graduate School of Economic and Social Sciences at the University of Mannheim for financial support during my doctoral studies. For our research on Chapter 3 we gratefully acknowledge funding from the German Research Foundation (DFG) through the project "Does a Minimum Wage Boost Automation and Outsourcing?" (GA 1813/5-1).

On a personal note, I thank my friends and family for being part of this journey. I am grateful to my parents for their support and for always believing that I can achieve my goals. Thank you to Mara, for joining our life. Finally, I want to thank my husband Matthias for all his personal support and his optimism.

Contents

List of Figures

List of Tables

General Introduction

This thesis consists of three self-contained chapters covering the research areas econometrics, applied economics and labor economics.

In many empirical economic studies panel data estimators have advantages. Panel data includes the same individuals over several time periods. Panel data estimation can provide unbiased results under mild assumptions. Often economists estimate the effect of a change of a variable x on the change of a variable y. A common difficulty in these applications is endogeneity due to omitted variables. These are unobserved variables that are correlated with both x and y. To give an important example: to estimate the the gender wage gap without bias, one has to control for all variables correlated with both gender and wages. Yet, there will always be unobserved factors for which we cannot rule out these correlations. An example are personality traits that could both up- or downward bias the estimation result. The linear fixed effects estimator allows for flexible forms of individual time-constant heterogeneity. Yet, in fixed effects regressions we cannot estimate the coefficients on time constant covariates. The reason is overparametrization. The time-constant heterogeneity in the fixed effect model is a focus in the first chapter of my dissertation. All three chapters in this thesis analyze different forms of heterogeneities which are discussed below. While chapter 2 analyzes heterogeneities in dynamic patterns over time, chapter 3 analyzes heterogeneous treatment effects after a minimum wage introduction.

The first chapter is joint work with Enno Mammen and Ralf A. Wilke. We propose a new estimation approach to identify latent groups within a linear fixed effects panel model. The approach allows for unobserved time-constant heterogeneity. We identify latent groups by combining unsupervised and supervised statistical learning techniques. This leads to a parameter reduction and enables us to include coefficients on time-constant parameters. We allow for correlation between the fixed effect and the covariates. The proposed approach works with large data structures (units and groups). It is applicable to practically relevant models. Compared to existing clustering approaches for panel models we impose

less restrictions on time-constant covariates (compare e.g. Bonhomme and Manresa, 2015) and allow for large datasets (compare e.g. Tutz and Oelker, 2017).

The second chapter analyzes regional dynamic patterns of the COVID-19 pandemic. The pandemic has affected regions in varying intensities. This is evident in several countries, including Germany (Dragano et al., 2021). Regional differences have also changed over time. I apply dynamic time warping, a flexible data driven approach to identify which districts have similar dynamic patterns. I link the dynamic time warping distances to measures of regional connectedness. Especially, I analyze whether highly connected regions have similar dynamic patterns.

A very important topic in the area of labor economics is the analysis of minimum wages. The topic is also highly relevant in the political debate. Especially in Germany that introduced a general minimum wage only in 2015 which affected more than 10% of employees at introduction (Statistisches Bundesamt, 2016). A large body of literature has researched the effect of minimum wages on employment. We focus on the firms' capital intensity. Capital investments have been researched only recently and with mixed evidence (see Harasztosi and Lindner, 2019; Dai and Qiu, 2022; Gustafson and Kotter, 2022). In Chapter 3, which is joint work with Christina Gathmann and Terry Gregory, we estimate the treatment effect of a first-time minimum wage introduction on the capital-labor ratio of firms. We focus on the introduction of industry specific minimum wages. Within our study period no general minimum wage was in place. This enables us to compare treated firms with untreated control firms.

The chapters cover different topics and apply different methodological approaches. A unifying focus is the analysis of heterogeneities. Chapter 1 and 2 focus on unobserved heterogeneities. Chapter 1 and 2 identify latent groups. Chapter 1 uses a linear fixed effect framework. Here a time-constant parameter characterizes heterogeneity. We identify latent groups such that individuals within a group share the same time-constant parameter. Chapter 2 analyzes dynamic patterns using time-series data. Here, heterogeneity has a dynamic component. While Chapter 1 has a methodological focus, Chapter 2 has an empirical focus. Chapter 3 analyzes the treatment effect of a minimum wage introduction on the capital-labor ratio. We estimate an overall effect and look at heterogeneities across different industries. Further, we distinguish firms with high and low capital-labor ratios prior to the minimum wage introduction.

I will now describe each of the three chapters in more detail.

Chapter 1. Data Driven Estimation of Group Structures in Individual Fixed Effects Models

This chapter, joint with Enno Mammen and Ralf A. Wilke, presents a new approach to finding latent groups in fixed effects in the linear panel model. It combines unsupervised clustering with an additional supervised regularisation step. The latter implies for the final estimates to possess the optimality properties of the LASSO. The approach works with large data structures (units and groups) and produces estimates for parameters on time-constant covariates. It is therefore applicable to practically relevant models which are not compatible with existing clustering approaches for panel models. The approach is compatible with the use of different clustering algorithms, we propose to use density based clustering (hdbscan) or k-Means. With the help of simulations we show that the suggested method performs well in finite samples. We show that our estimator is consistent and converges at rate $O_p(1/\sqrt{NT})$ when using density based clustering. The theory characterizes the density based clustering as a nonparametric density estimation. We provide an application where we compute the gender wage gap using administrative data from Germany. The results show that our approach gives sizeable different results than a Mundlak (1978) type estimator that rely on stricter assumptions.

Chapter 2. Regional Patterns of the COVID-19 Pandemic: An Application of Dynamic Time Warping

This paper analyzes regional dynamic patterns in the COVID 19 Pandemic. Applying a data mining method, dynamic time warping, I describe different geographic patterns regarding the Covid-19 pandemic in Germany. The method dynamic time warping has originally been used in the context of speech recognition (e.g. Sakoe and Chiba, 1978). The algorithm measures the distance between time series after aligning the time periods. The algorithm first decides which time period in one time series should be compared to which time period in the other time series and thereby accounts for time delays and differences in speed (e.g. Müller, 2007). Further, I link these dynamic distances to different regional characteristics and economic connections between German regions. Personal travel patterns as well as interregional trade are associated with similar dynamic patterns between the respective regions. These relationships are also present conditional on geographic distance. Results suggest that the importance of different economic network struc-

tures differs across the first, second and third wave. During the first wave, in which many businesses were closed, similar dynamic patterns are associated with private networks. During the third wave this is the case for both private and economic networks. In the second wave geographic distances plays a role.

Chapter 3. Do Minimum Wages Encourage Capital Deepening?

Much of the minimum wage literature has focused on employment or displacement effects. Most studies find no evidence for large negative employment effects (Manning, 2021). Also for the German national minimum wage recent studies suggest no or very small displacement effects (see Dustmann et al., 2022; Caliendo et al., 2018). These findings can point towards other important adjustment channels. Capital investments have been researched only recently. Some studies find capital deepening (see Harasztosi and Lindner, 2019; Dai and Qiu, 2022), while others find declining capital investments (Gustafson and Kotter, 2022). We investigate whether the first time introduction of a minimum wage spurs capital investments, encourages automation and outsourcing. To answer these questions, we analyze rich firm-level balance sheet data, the Dafne dataset provided by Bureau van Dijk. We exploit the first-time introduction of industry-specific minimum wages that were introduced in Germany prior to the national minimum wage in 2015. This setup has advantages. Because the investment in capital is a long term decision firms may react differently to a first time introduction than to incremental changes of a minimum wage. The institutional setup further enables us to contrast firms who are covered by an industry minimum wage with suitable control firms that operate in closely related industries but did not introduce a minimum wage. In our empirical approach, we combine a matching strategy with a flexible event study estimation. On average across industries we only find weak evidence for capital-labor substitutions. Yet, capital intensive industries do respond by increasing their capital intensity. The adjustment is driven by firms with low capital intensities prior to minimum wage introduction. Among firms entering after the minimum wage introduction we see similar adjustments, though not statistically significant. We complement the analyzes with descriptive evidence about the task content in the industries. The positive adjustments take place in industries with higher routine task shares.

Chapter 1

Data Driven Estimation of Group Structures in Individual Fixed Effects Models

Joint with Enno Mammen and Ralf A. Wilke.

1.1 Introduction

Panel data are characterised by high dimensionality due to having both cross-sectional (N units) and longitudinal (T time periods) dimensions. Panel analysis is appealing because it gives consistent estimates under weaker restrictions on the correlation between observables and unobservables than cross-sectional analysis. The leading example is the so-called linear fixed effects (FE) model, where observables can be arbitrarily correlated with time-constant unobservables. A disadvantage of the FE model is that it is overparametrised in the presence of time-constant covariates. While parameters on time-varying covariates are identifiable and are consistently estimated by the classical FE estimator, or by Mundlak (1978) type estimators (compare Wooldridge, 2019), the parameters on the time-constant covariates are not identifiable in the FE model and Mundlak (1978) type models substantially restrict their correlation with the FEs. Combining machine learning methods to regularise the space of fixed effects is tricky, as by increasing the number of units, the number of parameters in the model increases and therefore it is different from regularisation in a given parameter space. The overparametrisation also leads to multicollinearity, which causes problems for regularisation methods.

This paper suggests a new approach to regularising the space of FE in the

linear panel model by means of a combination of un- and supervised learning techniques. Our approach is attractive to practitioners for the following reasons: It works with a large number of units and the unknown number of groups can be endogenously determined. It gives estimates for parameters on time-constant covariates, which can be arbitrarily correlated with the FEs. Our approach is compatible with different regularisation methods, which makes it flexible and adaptable in practice. We show that our estimator is consistent and converges at rate $O_p(1/\sqrt{NT})$. The theory characterises the density-based clustering as a nonparametric density estimation problem that allows for an unknown number of groups and the existence of atoms, i.e. observations not belonging to any group. For estimation in practice we propose an estimator using density based clustering which includes heuristic density estimation and is therefore compatible with the provided theory.

Our estimation approach overcomes important shortcomings of existing approaches, which render them impracticable in many applications as they are not compatible with large sample sizes (Bonhomme and Manresa, 2015; Tutz and Oelker, 2017; Tutz and Schauberger, 2015) or correlations between time-constant covariates and FEs (Berger and Tutz, 2018; Bondell et al., 2010; Bonhomme et al., 2022; Fan and Li, 2012; Heinzl and Tutz, 2014; Li et al., 2018; Schelldorfer et al., 2011; Rohart et al., 2014; Su et al., 2016). While Bonhomme and Manresa (2015) allow for such correlations, they require time-constant covariates to take on sufficiently many values and sufficient variation in time-varying covariates. These restrictions mean that there must be time-varying covariates and the set of time-constant covariates must not only consist of a couple of dummy variables. Both are not required in our approach. Other advantages of our approach are that it is capable of endogenously determining the number of groups, it works well with large group numbers and large N without computation time becoming too extensive. The key restrictions of the relevant approaches are listed in Table 1.1. In addition, our approach benefits from an efficiency-ensuring final regularisation step that eliminates redundant structures in group FE patterns. We show the equivalence to a generalised LASSO problem, in particular a fused LASSO. Transforming the problem into a regular LASSO leads to large increases in computational efficiency and ensures that final estimates possess statistical optimality properties of the LASSO such as the Oracle property. Our combination of unsupervised and supervised machine learning methods therefore leads to new insights that can be obtained neither with conventional FE panel models, nor with the existing clustering approaches for linear FE panel models. We conduct a series of Monte

Carlo simulations to provide evidence of the approach producing reliable estimates in the case of small T ($T = 5$). It is shown that our approach works reasonably well in very short panels, although the quality of the estimates increases with T.

Table 1.1: Comparison of Panel FE models with group structure

Paper	Regularisation Method	time-constant covariates	G	large N
Bonhomme and Manresa (2015)	k-Means Clustering	(yes)	known &small	no
Tutz and Schauberger (2015)	Adaptive LASSO	yes	unknown &small	no
Tutz and Oelker (2017)	Adaptive LASSO	no	unknown &small	no
Berger and Tutz(2018)	Tree-Structured Clustering	yes	unknown &small	yes
Bonhomme, Lamadon and Manresa (2022)	k-Means Clustering	no	unknown	yes
Mammen,Wilke and Zapp (2022)	k-Means or Density-based Clustering Fused LASSO	yes	unknown	yes

Notes: Overview of Related Literature. Mammen, Wilke and Zapp (2022) denotes this chapter. G denotes the number of groups, N the number of units.

The paper is structured as follows. Section 1.2 presents the model and the statistical approach. Section 1.3 presents simulations results to investigate finite sample performance, while Section 1.4 presents the results from an application to labour market data. The last section summarises the main findings and derives some recommendations.

1.2 The Model

We consider the linear FE panel model

$$\begin{aligned} y_{it} &= \boldsymbol{W}_{it}\boldsymbol{\theta} + v_i + u_{it} \\ &= \boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma} + v_i + u_{it}, \end{aligned} \tag{1.1}$$

where $i = 1, ..., N$ is the unit and $t = 1, ..., T$ is the time period. $\boldsymbol{W}_{it}' = (\boldsymbol{X}_{it}, \boldsymbol{Z}_i)' \in \mathcal{W} \subset \mathbb{R}^K$ are observable covariates, where $\boldsymbol{X}_{it}' \in \mathcal{X} \subset \mathbb{R}^{K_1}$ are time-varying

covariates which may be continuous or discrete. $\boldsymbol{Z}_i' \in \mathcal{Z} \subset \mathbb{R}^{K_2}$ are time-constant discrete covariates. For simplicity and to be able to focus on the main approach, we outline the model for discrete $\boldsymbol{Z}_i$, although an extension to continuous is possible and outlined under model extensions. $\boldsymbol{Z}_i$ can take on finitely many values $\mathcal{Z}$. Only $y_{it}, \boldsymbol{W}_{it}$ are observed, while $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}) \in \mathbb{R}^{\mathrm{K}}$ is unknown. v_i is an unknown fixed effect and u_{it} is an unknown idiosyncratic error. The objective is to identify and estimate $\boldsymbol{\theta}$. Following general convention, we assume $E(u_{it}|\boldsymbol{W}_i, v_i) = 0$, where $\boldsymbol{W}_i = (\boldsymbol{W}_{i1}', ..., \boldsymbol{W}_{iT}')'$, i.e. $\boldsymbol{W}_{it}$ is strictly exogenous conditional to v_i. $v_i \in \mathbb{R}$ are not mean restricted, because we consider a variant of the model without a common intercept. The fixed effect is generated by the sum of a continuous and a discrete random variable. The discrete random variable can take on $g = 1, \ldots, G_1$ values which correspond to groups or clusters. Within a group g all observations have the same value for $v_i = q_g$. The continuous part adds additional noise to the fixed effect, which causes G_2 of the units not to belong to any cluster and are therefore atomic. Let G_1 be unknown but not random. $G_1 + G_2 = G < N - K_2$ is required for identifiability. Further, for each $Z \in \mathcal{Z}$ in the subpopulation with $\boldsymbol{Z}_i = Z$ at least two non-atomic groups have to present, and for one level of $\boldsymbol{Z}_i$ all G_1 non-atomic groups have to be present (compare Assumption (A7) below). The distances $q_l - q_{l+1} = v_i - v_j \forall i \in$ group l and j in group $l+1$ are assumed to be pairwise different for $1 \leq l \leq G_1 - 1$. There is no restriction on the correlation between v_i and $\boldsymbol{W}_{it}$, with the exception that they should not be perfectly collinear. This correlation makes the latter endogenous. It is important in most applications with data on human or economic activity to allow for endogeneity due to for example correlated omitted variables. The difficulty is that G_1, the group association for each i and the v_is are unknown. Evidently, if all i were atomic, there is an identification problem in the case the set of $\boldsymbol{Z}_i$ was non empty. The classical FE panel model assumes $G_2 = N$ atomic groups and therefore cannot identify $\boldsymbol{\gamma}$ and v_i but only the sum $\boldsymbol{Z}_i\boldsymbol{\gamma} + v_i$. This is because the model is overparametrised, leading to multicollinearity between the time-constant $\boldsymbol{Z}_i$ and v_i. Therefore, though the model permits for general forms of endogeneities, the interpretability of the results is unclear as only $\boldsymbol{Z}_i\boldsymbol{\gamma} + v_i$ can be identified and not the role of the components $\boldsymbol{Z}_i$. This is a severe limitation in applications, when the focus is on time-constant variables, such as geographic factors or gender. This paper suggests a new approach for identifying $\boldsymbol{\gamma}$ using regularisation of v_i in the presence of endogeneities.

1.2.1 Group FE Estimation: Informal Presentation

The point of departure is that $\boldsymbol{\gamma}$ cannot be estimated by the FE estimator due to overparametrisation. In this subsection we provide an intuitive presentation of the main ideas and steps of our approach to estimating $\boldsymbol{\gamma}$, which consists of three main steps. The exact step by step description of our estimation procedure is given in Appendix 1.A.

Step 1: Clustering

a) There is some evidence that FE estimation of model (1.1) provides well behaved estimates for $\boldsymbol{Z}_i\boldsymbol{\gamma} + v_i$ as N and T become large and if there is an underlying group structure Hahn and Moon (2010). In a classical individual level FE model, this term is the fixed effect. We denote the estimator of this individual level FE as $a_i = \boldsymbol{Z}_i\boldsymbol{\gamma} + v_i + e_i$, where e_i is the estimation error in a_i with $E(e_i|v_i, \boldsymbol{Z}_i) = 0$ and plim $e_i \to 0$ for $N, T \to \infty$. Simulations show that convergence in T is quick and not many periods are required. Consistent estimation of $\boldsymbol{\gamma}$ in the model for a_i by OLS is not possible because the relationship between $\boldsymbol{Z}_i$ and v_i is unrestricted and therefore $\boldsymbol{Z_i}$ is endogenous. There is, however, one identifying property that holds: $a_i = \mathbf{0}\boldsymbol{\gamma} + v_i + e_i = v_i + e_i$. This means that for the subpopulation in the model with $\boldsymbol{Z}_i = \mathbf{0}$, we can identify v_i, although $\boldsymbol{\gamma}$ is unknown. We apply statistical learning methods, such as density-based clustering, to cluster a_i with $\boldsymbol{Z}_i = 0$ into latent groups. Depending on the chosen clustering algorithm, the number of groups is endogenously determined (e.g. HDBSCAN, G_1 is estimated) or known (e.g. k-Means, G_1 is fixed). See Appendix 1.C for more details about the clustering procedures. A numerical illustration of HDBSCAN and k-Means is given in Appendix 1.I.

b) Step 1a is repeated for $|\mathcal{Z}|$ reparametrisations of the model for y_{it}, where $\boldsymbol{Z}_i$ is transformed such that it has a different set of reference observations with $\boldsymbol{Z}_i = 0$. This is possible as $\boldsymbol{Z}_i$ is discrete and a reparametrisation will change $\boldsymbol{\gamma}$ accordingly. In practice this means that we divide the set of N observations into subsets with the same value of $\boldsymbol{Z}_i$. On each of these subsets we perform the clustering step separately: For each of the different values for $\boldsymbol{Z}_i$, i.e. for each $Z \in \mathcal{Z}$ we obtain and cluster the estimates a_i with $a_i = c_Z + v_i + e_i$. This implies that after the clustering we know whether observations with the same $\boldsymbol{Z}_i$ are in the same or distinct groups but cannot compare the clustering outcome across levels of $\boldsymbol{Z}_i$ due to the unknown c_Z which are unknown location shifts.

Step 2: Mapping of Cluster Membership Variables

If in all reparametrisations of the model the clustering produces the same number of groups with the same distances between them, the correspondence between the groups across levels of $\boldsymbol{Z}_i$ simplifies to sorting clusters by size of the corresponding mean a_i. Otherwise an algorithm is required that creates a mapping. We assume that there is at least one realisation Z of $\boldsymbol{Z}_i$, a reference level, such that the subset of observations with $\boldsymbol{Z}_i = Z$ contains all non-atomic groups $g = 1, \ldots, G_1$ and the distances $q_l - q_{l+1} = v_i - v_j \forall i$ in group l and j in group $l+1$ are assumed to be pairwise different for $1 \leq l \leq G_1 - 1$. The identified group patterns in the distribution of v_i for this reference level subpopulation is then informative for the unconditional distribution. In particular, it identifies G_1, the ordering and distances between all latent groups/clusters. It is then possible to match the groups/clusters from all other possible reparametrisations (i.e. all other realisations of $\boldsymbol{Z}_i$) to these reference groups, using the ordering and distance information for the group structures. This mapping also permits for incomplete group structures in subpopulations other than the reference, e.g. there can be less groups for some realisations of $\boldsymbol{Z_i}$. The mapping algorithm is described in detail in Appendix 1.D.

Step 3: Dummy Variable Regression for the Regularised Model

After the correspondence between the reference groups and the groups in all reparametrised models is established, a vector of G dummy variables $\hat{\boldsymbol{D}}$ indicating estimated group membership is created and the following model is established:

$$y_{it} = \boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma} + \hat{\boldsymbol{D}}_i\boldsymbol{\alpha} + \tilde{u}_{it}, \tag{1.2}$$

where $\boldsymbol{\alpha}$ has $\hat{G}$ components for the fixed effects of each estimated group. $\tilde{u}_{it} = u_{it}$ iff $\hat{\boldsymbol{D}}_i\alpha = v_i$. This regression produces consistent estimates for previously unidentified components $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$. The estimates for β are more precise than in the conventional FE model due to restricting the fixed effects to take on $\hat{G}$ values.

$\hat{\boldsymbol{D}}_i$ may not be free of error in applications. That is it may not be the same as $\boldsymbol{D}_i$ based on the true groups. We provide a discussion of how estimation error in $\hat{\boldsymbol{D}}_i$ affects estimated $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ in 1.G. When the algorithm in Step 1 creates too many groups, an additional supervised regularisation in Step 3 will remove inefficiencies. We show in Appendix 1.E that the underlying problem is a generalised LASSO and that it can be transformed into a regular LASSO as in Tibshirani and Taylor (2011). Therefore, our approach benefits from existing computational advantages and the resulting estimates inherit desirable statistical properties of the LASSO such as

the Oracle property. The properties of the regular LASSO are well developed, see for example Tibshirani (1996) and Hastie et al. (2017). While our approach is a variant of the fused LASSO, an alternative shrinkage approach would be pairwise cross-smoothing as suggested by Heiler and Mareckova (2021).

1.2.2 Large Sample Properties

In our asymptotic approach we assume that N and T converge to infinity. More formally, we assume that $N \to \infty$ and that $T = T_N$ depends on N and fulfils $\lim_{N\to\infty} T_N = \infty$. We discuss asymptotics for our approach by characterising the density-based clustering as a nonparametric kernel density estimation problem.

In Assumption (A5) below we will state the model for the distribution of v_i. We assume that with a positive probability v_i is generated from a continuous distribution. Further, with positive probability v_i takes a value from the finite set $\{q_g : g \in \{1, \dots, G_1\}\}$, where q_g are some unknown real numbers. Thus we have a fraction of v_is spread over the real line and fractions of v_is equal to q_g for some $g \in \{1, \dots, G_1\}$. We call the first v_is "atoms" and we call the index sets $\{i : v_i = q_g\}$ "clusters". In our asymptotic setting we allow that the probabilities of both fractions do not converge to zero. Then we will have that the number of atoms as well as the number of cluster points are of order N.

We suppose that an estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is given which fulfils $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| = O_p((NT)^{-1/2})$, see Assumption (A1).

Put $a_i = \bar{y}_i - \bar{\mathbf{X}}_i \hat{\boldsymbol{\beta}}$, where $\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$, $\bar{\boldsymbol{X}}_i = T^{-1} \sum_{t=1}^{T} \boldsymbol{X}_{it}$.

We define the kernel density estimators

$$\hat{f}_b^Z(x) = \frac{1}{N_Z} \sum_{i=1}^{N} \mathbb{1}[\boldsymbol{Z}_i = Z] \frac{1}{b} K\left(\frac{a_i - x}{b}\right),$$

with $N_Z = \#\{i : \boldsymbol{Z}_i = Z\}$. We consider high level sets of $\hat{f}_b^Z$ and correct their boundaries

$$\begin{aligned} I_*^Z &= \{x : \hat{f}_b^Z(x) \geq c_2^b \frac{1}{b}\}, \\ I^Z &= \{x : |x - w| \leq c_3^b b \quad \exists w \in I_*^Z\} \text{ for some constants } c_2^b, c_3^b > 0. \end{aligned}$$

We will show that I^Z is a union of disjoint closed intervals

$$I^Z = I_1^Z \cup \ldots \cup I_{l(Z)}^Z \text{ with } l(Z) \leq G_1,$$

where G_1 is the number of clusters. Furthermore, with probability tending to one, each interval I_j^Z contains exactly $q_g + Z\boldsymbol{\gamma}$ for exactly one $1 \leq g \leq G_1$. We also write $I^{g,Z}$ for this interval. If there exists no j with $q_g + Z\boldsymbol{\gamma} \in I^{g,Z}$ we define $I^{g,Z} = \varnothing$.

Under our assumptions one can identify the value of g with $I_j^Z = I^{g,Z}$ with probability tending to one, see Assumption (A7).

For i with $a_i \in I^{g,\boldsymbol{Z}_i}$ for some $1 \leq g \leq G_1$ we denote this value of g by $g(i)$. For indices i with a_i contained in no interval $I^{g,\boldsymbol{Z}_i}$, we define $g(i) = G_1 + i$.

We now define the estimator $\hat{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ as the minimiser over $\boldsymbol{\gamma}$ of

$$\min_{\bar{a}_g : g \geq 1} \sum_{i=1}^{N} (a_i - \bar{a}_{g(i)} - \boldsymbol{Z}_i \boldsymbol{\gamma})^2.$$

It can be easily checked that

$$\hat{\boldsymbol{\gamma}} = \left(\sum_{i=1}^{N} (\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_{g(i)})' (\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_{g(i)}) \right)^{-1} \sum_{i=1}^{N} (\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_{g(i)})' (a_i - \bar{a}_{g(i)}),$$

where

$$\begin{aligned} \bar{\boldsymbol{Z}}_g &= \sum_{i=1}^{N} \boldsymbol{Z}_i \mathbb{1}[g(i) = g] / \sum_{i=1}^{N} \mathbb{1}(g(i) = g), \\ \bar{a}_g &= \sum_{i=1}^{N} a_i \mathbb{1}[g(i) = g] / \sum_{i=1}^{N} \mathbb{1}(g(i) = g). \end{aligned}$$

Note that

$$\hat{\boldsymbol{\gamma}} = \left(\sum_{i : 1 \leq g(i) \leq G_1} (\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_{g(i)})' (\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_{g(i)}) \right)^{-1} \sum_{i : 1 \leq g(i) \leq G_1} (\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_{g(i)})' (a_i - \bar{a}_{g(i)}).$$

We now state our result for the rate of convergence of $\hat{\boldsymbol{\gamma}}$.

Theorem 1 *Make assumptions (A1)-(A7) stated below and assume that* $T^{-1} = O(1/\sqrt{N})$. *Then it holds that*

$$\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} = O_p(1/\sqrt{NT}).$$

The proof of Theorem 1 is given in Appendix 1.F.

There can be two types of clustering errors in our asymptotic setting. First, atoms with a value of v_i in the neighbourhood of a cluster centre q_g are accidentally

assigned to the cluster. Second, cluster elements with large enough $\bar{u}_i$, with $\bar{u}_i = T^{-1}\sum_{t=1}^{T} u_{it}$, are accidentally classified as atoms. Both errors do not disappear and they remain to be present in the limit. The first error leads in particular to bias effects which are shown to be of second order in our proof of Theorem 1. The second type of errors leads to a loss of efficiency. Note that nevertheless our estimator of $\boldsymbol{\gamma}$ achieves the same rate $O_p(1/\sqrt{NT})$ as if all clusters were known. The assumption $T^{-1} = O(1/\sqrt{N})$ could be weakened if the relative number δ_N of atoms converges to 0 or, more explicitly if we replace α_0^Z and α_g^Z in Assumption (A5) by $\delta_N\alpha_0^Z$ or $\alpha_g^Z(1-\delta_N\alpha_o^Z)/(1-\alpha_0^Z)$, respectively. Then the assumption that $\delta_N T^{-1} = O(1/\sqrt{N})$ would suffice.

If T converges slower to $+\infty$ as $N^{-1/2}$ the asymptotic bias of $\hat{\boldsymbol{\gamma}}$ is not negligible because it is of order $T^{-3/2}$, a rate which is not of order $O(1/\sqrt{NT})$. The bias term would vanish if the density of $\bar{u}_i$ is symmetric which in general may not be the case. We assume that $\bar{u}_i$ is the sum of T i.i.d. variables and thus its distribution differs from a symmetric distribution by a distance of order $T^{-1/2}$. Theoretically, it is possible to weaken the assumptions further by applying an approach that corrects for bias. In particular, one can construct an estimator of $\boldsymbol{\gamma}$ that achieves the $O_p(1/\sqrt{NT})$ rate under the assumption $T^{-3/2} = O(1/\sqrt{N})$ or $\delta_N T^{-3/2} = O(1/\sqrt{N})$, respectively. We do not report on this approach because its practical success would heavily depend on the finite sample accuracy of the used higher order expansions.

Assumptions

Assumption A1 *There exists an estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ with $\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\| = O_p((NT)^{-1/2})$. The values of $\boldsymbol{X}_{it}$ lie in a bounded set: $\|\boldsymbol{X}_{it}\| \leq C^x$ a.s. for some $C^x > 0$.*

With this assumption there is no need to discuss estimation of $\boldsymbol{\beta}$. A possible choice for an estimator that fulfils this condition is the fixed effects estimator.

Assumption A2 *The tuples $(\boldsymbol{Z}_i, v_i, \bar{u}_i)$ are i.i.d. It holds that for $1 \leq i \leq N$ the variables $u_{i1}, \ldots, u_{iT}$ are i.i.d. with $Eu_{it} = 0$ and $E|u_{it}|^3 < +\infty$.*

We obtain for the distribution function F_N of $\sqrt{n}\bar{u}_i$ by an application of the Berry-Essene bound (Feller, 1971):

$$\sup_{u\in\mathbb{R}}|F_n(u) - \Phi(u/\sigma)| \leq C_F T^{-1/2}$$

for some $C_F > 0$ with $\sigma^2 = Eu_{it}^2$. Here Φ is the distribution function of the standard normal distribution. We will exploit below that the density of Φ is symmetric.

We now describe the distribution of $(\boldsymbol{Z}_i, v_i, \bar{u}_i)$.

Assumption A3 *$\bar{u}_i$ is independent of $(\boldsymbol{Z}_i, v_i)$.*

Assumption A4 *The $\boldsymbol{Z}_i$s have a finite support $\mathcal{Z} \subset \mathbb{R}^{d_Z}$. The linear span of $\mathcal{Z}$ is equal to $\mathbb{R}^{d_Z}$. For $Z \in \mathcal{Z}$ we put $p(Z) = P(\boldsymbol{Z}_i = Z)$. We suppose that p does not depend on N.*

We now describe the conditional distribution of v_i given $\boldsymbol{Z}_i$.

Assumption A5 *The conditional distribution of v_i given $\boldsymbol{Z}_i = Z \in \mathcal{Z}$ is equal to*

$$\alpha_0^Z S^Z + \sum_{g=1}^{G_1} \alpha_g^Z \delta_{q_g},$$

where δ_q denotes a mass point in q, where $q_1 > \ldots > q_{G_1}$ are points in [0,1] and $S^Z (Z \in \mathcal{Z})$ are probability measures on [0,1] with densities s^Z that allow for a continuous derivative. Furthermore, α_g^Z $(Z \in \mathcal{Z}, 0 \leq g \leq G_1)$ are real weights with $\alpha_g^Z \geq 0$, $\sum_{g=0}^{G_1} \alpha_g^Z = 1$ for all $Z \in \mathcal{Z}$, $\sup_{Z \in \mathcal{Z}} \alpha_g^Z > 0$ $\forall g$.

In (A5) we allow that $\alpha_g^Z = 0$ for some $(g, Z) \in \{0, \ldots, G_1\} \times \mathcal{Z}$. In (A5) we assume that the order of the number of atom units is the same as that for units in clusters. One could change and model the conditional distribution of v_i given $\boldsymbol{Z}_i = Z$ as

$$\delta_N \alpha_0 S^Z + (1 - \delta_N) \sum_{g=1}^{G_1} \alpha_g^Z \delta_{q_g},$$

where δ_N is a sequence with $\delta_N \to 0$. By doing so we can replace the assumption $T^{-1} = O(1/\sqrt{N})$ by $\delta_N T^{-1} = O(1/\sqrt{N})$, or in case we do assume that the density f_N of $\bar{u}_i$ is symmetric, see comment after the statement of (A2), by $\delta_N T^{-3/2} = O(1/\sqrt{N})$.

For the kernel K we assume:

Assumption A6 *The kernel K is a strongly unimodal symmetric density function and differentiable with derivative absolutely bounded by c_1^K. For the bandwidth b it holds $b = c_1^b/\sqrt{T}$ for some c_1^b. The constant c_2^b in the definition of I_*^Z is chosen small enough.*

Note that a density is strongly unimodal if its convolution with a unimodal density is always unimodal. This also implies that the density of the convolution of two strongly unimodal densities is also strongly unimodal. For a density strong unimodality is equivalent to log-concavity. In particular, normal densities are

strongly unimodal. Below we will make use of the fact that the convolution of the kernel K with a normal density is strongly unimodal and thus log-concave. For a discussion of strongly unimodal densities see Ibragimov (1956).

The following assumption simplifies identification of clusters and their centres.

Assumption A7 *There exists a* $Z_* \in \mathcal{Z}$ *with* $\alpha_g^{Z_*} > 0$ *for all* $1 \leq g \leq G_1$*. For all* $Z \in \mathcal{Z}$ *we assume that there exist* $g_1, g_2 \in \{1, \ldots, G_1\}$*,* $g_1 \neq g_2$*, depending on* Z *with* $\alpha_{g_1}^Z, \alpha_{g_2}^Z > 0$*. Furthermore, we suppose that the values of* $q_{g_1} - q_{g_2}$ *are pairwise different for* $1 \leq g_1 < g_2 \leq G_1$*.*

We conjecture that Assumption (A7) could be weakened but this would require more refined statistical methods and the application of more technical arguments in the mathematical analysis. Note that we identify clusters for each value of $\boldsymbol{Z}_i$ separately without making use of the link $(Z_{i_1} - Z_{i_2})\boldsymbol{\gamma}$. Including this information may motivate more effective approaches if $\mathcal{Z}$ contains more than 2 elements.

1.2.3 Comparison of Approaches

We now summarise how the various approaches differ in terms of restrictions. There are no restrictions on the correlation between v_i and $\boldsymbol{Z}_i$ in the FE model. Moreover, v_i can take on N values. This corresponds to that all units are allowed to be atoms. $\boldsymbol{\gamma}$ is not identified without restricting G or the relationship between the fixed effect and the observables. The model by Mundlak (1978) assumes that the fixed effect is a function of the time average of time-varying covariates ($\bar{\boldsymbol{X}}_i$) and that its residual variation is not correlated with $\boldsymbol{Z}_i$. The model produces inconsistent estimates if there is anything in v_i that is related to $\boldsymbol{Z}_i$ conditional on $\bar{\boldsymbol{X}}_i$. The model does not restrict the range of values that v_i can take, therefore it only restricts the correlation structure of the observables with v_i.

Our suggested approach does not restrict the correlation structure between observables and v_i but for identifiability G, the number of values that v_i can take. In the case of k-Means clustering it is assumed that this is a known small number. In the case of density-based clustering, the number of points, G, is unknown and can be large. For identifiability, in each realisation of $\boldsymbol{Z}_i$ (or more specifically for each distinct realisation of the combination of the K_2 time-constant variables together) at least two non-atomic clusters must be present in the dataset. However, our theory explicitly allows that groups are not present in some realisations of $\boldsymbol{Z}_i$ (compare Assumption (A5)). Assumption (A7) ensures that there is one realisation Z that contains all non-atomic groups. We conjecture that this latter assumption

could be relaxed both in the theory as well as in the practical estimation. In regard to theory, this would require both more refined statistical methods and the application of more technical argumentation in the mathematical analysis. The different groups across realisations of $\boldsymbol{Z}_i$ are computed endogenously in our estimator when using density-based clustering as in the provided theory. Using k-Means requires the clustering of observations into the same number of groups across all Z levels unless specific knowledge is available to the researcher regarding those different numbers of groups. It is therefore possible to characterise the various models. When using density-based clustering, no specific assumption is made on the number of groups and groups are allocated by a mechanism that has a nonparametric spirit. Given that there are parametric components in Model (1.1), we characterise this approach as semiparametric. In contrast, when using k-Means clustering, the model is closer to a parametric model as the number of groups is fixed and known. Further, the k-Means approach assumes that the clustering structure can be estimated by estimating the cluster means (Campello et al., 2020). The Mundlak model is parametric as the number of parameters is fixed and small. In Section 1.3 we show with simulations that as long as the restrictions are satisfied, the more restrictive models are more efficient, while they are biased when restrictions are violated. This is the usual trade-off one faces in terms of bias and efficiency. On the grounds of these considerations it would be of interest to formulate inference approaches that test for the validity of restrictions. For example, a Hausman type test could be done to test for the validity of the additional restrictions of the k-Means clustering in comparison to density-based clustering.

1.2.4 Remarks and Extensions

The following two remarks should be of use for practitioners who apply our approach:

- In step 1 it does not matter what is the initial parametrisation of model (1.1), i.e. what is $\boldsymbol{Z}_i = \mathbf{0}$. After the group membership v_i is determined, Step 3 can be done for any re-parametrisation of the component $\boldsymbol{Z}_i\boldsymbol{\gamma}$.

- $P(\boldsymbol{Z}_i = Z|g_i = g)$ can become low for some values of Z in the case of strong correlation between v_i and $\boldsymbol{Z}_i$. In this case a large data set may be required for the algorithm to detect a cluster.

- While from a theoretical point of view, $\boldsymbol{Z}_i$ can be high dimensional, there

are practical constraints as the clustering step 1 has to be done conditional for all values of $\boldsymbol{Z}_i$. The applied researcher is advised to include only low dimensional $\boldsymbol{Z}_i$ of key interest. The remaining time-constant variables will be simply absorbed by v_i.

There are several practically relevant extensions to our model that we omitted to focus on the main idea of our approach:

Multi-level models. Linear multi level models are routinely applied in a wide range of applications. Our model can be extended to multi-level fixed effects, e.g. $v_i + f_j$ in the case of two levels. They comprise of, for example, a regional or firm component f_j in addition to v_i. Higher dimensional density clustering methods can be used for regularisation in Step 1.

Continuous $\boldsymbol{Z}_i$. In the case $\boldsymbol{Z}_i$ contains one or multiple continuous covariates, we suggest a pragmatic approximation by specifying the partial relationship of the continuous time-constant covariates and y_{it} as piecewise constant model (interval dummies).

Continuous v. The fixed effects could be continuously distributed with unknown distribution. This is likely the case in many empirical applications. Forcing them into groups, will lead to an approximation error. The problem is similar to that considered in Bonhomme et al. (2022) who show that incorrect grouping of similar units will not or will only slightly bias estimates. Our simulation results in Section 1.3 confirm that incorrect grouping of similar units will only lead to small inconsistencies in coefficients of interest.

Further regularisation step to group atomic units. The clustering in Step 2 of our procedure typically produces atoms in applications. These are units that are not clustered with any other unit. In addition to the supervised regularisation to combine groups as outlined in Appendix 1.E, it is possible to test whether atoms are different from groups or other atoms. In this case they can be combined or merged into existing groups to further reduce the dimensionality of the model. The starting point is that a dummy variable regression model as in equation (1.2) after the generalised LASSO (outlined in Appendix 1.E), will give estimated fixed effects for groups and atoms. On the grounds of these estimates it is possible to determine the nearest neighbours for each atom. Using a Wald test or a t-test based on a reparametrised model it is possible to test the null that the FE of the

atom and its nearest neighbour are the same. If the null is not rejected, the two are to be merged into one group.

Inference. After the transformation into a regular LASSO as outlined in Appendix 1.E, the distribution of the transformed parameters is unfortunately not directly informative about the distribution of $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}})$. What can be conveniently conducted is post LASSO inference. In this case, groups are merged as indicated by the LASSO step and Model (1.2) is estimated with a dummy variable regression using the reduced group structure.

For honest inference it would be important to take the uncertainty of the regularisation steps into account. Chatterjee and Lahiri (2011) and Chatterjee and Lahiri (2013) suggest residual based bootstrap methods for high dimensional linear regression models that are valid for sparse estimators. Our estimation procedure additionally involves an unsupervised clustering step, but it would be of interest to develop a residual based bootstrap procedure that produces valid standard errors and p-values.

1.3 Simulations

We conduct a series of Monte Carlo simulations to investigate the numerical performance of the approaches described in Section 1.2. Table 1.2 summarises the 5 designs M1-M5 that we simulate. They differ in terms group structures of fixed effects and correlation structures between observables and the FEs and are aligned to designs in the related literature. Designs M1 and M2 follow Berger and Tutz (2018) and Tutz and Oelker (2017) and are characterised by five distinct latent groups, high correlation between the time-varying covariate and the group intercept and high error variance. Setting M3 bases on Bonhomme et al. (2022, Supplementary Appendix S3), where the latent groups are only approximately present and clustering is a means of discretisation. We use a linear model and include a time-constant bivariate variable $\boldsymbol{Z}_i$ in contrast to Bonhomme et al. (2022, Supplementary Appendix S3) such that the DGP fits to our approach. In the mixed design (M4) both approaches are combined. Design M5 combines varying group sizes with larger differences between group intercepts. In all design we include a bivariate time-constant covariate. In terms of Assumption (A5) M1 is characterised by $\alpha_0^Z = 0$ for all Z and $\alpha_g^Z = 1/G = 1/G_1$ for all G clusters and all Z levels. In design M5 $\alpha_0^Z = 0$ and α_g^Z varies in g but not in Z while in M2 α_g^Z varies both across groups and Z. M3 and M4 include atoms, i.e. $\alpha_0^Z \neq 0$. In

setting M3 even all observations are atoms: $a_0^Z = 1$ for all Z. We note that this extreme case of no group structure present does not fulfil Assumption (A7).

Table 1.2: Simulation Designs

Design	**M1**	**M2**	**M3**	**M4**	**M5**
Group Structure					
adapted from	B&T(2018) & T&O(2017)	B&T(2018) T&O(2017)	B,L&M(2022)	Mixed	
G	5	5	N	$G_1 = 5$ $G_2 = N/2$	5 varying sizes
N	500	500	500	1000	1000
T	20	20	20	20	20
Fixed Effect v_i					
drawn from	$N(1,2)$	$N(1,2)$	$N(0,1)$	$N/2 \sim N(1,2)$ $N/2 \sim N(0,1)$	$U(G_5)$
discretised	5 quantile means	5 quantile means	none	N/2: 5 q. means $N/2$: none	yes
Time-constant covariate $\boldsymbol{Z}_i$	$B(0.5)$	$P(\boldsymbol{Z}_i = 1\|v_i) = P_2$	$B(0.5)$	$B(0.5)$	$B(0.5)$
Time-varying x_{it} **covariate**	$0.4v_i + 0.6N(0,1)$	$N(0,1)$	$N(0,1) + v_i$	$0.4v_i + 0.6N(0,1)$	$N(0,1)$
β, γ	2,2	2,2	1,1	2,2	2,2
Correlation structure	$cor(v_i, x_{it})$ ≈ 0.8	$cor(v_i, \boldsymbol{Z}_i)$ > 0	$cor(v_i, x_{it})$ ≈ 0.7	$cor(v_i, x_{it})$ ≈ 0.8	none
Error term u_{it}	$N(0,3)$	$N(0,3)$	$N(0,1)$	$N(0,3)$	$N(0,3)$

Notes: B&T(2018): Berger and Tutz (2018),T&O(2017): Tutz and Oelker (2017), B,L&M (2022): Bonhomme et al. (2022). Vector $G_5 = [[-15,-14],[-2,-1.5],[1.5,2.5],[6,8.5],[13.5,14.5]]$, Vector $P_2 = (0.35, 0.45, 0.55, 0.55, 0.65)$

We apply different variants of our estimation approach in order to compare their performances. In particular, we use different clustering techniques such as k-Means and HDBSCAN with and without the LASSO regularisation. In the LASSO step we use different criteria for determining the tuning parameter: BIC, Cross Validation and General Cross Validation. We also compute Post LASSO after Cross Validation but do not report the results here for reasons of brevity, in most settings Cross Validation performs better. HDBSCAN is computed using the R package dbscan described in Hahsler et al. (2019). $MinPts$ (compare appendix 1.C) is set to 5 in M3, 7 in M1,M4 and 10 in M2,M5. k-Means is computed with different choices of k, including too small, too large and the correct number of groups to investigate how results are affected by misspecification. In setting M3, where all observations are modelled as atoms we use larger k values. All k-Means computations apply 100 iterations and 1000 random starting values. As a baseline, we compare our estimators to the Mundlak estimator and a pooled OLS regression. The methods in the cited literature (compare Table 1.1) are not applicable to our designs. Of these approaches only Bonhomme and Manresa (2015) and Tutz and

Table 1.3: Simulation Results

	β			γ		
	Bias	MAD	MSE	Bias	MAD	MSE
M1						
POLS	1.5196	1.5196	2.3117	-0.0010	0.0718	0.0080
Mundlak	0.0014	0.0407	0.0025	-0.0000	0.0505	0.0041
k-Means						
k-Means, 3	0.3427	0.3427	0.1199	-0.0100	0.1538	0.0393
k-Means, 5	0.1206	0.1207	0.0171	-0.0024	0.2426	0.0915
k-Means, 10	0.0400	0.0517	0.0041	-0.0197	0.2549	0.0983
HDBSCAN	0.0360	0.0535	0.0055	-0.0227	0.3484	0.2019
HDBSCAN with LASSO						
Cross Validation	-0.0076	0.0462	0.0041	-0.0168	0.2828	0.1291
Gen Cross Val	0.0339	0.0526	0.0053	-0.0225	0.3434	0.1960
BIC	0.0335	0.0524	0.0053	-0.0224	0.3431	0.1957
M2						
POLS	0.0022	0.0735	0.0086	3.7064	3.7064	14.3185
Mundlak	0.0013	0.0235	0.0009	3.7041	3.7041	14.3027
k-Means						
k-Means, 3	0.0032	0.0339	0.0018	4.6190	4.6235	23.3694
k-Means, 5	0.0013	0.0224	0.0008	-0.0011	0.0474	0.0035
k-Means, 10	0.0013	0.0237	0.0009	0.3208	0.3887	0.8949
HDBSCAN	0.0013	0.0224	0.0008	0.0175	0.0661	0.0903
HDBSCAN with LASSO						
Cross Validation	-0.2037	0.2037	0.0429	0.1251	0.1281	0.1015
Gen Cross Val	-0.0368	0.0399	0.0022	0.0376	0.0686	0.0906
BIC	-0.0368	0.0399	0.0022	0.0376	0.0686	0.0906
M3						
POLS	0.4976	0.4976	0.2479	0.0033	0.0404	0.0025
Mundlak	-0.0001	0.0082	0.0001	0.0011	0.0217	0.0007
k-Means						
k-Means, 5	0.0688	0.0688	0.0049	0.0126	0.1854	0.0531
k-Means, 20	0.0092	0.0116	0.0002	-0.0030	0.1628	0.0396
k-Means, 100	0.0046	0.0093	0.0001	0.0093	0.1004	0.0163
HDBSCAN	0.0068	0.0103	0.0002	-0.0051	0.1664	0.0457
HDBSCAN with LASSO						
Cross Validation	-0.0044	0.0101	0.0002	-0.0032	0.1537	0.0387
Gen Cross Val	0.0056	0.0097	0.0002	-0.0048	0.1648	0.0448
BIC	0.0056	0.0097	0.0002	-0.0048	0.1648	0.0448
M4						
POLS	1.5981	1.5981	2.5552	-0.0001	0.0493	0.0038
Mundlak	-0.0021	0.0282	0.0013	0.0013	0.0347	0.0019
k-Means						
k-Means, 3	0.4994	0.4994	0.2508	0.0106	0.1580	0.0390
k-Means, 5	0.2263	0.2263	0.0525	-0.0230	0.2522	0.0987
k-Means, 10	0.0738	0.0744	0.0068	-0.0010	0.3168	0.1588
HDBSCAN	0.0163	0.0330	0.0017	-0.0014	0.3023	0.1450
HDBSCAN with LASSO						
Cross Validation	0.0012	0.0312	0.0016	-0.0032	0.2618	0.1089
Gen Cross Val	0.0162	0.0330	0.0017	-0.0016	0.2985	0.1412
BIC	0.0161	0.0330	0.0017	-0.0018	0.2982	0.1409
M5						
POLS	0.0070	0.0570	0.0052	0.0012	0.5056	0.3958
Mundlak	0.0007	0.0171	0.0005	0.0017	0.5056	0.3956
k-Means						
k-Means, 3	0.0025	0.0215	0.0007	0.0201	0.2617	0.3107
k-Means, 5	0.0005	0.0168	0.0004	0.0008	0.0333	0.0017
k-Means, 10	0.0006	0.0171	0.0005	0.0513	0.4351	0.6393
HDBSCAN	0.0005	0.0168	0.0005	0.0147	0.1152	0.4071
HDBSCAN with LASSO						
Cross Validation	-0.1770	0.1770	0.0323	0.0131	0.1099	0.3546
Gen Cross Val	-0.0394	0.0402	0.0020	0.0144	0.1139	0.3997
BIC	-0.0395	0.0403	0.0020	0.0145	0.1139	0.3997

Notes: Means of 500 simulations. Simulation designs are defined in Table 1.2.

Schauberger (2015) allow for the inclusion of time-constant covariates. Berger and Tutz (2018) only allow for time-constant variables in a random effects framework. However, Bonhomme and Manresa (2015) in general require that the time-constant variables have more distinct values (support points) than the number of clusters. Therefore an indicator variable as in our designs is not feasible. The estimators proposed by Tutz and Schauberger (2015) can only be computed for smaller sample sizes.

In setting M2 we simulate a correlation between the group intercept v_i and the time-constant variable $\boldsymbol{Z}_i$. The probability for $\boldsymbol{Z}_i = 1$ varies across different groups. The Mundlak estimator is expected to perform worse in this setting compared to the settings M1, M3 and M4. In the mixed design (M4) both M1/M2 and M3 are combined: $N/2$ units are assigned into 5 distinct groups $N/2$ units are independent $N(0, 1)$ draws and therefore atoms. In setting M4 the mixture of clear groups and atoms creates intervals with different densities, there HDBSCAN is known to have advantages. k-Means on the other hand can have advantages in settings where the clustering is a tool for discretisation as in design M3 and to some extent in design M4. In design M5 the groups have random heterogeneous group sizes with which the k-Means algorithm is known to struggle and it is therefore expected to perform worse in design M5. Further, the increased distance between the group intercepts should increase the bias induced by choosing an incorrect group number.

v_i is modelled as follows: In M1/M2 each v_i is a realisation of a $N(1, 2)/N(1, 10)$ random variable, that is subsequently discretised into 5 groups. First observations are binned into 5 quantiles and the quantile means are used as final group intercepts. In design M3 the FEs are independent random draws from a $N(0, 1)$ distribution without discretization. Design M4 combines both approaches. In design M5, for four groups the share of the entire population is drawn with a uniform distribution in the interval [0.1,0.25]. The fifth group is formed by the residual share. The group intercepts are drawn from five different uniform distributions to ensure more spacing between the groups, the intervals used for the uniform distributions are: $G_5 = [[-15, -14], [-2, -1.5], [1.5, 2.5], [6, 8.5], [13.5, 14.5]]$. In all designs we model one bivariate time-constant covariate and one continuous time-varying covariate, initially drawn from a binomial or standard normal distribution respectively (see column 6 and 7, Table 2). The correlation structures (row **Correlation Structure**) are induced by tying $\boldsymbol{Z}_i$ or x_{it} to v_i. In M1 and M4, x_{it} is a convex combination of the normally distributed draw (factor 0.6) and the group intercept (factor 0.4), leading to a correlation of ≈ 0.8. In M3, the group intercept is added to the initial draw leading to a correlation of ≈ 0.7. In M2, the probability of $\boldsymbol{Z}_i = 1$

depends on v_i. Ordering the probability vector in ascending order of the respective group intercepts the vector is $P_2 = P(\boldsymbol{Z}_i = 1|v_i) = (0.35, 0.45, 0.55, 0.55, 0.65)$. γ and β (as in equation 1.1, section 2) are set as in the cited literature, namely: $\gamma = \beta = 2$ in the settings M1,M2,M4,M5 and $\gamma = \beta = 1$ in setting M3 following Bonhomme et al. (2022, Supplement S3). We use T=20 as in Bonhomme et al. (2022, Supplement S3). In Appendix 1.H we also provide results for $T = 5$. The idiosyncratic error u_{it} is independently distributed as in the cited literature.

We simulate the 500 samples for each design and report bias, MAD and MSE for the various approaches in Table 1.3. The results confirm our suggested approach performs well. Whether HDBSCAN or k-Means clustering give superior results depends on the design and the chosen k. Given that G is normally unknown in applications, there is always the risk of assuming the wrong k. There is no clear pattern for the MSE, whether G is assumed to be too little or too great. The group intercepts have a larger distance in settings M2 and M5. Choosing an incorrect number increases the error by a larger factor than for example in setting M1. Specifying a too small k leads to worse performances for coefficients on both the time-varying and the time-constant covariates. Setting k too large leads to a larger MSE for coefficient on the time-constant covariate. Due to its nonparametric nature the MSE for HDBSCAN tends to be larger than for k-Means if there are any sizeable differences. The LASSO step improves the results with the HDBSCAN clustering, although not always. An exception is setting M3 without distinct group structure. Cross validation outperforms BIC and general cross validation in most settings.

Further simulations are presented in Appendix 1.H. They include variants of design M2 with varying combinations of fixed effects and error terms. A larger variance of fixed effects and a smaller error variance both improve estimation results with HDBSCAN and HDBSCAN with LASSO. We explain this by a clearer and more distinct group structure and more precise estimation of fixed effects. We also show results for designs M1, M4 and M5 with $T = 5$. While the errors are larger than for $T = 20$, as expected, our approach is shown to work reasonably well in very short panels. We also provide a graphical representation of the clustering step in Appendix 1.I.

1.4 Application

We apply the proposed methods to labour market data and estimate the gender wage gap. Thereby we demonstrate the applicability to large scale data structures

that are commonly used for empirical economic research. We extract a sample from the Sample of the Integrated Employment Biographies 1975-2014 (SIAB) of the Institute for Employment Research (IAB), Germany. These data contain information from various linked administrative social security registers. SIAB is a 2% random sample of the workforce in Germany that contributed to social insurance in the period 1975–2014. Among other things the SIAB contains daily information about periods of dependent employment and wages with basic information about the individual (such as gender, age and education) and the employing firm (such as business sector). SIAB is available as a scientific use file for independent research. For more information on the data see Ganzer et al. (2017). We extract a yearly panel of wages on the 30th of June for the years 2006-2013. Our sample contains employees aged 16-65, that are subject to social insurance contributions, including those in vocational training. If an employee has a part-time and a full-time job we only consider the full-time job. Further, we only consider the job with the highest salary. In addition to the provided variables, we compute others based on the individual employment history to include tenure (time with the current employer) and additional labour market experience (in addition to current tenure). After some data cleansing, we are left with a balanced panel of 241,076 individuals with 1,928,608 person-year observations. In our model we use one time-constant covariate ($female$), 14 time-varying covariates, 7 year dummies and an intercept, whenever adequate. The analysis of the partial effect of gender and education on wages is popular in empirical economic research and is the reference example in leading econometrics textbooks (e.g. Wooldridge, 2010).

We compare results of the following models:

- Pooled OLS model, where $\boldsymbol{X}_{it}$ and $\boldsymbol{Z}_i$ are contemporaneously exogenous and therefore not allowed to be correlated with v_i.
- Mundlak model, which allows for arbitrary correlation between v_i and $\boldsymbol{X}_{is}$ and some correlation between v_i and $\boldsymbol{Z}_i$ if it is through $\bar{\boldsymbol{X}}_i$, the within time average of the time-varying covariates.
- Our regularisation approach with HDBSCAN as clustering step.
- Our regularisation approach with k-Means as clustering step. We work with 5 clusters and with 55 clusters to illustrate the role of the number of clusters.
- Our regularisation approach with HDBSCAN, followed by a LASSO to regularise group membership further as outlined in Appendix 1.E.

We use R V4.0.2 for the analysis on Windows Server 2019 with 96GB RAM. Our suggested clustering methods run quickly and give results within several hours, though we encounter memory problems in the clustering steps and when running the grouped fixed effects regression (1.2) of step 3 when groups are created by HDBSCAN. Because of the large number of atoms (individuals that are not assigned to any group), this regression easily contains 10,000s of dummy variables. Despite the use of big data packages such as biglm (Lumley, 2020) we must restrict the analysis to a randomly chosen 77,500 individuals. The final LASSO step to reduce the groups numbers turned out to require even more memory. For this reason, this last step is only estimated on a smaller sample of 7,500 individuals. This gives some insight as to how much the last supervised regularisation step contributes to dimension reduction. In practice, the final LASSO step is only applicable to large scale data when high performance computing facilities are available. For this step, R requests more than 2800GB of RAM in the case of 77,500 individuals. The running time for the large sample is approximately 3 days to obtain the results in Table 1.4, where the HDBSCAN based model takes around 2 of these days. The results for the smaller sample are obtained within a couple of hours. We report cluster robust standard errors if not otherwise stated using lm.cluster (Robitzsch et al., 2020), where clustering is done at the individual level. For our suggested approaches we report post-clustering standard errors. For the fused LASSO, we only report point estimates to avoid further computational challenges. It would of course be possible with little difficulty to compute post LASSO standard errors.

The estimation results for the various models are displayed in Table 1.4 and in Table 1.5 for the smaller sample. The results in Table 1.4 show that using Mundlak regression instead of POLS leads to considerable changes in many coefficients, including the work history, part-time, certain business sectors and education. Such an observation is frequent in empirical work as POLS is only consistent if the regressors are not correlated with any component in the error term, while the Mundlak model allows for such correlation via the means of the time-varying covariates. The application of our method with HDBSCAN and k-Means with a larger number of groups gives often similar results as already seen in the simulations. k-Means with a small number of groups is also similar, although there are some economically meaningful differences for several variables such as part-time, several business sectors and higher education. Similar to those findings in the simulations, this can be interpreted as evidence suggesting that an insufficient number of groups has been selected. When comparing the Mundlak results with the results of our methods, we see that the estimated effect of the time-constant variable *female* in

particular is quite different when Mundlak is used. Even though the estimates with our methods are not identical they are much more similar and negative. Our method with HDBSCAN clustering suggests a gender wage gap of 32%, while it is only 18% when the Mundlak model is used. Interestingly, while the Mundlak model suggest that POLS is downward biased for this variable, the results with our methods suggest that the direction of the bias is actually in the opposite direction. This illustrates that the Mundlak model can lead to incorrect conclusions when the correlation of the observables with the fixed effects is not only through the means of the time-varying observables. However, most of the coefficients on the time-varying variables do not differ economically between our methods and Mundlak with the exception of age and part-time. In conjunction with the simulation results, Table 1.5 confirms that the additional LASSO step leads only to small changes in results. In our application it is because only a small number of group FE are being regularised (6 after HDBSCAN and 1 after k-Means). The main benefit of the LASSO step therefore seems to be that the resulting estimates have statistical optimality properties. Thus, it can be used to check whether the clustering method is working well.

Our example here shows that the application of statistical learning methods in panel analysis is possible for larger data sets. Our results demonstrate that our suggested methods produce sizeably different estimates than the classical panel models under stronger restrictions. This is particularly true in the case of the coefficient on the time-constant covariate that benefits most from the weaker restrictions of our methodology. Our application has also shown that an analysis with 620,000 person year observations is possible on a computer with 96GB RAM, although the last regularising LASSO step requires too much memory. Note that our application cannot definitively answer the question of the size of the gender wage gap. This is because the dependent variable is daily and not hourly wages. The variable *part-time* provides some information about the number of hours worked, but only represents an indicator for reduced working time without precisely controlling for hours worked. Further, the reported variable might be incomplete. To conclude, our estimates point to considerably lower daily wages for females.

Table 1.4: Estimated coefficients of wage regression model.

	POLS	**Mundlak**	**HDBSCAN**	**k-Means(55)**	**k-Means(5)**
$\mathbf{Z}_i$					
$female$	-0.2115***	-0.1829***	-0.3190***	-0.3647***	-0.3347***
	(0.0034)	(0.0035)	(0.0005)	(0.0007)	(0.0016)
$\mathbf{X}_{it}$					
tenure	0.0216***	-0.0213***	-0.0199***	-0.0197***	-0.0145***
	(0.0003)	(0.0032)	(0.0000)	(0.0000)	(0.0001)
additional experience	0.0187***	-0.0207***	-0.0194***	-0.0193***	-0.0144***
	(0.0003)	(0.0032)	(0.0000)	(0.0000)	(0.0001)
age	-0.0087***	-0.0089***	0.0498***	0.0497***	0.0415***
	(0.0002)	(0.0002)	(0.0000)	(0.0000)	(0.0001)
$part-time$	-0.4289***	-0.1270***	-0.1569***	-0.1748***	-0.2101***
	(0.0042)	(0.0032)	(0.0006)	(0.0010)	(0.00019)
trainee	-1.0791***	-1.0254***	-1.0204***	-1.0163***	-1.0225***
	(0.0095)	(0.0074)	(0.0019)	(0.0057)	(0.0066)
business sector (ref: production)					
agriculture	-0.2787***	-0.1205***	-0.1154***	-0.1070***	-0.1290***
	(0.0117)	(0.0155)	(0.0017)	(0.0022)	(0.0053)
gastronomy	-0.4663***	-0.2264***	-0.2275***	-0.2290***	-0.2681***
	(0.0112)	(0.0194)	(0.0016)	(0.0026)	(0.0053)
construction	-0.2291***	-0.0612***	-0.0540***	-0.0490***	-0.0752***
	(0.0051)	(0.0081)	(0.0009)	(0.0010)	(0.0026)
trade	-0.1221***	-0.0561***	-0.0561***	-0.0563***	-0.0681***
	(0.0042)	(0.0058)	(0.0006)	(0.0008)	(0.0018)
services	-0.0266***	-0.1200***	-0.1125***	-0.1105***	-0.0986***
	(0.0035)	(0.0050)	(0.0005)	(0.0007)	(0.0016)
education/social/health	-0.0220***	-0.0860***	-0.0795***	-0.0748***	-0.0739***
	(0.0043)	(0.0112)	(0.0007)	(0.0001)	(0.0020)
public institutions	0.0302***	-0.0580***	-0.0466***	-0.0407***	-0.0382***
	(0.0045)	(0.0133)	(0.0008)	(0.0010)	(0.0023)
education (ref: none)					
higher education	0.5727***	0.0331***	0.0375***	0.0393***	0.1084***
	(0.0038)	(0.0028)	(0.0007)	(0.0010)	(0.0020)
vocational education	0.1062***	0.0136***	0.0139***	0.0151***	0.0296***
	(0.0027)	(0.0011)	(0.0004)	(0.0006)	(0.0012)
$N = 77,500$, $T = 8$					
Clustering					
individuals with $Z = 0$: 45,974, $Z = 1$: 31,526					
cluster (0/1)			131/134	55/55	5/5
atoms (0/1)			14,172/9,769	0/0	0/0

Notes: *p<0.1; **p<0.05; ***p<0.01. Cluster robust standard errors in parentheses. Non-robust for HDBSCAN. Post-clustering standard errors for HDBSCAN and k-Means. Intercept and year dummies not reported. Averages of $\mathbf{x}_{it}$ not reported (Mundlak).

Table 1.5: Estimated coefficients of wage regression model (smaller sample).

	Mundlak	**HDBSCAN**	**HDBSCAN +fLASSO**	**k-Means**	**k-Means +fLASSO**
$\mathbf{Z}_i$					
female	-0.1664***	-0.3881***	-0.3746	-0.2564***	-0.2529
	(0.0114)	(0.0025)		(0.0030)	
$\mathbf{X}_{it}$					
tenure	-0.0020	-0.0010***	0.0000	-0.0016***	0.0000
	(0.0098)	(0.0002)		(0.0002)	
additional experience	-0.0021	-0.0011***	-0.0001	-0.0016***	-0.0003
	(0.0098)	(0.0001)		(0.0002)	
age	-0.0089***	0.0316***	0.0292	0.0316***	0.0294
	(0.0007)	(0.0002)		(0.0002)	
$part-time$	-0.1349***	-0.1678***	-0.1829	-0.1862***	-0.1977
	(0.0106)	(0.0037)		(0.0040)	
trainee	-1.0206***	-1.0176***	-1.0213	-1.0156***	-1.0215
	(0.0229)	(0.0190)		(0.0182)	
business sector (ref: production)					
agriculture	-0.0931*	-0.0830***	-0.0947	-0.0709***	-0.0770
	(0.0562)	(0.0077)		(0.0082)	
gastronomy	-0.2559***	-0.2664***	-0.2793	-0.2581***	-0.2662
	(0.0733)	(0.0106)		(0.0101)	
construction	-0.0693**	-0.0587***	-0.0731	-0.0576***	-0.0666
	(0.0318)	(0.0042)		(0.0046)	
trade	-0.0467***	-0.0476***	-0.0536	-0.0473***	-0.0503
	(0.0180)	(0.0028)		(0.0031)	
services	-0.0951***	-0.0905***	-0.0856	-0.0850***	-0.0801
	(0.0166)	(0.0026)		(0.0028)	
education/social/health	-0.0630*	-0.0561***	-0.0558	-0.0531***	-0.0520
	(0.0348)	(0.0030)		(0.0035)	
public institutions	-0.0117	-0.0010	-0.0005	0.0060*	0.0065
	(0.0430)	(0.0032)		(0.0035)	
education (ref: none)					
higher education	0.0445***	0.0407***	0.0743	0.0443***	0.0715
	(0.0113)	(0.0039)		(0.0041)	
vocational education	0.0190***	0.0193***	0.0251	0.0209***	0.0236
	(0.0037)	(0.0022)		(0.0024)	
$N = 7,500$, $T = 8$					
Clustering					
number of individuals with $Z = 0 : 4,359$, $Z = 1 : 3,141$					
cluster (0/1)		57/56		55/55	
atoms (0/1)		1,314/884		0/0	
group FE		2,255	2,249	55	54

Notes: *p<0.1; **p<0.05; ***p<0.01. Cluster robust standard errors in parentheses. Post-clustering standard errors for HDBSCAN and k-Means. Intercept and year dummies not reported. Averages of $\mathbf{x}_{it}$ not reported (Mundlak).

1.5 Summary

We introduce a new approach that incorporates unsupervised and supervised learning techniques for the estimation of linear FE panel models. In particular, various statistical regularisation methods are used to regularise the space of fixed effects. By allowing both time-varying and time-constant regressors to be correlated with fixed effects, our method gives estimates for both time-constant and time-varying variables. It complements existing approaches to estimation of panel models by means of statistical learning techniques by allowing for endogeneity of the number of groups, by being applicable to larger data structures and by giving coefficients on time-constant covariates. We provide asymptotic theory for the estimator of the parameters on the time-constant covariates and show that it converges in probability at rate $\sqrt{NT}$. Our simulations confirm that our methods work as expected and yield low MSE and bias. Our application to the estimation of the gender wage gap confirms that a practically relevant different estimate is obtained when our methods are used.

Appendix

1.A Estimation Approach

This appendix describes step by step the estimation procedure. Notation is either introduced in this appendix or taken from the main text.

1. We begin by fitting a linear model with fixed effects at the individual level:

$$y_{it} = \boldsymbol{X}_{it}\beta + \boldsymbol{Z}_i\boldsymbol{\gamma} + v_i + \epsilon_{it} \tag{1.3}$$

   by using a within regression.

2. We use the regression results to retrieve the estimated fixed effects a_i. ($a_i = \boldsymbol{Z}_i\boldsymbol{\gamma} + v_i + e_i$, with e_i an estimation error, compare the introduction of this notation in section 1.2.1, Step 1: Clustering)

3. Let $\mathcal{Z}$ be the set of all distinct values of $\boldsymbol{Z}_i$. For each $Z \in \mathcal{Z}$: We cluster all individuals with $\boldsymbol{Z}_i = Z$ using their corresponding values for a_i.[1]

4. Let C_{Z_*} be the assigned cluster membership variable for all individuals with $\boldsymbol{Z}_i = Z_*$. We define the reference level Z_0 of the variable $\boldsymbol{Z}_i$ as $Z_0 \in \mathcal{Z}$: $max(C_{Z0}) > max(C_{\tilde{Z}}) \forall \tilde{Z} \in \mathcal{Z}, \tilde{Z} \neq Z_0 \in \mathcal{Z}$. This means the reference level is the level of $\boldsymbol{Z}_i$ for which the maximum number of clusters was estimated. The cluster membership variable contains labels corresponding to all distinct clusters, starting at 1 and counting upwards. Therefore the maximum of the vector corresponds to the number of distinct non-atomic clusters. Non-clustered "atomic" individuals are labelled as zero and do not enter the estimated number of clusters. The number of identified clusters in the reference level is denoted $\hat{G}_1$, the estimated value for G_1.

[1] HDBSCAN is computed using R package dbscan (Hahsler et al., 2019), k-Means with base R (R Core Team, 2021), see Appendix 1.C for further details on the clustering algorithms.

5. We use the clustering in the subsamples to create a uniform cluster variable in the whole sample. First, the clusters identified in the reference level $\boldsymbol{Z}_0$ are sorted and relabelled in ascending order of mean(a_i). Let m_0 denote the vector of sorted means, where each element in the vector corresponds to a distinct cluster.

   Then, for each $Z \in \mathcal{Z}$:

   5.a $\forall c \in \mathcal{C} = \{C_{Zi} | \boldsymbol{Z}_i = Z\}$, i.e. $\mathcal{C}$ is the set of all cluster labels corresponding to individuals with $\boldsymbol{Z}_i = Z$, : compute the corresponding mean of the estimated individual fixed effects $m_c = \text{mean}(a_i) : \boldsymbol{Z}_i = Z \;\&\; C_{Zi=c}$. Store the computed means in the first column of a matrix with the corresponding cluster labels in a second column. Order the matrix rows in ascending order of the means, denote the resulting matrix as $\boldsymbol{M}_Z$.

   5.b Compute the set D of all possible draws (combinations) of $|\mathcal{C}|$ elements out of $\hat{G}_1$ elements, where $|.|$ denotes the cardinality of a set. For each $d \in D$: compute steps 1-3.

   1. Compute the subvector m_{0d} of m_0 containing all elements indexed with elements contained in d.

   2. Compute the vectors $diff_{0d}$ with $diff_{0d(i)} = m_{0d(i)} - m_{0d(i+1)}$,
   $diff_Z$ with $diff_{Z(i)} = \boldsymbol{M}_{Z(,1)(i)} - \boldsymbol{M}_{Z(,1)(i+1)}$ and

   3. $f_d = \sum |diff_Z - diff_{0d}|$, where $\boldsymbol{M}_{Z(,1)(i)}$ denotes the i-th element of vector $\boldsymbol{M}_{Z(,1)}$ and $\boldsymbol{M}_{Z(,1)}$ the first column of the matrix $\boldsymbol{M}_Z$ and $\sum |x|$ the sum of all absolute values of elements in a vector x.

   Choose $\hat{d} \in D$ such that $f_{\hat{d}} < f_d \quad \forall d \in D, d \neq \hat{d}$. This combination out of all combinations is chosen as the clusters of the reference level into which the clusters of the level Z are grouped into.

   Relabel the cluster assignment of Z: the cluster label stored in $\boldsymbol{M}_{Z(i,2)}$ is relabelled as the i-th element of $\hat{d}$.

   The number of combinations can take on very large values, when the number of estimated groups (substantially) differ between the realisations of $\boldsymbol{Z}_i$. Therefore, we make use of an iterative strategy: only one combination is computed at a time (using R package arrangements Lai (2020)) and a difference f_d is saved only if it is smaller than all previous differences.

6. Assign to each atomic cluster an individual specific label. Starting at $\hat{G}_1 + 1$ up to $\hat{G}_1 + \hat{G}_2$

7. Regularise the model with a generalised LASSO:

   7.a Set up matrix $\tilde{\boldsymbol{Q}} = [\boldsymbol{Q}', \boldsymbol{A}']'$ and matrix $\tilde{\boldsymbol{W}} = [\boldsymbol{G}_1, \boldsymbol{W}, \boldsymbol{G}_2]$

   7.b Compute $\tilde{\boldsymbol{W}}\tilde{\boldsymbol{Q}}^{-1}$, set up $\tilde{\boldsymbol{W}}_1 = \tilde{\boldsymbol{W}}[1 : N * T, 1 : \hat{G}_1]$, $\tilde{\boldsymbol{W}}_2 = \tilde{\boldsymbol{W}}[1 : N * T, 1 + \hat{G}_1 : \hat{G}_1 + \hat{G}_2 + K_1 + K_2]$. This means that $\tilde{\boldsymbol{W}}_1$ contains the first $\hat{G}_1$ columns of $\tilde{\boldsymbol{W}}$ and $\tilde{\boldsymbol{W}}_2$ the remainder of the columns of $\tilde{\boldsymbol{W}}$.

   7.c compute $\boldsymbol{P} = \tilde{\boldsymbol{W}}_2(\tilde{\boldsymbol{W}}_2'\tilde{\boldsymbol{W}}_2)^{-1}\tilde{\boldsymbol{W}}_2'$ and $\boldsymbol{y}_p = (\boldsymbol{I} - \boldsymbol{P})y$, $\tilde{\boldsymbol{W}}_{1p} = (\boldsymbol{I} - \boldsymbol{P})\tilde{\boldsymbol{W}}_1$, where $\boldsymbol{I}$ denotes the identity matrix.

   7.d Compute the LASSO path with $\boldsymbol{y}_p$ as response vector and $\tilde{\boldsymbol{W}}_{1p}$ as input matrix.

   7.e Choose the optimal tuning parameter for the LASSO estimator:
   We apply three different criteria: 10-fold cross validation (CV), generalised cross validation (GCV) and Bayesian Information Criterion (BIC). In the cross validation the 10 random subsets of the data are created using the dimension N of individuals only such that all $\boldsymbol{T}$ observations of one specific individual are in the same subset. As optimal coefficient vector the most regularised model is chosen such that the CV error conditional on the coefficient vector is within one standard error of the minimum. Regarding BIC and GCV we implement the expressions defined in Hastie et al. (2017), see p. 244, formula 7.52 for GCV and p.233, formula 7.36 for BIC. This leads to an optimal parameter vector $\boldsymbol{\varphi_1}$.

   7.f Transform the parameter vector back to match the response $\boldsymbol{y}$ and input matrix $\tilde{\boldsymbol{W}}$.
   Compute $\hat{\boldsymbol{\varphi_2}} = (\tilde{\boldsymbol{W}}_2'\tilde{\boldsymbol{W}}_2)^{-1}\tilde{\boldsymbol{W}}_2'(\boldsymbol{y} - \tilde{\boldsymbol{W}}_1)\hat{\boldsymbol{\varphi}}_1$ and $\tilde{\boldsymbol{\lambda}} = \tilde{\boldsymbol{D}} * (\hat{\boldsymbol{\varphi}}_{\boldsymbol{1}}, \hat{\boldsymbol{\varphi}}_{\boldsymbol{2}})$

8. Different Option: Without the LASSO step directly after step 7 estimate the linear model:

$$y_{it} = \boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma} + v_{\hat{g}(i)} + u_{it}, \tag{1.4}$$

   where $\hat{g}(i)$ denotes the cluster individual i was clustered into.

9. Option Post Lasso: We compute step 8 using cross validation. This leads to a shrunken vector $v_{\hat{g}(i)cv}$. Then we estimate by OLS:

$$y_{it} = \boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma} + v_{\hat{g}(i)cv} + u_{it}. \tag{1.5}$$

1.B Computation

We use the following R packages in the computation: dbscan (Hahsler et al., 2019), glmnet (Friedman et al., 2010), biglm (Lumley, 2020), plyr (Wickham, 2011), dplyr (Wickham et al., 2021), arrangements (Lai, 2020) , plm (Croissant and Millo, 2008), aricode (Chiquet et al., 2020), miceadds (Robitzsch et al., 2020), haven (Wickham and Miller, 2021), car (Fox and Weisberg, 2019), cluster (Maechler et al., 2021), VeryLargeIntegers (Cuadrado, 2020).

For plots and tables we further use ggplot2 (Wickham, 2016), cowplot (Wilke, 2020), xtable (Dahl et al., 2019). Data preparation and cleaning regarding the SIAB dataset (compare Section 1.4) is computed in Stata.

1.C Clustering

Our aim is to derive latent patterns of heterogeneity. For this we use the estimated fixed effects retrieved from a within estimation as described in section 1.2. Naturally, these are real valued and thus can be ordered but the fixed effects/ individuals are not labelled with any group membership. Finding structures in unlabelled datasets is generally referred to as clustering. The general aim thereby is to structure individuals into groups ("clusters") such that individuals within the clusters are more similar/ close than those belonging to different clusters. There can be additional aims such as clear distinction of clusters, a convex structure or a hierarchical (i.e. nested) ordering of the cluster structure. These are addressed by a variety of clustering algorithms. Clustering algorithms require a notion of similarity and dissimilarity, i.e. specifying a distance measure, in our case the Euclidean distance is the natural choice.

Both widely used and computationally efficient is the k-Means algorithm which is a so called prototype-based clustering method. The k-Means algorithm, going back to MacQueen (1967) and Lloyd (1982), assumes that the data is partitioned into a number k of convex clusters. k has to be specified ex ante. Given a p-dimensional dataset, the algorithm starts with an initialization of k points as prototypes and in each iterative step: 1. each point $i \in \mathbb{R}^p$ in the dataset is clustered into the group represented by the nearest prototype with respect to Euclidean distance. 2. The prototypes are updated: the mean of each cluster is chosen as a new prototype. Step 1 and 2 are repeated until convergence. By choosing the cluster means as updated prototypes the algorithm minimises the intra-cluster variance. The random initialization of prototypes and its convergence

to local minima (Reddy and Vinzamuri, 2014, p. 91) make the algorithm non-deterministic. Therefore it is common to initiate the algorithm a large number of times. There are proposals to estimate the number of clusters, see Reddy and Vinzamuri (2014, p.92) for a list of prominent examples. Still, the requirement to specify the number of groups is an important disadvantage of k-Means. Changing the number of clusters can change the clustering relations of points significantly and in a non-schematic way. Reducing the number of clusters does not necessarily lead to a nested structure (Hastie et al., 2017, p. 514). The Euclidean distance with respect to the cluster means makes the algorithm sensitive to outliers. Put into a statistical perspective, k-Means can be interpreted as the estimation of the means of k underlying Gaussian distributions (Campello et al., 2020). Because of its popularity, computational efficiency and use in related literature (Bonhomme and Manresa, 2015; Bonhomme et al., 2022) we apply the k-Means algorithm as a comparison in our simulation.

A second type of clustering methods are density-based. Clusters are defined as high-density regions, that end at surrounding low density regions (Campello et al., 2020). Compare also Appendix 1.I for a numerical illustration of density-based clustering in the context of our simulation. These approaches are non-parametric as no implicit assumptions are made regarding the underlying distributions.Density-based clustering can detect clusters of arbitrary shapes Ester (2014, p.111). Further, they endogenously determine the number of clusters. Campello et al. (2013, 2015) propose the algorithm HDBSCAN, its basis is DBSCAN* (that is a small refinement of DBSCAN by Ester et al. (1996) which belongs to the most well-known density-based clustering algorithms). Whether a region in the data is high-density according to DBSCAN*, and intuitively speaking defines a cluster, is defined by a minimum distance parameter ϵ and a minimum number of points parameter $MinPts$: Points in high-density regions, so-called core points, are surrounded by at least $MinPts$ points within a distance of ϵ. Noise points, all points that are not core points, are not part of a cluster and considered as "atomic" points. Core points lie in the same cluster if they are connected by a chain of core points with each distance being

smaller than ϵ. In formulas this is defined by:

x in a dataset X is a **core point** w.r.t ϵ and $MinPts \Leftrightarrow |N_\epsilon(x)| \geq MinPts$.

y in a dataset X is a noise point $\Leftrightarrow |N_\epsilon(y)| < MinPts$.

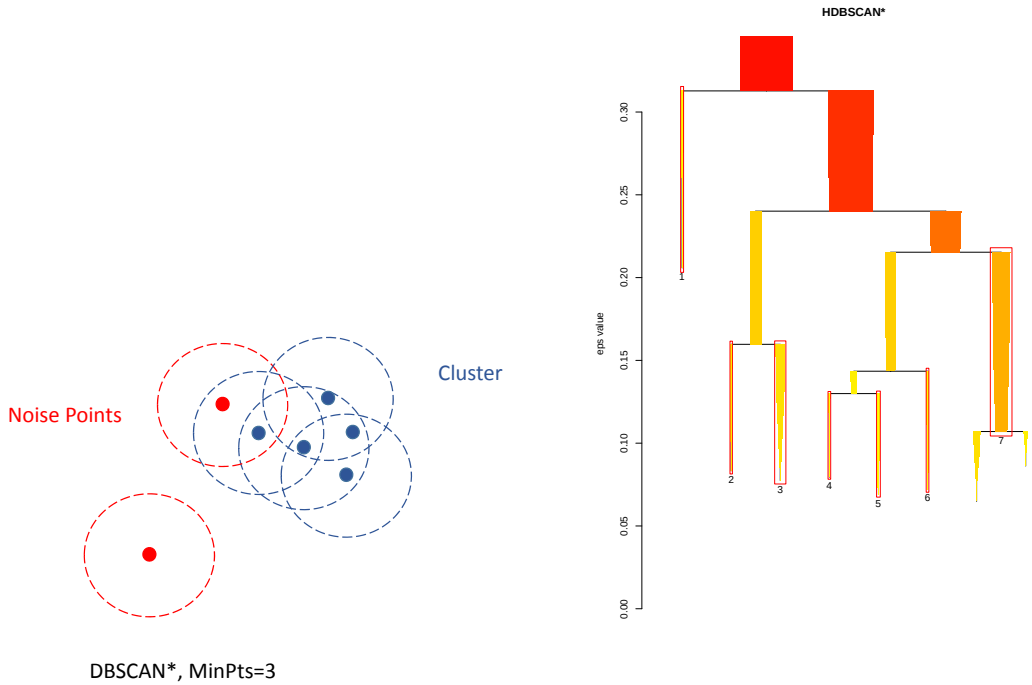
Two core objects x and $y \in X$ are **ϵ-reachable** if $x \in N_\epsilon(y)$ and $y \in N_\epsilon(x)$.

A **cluster** C w.r.t. ϵ and $MinPts$ is a non-empty maximal subset of X such that every pair of objects in C is density-connected.

This is Definition 1-Definition 4 in Campello et al. (2013, p. 162). Hereby $|.|$ denotes the cardinality of a set, X the dataset on which the clustering is computed and $N_\epsilon(x) = \{y \in X | d(x, y) < \epsilon\}$.

Campello et al. (2013) develop the algorithm going back to Ester et al. (1996) further to HDBSCAN by embedding it in a hierarchical clustering structure. Thereby, they also allow for different density thresholds, i.e. ϵ can vary across clusters within the dataset.This also implies that no ϵ parameter must be predefined by the researcher: The algorithm computes the different clustering outcomes for all possible ϵ values. For $\epsilon \to 0$ all data points will be atoms. For $\epsilon \to \infty$ all data points will be put into one large cluster. Between those extremes lies a nested clustering hierarchy, a tree structure. HDBSCAN identifies all ϵ values where changes in the clustering occur and spans the whole hierarchical clustering tree. Then a simplified tree is built by identifying the ϵ thresholds where "significant clustering changes" occur. These are defined as a split of one cluster into two non-atomic clusters or the disappearance of a non-atomic cluster. Finally out of this simplified clustering hierarchy a final clustering outcome is chosen. This is the result of an optimization that finds the most stable clusters with respect to changes in ϵ , i.e. clusters that are present in the hierarchy over the longest interval of ϵ, with the additional condition that each data point is in exactly one cluster or a noise point.

Figure 1.C.1: DBSCAN* and HDBSCAN



Notes: Left Picture: own illustration based on illustrations in Ester et al. (1996). Stylised Illustration of DBSCAN* in $\mathbb{R}^2$ with MinPts=3. Right Picture: own illustration created with R package dbscan. Simplified Tree of HDBSCAN in Simulation Setting M2, MinPts=10. The vertical axis plots different ϵ values.

1.D Mapping of Cluster Membership Variables

After the clustering has been carried out on all observed levels of the variable $\boldsymbol{Z}_i$, the estimated cluster membership variables have to be merged into one variable. First, it has to be noted that due to the unsupervised nature of the clustering approaches taken, the values/labels assigned to each cluster (e.g. 1,2,...) are non-informative of the relative size of a_i of the elements in the cluster. Labels are only informative about distinguishing different clusters. Therefore, we first compute for all $Z \in \mathcal{Z}$ the vector of means of the estimated clusters to establish an relationship between the cluster labels and the size of the estimated fixed effects. Then we relabel the clusters ordered by size within each level of $\boldsymbol{Z}_i$.

From the theoretical considerations we know that the estimated fixed effects a_i capture the coefficient $\boldsymbol{\gamma}$ on $\boldsymbol{Z}_i$ and the group specific intercept, due to the relationship:

$$a_i = \boldsymbol{\gamma}\boldsymbol{Z}_i + v_i + e_i. \tag{1.6}$$

$\boldsymbol{Z}_i$ can take a finite number of different realisations: $|\mathcal{Z}|$ and accordingly $\boldsymbol{\gamma} \in \mathbb{R}^{|\mathcal{Z}|-1}$, in the case of a binary $\boldsymbol{Z}_i$, $\boldsymbol{\gamma} \in \mathbb{R}$.

The individuals form across $Z \in \mathcal{Z}$ the same groups, but it is possible that certain groups are not represented for some values of $Z \in \mathcal{Z}$, compare Assumption (A7) in 1.2.2. Only for one reference level all groups have to be present. In the simple case that for one realisation of $\boldsymbol{Z}_i$ the same number of clusters were identified as for the reference level the mapping reduces to a sorting by size of the means of the estimated fixed effects in the respective clusters and an adequate relabelling of the cluster membership variable to correspond with the reference level.

Assume that for one level $Z \in \mathcal{Z}$, in the following 1, $|C_{Z1}|$ clusters are identified and in the reference level, $\boldsymbol{Z}_i = 0$,: $|C_{Z0}|$ clusters are identified with $|C_{Z0}| > |C_{Z1}|$. There are $\binom{|C_{Z0}|}{|C_{Z1}|}$ possible mappings between the cluster labels, i.e. possible ways to sort the clusters corresponding to Z_1 into the clusters corresponding to Z_0. Here, $\binom{n}{k}$ denotes the binomial coefficient. Because we order the clusters by the size of the estimated fixed effects we do not have to consider permutations. If for example $|C_{Z0}| = 3$ and $|C_{Z1}| = 2$ than there are three possibilities to which two groups of Z_0 the groups of z_1 correspond: $\{1, 2\}, \{1, 3\}, \{2, 3\}$. The estimated intercepts a_i of individuals with $Z_i = 1$, are - conditional on membership in the same group - shifted by $\boldsymbol{\gamma}$ (or the corresponding element in the vector $\boldsymbol{\gamma}$) plus an estimation error compared to individuals with $\boldsymbol{Z}_i = 0$. Therefore the differences of estimated individual fixed effects between groups will be approximately constant across different realisations of $\boldsymbol{Z}_i$:

$$a_i|_{g(i)=2,\boldsymbol{Z}_i=0} - a_i|_{g(j)=1,\boldsymbol{Z}_j=0} = v_2 + \hat{\epsilon}_i - v_1 - \hat{\epsilon}_j$$
$$a_k|_{g(k)=2,\boldsymbol{Z}_k=1} - a_l|_{g(l)=1,\boldsymbol{Z}_l=1} = v_2 - v_1 + \hat{\epsilon}_k + \boldsymbol{\gamma} - \hat{\epsilon}_l - \boldsymbol{\gamma} = v_2 - v_1 + \hat{\epsilon}_k - \hat{\epsilon}_l$$

Therefore we compute for all $\binom{|C_{Z0}|}{|C_{Z1}|}$ possible combinations d: the vector of the chosen clusters in the reference group and the corresponding differences of adjacent cluster means, i.e. the differences between the corresponding clusters in the reference group. We store this in a vector $diff_{0d}$. We also compute the differences of means for the clusters in $\boldsymbol{Z}_i = 1$ in a vector $diff_{Z1}$. Our chosen mapping is defined by the minimum of the absolute value of $f_d = diff_{Z1} - diff_{0d}$ (compare 5.b in the Algorithm description). That means the mapping is chosen for which the estimated clusters are grouped with the same spacing pattern as in the reference level but we allow for the fact that any type of clusters could be left out, not only the smallest or largest ones. As the number of combinations can take on very large values, when the number of estimated groups differ between the realisations of $\boldsymbol{Z}_i$, we make use of an iterative strategy by applying Lai (2020): only the next combination is computed at a time and a difference f_d is saved only if it is smaller than all previous differences.

1.E Regularisation of Redundant Groups

Whenever the algorithm in Step 1 splits one group into several groups, this overparametrisation should be eliminated in Step 3 to guarantee efficiency. When using k-Means $\hat{G}$ has to be prespecified, when using HDBSCAN it is determined endogenously but can be influenced by the input parameter *MinPts*, compare Appendix 1.C. This corresponds to finding whether $\hat{\boldsymbol{D}}_i$ is of greater length than G. Given that the position and ordering of each subgroups are known from Step 2, the regularisation corresponds to a fused LASSO. The corresponding optimisation problem is:

$$\min_{\tilde{\boldsymbol{\lambda}} \in \mathbb{R}^{K_1+K_2+\hat{G}}} \frac{1}{2}\|\boldsymbol{y} - \tilde{\boldsymbol{W}}\tilde{\boldsymbol{\lambda}}\|_2^2 + \eta \sum_{g=1}^{\hat{G}_1-1} |\tilde{\lambda}_{g+1} - \tilde{\lambda}_g|, \tag{1.7}$$

where $\boldsymbol{y}$ is stacked $N*T \times 1$, $\eta \geq 0$ is a tuning parameter and $\tilde{\boldsymbol{W}} = [\boldsymbol{G}_1, \boldsymbol{W}, \boldsymbol{G}_2] \in \mathbb{R}^{(N*T)\times(K_1+K_2+\hat{G})}$. $\tilde{\boldsymbol{W}}$ contains the stacked $\hat{\boldsymbol{D}}$ and $\boldsymbol{W}$ matrices, arranged in a specific column order. The vectors in the $N*T \times \hat{G}_1$ matrix $\boldsymbol{G}_1$ indicate the membership of individuals in the non-atomic groups and the $N*T \times \hat{G}_2$ matrix $\boldsymbol{G}_2$ indicates the atomic groups respectively. Further, we order the groups (i.e.

columns) in $\boldsymbol{G}_1$ by the mean estimated fixed effect. To ensure comparability when computing the ordering we only use the fixed effects of the individuals in the reference group with respect to Z, i.e. the reference level of Z which is assumed to contain all non-atomic groups. Using the same order as in $\tilde{\boldsymbol{W}}$, $\tilde{\boldsymbol{\lambda}} \in \mathbb{R}^{K_1+K_2+\hat{G}}$ is the rearranged $\boldsymbol{\lambda}$. Problem (1.7) is a variant of the so-called fused LASSO, as only the coefficients of the non-atomic groups shrink towards each other. The coefficients on both time-variant and time-constant covariates are not regularised. Also, we do not regularise the group coefficients towards zero. We define a matrix $\boldsymbol{Q} \in \mathbb{R}^{(\hat{G}_1-1)*(K_1+K_2+\hat{G})}$ as:

$$\boldsymbol{Q} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & & \\ 0 & 0 & \dots & 0 & -1 & 1 & 0 & \dots \end{bmatrix} \tag{1.8}$$

where only the first $\hat{G}_1$ columns of $\boldsymbol{Q}$ contain non-zero elements. By using $\boldsymbol{Q}$ it is possible to see that Equation (1.7) is equivalent to:

$$\min_{\tilde{\boldsymbol{\lambda}} \in \mathbb{R}^{K_1+K_2+\hat{G}}} \frac{1}{2} \|\boldsymbol{y} - \tilde{\boldsymbol{W}}\tilde{\boldsymbol{\lambda}}\|_2^2 + \eta \|\boldsymbol{Q}\tilde{\boldsymbol{\lambda}}\|_1, \tag{1.9}$$

which defines a generalised LASSO problem as discussed by Tibshirani and Taylor (2011).

While the LARS algorithm can be used to find the solution for the regular LASSO, the fused or generalised LASSO is computationally demanding, in particular if there are many groups. It is unfortunately not straightforward to transfer results for a regular LASSO to a generalised LASSO including the computation of degrees of freedom, choice of optimal tuning parameters and p-values. Tibshirani and Taylor (2011) show, however, that generalised LASSO problems can be written as a regular LASSO problem under a mild restriction by applying a known transformation. We adopt their approach such that more efficient software implementations can be used and to simplify the problem.

The condition that the link to a regular LASSO exists is satisfied in our context because the matrix $\boldsymbol{Q}$ in Problem (1.9) has full row rank. Following Tibshirani and Taylor (2011) we extend the matrix $\boldsymbol{Q}$ to $\tilde{\boldsymbol{Q}} = [\boldsymbol{Q}', \boldsymbol{A}']'$, where $\boldsymbol{A}$ is $K_1 + K_2 + 1 \times K_1 + K_2 + \hat{G}$ and comprises of $\hat{G}_1$ column vectors of zeros and a block diagonal matrix plus the last row of the matrix being a vector of $\hat{G}_1$ 1s and

$K_1 + K_2 + \hat{G}_2$ zeros:

$$\boldsymbol{A} = \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \dots & \vdots & \vdots & \dots & \ddots & \dots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 \\ 1 & \dots & 1 & 0 & 0 & \dots & 0 & 0 \end{bmatrix} \tag{1.10}$$

$\tilde{\boldsymbol{Q}}$ is $K_1 + K_2 + \hat{G} \times K_1 + K_2 + \hat{G}$ invertible and the rows in $\boldsymbol{A}$ are orthogonal to $\boldsymbol{Q}$. Therefore, the conditions on $\boldsymbol{A}$ defined in Tibshirani and Taylor (2011) are satisfied. By applying a transformation to the Problem in (1.9), we obtain

$$\min_{\boldsymbol{\varphi} \in \mathbb{R}^{K_1+K_2+\hat{G}}} \frac{1}{2} \|\boldsymbol{y} - \tilde{\boldsymbol{W}} \tilde{\boldsymbol{Q}}^{-1} \boldsymbol{\varphi}\|_2^2 + \eta \|\boldsymbol{Q} \boldsymbol{\varphi}\|_1, \tag{1.11}$$

with $\boldsymbol{\varphi} = \tilde{\boldsymbol{Q}} \tilde{\boldsymbol{\lambda}} = (\boldsymbol{\varphi}_1', \boldsymbol{\varphi}_2')'$, where $\boldsymbol{\varphi}_1$ contains the first $\hat{G}_1 - 1$ elements of $\boldsymbol{\varphi}$. Problem (1.11) is a regular LASSO with the exception that the penalty shrinks differences in a subset of parameters. Using an orthogonalisation Tibshirani and Taylor (2011) show that there is actually equivalence to a regular LASSO. Let $\tilde{\boldsymbol{W}} \tilde{\boldsymbol{Q}}^{-1} \boldsymbol{\varphi} = \tilde{\boldsymbol{W}}_1 \boldsymbol{\varphi}_1 + \tilde{\boldsymbol{W}}_2 \boldsymbol{\varphi}_2$, where $\tilde{\boldsymbol{W}}_1$ contains the first $\hat{G}_1 - 1$ columns of $\tilde{\boldsymbol{W}} \tilde{\boldsymbol{Q}}^{-1}$. Problem (1.11) corresponds then to

$$\min_{\boldsymbol{\varphi_1} \in \mathbb{R}^{\hat{G}_1 - 1}} \frac{1}{2} \|(\boldsymbol{I} - \boldsymbol{P}) \boldsymbol{y} - (\boldsymbol{I} - \boldsymbol{P}) \tilde{\boldsymbol{W}}_1 \boldsymbol{\varphi}_1\|_2^2 + \eta \|\boldsymbol{\varphi}_1\|_1, \tag{1.12}$$

with $\boldsymbol{P} = \tilde{\boldsymbol{W}}_2 (\tilde{\boldsymbol{W}}_2' \tilde{\boldsymbol{W}}_2)^{-1} \tilde{\boldsymbol{W}}_2'$, the projection onto the column space of $\tilde{\boldsymbol{W}}_2$ and $\boldsymbol{I}$ the identity matrix. The LARS algorithm can be applied to Problem (1.12) for estimating $\boldsymbol{\lambda}$, which is just a differently ordered $\tilde{\boldsymbol{\lambda}}$. This is achieved by back-transforming the estimated coefficients through pre-multiplication with the matrix $\tilde{\boldsymbol{Q}}^{-1}$, i.e. $\tilde{\boldsymbol{Q}}^{-1} \hat{\boldsymbol{\varphi}}$. $\hat{\boldsymbol{\varphi}}_2$ is obtained by a linear regression of $\boldsymbol{y} - \tilde{\boldsymbol{W}}_1 \hat{\boldsymbol{\varphi}}_1$ on $\tilde{\boldsymbol{W}}_2$.

1.F Proof of Theorem 1

We start by showing that, with probability tending to one,

(a) I^Z is a union of disjoint closed intervals $I^Z = I_1^Z \cup \cdots \cup I_{l(Z)}^Z$ with $l(Z) \leq G_1$.

Furthermore we will show that, with probability tending to one,

(b) each interval I_j^Z contains $q_g + Z\boldsymbol{\gamma}$ for exactly one $1 \leq g \leq G_1$. As said, we also write $I^{g,Z}$ for this interval.

For a proof of these claims define:

$$\tilde{f}_b^Z(x) = \frac{1}{N_Z} \sum_{i=1}^{N} \mathbb{1}(\boldsymbol{Z}_i = Z) \frac{1}{b} K\left(\frac{\boldsymbol{Z}_i\boldsymbol{\gamma} + v_i + \bar{u}_i - x}{b}\right).$$

It can be easily checked that

$$(c) \qquad \sup_{Z \in \mathcal{Z}, x} |\tilde{f}_b^Z(x) - \hat{f}_b^Z(x)| = O_p\left(\frac{\sqrt{T}}{\sqrt{N}}\right).$$

For a proof of this statement one makes use of

$$\sup_{Z \in \mathcal{Z}, x} \frac{1}{N_Z} \sum_{i=1}^{N} \mathbb{1}(\boldsymbol{Z}_i = Z) \mathbb{1}(|v_i + \bar{u}_i - x|) \leq Cb) = O_p(b\sqrt{T}) = O_p(1)$$

for $C > 0$,

$$\sup_{1 \leq i \leq N} \left| K\left(\frac{\boldsymbol{Z}_i\boldsymbol{\gamma} + v_i + \bar{u}_i - x}{b}\right) - K\left(\frac{a_i - x}{b}\right) \right| = O_p\left(\frac{1}{\sqrt{N}}\right).$$

For a proof of (a) choose $\boldsymbol{\delta} > 0$ with $q'_g - q_g > 2\boldsymbol{\delta}$ for all $g' \neq g$. Because of (A2),(A5) and (c), with probability tending to one it holds that $\hat{f}_b^Z(x) = O_p(1)$, uniformly for x not in an interval

$$I_{g,\delta}^Z = [q_g + Z\boldsymbol{\gamma} - \delta, q_g + Z\boldsymbol{\gamma} + \delta], (1 \leq g \leq G_1, Z \in \mathcal{Z}).$$

Choose $g_* \in \{1, \ldots, G_1\}, Z_* \in \mathcal{Z}$ with $\alpha_{g_*}^{Z_*} > 0$. We now show that:

$$(d) \qquad I^{Z_*} \cap I_{g_*,\delta}^{Z_*} \text{ is a closed interval.}$$

Note that (d) implies (a) and (b). To simplify notation we assume that $q_{g_*} + Z_*\boldsymbol{\gamma} = 0$ and that $c_1^b = 1$. Then we have that $b = 1/\sqrt{T}$. For the proof of (d) we define independent random variables

$$V^Z(i) \ (Z \in \mathcal{Z}, 1 \leq i \leq N)$$

with

$$P(V^Z(i) = g) = \alpha_g^Z, (g = 0, \ldots, G_1).$$

Given $\boldsymbol{Z}_i = Z, V^Z(i) = g$, put $v_i^* = q_g$ if $1 \leq g \leq G_1$ and v_i^* conditionally

distributed according to S^Z.

Note that, given $\boldsymbol{Z}_i, \bar{u}_i$, the variable v_i^* has the same conditional distribution as v_i. W.l.o.g. we assume $v_i^* = v_i$. For $x \in I_{g,\delta}^Z$ we have with probability tending to one,

$$\hat{f}_b^Z(x) = \hat{f}_{b,0}^Z(x) + \hat{f}_{b,g}^Z(x)$$

with

$$\hat{f}_{b,v}^Z(x) = \frac{1}{N_Z} \sum_{i=1}^N \mathbb{1}(\boldsymbol{Z}_i = Z, V^Z(i) = v) \frac{1}{b} K\left(\frac{a_i - x}{b}\right)$$

for $0 \le v \le G_1$. Put

$$\tilde{f}_{b,v}^Z = \frac{1}{N_Z} \sum_{i=1}^N \mathbb{1}(\boldsymbol{Z}_i = Z, V^Z(i) = v) \frac{1}{b} K\left(\frac{Z\boldsymbol{\gamma} + v_i + \bar{u}_i - x}{b}\right).$$

Uniformly for $x \in I_{g,\delta}^Z$ it holds that

$$\begin{aligned} &\text{(e)} \qquad \hat{f}_{b,0}^Z(x) - \tilde{f}_{b,0}^Z(x) = O_p(1/\sqrt{N}), \\ &\text{(f)} \qquad \hat{f}_{b,g}^Z(x) - \tilde{f}_{b,g}^Z(x) = O_p(\sqrt{T}/\sqrt{N}). \end{aligned}$$

Expansions (e), (f) follow similarly as (c). With $x_* = x\sqrt{T}$, $\bar{u}_{*,i} = \bar{u}_i\sqrt{T}$ we get for all constants $C > 0$ uniformly for $|x_*| \le C$ that

$$\begin{aligned} \tilde{f}_{b,g_*}^{Z_*}(x) &= \tilde{f}_{b,g_*}^{Z_*}(x_*/\sqrt{T}) \\ &= \frac{1}{Nb} \sum_{i=1}^N \mathbb{1}(\boldsymbol{Z}_i = Z_*, V^{Z_*}(i) = g_*) K\left(\frac{\bar{u}_i - x}{b}\right) \\ &= \frac{\sqrt{T}}{N} \sum_{i=1}^N \mathbb{1}(\boldsymbol{Z}_i = Z_*, V^{Z_*}(i) = g_*) K\left(\bar{u}_{*,i} - x_*\right) \\ &= \sqrt{T}\left(\Delta_N^1(x_*, Z_*) + \Delta_{N,T}^2(x_*, Z_*) + O_p(1/\sqrt{N})\right), \end{aligned}$$

where with the standard normal density φ

$$\begin{aligned} \Delta_N^1(x_*, Z_*) &= p(Z_*)\alpha_{g_*}^{Z_*} \int K(u - x_*) \frac{1}{\sigma} \varphi(u/\sigma)\, du = O(1), \\ \Delta_{N,T}^2(x_*, Z_*) &= p(Z_*)\alpha_{g_*}^{Z_*} \int K(u - x_*) \left(\frac{1}{\sigma}\varphi(u/\sigma)\, du - F_N(\, du)\right) = O(T^{-1/2}). \end{aligned}$$

Furthermore, we get for all constants $C > 0$ uniformly for $|x_*| \leq C$ that

$$\begin{aligned}
&\tilde{f}_{b,0}^{Z_*}(x) + \tilde{f}_{b,0}^{Z_*}(-x) \\
&\quad = \frac{1}{Nb}\sum_{i=1}^{N} \mathbb{1}(\boldsymbol{Z}_i = Z, V^{Z_*}(i) = 0) \\
&\qquad \times \left\{ K\left(\frac{v_i - q_{g_*} + \bar{u}_{*,i}/\sqrt{T} - x}{b}\right) + K\left(\frac{v_i - q_{g_*} + \bar{u}_{*,i}/\sqrt{T} + x}{b}\right)\right\} \\
&\quad = p(Z_*)\alpha_0^{Z_*} \int \left\{ K\left(\frac{v - q_{g_*} + v_i/\sqrt{T} - x}{b}\right) + K\left(\frac{v - q_{g_*} + v_i/\sqrt{T} + x}{b}\right)\right\} \\
&\qquad \times \frac{1}{b} F_N(du) s^{Z_*}(v) dv + O_p(1/\sqrt{Nb}) \\
= \;& \Delta_{N,T}^3(x_*, Z_*) + O_p(N^{-1/2}T^{1/4}),
\end{aligned}$$

where

$$\begin{aligned}
\Delta_{N,T}^3(x_*, Z_*) &= p(Z_*)\alpha_0^{Z_*} \int K(w)(s^{Z_*}(q_{g_*} + bw - u/\sqrt{T} + x_*/\sqrt{T}) \\
&\qquad + s^{Z_*}(q_{g_*} + bw - u/\sqrt{T} - x_*/\sqrt{T}))F_N(\,du)\,dw \\
&= O(1).
\end{aligned}$$

Finally, we get for all constants $C > 0$ uniformly for $|x_*| \leq C$ that

$$\tilde{f}_{b,0}^{Z_*}(x) - \tilde{f}_{b,0}^{Z_*}(-x) \;=\; \Delta_{N,T}^4(x_*, Z_*) + O_p(N^{-1/2}T^{1/4}),$$

where

$$\begin{aligned}
\Delta_{N,T}^4(x_*, Z_*) &= p(Z_*)\alpha_0^{Z_*} \int K(w)(s^{Z_*}(q_{g_*} + bw - u/\sqrt{T} + x_*/\sqrt{T}) \\
&\qquad - s^{Z_*}(q_{g_*} + bw - u/\sqrt{T} - x_*/\sqrt{T}))\, F_N(du)\, dw \\
&= p(Z_*)\alpha_0^{Z_*} T^{-1/2} \partial s^{Z_*}(q_{g_*}) 2x_* + o(T^{-1/2}) \\
&= O(T^{-1/2}).
\end{aligned}$$

Here ∂s^{Z_*} denotes the derivative of s^{Z_*}. We now consider $x_{*,-} < 0 < x_{*,+}$, where these values are solutions of the equations

$$\hat{f}_b(x_{*,-}/\sqrt{T}) = c_2^b \frac{1}{b} = \hat{f}_b(x_{*,+}/\sqrt{T}).$$

Note that $x_{*,-}$ and $x_{*,+}$ may not be uniquely defined by the equations. But one can check that the following considerations apply for all choices of $x_{*,-}$ and $x_{*,+}$.

For $x_{*,\pm} \in \{x_{*,-}, x_{*,+}\}$ we get that

$$c_2^b = \frac{1}{\sqrt{T}} \hat{f}_b\left(x_{*,\pm}\right) = H(\sqrt{T} x_{*,\pm}) + O\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \text{ where}$$
$$H_{N,T,Z_*}(x_*) = H(x_*) = \Delta_N^1(x_*, Z_*) + \Delta_{N,T}^2(x_*, Z_*) + \frac{1}{2\sqrt{T}} \Delta_{N,T}^3(x_*, Z_*).$$

We compare $x_{*,+}$ and $x_{*,-}$ with $x_{*,+}^j > 0 > x_{*,-}^j$ $(1 \le j \le 3)$, where $x_{*,\pm}^j \in \{x_{*,-}^j, x_{*,+}^j\}$ solves

$$\begin{aligned}
&\Delta_N^1(x_{*,\pm}^1, Z_*) = c_2^b, \\
&\Delta_N^1(x_{*,\pm}^2, Z_*) + \Delta_{N,T}^2(x_{*,\pm}^2, Z_*) = c_2^b, \\
&\Delta_N^1(x_{*,\pm}^3, Z_*) + \Delta_{N,T}^2(x_{*,\pm}^3, Z_*) + \frac{1}{2\sqrt{T}} \Delta_{N,T}^3(x_{*,\pm}^3, Z_*) = c_2^b.
\end{aligned}$$

For a study of $x_{*,\pm}^1$ note that $x_* \to J(x_*) = \int K(v - x_*) \frac{1}{\sigma} \varphi(v/\sigma)\, dv$ is a log-concave function. At this point we assume that $p(Z_*) \alpha_{g_*}^{Z_*} J(0) > c_2^b$. For this reason we assume in Assumption (A6) that c_2^b is small enough. We now use that $\log J$ is concave. This gives for $\delta > 0$ small enough that for $0 < x_1 < x_2$ with

$$\log c_2^{*,b} + \delta \ge \log J(x_1) > \log J(x_2) \ge \log c_2^{*,b} - \delta$$

for $c_2^{*,b} = c_2^b / (p(Z_*) \alpha_{g_*}^{Z_*} J(0))$ it holds that

$$x_2 - x_1 \le \frac{\log J(x_1) - \log J(x_2)}{\log J(0) - \log c_2^{*,b} - \delta}\ x_\delta,$$

where x_δ is the solution of

$$\log J(x_\delta) = \log c_2^{*,b} + \delta.$$

From this inequality we conclude that

$$\begin{aligned}
x_{*,\pm}^j - x_{*,\pm}^3 &= O(1/\sqrt{T}) + O_p(1/\sqrt{N}), \text{ for } j \in \{1, 2\}, \\
x_{*,\pm}^3 - x_\pm &= O(1/T) + O_p(1/\sqrt{N}).
\end{aligned}$$

Note also that, because of

$$\Delta_N^1(x_*, Z_*) = \Delta_N^1(-x_*, Z_*)$$

we have that $x^1_{*,-} = -x^1_{*,+}$. We conclude that I^{g_*,Z_*} is equal to $\left[\frac{x_{*,-}}{\sqrt{T}} - c_3^b b, \frac{x_{*,+}}{\sqrt{T}} + c_3^b b\right]$. Note that the centre of this interval $\frac{1}{2\sqrt{T}}(x_{*,+}+x_{*,-})$ is of order $O(T^{-1})+O_p(1/\sqrt{NT})$ and that its length is of order $\frac{1}{\sqrt{T}}(x_{*,+} - x_{*,-}) = O(1/\sqrt{T}) + O_p(1/\sqrt{NT})$.

Remind that for simplifying notation we have assumed that $Z_*\boldsymbol{\gamma} + q_{g_*} = 0$. For general Z, g we have that $I^{q,Z}$ is an interval with midpoint $Z\boldsymbol{\gamma} + q_g + O(T^{-1}) + O_p(1/\sqrt{NT})$ and length $O(1/\sqrt{T}) + O_p(1/\sqrt{NT})$, which shows (d) and thus also (a) and (b).

At this point we would like to mention that the term $O(T^{-1})$ for the rate of the midpoint of the intervals is caused by the term $\Delta^2_{N,T}(x_*, Z_*)$. In principal one could apply a bias correction of this term based on an estimate of the skewness of the errors u_{it}. Because of symmetry of the third term $\Delta^3_{N,T}(x_*, Z_*) = \Delta^3_{N,T}(-x_*, Z_*)$ this would result in an error of order $O(T^{-3/2}) + O_p(1/\sqrt{NT})$ for the midpoint of the intervals. We do not pursue this idea here and we do not construct bias corrected estimates of the midpoints of the intervals because their success heavily depends on the finite sample accuracy of Edgeworth expansions which may be doubted. Furthermore it requires that the error variables have the same skewness which may not be true in many applications. We only mention shortly below the resulting order of convergence for the estimator of $\boldsymbol{\gamma}$.

We now make use of our considerations to discuss the rate of convergence of the estimator $\hat{\boldsymbol{\gamma}}$. Using the results from above we conclude that

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(\boldsymbol{Z}_i = Z, g(i) = g, V^Z(i) = 0)(v_i + \bar{u}_i - q_g) \\ &\qquad = O\left(T^{-1/2}N^{-1/2}\right). \end{aligned} \tag{1.13}$$

For getting this bound we note first that for constants $c > 0$ and for $1 \leq g \leq G_1$ one gets that the number of atom points (i.e. $V^Z(i) = 0$) in the interval $[q_g - c/\sqrt{T}, q_g + c/\sqrt{T}]$ is of order $O_P(N/\sqrt{T})$. Furthermore, conditionally given $\boldsymbol{Z}_i = Z, V^Z(i) = 0$, the random variables $v_i + \bar{u}_i - q_g$ have a conditional expectation of order $O(1/T)$ and a conditional standard deviation of order $O(1/\sqrt{T})$. This

gives that

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(\boldsymbol{Z}_i = Z, V^Z(i) = 0)(v_i + \bar{u}_i - q_g)\mathbb{1}_{v_i+\bar{u}_i \in [q_g - c/\sqrt{T}, q_g + c/\sqrt{T}]} \\
&\quad = O_P\left(N^{-1}(N/\sqrt{T})T^{-1} + N^{-1}\sqrt{N/\sqrt{T}}(1/\sqrt{T}\right) \\
&\quad = O\left(T^{-3/2} + N^{-1/2}T^{-3/4}\right) \\
&\quad = O\left(T^{-1/2}N^{-1/2}\right),
\end{aligned}$$

where the condition $T^{-1} = O(N^{-1/2})$ has been used. Under the above discussed bias correction we expect that at this point as at other points of the proof the much weaker condition $T^{-3/2} = O(N^{-1/2})$ would suffice. Now, $g(i) = g$ is equivalent to the condition that $v_i + \bar{u}_i \in [q_g - c/\sqrt{T} + \Delta_1, q_g + c/\sqrt{T} + \Delta_2]$, where c is an appropriately chosen constant and where Δ_1 and Δ_2 are random variables of order $O_P(T^{-1/2}N^{-1/2} + T^{-1})$. Thus we have $O_P(N(T^{-1/2}N^{-1/2} + T^{-1}))$ values of $v_i + \bar{u}_i$ between $q_g - c/\sqrt{T}$ and $q_g - c/\sqrt{T} + \Delta_1$ or between $q_g - c/\sqrt{T}$ and $q_g - c/\sqrt{T} + \Delta_2$. This gives

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(\boldsymbol{Z}_i = Z, g(i) = g, V^Z(i) = 0)(v_i + \bar{u}_i - q_g) \\
&\quad = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(\boldsymbol{Z}_i = Z, V^Z(i) = 0)(v_i + \bar{u}_i - q_g)\mathbb{1}_i \\
&\quad = O_P\left(T^{-1/2}(N(T^{-1/2}N^{-1/2} + T^{-1}))\right) \\
&\quad = O_P\left(T^{-1/2}N^{-1/2}\right),
\end{aligned}$$

where $\mathbb{1}_i$ is the indicator function of the event that $v_i + \bar{u}_i$ lies between $q_g - c/\sqrt{T}$ and $q_g - c/\sqrt{T} + \Delta_1$ or between $q_g - c/\sqrt{T}$ and $q_g - c/\sqrt{T} + \Delta_2$. This shows (1.13).

With the help of (1.13) we now get that

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{1\leq g(i)\leq G_1}(\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{g(i)})'(\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{g(i)})(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})\\
&=\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{1\leq g(i)\leq G_1}(\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{g(i)})'(a_i-\bar{a}_{g(i)}-(\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{g(i)})\boldsymbol{\gamma})\\
&=\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{1\leq g(i)\leq G_1}(\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{g(i)})'(v_i+\bar{u}_i-\bar{v}_{g(i)}-\bar{u}_{g(i)})+O_p(1/\sqrt{NT})\\
&=\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{1\leq g(i)\leq G_1,V^{Z_i}(i)\neq 0}(\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{g(i)})'(q_{g(i)}+\bar{u}_i-\bar{v}_{g(i)}-\bar{u}_{g(i)})+O_p(1/\sqrt{NT})\\
&=\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{1\leq g(i)\leq G_1,V^{Z_i}(i)\neq 0}(\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{g(i)})'(\bar{u}_i-\bar{u}_{g(i)})+O_p(1/\sqrt{NT})\\
&=O_p(1/\sqrt{NT}),\ \text{where}\\
&\bar{u}_g=\frac{\sum_{i=1}^{N}\bar{u}_i\mathbb{1}(g(i)=g)}{\sum_{i=1}^{N}\mathbb{1}(g(i)=g)},\\
&\bar{v}_g=\frac{\sum_{i=1}^{N}v_i\mathbb{1}(g(i)=g)}{\sum_{i=1}^{N}\mathbb{1}(g(i)=g)}.
\end{aligned}
$$

For the statement of the theorem it remains to show that the smallest eigen value of
$\Sigma_N = \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{1\leq g(i)\leq G_1}(\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{g(i)})'(\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{g(i)})$ is bounded away from 0. This can be done by choosing $g_Z \in \{1,...,G_1\}$ with $\alpha^Z_{g_Z}>0$. One can show that with probability tending to one for $\delta>0$ small enough

$$
\begin{aligned}
\Sigma_N &\geq \frac{1}{N}\sum_{i=1}^{N}\sum_{Z\in\mathcal{Z}}\mathbb{1}_{\boldsymbol{Z}_i=\boldsymbol{Z},V^Z(i)=g_Z,g(i)=g_Z}(\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{g(i)})'(\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{g(i)})\\
&\geq \delta\mathbb{E}\left[(\boldsymbol{Z}_i-\mathbb{E}\left[\boldsymbol{Z}_i\right])'(\boldsymbol{Z}_i-\mathbb{E}\left[\boldsymbol{Z}_i\right])\right]+o_P(1),
\end{aligned}
$$

where $A\leq B$ for two quadratic matrices means that $B-A$ is positive semidefinite. One can now use Assumption (A4) to bound the smallest eigenvalue of this matrix from below. This concludes the proof of the theorem. □

1.G Consequences of Incorrect Subgrouping

We provide large sample results in Section 1.2.2 and show consistency of our estimator for $\boldsymbol{\gamma}$. Using density based clustering, the estimator reaches the same convergence rate as if group membership was known ex ante. Given that any data set is finite, it is of importance to study possible errors that occur in the clustering

step. Given a finite dataset a_i will contain an estimation error. In this Appendix we provide a non-technical discussion to compare different estimators. For a more technical discussion with respect to density based clustering see Section 1.2.2. The clustering algorithm makes two types of errors: atoms with values of v_i in close neighbourhood of a non-atomic cluster may be grouped into this cluster. Cluster points with corresponding large average error terms $\bar{u}_i = T^{-1} \sum_{t=1}^{T} u_{it}$ can be considered as atoms. The latter will increase the number of estimated parameters in the final model and decrease efficiency. The former error leads to a bias, because there are units with different v_i that are assigned to the same cluster. In the limit both errors do still exist. Nevertheless the estimator converges with $O_P(1/\sqrt{NT})$ i.e. with the same rate as if the true group structure would be known. The main result of our proof bases on the assumption that T apporaches infinity with rate $N^{-1/2}$. In practice it is important that the estimator works well in short to medium panels. Typically the number of observations in a dataset is much larger than the number of available time periods. In our simulations we show that the estimator produces reliable estimates in finite samples, we also provide simulation results for very short panels ($T = 5$). In general, the requirements for T can be relaxed if we are willing to make stricter assumptions regarding the error term: i.e. a symmetric density of $\bar{u}_i$ (see Section 1.2.2 for more details and specifically Assumptions (A2) and (A3) for the proposed more general assumptions on $\bar{u}_i$). Importantly there is also a relation between the requirements for T and the existence of atoms. The rate of T approaching infinity can be relaxed if we assume that the relative number of atoms approaches 0 as N approaches infinity. This corresponds to a stricter or less general Assumption (A5). In the limit we will not group two non-atomic clusters into one cluster. This is however true for density based clustering but not necessarily for other clustering approaches. In the k-Means clustering approach the number of clusters has to be specified ex ante. If it is unknown to the researcher two types of errors are possible: If G is specified too small clusters will be grouped together although the corresponding observations have different values of v_i. This will lead to biased estimation. G can also be specified too large. If clusters are formed by splitting up true clusters this will only affect efficiency. If additional clusters are formed "between" two existing clusters by combining observations both, this will also lead to a bias. The presence of atoms is not incorporated in the k-Means approach: atoms will be grouped into one of the clusters. We test the finite sample performance of both density based clustering and k-Means in our simulations in Section 1.3. In Appendix 1.I we provide graphs that illustrate the cluster assignment in settings for density based clustering an k-Means with and

without atoms and for different values of G in k-Means.

1.H Additional Simulation Results

1.H.1 Additional Designs

The following tables shows variants of simulation design M2 from the main text with different distribution of the group intercepts v_i and idiosyncratic errors u_{it}. HDBSCAN with and without LASSO performs best, when the group differences are larger. In these settings it leads to large errors when k-Means is computed with a too small k. Bonhomme et al. (2022) suggest that a too small k is leading to omitted variable bias. Too large k is also leading to errors, but by a much smaller magnitude. Both Mundlak and Pooled OLS lead to biased results, especially in the settings with larger group intercepts and differences. HDBSCAN performs worse in the setting M2A where errors are larger and group intercepts relatively small. This might indicate that it is sensitive to biased estimation of fixed effects rather than to an included correlation structure.

1.H.2 Smaller Time Dimension

Table 1.H.3 displays the results for Monte Carlo simulations with T=5, all other parameters are kept as in Table 1.2.

1.H.3 Clustering Evaluation

The effect of grouping individuals from different groups into the same cluster on the estimation error will depend on the difference of their true underlying

Table 1.H.1: Simulation Designs M2

Design	Group Structure adapted from	G	N	T	Fixed Effect v_i drawn from	discretised	error u_{it}
M2A	B&T(2018) & T&O(2017)	5	500	20	$N(1,2)$	5 quantile means	$N(0,3)$
M2	B&T(2018) & T&O(2017)	5	500	20	$N(1,10)$	5 quantile means	$N(0,3)$
M2B	B&T(2018) & T&O(2017)	5	500	20	$N(1,10)$	5 quantile means	$N(0,1)$
M2C	B&T(2018) & T&O(2017)	5	500	20	$N(1,2)$	5 quantile means	$N(0,1)$

Notes: B&T(2018): Berger and Tutz (2018),T&O(2017): Tutz and Oelker (2017), $P_2 = (0.35, 0.45, 0.55, 0.55, 0.65)$, x_{it},$\boldsymbol{Z}_i$,γ,β defined as in Table 2, Design M2.

Table 1.H.2: Simulation results M2

	β			γ		
	Bias	MAD	MSE	Bias	MAD	MSE
M2A						
POLS	0.0015	0.0272	0.0011	0.7406	0.7406	0.5755
Mundlak	0.0013	0.0235	0.0009	0.7402	0.7402	0.5749
k-Means						
k-Means, 3	0.0007	0.0240	0.0009	0.3412	0.3557	0.1755
k-Means, 5	0.0013	0.0235	0.0009	0.3590	0.4035	0.2343
k-Means, 10	0.0013	0.0234	0.0009	0.4250	0.4452	0.2740
HDBSCAN	0.0015	0.0236	0.0009	0.3085	0.7462	0.9394
HDBSCAN with LASSO						
Cross Validation	-0.1951	0.1951	0.0397	0.4277	0.6624	0.6870
Gen Cross Val	-0.0080	0.0247	0.0009	0.3143	0.7414	0.9237
BIC	-0.0081	0.0247	0.0009	0.3144	0.7415	0.9237
Cross Val Post Lasso	0.0019	0.0238	0.0009	0.3071	0.7478	0.9406
M2						
POLS	0.0022	0.0735	0.0086	3.7064	3.7064	14.3185
Mundlak	0.0013	0.0235	0.0009	3.7041	3.7041	14.3027
k-Means						
k-Means, 3	0.0032	0.0339	0.0018	4.6190	4.6235	23.3694
k-Means, 5	0.0013	0.0224	0.0008	-0.0011	0.0474	0.0035
k-Means, 10	0.0013	0.0237	0.0009	0.3208	0.3887	0.8949
HDBSCAN	0.0013	0.0224	0.0008	0.0175	0.0661	0.0903
HDBSCAN with LASSO						
Cross Validation	-0.2037	0.2037	0.0429	0.1251	0.1281	0.1015
Gen Cross Val	-0.0368	0.0399	0.0022	0.0376	0.0686	0.0906
BIC	-0.0368	0.0399	0.0022	0.0376	0.0686	0.0906
Cross Val Post Lasso	0.0013	0.0224	0.0008	0.0176	0.0663	0.0903
M2B						
POLS	0.0014	0.0692	0.0078	3.7070	3.7070	14.3176
Mundlak	0.0004	0.0078	0.0001	3.7046	3.7046	14.3016
k-Means						
k-Means, 3	0.0024	0.0266	0.0011	4.6768	4.6822	23.7525
k-Means, 5	0.0004	0.0075	0.0001	-0.0004	0.0158	0.0004
k-Means, 10	0.0006	0.0088	0.0001	0.3224	0.3462	0.8951
HDBSCAN	0.0005	0.0075	0.0001	0.0080	0.0240	0.0350
HDBSCAN with LASSO						
Cross Validation	-0.0764	0.0764	0.0060	0.0484	0.0490	0.0364
Gen Cross Val	-0.0376	0.0376	0.0015	0.0281	0.0318	0.0354
BIC	-0.0376	0.0376	0.0015	0.0281	0.0318	0.0354
Cross Val Post Lasso	0.0005	0.0075	0.0001	0.0080	0.0241	0.0350
M2C						
POLS	0.0006	0.0161	0.0004	0.7412	0.7412	0.5729
Mundlak	0.0004	0.0078	0.0001	0.7407	0.7407	0.5723
k-Means						
k-Means, 3	0.0007	0.0091	0.0001	0.8453	0.8461	0.8162
k-Means, 5	0.0004	0.0075	0.0001	0.0004	0.0179	0.0005
k-Means, 10	0.0004	0.0078	0.0001	0.0738	0.0988	0.0375
HDBSCAN	0.0004	0.0076	0.0001	0.0033	0.0252	0.0046
HDBSCAN with LASSO						
Cross Validation	-0.0679	0.0679	0.0048	0.0407	0.0436	0.0059
Gen Cross Val	-0.0105	0.0119	0.0002	0.0093	0.0256	0.0046
BIC	-0.0105	0.0119	0.0002	0.0093	0.0256	0.0046
Cross Val Post Lasso	0.0004	0.0076	0.0001	0.0035	0.0252	0.0046

Notes: Simulation Designs are defined in Table 1.H.1. Means of 500 simulations.

Table 1.H.3: Simulation results, T=5

	β			γ		
	Bias	MAD	MSE	Bias	MAD	MSE
M1						
POLS	1.5208	1.5208	2.3185	0.0011	0.1185	0.0210
Mundlak	-0.0028	0.0929	0.0137	0.0001	0.1091	0.0182
k-Means						
k-Means, 3	0.3164	0.3165	0.1101	-0.0087	0.2660	0.1124
k-Means, 5	0.1290	0.1412	0.0287	-0.0093	0.3682	0.2157
k-Means, 10	0.0363	0.0954	0.0146	-0.0257	0.3878	0.2308
HDBSCAN	0.0261	0.0970	0.0171	0.0197	0.4604	0.3490
HDBSCAN with LASSO						
Cross Validation	-0.0933	0.1258	0.0238	0.0256	0.3930	0.2536
Gen Cross Val	0.0210	0.0963	0.0168	0.0200	0.4566	0.3432
BIC	0.0200	0.0962	0.0167	0.0204	0.4555	0.3418
Cross Val Post Lasso	0.0835	0.1476	0.1063	0.0403	0.4533	0.3438
M4						
POLS	1.5987	1.5987	2.5585	-0.0040	0.0854	0.0110
Mundlak	0.0073	0.0608	0.0059	-0.0050	0.0794	0.0095
k-Means						
k-Means, 3	0.4330	0.4330	0.1916	-0.0132	0.2184	0.0775
k-Means, 5	0.1974	0.1974	0.0439	-0.0137	0.3203	0.1597
k-Means, 10	0.0664	0.0804	0.0099	-0.0393	0.3463	0.1931
HDBSCAN	0.0223	0.0632	0.0064	0.0118	0.3397	0.1841
HDBSCAN with LASSO						
Cross Validation	-0.0437	0.0715	0.0078	0.0103	0.2957	0.1410
Gen Cross Val	0.0192	0.0623	0.0062	0.0117	0.3366	0.1811
BIC	0.0180	0.0619	0.0062	0.0111	0.3354	0.1797
Cross Val Post Lasso	0.2710	0.2990	0.5435	0.0400	0.3178	0.1666
HDBSCAN	0.0223	0.0632	0.0064	0.0118	0.3397	0.1841
M5						
POLS	-0.0095	0.1132	0.0207	0.0363	0.4888	0.3698
Mundlak	-0.0006	0.0405	0.0025	0.0363	0.4898	0.3714
k-Means						
k-Means, 3	-0.0028	0.0447	0.0031	0.0012	0.2574	0.1368
k-Means, 5	-0.0008	0.0386	0.0022	0.0084	0.1250	0.0653
k-Means, 10	0.0000	0.0404	0.0025	0.0248	0.5234	0.5902
HDBSCAN	-0.0014	0.0400	0.0025	-0.1102	0.6857	2.5903
HDBSCAN with LASSO						
Cross Validation	-0.2703	0.2703	0.0780	-0.1033	0.6447	2.2937
Gen Cross Val	-0.0404	0.0528	0.0041	-0.1097	0.6815	2.5582
BIC	-0.0409	0.0531	0.0042	-0.1100	0.6815	2.5563
Cross Val Post Lasso	0.0301	0.0706	0.3161	-0.0689	0.7182	2.7161

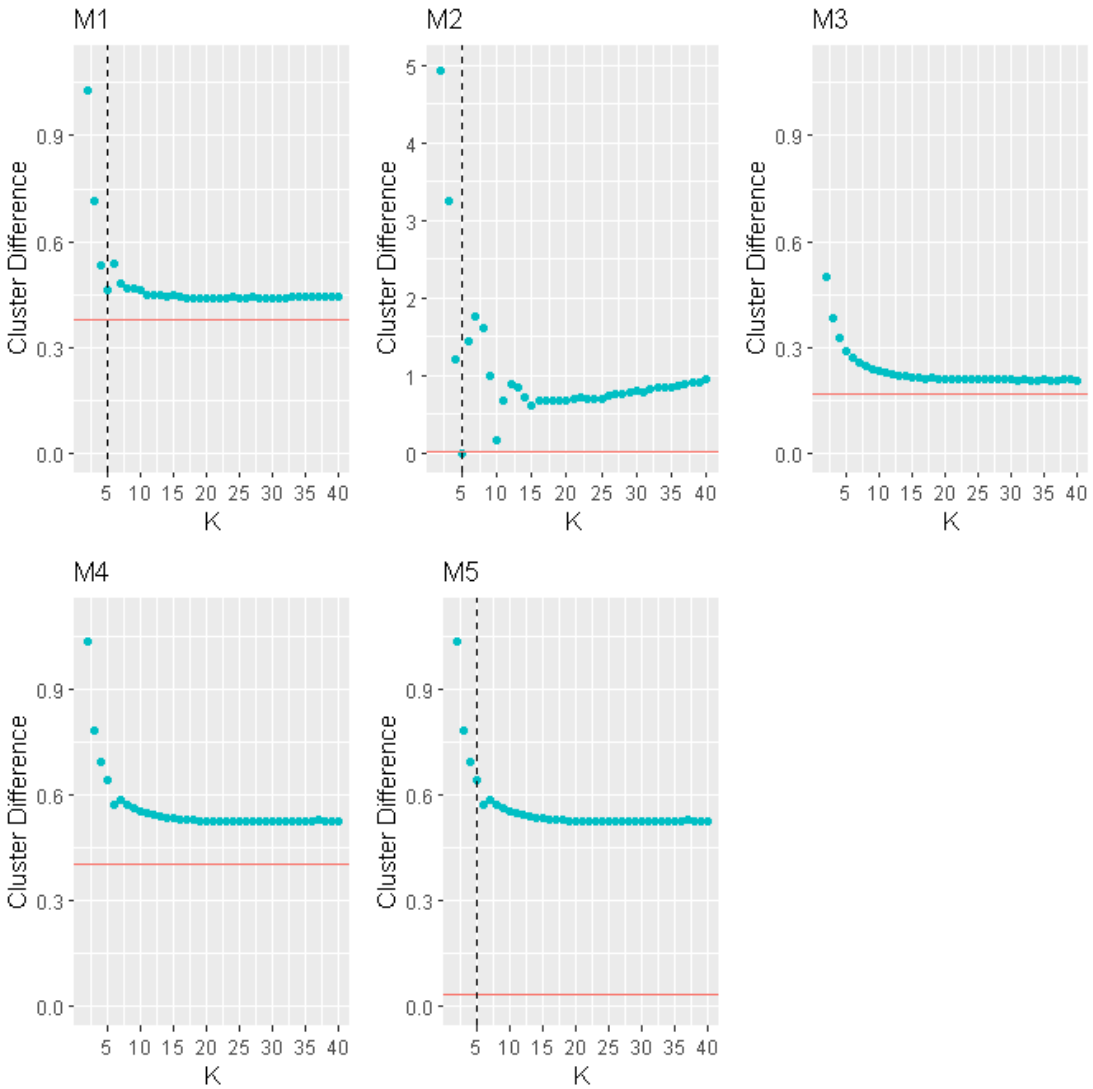
Notes: Simulations as defined in Table 2 with T=5. Means of 500 simulations.

group intercepts. Therefore we compute the difference between an individuals true intercepts and the mean true intercept of all individuals in the same cluster. Specifically, let $c_i \in 1, ..., C$ be the cluster where individual i was grouped into, c_I the set of all individuals grouped into this cluster and v_i i's true group intercept. Then we compute

$$CD = \frac{1}{N}(\sum_{c \in C}(\sum_{i \in c_I} |v_i - \frac{1}{|c_I| - 1} \sum_{j \in c_I, j \neq i} v_j|)). \tag{1.14}$$

Table 1.H.4 displays this measure for different clustering methods and across the different settings defined in Table 1.2. Further values of k are plotted in Figure 1.H.1. Table 1.H.5 displays additional information regarding the estimated group structures in the HDBSCAN step and the LASSO step after HDBSCAN.

Figure 1.H.1: Within Cluster Differences



Notes: The blue scatterplot displays CD as defined in equation 1.14 for different values of k in k-Means across the simulation settings as defined in Table 1.2. For each setting the value for HDBSCAN with cross validation as described in Section 1.3 is plotted as the orange line, compare Table 1.H.4. The dashed line denotes the true groups in the settings with small number of true groups. Data source: simulations.

Table 1.H.4: Clustering Bias

	M1	M2	M3	M4	M5
HDBSCAN with LASSO					
Cross Validation	0.3801	0.0114	0.1667	0.4047	0.0338
Gen Cross Val	0.3795	0.0114	0.1665	0.4035	0.0338
BIC	0.3795	0.0114	0.1665	0.4035	0.0338
HDBSCAN	0.3795	0.0114	0.1665	0.4035	0.0338
k-Means					
k-Means, 3 (M3: K=5)	0.7156	3.2451	0.2919	0.7868	1.7834
k-Means, 5 (M3: K=20)	0.4634	0.0004	0.2096	0.6394	0.0091
k-Means, 10 (M3: K=100)	0.4639	0.1549	0.2160	0.5560	0.1908

Notes: Displays the measure CD defined in equation (1.14). Means of 500 simulations. Simulation designs are defined in Table 1.2.

Table 1.H.5: Estimated Group Structure

	No Groups		No Atoms		No Groups Regularized		
	$Z=0$	$Z=1$	$Z=0$	$Z=1$	CV	GCV	BIC
M1	12.168	12.122	57.572	56.620	1.144	0.068	0.116
M2	5.058	5.042	0.808	0.624	0.008	0	0
M3	19.620	19.452	50.758	52.456	0.634	0.272	0.284
M4	24.742	24.700	118.048	117.154	2.290	0.540	0.592
M5	5.448	5.482	9.336	9.808	0.148	0.010	0.016

Notes: Estimated Group structures by HDBSCAN and HDBSCAN with LASSO. Means across 500 simulations. Simulation designs are defined in Table 1.2.

1.I Illustration of Clustering Algorithms

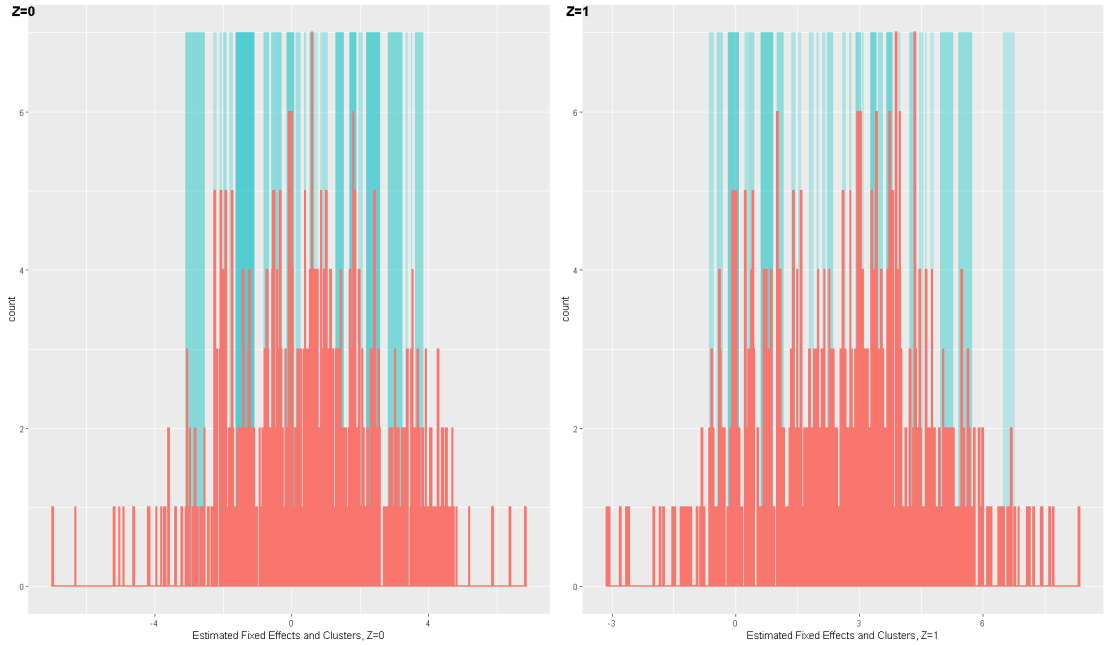
1.I.1 Illustration of HDBSCAN

Figure 1.I.1 illustrates the clusters computed by the HDBSCAN algorithm for simulation design M4, the first of 500 iterations is used as an example. The histogram displays the distribution of the computed fixed effects for two realisations of Z. The blue intervals display the regions of non-atomic clusters. Observations outside of these regions are labelled as atoms. Figure 1.I.2 displays the analogous picture for the first Monte Carlo realisation of Design M5.

1.I.2 Illustration of k-Means

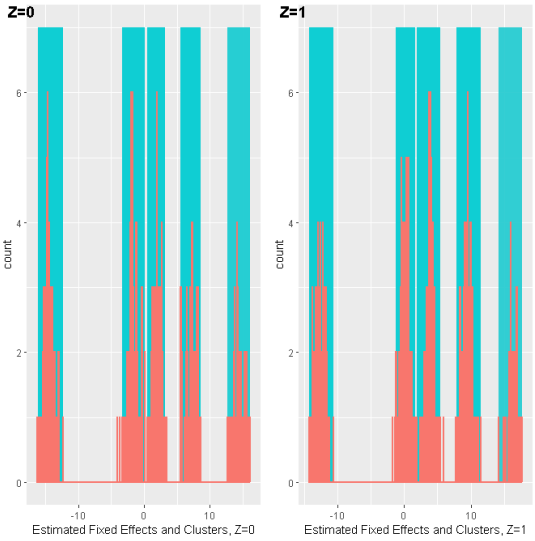
Figure 1.I.1 illustrates the clusters computed by the k-Means algorithm for simulation design M4, the first of 500 Monte Carlo iterations is used as an example. The histogram displays the distribution of the computed fixed effects for two realisations of Z. The coloured intervals display the regions of all k clusters, each cluster is

Figure 1.I.1: Estimated Fixed Effects and HDBSAN Clusters Simulation M4



Notes: Histograms of estimated fixed effects. The blue regions indicate the intervals of non-atomic clusters computed by HDBSCAN. Dataset: first Monte Carlo realisation of Simulation Design M4 as defined in Table 1.2.

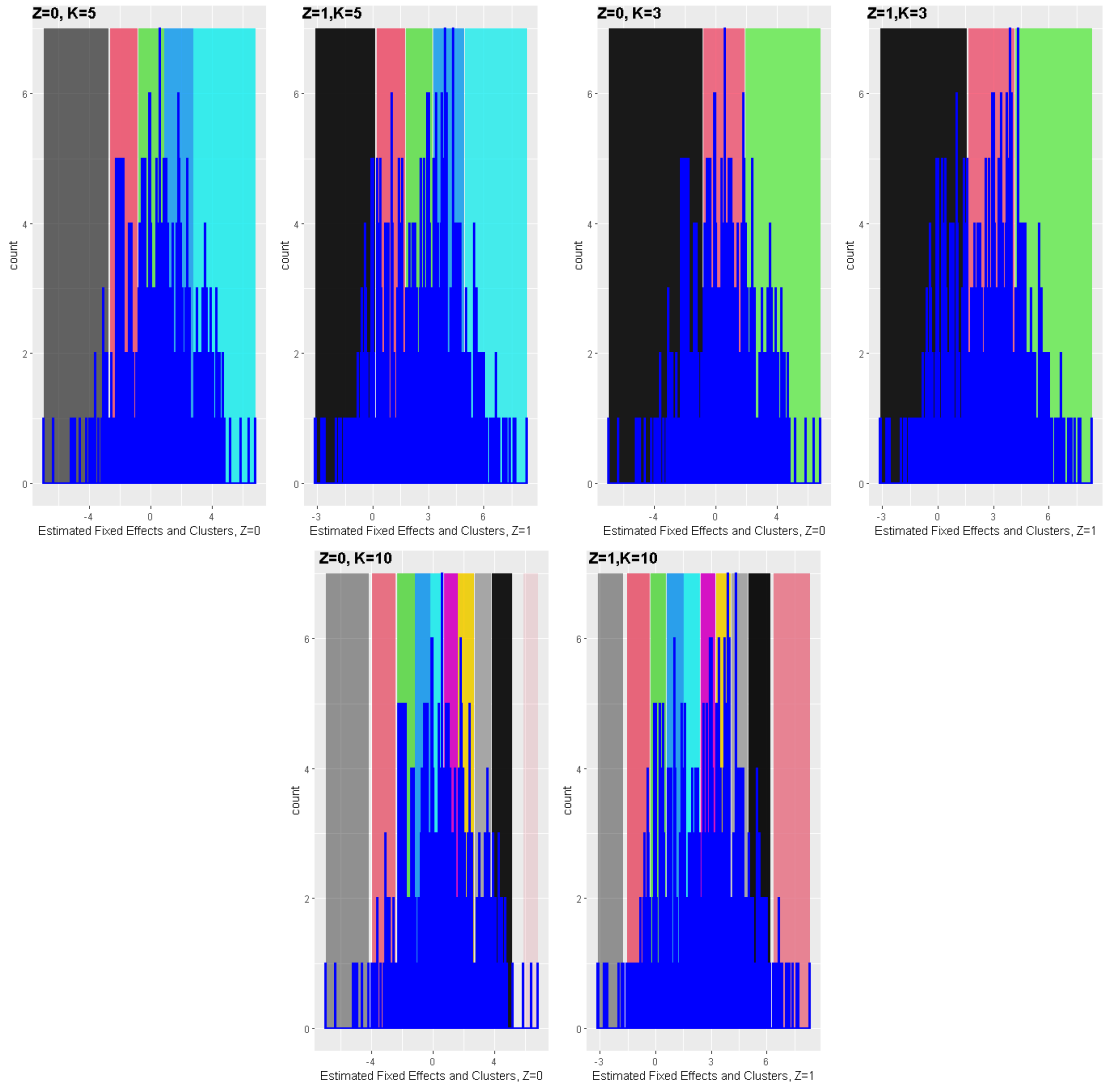
Figure 1.I.2: Estimated Fixed Effects and HDBSAN Clusters Simulation M5



Notes: Histograms of estimated fixed effects. The blue regions indicate the intervals of non-atomic clusters computed by HDBSCAN. Dataset: first Monte Carlo realisation of Simulation Design M5 as defined in Table 1.2.

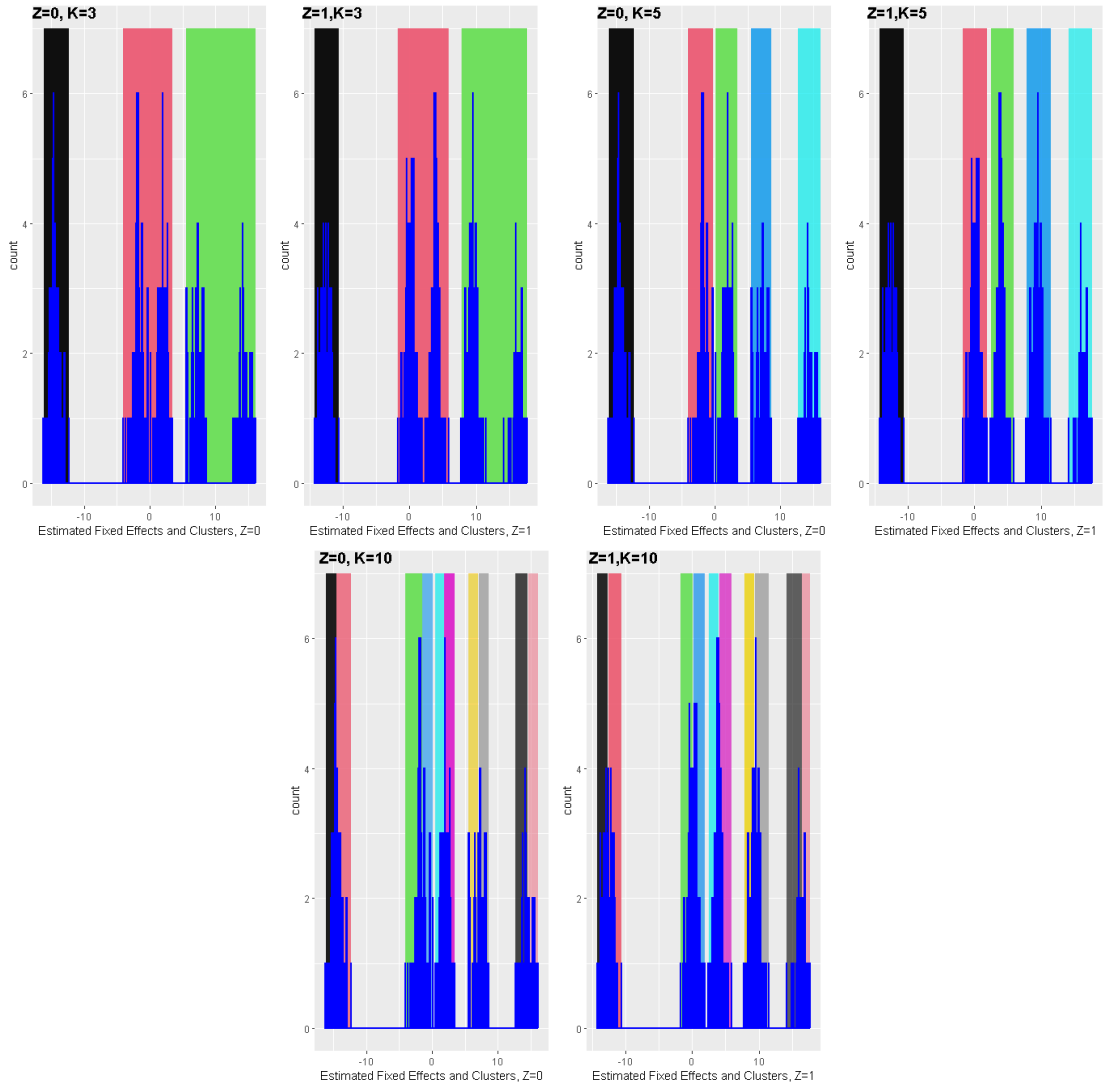
illustrated with a different colour. Figure 1.I.4 displays the analogous picture for the first Monte Carlo realisation of Design M5.

Figure 1.I.3: Estimated Fixed Effects and k-Means Clusters Simulation M4



Notes: Histograms of estimated fixed effects. The coloured regions indicate the intervals of k different clusters computed by k-Means. Dataset: first Monte Carlo realisation of Simulation Design M4 as defined in Table 1.2.

Figure 1.I.4: Estimated Fixed Effects and k-Means Clusters Simulation M5



Notes: Histograms of estimated fixed effects. The coloured regions indicate the intervals of k different clusters computed by k-Means. Dataset: first Monte Carlo realisation of Simulation Design M5 as defined in Table 1.2.

Chapter 2

Regional Patterns of the COVID-19 Pandemic: An Application of Dynamic Time Warping

2.1 Introduction

Pandemics evolve in waves. The COVID-19 pandemic has shown that regions within a country can be hit very differently, both in Germany as well as in other countries. These differences can change over time. An important example is income. Waves started in high income regions (e.g. Berkessel et al., 2021). Yet, waves persisted longer in disadvantaged neighbourhoods (De Ridder et al., 2021). Overall, socioeconomic disadvantaged regions experienced higher number of infections (Chang et al., 2021). A possible explanation are different mobility patterns (Chang et al., 2021). Similarities in the evolution can give a hint on underlying drivers of outbreak patterns. It can point to vulnerable regions and guide policy responses (Chang et al., 2021). Further, it can indicate how outbreaks spread between regions. In this paper I propose to measure similarities in the dynamic patterns regarding the spread of COVID-19 over time by a flexible data driven method, dynamic time warping. Thus I detect common patterns in different regions without comparing the time series at the same points in time. I link the similarities in dynamic patterns to economic connectedness measures and regional characteristics. Especially, I analyze commuter flows, trade flows, travel patterns and social media connections. I note that these variables were measured prior to the start of the

pandemic such that they reflect established networks rather than contact patterns during the pandemic. Specifically, I want to answer the following questions: (1) Which dynamic patterns have evolved over time and how are those geographically distributed? (2) Do regions with stronger economic or social connections share more similar dynamic patterns than less connected regions? (3) Which measures of connectedness have most pronounced correlations and do these results vary over time?

To answer these questions I apply a data-driven algorithm, dynamic time warping (DTW), on daily regional COVID-19 data from Germany and compute which districts had similar or distinct dynamic patterns. The main idea of DTW is to first align the time series. It computes which time point in one time series corresponds to which time point in the other time series. Then the distance of the aligned time series is calculated. A major advantage of DTW is that it accounts for differences in speed and amplitude and time delays. Still, I use an implementation in which regions with smaller time lags are considered more similar. By implementing the algorithm on standardized time series I net out the effect of a pure shift in levels. The algorithm was initially used in the context of speech recognition (e.g. Sakoe and Chiba, 1978) and has been widely applied in many contexts (Rakthanmanon et al., 2012). In the context of economics, however, to the best of my knowledge, there are only a few applications. Exceptions are Mastroeni et al. (2021) who research (de-)coupling of oil price measures, Franses and Wiemann (2020) who cluster US states' business cycles, Raihan (2017) who assesses the methods' performance in predicting US recessions and Wang et al. (2012) who analyze time-series patterns in international foreign exchange markets.[1]

I combine the DTW computation with clustering. This is defined as finding groups in data such that observations within a group ("cluster") are similar and observations from different groups are dissimilar. Thereby, I define clusters of districts with small DTW distances and similar dynamic patterns within the group. I distinguish five distinct clusters, two larger and three smaller clusters. A clear East-West pattern is visible. This reflects a small first wave and severe third and especially second wave in East Germany in comparison to West Germany. The smaller clusters are distributed across Germany without a concentration in specific regions. The geographic distribution points towards similar dynamic patterns in adjacent districts. Yet very large differences in adjacent districts also exist.

The results suggest that more connected regions have more similar dynamic

[1]Mastroeni et al. (2021) further provide a small number of references that use DTW in finance research.

patterns. These relationships are also present conditional on geographic distance. Private travel and trade flows have a negative correlation with DTW distance. A one standard deviation increase in (log) private travel flows between two districts increases the probability for the two districts to have a small DTW distance by 7.5%. An analysis of the DTW distances for three waves separately suggests that during the first wave, in which many businesses were closed, similar dynamic patterns correspond to stronger private networks, while in the third wave both private and economic networks were important. For the second wave the network measures do not significantly reduce the DTW distance, but both geographic distance and difference in the driving distance to international airports increase the DTW distance. This corresponds to the second wave starting after the holiday season. The relationship with social media connections and federal state is less clear. According to a variable importance analysis using random forests, personal trips is the most important variable for explaining DTW distances. The connectivity measures business travel, geographic distance and commuters are also among the most important variables. The most important variables regarding the overall cumulative cases per district are variables regarding the employment structure and sectoral composition, income variables and density related variables. Aggregated network measures however, have a low estimated variable importance. This could imply that the connectivity of a district is related to the timing of the waves but not strongly connected to the overall intensity. I note that the analysis does not represent exogenous shocks and is descriptive.

My research contributes to the literature relating the spread of viruses to economic networks. Oster (2012) finds that an increase in exports leads to an increase in HIV infections for the case of Sub-Saharan Africa. Further, she finds larger effects for regions with more established road networks. The role of booms and downturns is researched by Adda (2016) who points out that booms increase the spread of epidemics and links this to higher personal contacts induced by a rise in travelling. Further, Adda (2016) finds that interregional trade intensifies inter-regional spread of viral diseases. In contrast to these papers I measure dynamic patterns of disease spread in a flexible way and not within a parametric panel estimation. Further, I compare a larger set of different measures of economic connectedness. Especially, I can differentiate between means of travel such as commuting, business travel and private travel. Knittel and Ozaltun (2020) point out the significant and robust correlation of commuter shares and public transport use with COVID-19 death rates. The role of these variables is supported by several other papers such as Jo et al. (2021) who find a positive association of COVID-19

cases and connectedness calculated based on transport use data, and Harris (2020) who underlines the role of public transport for virus transmission in the case of New York. Adda (2016) points out the role of transportation regarding other infectious diseases. Kuchler et al. (2021) find that the spread of COVID-19 correlates with social media connections via Facebook.

Further, I add to a very large and growing body of (economic) literature on the COVID-19 pandemic. Methodologically my analysis is related to Schuppert et al. (2021). They cluster German districts with a DTW based similarity measure and identify the most relevant cluster for each state. However, their main focus is different as they apply the clustering to analyze lockdown impacts on state level. Two further examples that research inter-regional or international COVID-19 dynamics with dynamic time warping are Rojas-Valenzuela et al. (2021) and Stübinger and Schneider (2020) who set up predictive models. Stübinger and Schneider (2020) use the computed lead and lag relationships in a cross-country framework. Rojas-Valenzuela et al. (2021) also use DTW within an intra-country clustering framework for the US.

Alternative methodological approaches to detect latent groups in time series or panel data are model-based clustering. In the context of economics Frühwirth-Schnatter et al. (2016) and Frühwirth-Schnatter et al. (2018) apply Markov-Chain clustering on the effect of first birth on mothers' long-run career outcomes and on the effect of plant closures on labour market outcomes of workers. Heterogeneous coefficient models such as the classifier Lasso (Su et al., 2016) are a further approach to detect latent groups. Reviews of time series clustering methods are given by Aghabozorgi et al. (2015) and Liao (2005).

My analysis is also related to a literature that links differences in the COVID-19 pandemic across countries or regions within a country to pre-pandemic characteristics measured at the regional level.[2] There are studies that explore the regional relation of COVID-19 with a specific variable in detail. Examples include social capital (Bartscher et al., 2021) and tuberculosis vaccination (Bluhm and Pinkovskiy, 2021). These studies typically have a causal design. A different approach is to relate the regional pandemic situation to a set of regional characteristics and compare their importance, such as Ehlert (2021) who estimates spatial regressions for the case of Germany. Doblhammer et al. (2022, 2021) take a variable selection approach to identify correlations between regional characteristics and regional COVID-19

[2]This literature on pre-pandemic characteristics is distinct from the large and important literature on (regional) interventions and their effects on COVID-19 infections which I do not discuss here.

cases for subsequent 14 day intervals independently. Thereby the authors combine the analysis of dynamic differences and overall intensity. Their findings suggest that the pandemic - in both the first and second wave - started in high income/ high socioeconomic status (SES) regions and later was predominantly present in low income/low SES regions. Geographical connectedness such as airports or borders only played a role at the beginning of a wave. The authors use a huge set of regional indicators and apply machine learning variable selection tools. Compared to Doblhammer et al. (2022) and Doblhammer et al. (2021) I assess the dynamic structure in a flexible way instead of looking at subsequent 14 day intervals. Further, I differentiate between intensity and dynamic structure instead of looking at a combined measure. A switching role of income is also found by Berkessel et al. (2021) for both the Spanish flu and the COVID-19 pandemic. They relate their results to more diverse networks of individuals with higher economic status but later on also higher possibilities for social distancing. Mogi and Spijker (2021) find that the association of COVID-19 cases and population density becomes important only later in the pandemic while socioeconomic factors play a role from the beginning onwards. Ehlert (2021) on the other hand finds for the case of Germany no significant correlation of income and COVID-19 cases on the regional level. Reviewing literature on urbanity and virus spread as well as urbanity and COVID-19 Goujon et al. (2021) state that the relationship of population density and virus spread is still not sufficiently answered in the literature as the results remain inconclusive. Their study analyzes European regions and finds a positive correlation of COVID-19 infections and population density in the first wave but also a decrease of regional disparities over time. The importance of regional factors is stressed by Decoster et al. (2021) who find an effect of regional variables even in an individual level analysis. Summing up, the literature finds a significant role for regional factors while sign and magnitude of effects or associations remain inconclusive in several cases or might depend on the specific circumstances studied.

Regional differences can also be driven by varying compliance to social distance measures. Papageorge et al. (2021) conduct a survey that reveals higher compliance by individuals with higher income. In part they refer these differences to higher possibilities of teleworking and larger living space. Further papers regarding differences in compliance behaviour are reviewed in Brodeur et al. (2021).

The remaining paper is structured as follows: section 2.2 describes the method Dynamic Time Warping, section 2.3 the datasets used and the institutional background, 2.4 discusses the results, also graphically, and section 2.5 links the dynamic patterns to different economic connectedness measures and differences in regional

characteristics while section 2.6 contrasts these with a geographical analysis of cumulative cases. Finally, section 2.7 concludes.

2.2 Dynamic Time Warping

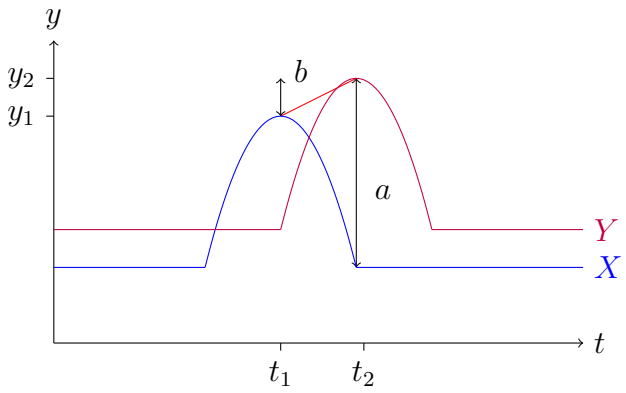
Dynamic Time Warping (DTW)[3] is a method to compute the distance between time series as well as their relative alignment in time, for example whether one follows the other with a specific lag pattern. It is designed to detect the similarity between time series that move along a similar trajectory but with a different pace. The method is flexible to changes of the relative speed along the time horizon. Therefore it can also detect and align similar shapes with different starting points. DTW was originally proposed for speech recognition (e.g. Sakoe and Chiba, 1978). Since then it has been widely applied in time series data mining as well as in classification and clustering of other indexed data sets such as images.

Figure 2.2.1 shows a simple example. Two time series X and Y are plotted. When the time series are compared with the Euclidean distance, the distance is computed at each point separately, e.g. a distance of length a at point t_2. The DTW algorithm first aligns the two series and tries to match X at point $t1$ with Y at point $t2$, this gives a distance of $y2 - y1 = b$ for these points. Figure 2.2.2 shows an example of a complete computed alignment for two time series. The left part of the figure plots two time series, the gray dashed lines connect the aligned points. The right figure plots both time series (small figures next to the axes) and the aligned time indices on a x-y scale (large figure). The time series plotted below the x-axis is called Query, the time series plotted next to the y-axis is called Reference. A point below the diagonal shows that at this pair of time indices the Reference leads the Query: a lower time index of the Reference (y value) is aligned to a higher time index of the Query (x value). For a point above the diagonal the reverse is true. At a point on the diagonal the same time indices are aligned. For the overall distance, the distance of all aligned pairs of points will be summed up.

In the following I will describe the main ideas and computational steps. Further, I will discuss some proposed extensions from the literature to the basic DTW distance computation that I consider to be especially useful when comparing econometric time series.

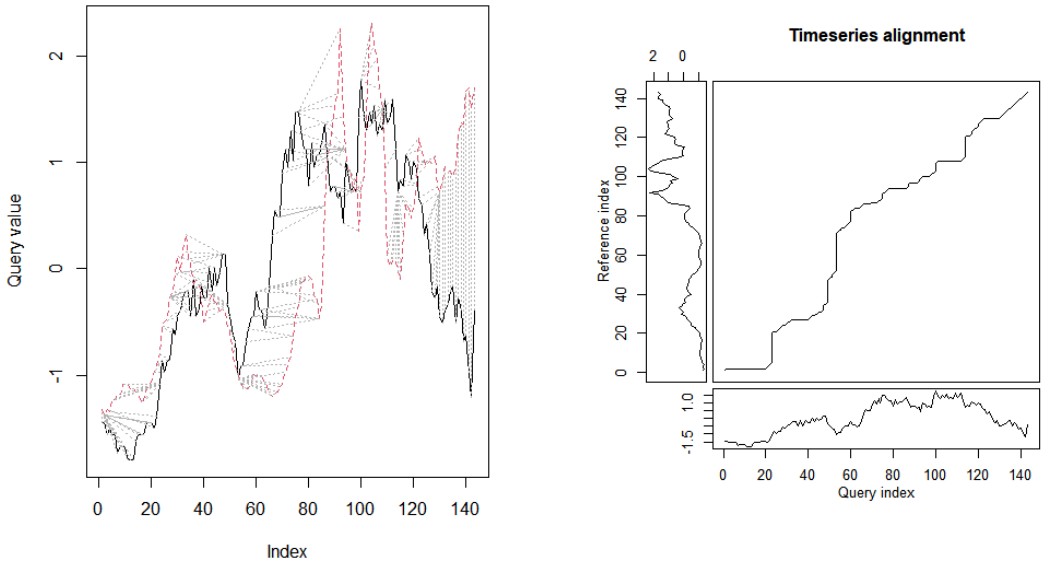
[3]A description of DTW can be found in related textbooks and articles, this exposition is based on Giorgino (2009), Kotsakos et al. (2014) and Müller (2007), the notation mainly follows Giorgino (2009).

Figure 2.2.1: Stylized Example



Notes: Example of two time series that follow a similar trajectory with different timing.

Figure 2.2.2: Alignment and Warping Curve



Notes: Own illustrations created with R package dtw (Giorgino, 2009). Exemplaric Alignment and Warping Curve. Left figure: Two time series are plotted, the gray dashed lines indicate which time indices are aligned. Right figure: The two time series and the corresponding warping path (center). The points on the warping path indicate which time indeces are aligned. Query time series: black line (left figure) and x-axis (right figure). Reference time series: Red dashed line (left figure) and y-axis (right figure). Data sources: Robert Koch Institut, Federal Statistical Office, own calculation.

Algorithm. In the following: Let X and Y be two time series indexed with $t \in 1, 2, ...T$[4] and let x_t and y_t be the corresponding realizations of the series at time t.

First, DTW computes a warping path between the time indices of the two time series. The warping path describes which time index of X is aligned to which time index of Y. The information is stored as tuples of the time indices in the warping curve $\phi(.)$:

$$\phi(k) = (\phi_x(k), \phi_y(k)). \tag{2.1}$$

Here k denotes an index from 1 to L, the length of the warping path, and $\phi_x(k), \phi_y(k) \in \{1, \ldots, T\}$. For each k, $\phi_x(k)$ denotes the index of the element in time series X that is aligned to the element with index $\phi_y(k)$ in time series Y. The algorithm aligns those time indices at which the time series take on the most similar values, subject to certain conditions: In general, points in X and Y can be aligned to more than one point in the respective other series. This corresponds to detecting speed differences in trajectories. Therefore, the values in ϕ_x and ϕ_y can contain repeated values. However, it is assumed that both series move in the same time dimension, which for example avoids loops. This translates into a monotonicity criterium:

$$\phi_x(k+1) \geq \phi_x(k), \tag{2.2}$$

$$\phi_y(k+1) \geq \phi_y(k). \tag{2.3}$$

Further, all elements in X and Y must be matched to at least one element in the other series. That means each index value of the time series has to appear at least once in the warping curve. This is ensured by the step size condition:

$$\phi(k+1) - \phi(k) =\in \{(1,0), (0,1), (1,1)\} \tag{2.4}$$

Together with the monotonicity criterion this implies that x_1 is matched to y_1 and x_T is matched to y_T. The warping curve contains the information which points in the two series are closest and through these "pairwise correspondences" (Giorgino, 2009, p.3) indicates how one series has to be stretched or compressed to

[4]To simplify notation, X and Y have the same length T, the algorithm can handle time series of unequal length.

most closely resemble the other.

Besides the monotonicity criterion further constraints can be imposed. The Sakoe-Chiba band (Sakoe and Chiba, 1978) limits the allowed compression/stretch of the series by imposing that only elements that lie within a time difference of maximum T_0 can be matched:

$$|\phi_x(k) - \phi_y(k)| \leq T_0. \tag{2.5}$$

This means that the warping curve (compare Figure 2.2.2, right-hand side) cannot lie more than T_0 off the diagonal. The Itakura parallelogram (Itakura, 1975) allows only small time distortions in the beginning and end of the time series and larger distortions in between.

After the alignment is computed, the Euclidean distance between each matched pair of points, i.e. each tuple in the warping curve is computed and summed up (and potentially normalized). Therefore the Dynamic Time Warping distance can be seen as a generalization of the Euclidean distance of time series (Aggarwal and Reddy, 2014) with the latter having a warping curve of $\phi_x(k) = \phi_y(k) = k$. Other distance measures than the Euclidean distance are possible. The Euclidean distance is a very common choice which I use in the computations. In the following definitions of step patterns I directly use the Euclidean distance for simplicity.

The warping curve is the result of an optimization problem that minimizes the overall summed distance. The summation of the distances can follow different recursive *step patterns*. Thereby it is for example possible to penalize time distortions in the alignments. Two well known and highly used choices, symmetric 1 and symmetric 2 are defined in the following. *symmetric 2* is defined by the following recursive formula:

$$d((x_1, ...x_n), (y_1, ..., y_m)) = \sqrt{(x_n - y_m)^2} + \min \begin{cases} 2 * d((x_1, ...x_{n-1}), (y_1, ..., y_{m-1}), \\ d((x_1, ...x_{n-1}), (y_1, ..., y_m)), \\ d((x_1, ...x_n), (y_1, ..., y_{m-1})). \end{cases} \tag{2.6}$$

This pattern does not penalize any form of stretching or compressing in the alignments which is achieved by multiplying the distance corresponding to the diagonal step with a factor two.

In contrast, the step pattern *symmetric 1* assigns a lower distance to alignments

that involve less stretching/compression and contain more diagonal steps:

$$d((x_1,...x_n),(y_1,...,y_m)) = \sqrt{(x_n - y_m)^2} + \min \begin{cases} \{d((x_1,...x_{n-1}),(y_1,...,y_{m-1}), \\ d((x_1,...x_{n-1}),(y_1,...,y_m)), \\ d((x_1,...x_n),(y_1,...,y_{m-1}))\}. \end{cases} \tag{2.7}$$

Because I want to allow for compression/ stretching but consider time series as more similar that are less shifted in time I use step pattern symmetric 1 in the computations. Müller (2007) discusses that instead of the parameter 2 any real-numbered weighting can be imposed and refers to the definitions of symmetric 1 as the "classic" algorithm. This definition is also used in other literature such as Keogh and Ratanamahatana (2005).

I use the overall summed up DTW distance as defined and refered to as DTW distance for example by Kotsakos et al. (2014) and Müller (2007). It is also possible to compute the normalized DTW distance which refers to the average per step distance along the warping path. The latter is especially useful when comparing time series of different length (Franses and Wiemann, 2020; Giorgino, 2009). I note that DTW is not a metric distance as the triangle inequality does not hold (e.g. Kotsakos et al., 2014).

Limitations. A potential dysfunctionality of DTW is that it can produce so-called “singularities”: If two time series X and Y differ in their Y-axis values this can lead to the alignment of one time point in X to a large number of time points in Y (this is for example discussed in Keogh and Pazzani (2001) and in Deriso and Boyd (2019)). Differences in the Y-axis can therefore lead to “spurious warping” (Keogh and Pazzani, 2001) even if the time dimension was already correctly aligned in the two time series. I approach this problem by applying the algorithm on standardized time series, which is strongly recommended by Mueen and Keogh (2016) and Rakthanmanon et al. (2012). Further, I use step pattern symmetric 1 which penalizes non-diagonal steps in the warping path. A further potential restriction is the high computational complexity of DTW. Solutions to this have been proposed. A highly cited approach is proposed by Keogh and Ratanamahatana (2005). In my application the relatively small dataset does not require such techniques.

Clustering. Clustering refers to finding meaningful groups in a dataset that is not labelled, i.e. data that has no group variable ex ante. The aim is to define groups (clusters) such that observations within a cluster are similar and points in distinct clusters dissimilar. In this setting the observations are time series and the DTW distance is used to quantify their similarity. When DTW is used for clustering it is important to use a step pattern that leads to a symmetric distance matrix.

I apply a hierarchical agglomerative algorithm. First the whole hierarchical clustering tree is computed: each observation/ time series starts as a single cluster and in each step the two clusters with minimum distance are merged. The tree represents the set of all possible clustering outcomes. The cluster merge steps require a definition of the distance between two clusters, i.e. how to aggregate distances from observations within the clusters to distances of the clusters. In order to form coherent clusters with similar observations I choose the complete linkage criterion: the distance between two clusters A with observations $i = 1, .., I$ and B with observations $j = 1, ..., J$ is the maximal distance between observations:

$$d(A, B) = \max_{i \in A, j \in B} \text{distance}(i, j) \tag{2.8}$$

In general hierarchical algorithms are deterministic and have an intuitive graphical representation. A dendrogram plots the whole clustering hierarchy together with the corresponding inter-cluster distances. When the dendrogram is "cut" at a certain height, this represents a specific clustering outcome of the dataset.

The aim is to find clusters such that the intra cluster distance is low and the distance between clusters is high. At the same time a too granular clustering reduces interpretability. A specific clustering, corresponding to a specific vertical cut in the dendrogram, can be interpreted as present in the dataset, if the next higher clustering level with fewer clusters is at a substantial bigger height difference compared to the next lower clustering level with more clusters (Hastie et al., 2017, p. 521). This means that further merging of clusters results in a substantial increase of intra cluster differences. Below the cut differences in height should be small: a more granular clustering does not substantially add to intra cluster similarity. At the same time there is no single clear cut decision criterion. This lies in part in the unsupervised nature of clustering.

2.3 Data and Institutional Background

2.3.1 Datasets

COVID-19 Data. Data on registered COVID-19 infections per day in German regions is provided by the Robert Koch Institut (RKI) and downloadable from ESRI Deutschland (Robert Koch Institut, 2021a)[5]. The RKI is the federal governmental public health institution. Among its tasks are research, prevention and control of infectious diseases[6]. Information on COVID-19 infections in Germany is reported for 412 regions separately. Data is reported for 400 districts (equivalent to NUTS 3, all except for Berlin) and for 12 subregions of Berlin separately. The dataset contains all registered infections in Germany. It is updated daily with new registered cases and in case of known reporting errors. For each registered infection the dataset contains information on the date the case was communicated to the local health authority.

Key Variable. I compute the daily number of new registered cases per 100,000 inhabitants in each district. For the main analysis I use the **7 day incidence rate**: the sum of new cases during the last 7 days per 100,000 inhabitants (i.e. actual day up to actual day -6). This measure was frequently used in German media coverage, political communication and is a main figure reported by the RKI. Absolute numbers of new registered cases per day have since the start of the pandemic systematically fluctuated over weekdays with fewer reported cases on Sundays and Mondays. One important reason is fewer testing during the weekend and fewer reporting by the local health authorities to the RKI[7]. By using 7 day incidence rates this fluctuations are smoothened.

Time Window. I start the analysis with the first registered case in Germany: on 2020/01/02 and use data up to 2021/06/30. This corresponds approximately to the end of the 3rd wave. When analyzing the three waves separately, I follow the classification by Schilling et al. (2021b) and Schilling et al. (2021a) for the first and second wave: 2nd March 2020 - 17th May 2020 as wave 1 (calendar week 10-20), 28th September 2020 - 18th February as wave 2. For the third wave, that was not

[5] `https://www.arcgis.com/home/item.html?id=dd4580c810204019a7b8eb3e0b329dd6`, downloaded September 17, 2021.

[6] compare RKI website: `https://www.rki.de/EN/Content/Institute/institute_node.html`

[7] See for example the RKI COVID-19 FAQ website: `https://www.rki.de/SharedDocs/FAQ/NCOV2019/gesamt.html`, accessed on August 11, 2021

included in their classification, I use the time window 19th of February 2021 - 30th June 2021. Figure 2.3.1 plots the 7 day incidence rate aggregated for Germany.

Standardization. Before computing the DTW distances I standardize the regional 7 day incidence rate. I compute $I_{zit} = \frac{I_{it}-\mu_i}{sd_i}$, where I_{zit} denotes the standardized incidence rate in region i and day t, I_{it} the raw incidence rate, μ_i the mean incidence rate in region i across time and sd_i the corresponding empirical standard deviation. This standardization aims at avoiding singularities in the DTW alignment (compare section 2.2). Standardization is strongly recommended by Mueen and Keogh (2016) and Rakthanmanon et al. (2012).

Regional Characteristics. Further, I use a rich set of pre-pandemic regional covariates from the INKAR database (BBSR Bonn, 2021) – an online database for regional data provided by the Federal Institute for Research on Building, Urban Affairs and Spatial Development. For these variables the most recent information is used that was available in August 2021. In all cases this is prior to 2020. For a full list of variables, corresponding years and definitions see Appendix Table 2.C.1 and the corresponding section.

Commuting Data. Commuter flow data from district to district is provided by the Federal Employment Agency (Statistik der Bundesagentur für Arbeit, 2020). A commuter is defined as a person with different districts reported for living and working - where the reporting is done by the employer for official social security records. Only employees in the mandatory social security scheme are included, this includes apprentices and excludes employees earning less than 450 euros, self-employed and civil servants. All values of one or two commuters are anonymized and set to zero. I therefore impute all values of zero commuters with a value of one.

I use data on **Travel Flows and Trade Flows** from the Federal Ministry of Transport and Digital Infrastructure (Bundesministeriums für Verkehr und digitale Infrastruktur (BMVI). Berlin, 2014) to measure the economic connectedness between districts. Both datasets are provided for the year 2010 and a projection for the year 2030, I use information for the year 2010. I measure trade flows in tons of goods transported between two districts, aggregated over different good categories. Further, the data on trade flows differentiates in some cases between the overall route and a main route in between such that different number of tons

are transported on main route and overall route. Therefore, I focus on the cases in which main route and overall route coincide. This represents ca. 98% of all tons transported. Travel data is differentiated into six different categories: commuting to work, trips to educational institutions, running errands such as shopping or doctor visits, business trips, holiday trips (5 days and more) and other personal trips. I treat those six categories separately but aggregate data across different means of transportations. The datasets treat airports and ports as separate units which I assign to the district in which these are located. I do not use data on commuting trips from the BMVI data but the more recent commuting data from the IAB. Further, I reassign the data into the district systematic of 2019/2020. 6 districts were rearranged into more than one new district, I use the new district where the majority of the population is reassigned to. This is in all cases more than 80%.

Social Media Connections. I use data on regional connections via Facebook friendships: The Social Connectedness Index (SCI) is a measure for the probability that Facebook users from district i and j are facebook friends. Specifically, the Social Connectedness Index (SCI) is defined as the number of Facebook friendships of users in region i with users in region j divided by the product of Facebook users in region i and j:

$$SCI_{i,j} = \frac{FBconnections_{i,j}}{FBusers_i * FBusers_j}$$

Anonymized and regionally aggregated data is made publicly available by Meta (Facebook Data for Good Program, 2021), for a description see also the accompanying paper Bailey et al. (2018). Data is only available as a scaled transformation: the scaled SCI is rounded to integers and scaled to have a maximum value of 1,000,000,000 and a minimum value of 1. I log transform the variable scaled SCI in all computations. The dataset contains Facebook friendships from the year 2016. Regional identifiers are updated, I downloaded the dataset on October 29, 2021.

I use data on **regional population** as of 31st December 2019 by district provided by the Federal Statistical Office (Destatis) (Statistisches Bundesamt, 2020) and data on the number of inhabitants of 12 different Berlin subregions as of 31st December 2019 by the Statistical Office of Berlin Brandenburg (Amt für Statistik Berlin Brandenburg, 2020).

Geographic Data. Geographic data on German districts and states as of December 31, 2019 is provided by the Federal Agency for Cartography and Geodesy (BKG) (GeoBasis-DE / BKG, 2021b,a). As geographic distances I compute both polygon distance - the smallest distance between two districts, which implies a distance of zero for neighbouring districts, and the distance between district centres. The two measures are nearly perfectly correlated, therefore I only use polygon distance.

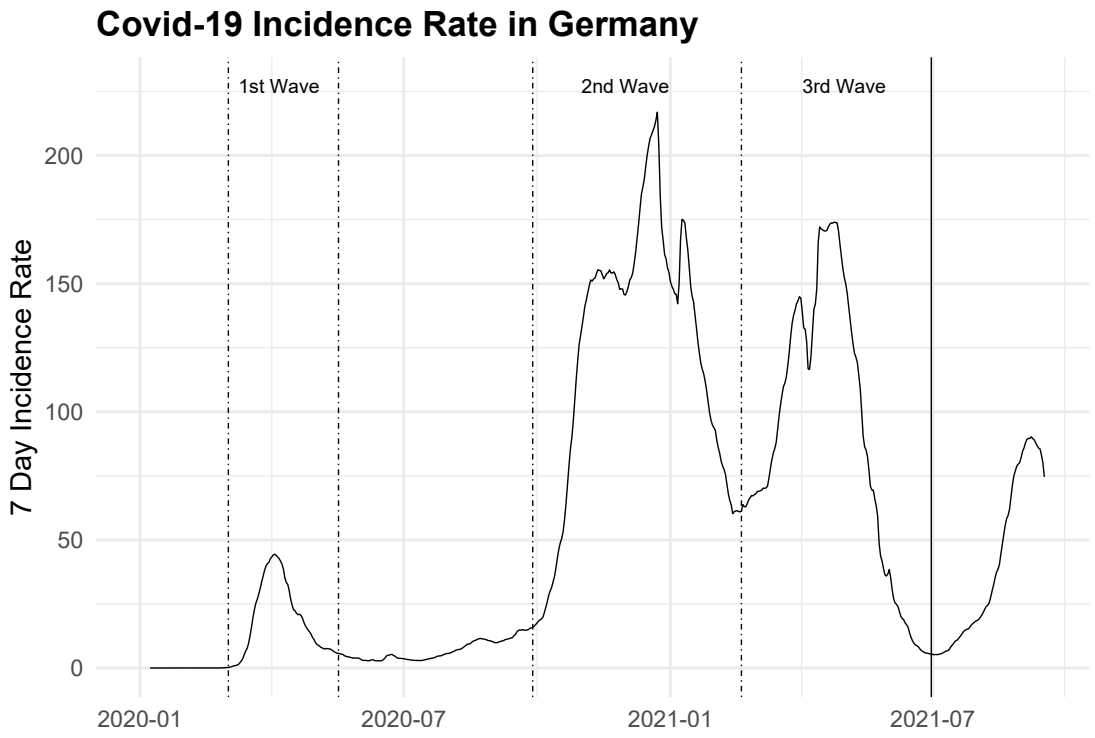
Data Limitations. Ideally all, also unobserved, infections would be included in the analysis while of course only observed infections are available. It is known that test capacities in Germany were systematically increased during the pandemic. This process might lower the difference between total and observed infections over time. I have no specific knowledge that or how the number of undetected infections varies systematically across German districts. The dataset contains information of positive PCR tests which have to be registered. Results of so called rapid tests are not included. Rapid tests became available in Germany during autumn 2020 and available for everyone in pharmacies and supermarkets in spring 2021. Over the period of my analysis individuals with positive rapid tests were supposed to validate the result with a professionally carried out PCR test. Test capacities/ number of tests carried out fluctuate systematically over the weekdays. This problem is leveraged by computing the 7 day incidence rates (compare above).

Regional characteristics as well as the economics connectedness measures are not reported for the twelve Berlin subregions. Therefore those 12 regions are excluded from the respective parts of the analysis which is carried out on the remaining 400 districts.

2.3.2 Institutional Background

A chronological overview of the first year of the COVID-19 pandemic until February 2021 in Germany is given by Schilling et al. (2021a) and Schilling et al. (2021b) which I briefly summarize in the following paragraph: The first COVID-19 case was registered in Germany on the 27th January 2020. During the first weeks a large proportion of infected persons were young and infections were corresponding to holiday trips abroad and carnival festivities. Calendar week 10 can be interpreted as the start of the first wave, infections were then to much higher proportion unrelated to international travel (Schilling et al., 2021a). Containment measures were implemented in the first wave and a lockdown started in calendar week 13,

Figure 2.3.1: Time Series COVID-19 Incidence Rate in Germany



Note: Own illustration based on Schilling et al. (2021b,a) for the definition of wave 1 and 2. COVID-19 7 day incidence rate in Germany, accumulated over all districts. Data source: RKI, Destatis.

the maximum of national reported incidence rate in the first wave was 43 (Schilling et al., 2021b). Calendar week 20 marks the end of the first wave (Schilling et al., 2021b,a). After the first wave ended mid May 2020 the incidence remained low during the summer with local outbreaks as exceptions. The second wave started in calendar week 40/2020 and was characterized by much higher incidence rates than the first. During a partial lockdown in November gastronomy and hotels closed. Starting in December further restrictions were imposed including school closures and closures of shops. The peak of the second wave was reached in the second last week of 2020 with a national incidence of 210.

After a short plateau in February the incidence rose sharply again into a third wave. The third wave ended around 30th June 2021 (compare Figure 2.3.1) which is the end of the time frame studied here.

Bauer and Weber (2021) provide a database of German containment measures.[8] The majority of measures were introduced at the state level. At the end of the

[8]The database can be accessed via `https://www.iab.de/de/daten/corona-eindaemmungsmassnahmen.aspx`

study period there were in all 16 states still at least some containment measures in place. von Bismarck-Osten et al. (2022) point out that regulations were put in place at state level but the management of dealing with infections in schools and quarantine was decided at the local health office (which divide Germany into 375 distinct areas).

Vaccinations started at the end of December 2020 in whole Germany for highly vulnerable persons, medical staff working at COVID-19 units and persons with contact to highly vulnerable persons. It continued with a priority system and since approximately summer 2021 vaccination possibilities have been available for all adults. At the end of the time period stuidied, the RKI estimated that 36,5% of the population were fully vaccinated (Robert Koch Institut, 2021b).

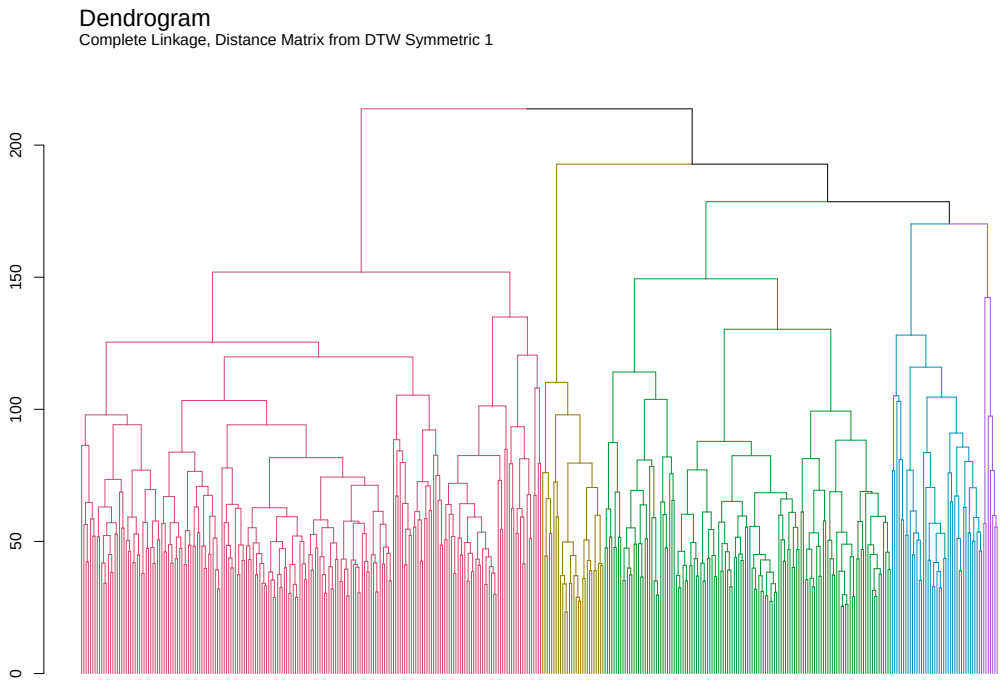
2.4 DTW and Clustering Results

Clustering. I compute the DTW distances with step pattern symmetric 1 using all available 412 regions and the whole time period. Then I analyze the presence of groups using hierarchical clustering and complete linkage as described in section 2.2. DTW distances are computed using the R package dtw (Giorgino, 2009). Figure 2.4.1 shows the corresponding dendrogram. The dendrogram starts at the bottom where each district represents a distinct cluster and ends at the top where all $N = 412$ districts are merged into one cluster. The whole tree represents $N - 1 = 411$ cluster merges where each time the two closest clusters are combined into one. Each merge of two daughter nodes/clusters is represented by a horizontal line. The height of this horizontal line (y-axis in the dendrogram) is the distance between the two clusters that are merged. Due to the complete linkage criterion the dendrogram directly reveals the largest pair-wise distance between two time series by the height of the highest merge/horizontal line.

In my interpretation, the dendrogram (Figure 2.4.1) shows a clustering structure and points to the presence of 5 clusters. I use the clustering outcome, especially the number of clusters and cluster labels to compare the different dynamic patterns. The regressions are based on the DTW distances directly. A final determination of the number of clusters could be further justified by additional analyses.

Time Series Patterns. Figure 2.4.2 plots the time series of 5 distinct clusters. In all clusters the three waves can be distinguished. The main difference between the clusters is the different relative intensity of the three waves. The exact timing of the three waves differs also within clusters – this is allowed by the DTW distance

Figure 2.4.1: Dendrogram Hierarchical Clustering of DTW Distances



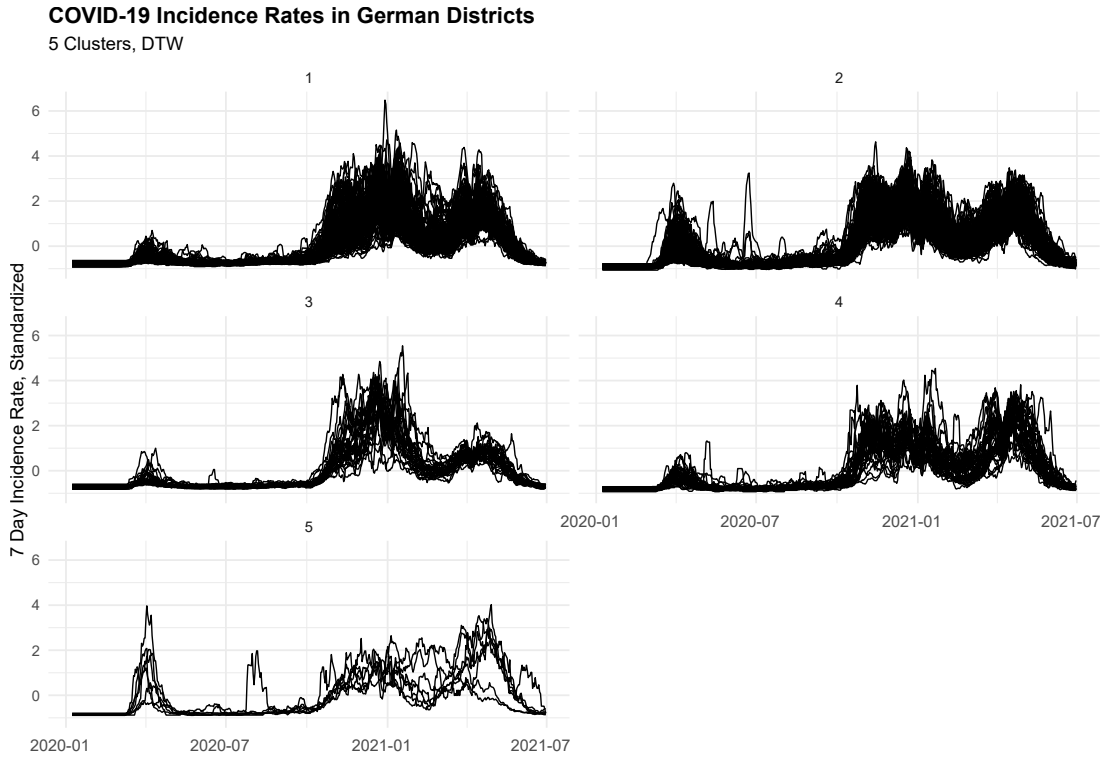
Note: Dendrogram for hierarchical clustering based on DTW distance. Time series used are COVID-19 7 day incidence rates in 400 German districts and 12 regions in Berlin. For a clustering outcome of five clusters the observations are colured differently according to cluster membership. Data source: RKI.

used.

Cluster 5 stands out as the first wave is here more severe than in other clusters. Further, some regions in this cluster have a long combined second and third wave. Cluster 1, 3 and 4 share a high second wave and a relatively small first wave. The third wave increases in intensity in cluster 1,3 and 4 with cluster 4 having a higher third than second wave. In cluster 2 the first wave is much higher compared to the other clusters, while it is still much shorter than wave 2 and 3. Also the timing of the first wave differs across districts in cluster 2. In this cluster wave 2 is the highest but wave 2 and 3 are of comparable height. The size of the clusters differ with cluster 2 containing about half of all districts, see Table 2.4.1 for exact values.

Because of the standardization all comparisons of intensities across time are within districts. Across districts one can infer relative comparisons: whether in region i wave 1 was more severe than wave 2, while in region j it might be the opposite. I also plot the time series of non-standardized incidence rates, see Figure 2.D.1. The comparisons shows that by standardizing mainly the differences regarding the peaks are decreased while the dynamic patterns in the distinct clusters are still visible.

Figure 2.4.2: Clustered Time Series



Note: Time Series of standardized COVID-19 7 day incidence rates in 400 German districts and 12 Berlin regions. Observations are assigned into 5 clusters. Time window: 2020/01/02-2021/06/30. The clustering is computed with hierarchical clustering based on DTW distance. Data source: RKI, Destatis, Statistical Office Berlin Brandenburg, own computations.

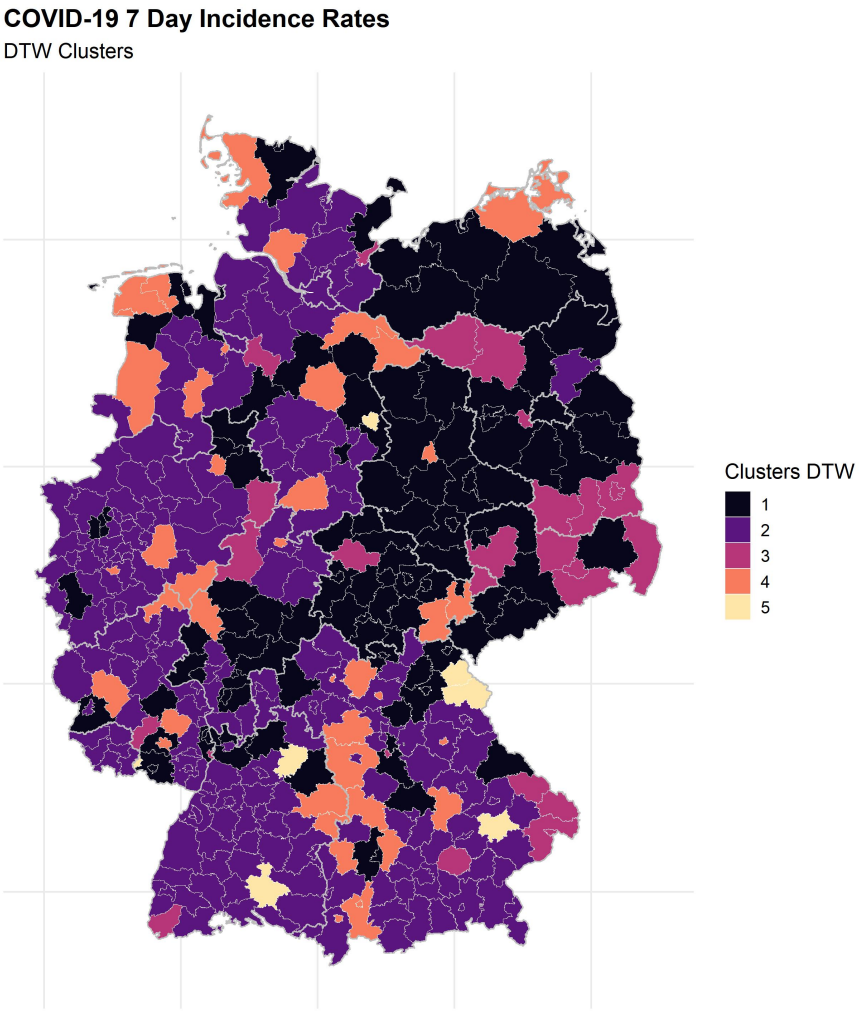
Geographic Patterns. Figure 2.4.3 plots the corresponding geographic distribution of clusters. A clear East-West pattern is visible with Cluster 1 being the predominant cluster in East and 2 the predominant cluster in West Germany. This reflects a relatively small first wave in East Germany compared to West Germany. The smaller clusters 3,4, and 5 are distributed across Germany without a concentration in specific regions. However, clusters 3 and 4 form small groups of 2 or more adjacent districts in many cases. This points towards similar dynamic patterns in adjacent districts. Cluster 5 on the other hand which contains only 7 districts with a very distinct pattern is mostly scattered across the map. Very large differences in adjacent districts also exists. Figure 2.4.4 plots examples of time series of pairs with very small and very large differences respectively. The largest cluster 2 is also the predominant cluster among urban districts with a share of more than two thirds. Across rural districts the clusters are more evenly distributed (compare Table 2.4.1). Notably the different geographic patterns do not follow the geographic state boundaries (illustrated by grey lines in all maps). This is interesting as the state level was the main level of political differences in mitigation and containment policies (Bauer and Weber, 2021).

Table 2.4.1: Urban-Rural Distribution of Clusters

Cluster	1	2	3	4	5
Rural	77	69	22	29	6
Urban	43	137	5	12	1
Sum	120	206	27	41	7

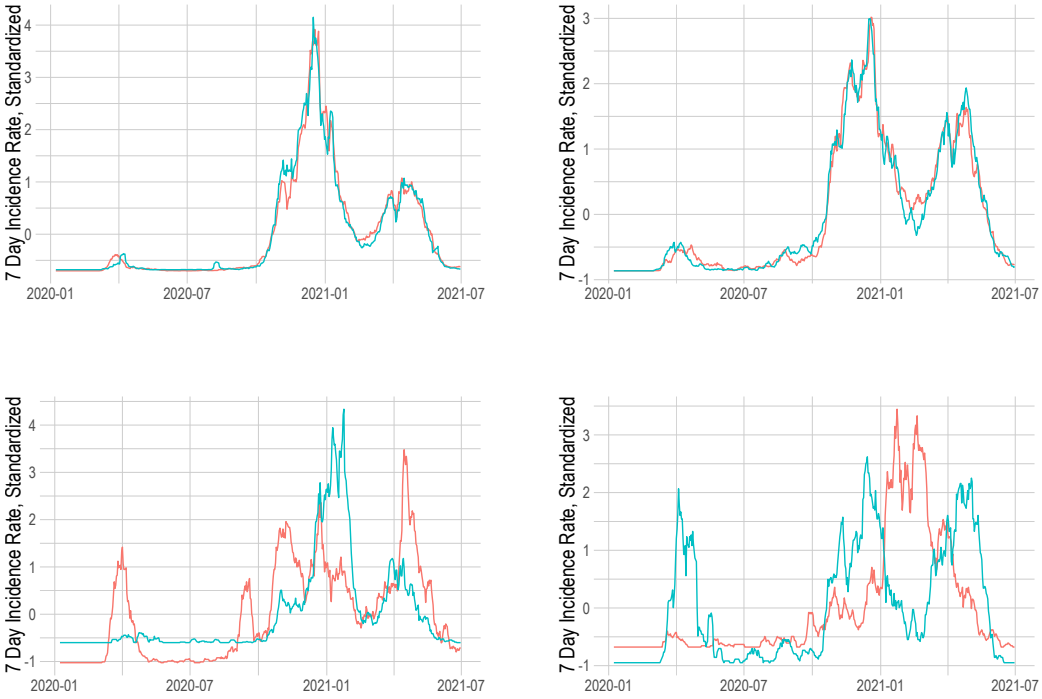
Note: Distribution of Cluster assignment with 5 clusters. Clusters are computed with hierarchical clustering based on DTW distance. Time series used are COVID-19 standardized 7-day incidence rates. 400 German districts and 12 Berlin regions. Data source: RKI.

Figure 2.4.3: Geographic Cluster Distribution



Note: Map of clustering outcome with 5 clusters. Hierarchical clustering based on DTW distance. Time series used are COVID-19 7 day incidence rates in 412 German regions: 400 districts and 12 Berlin regions. For Berlin the map shows the clustering outcome with the highest frequency: 10 regions are in cluster 1, 1 in cluster 2, 1 in cluster 3. Thicker grey lines denote state borders. Data sources: RKI, Destatis, own computation, geographical data: BKG.

Figure 2.4.4: Examples of Large and Small Distances



Note: Time Series graphs of four district pairs with very small (upper graphs) and very large (bottom graphs) Dynamic Time Warping distances. The plotted variable is the standardized COVID-19 7 day Incidence rate in the respective districts. These Time Series were used to calculate the DTW distances. The districts are Top Left: Dresden, city (orange), Görlitz (green); Top Right: Main-Kinzig-Kreis (orange), Mannheim, city (green); Bottom Left: Würzburg, city (orange), Ostprignitz-Ruppin (green); Bottom Right: Flensburg, city (orange), Straubing, city (green). Data source: RKI, Destatis, own computations.

2.5 Distance, Networks and Regional Characteristics

2.5.1 Nearest Neighbours and Large Differences

In this section I analyze which economic connections and regional differences correspond to small or large DTW distances between the respective districts. A first descriptive way of approaching the regional associations is to look at pairs of districts with the smallest distances: nearest neighbour. For each district i, I identify the district NN_i which is closest in terms of the DTW distance:

$$NN_i = \min_j d_{DTW}(i, j), \tag{2.9}$$

where $d_{DTW}(i, j)$ denotes the DTW distance between district i and district j. A comparison of the sample of nearest neighbours with the complementary sample of all other district pairs shows an increased frequency of location in the same state, higher commuter flows, higher social media connections via facebook and smaller geographic distance, while region type: urban/rural seems to play only a minor role (compare Table 2.5.1 and Figure 2.5.1).

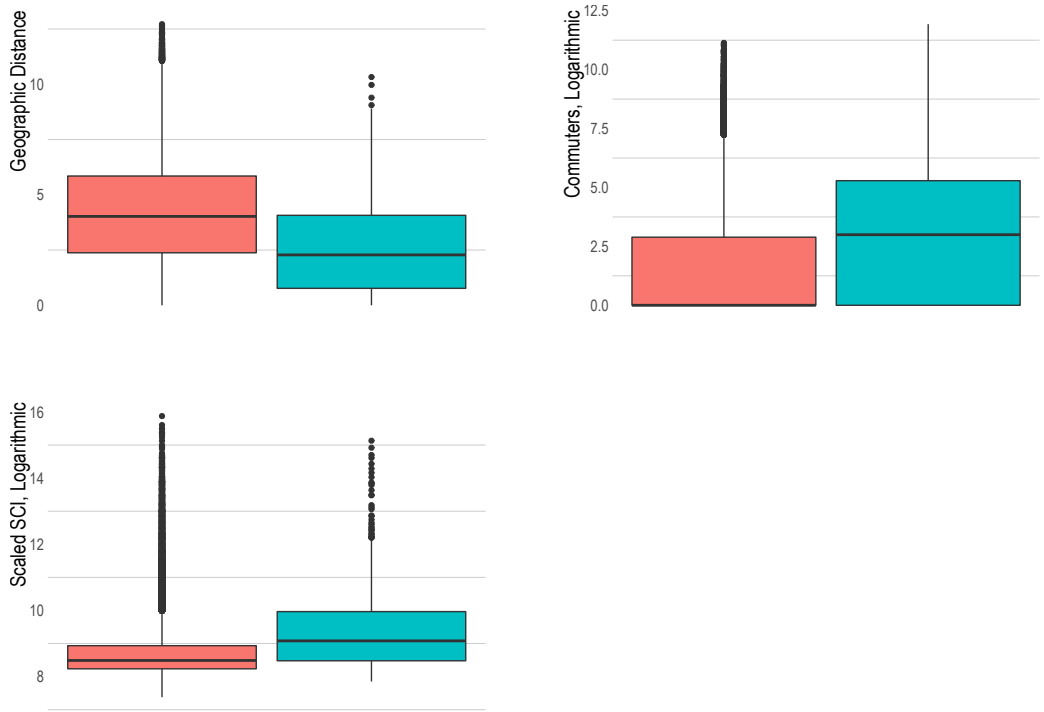
In a second step I look at these variables along the whole distribution of the DTW distance. Figure 2.5.2 plots box plots for geographic distance, commuters and scaled SCI for each decile of the DTW distribution. The graphs show that the unconditional relationship of geographic distance and DTW distance seems to be nearly linear across the distribution. The median and the third quartile of scaled SCI is higher for the 10% of district pairs with the smallest DTW distances. For the other deciles there seem to be only small differences. For number of commuters there is a decrease along the distribution of the DTW distances.

Table 2.5.1: DTW Distances: Nearest Neighbours

	Nearest Neighbours			Complementary Sample		
Variable	Mean	Median	Std Dev	Mean	Median	Std Dev
Log Commuter	3.22	3.04	3.348	1.40	0	2.062
Geographic Distance	2.674	2.270	2.063	4.212	4.020	2.385
Scaled SCI, Log	9.56	9.08	1.53	8.74	8.49	0.845
Same State	0.295	0	0.457	0.119	0	0.324
Region Type	0.651	1	0.477	0.498	0	0.50
N	359			79441		

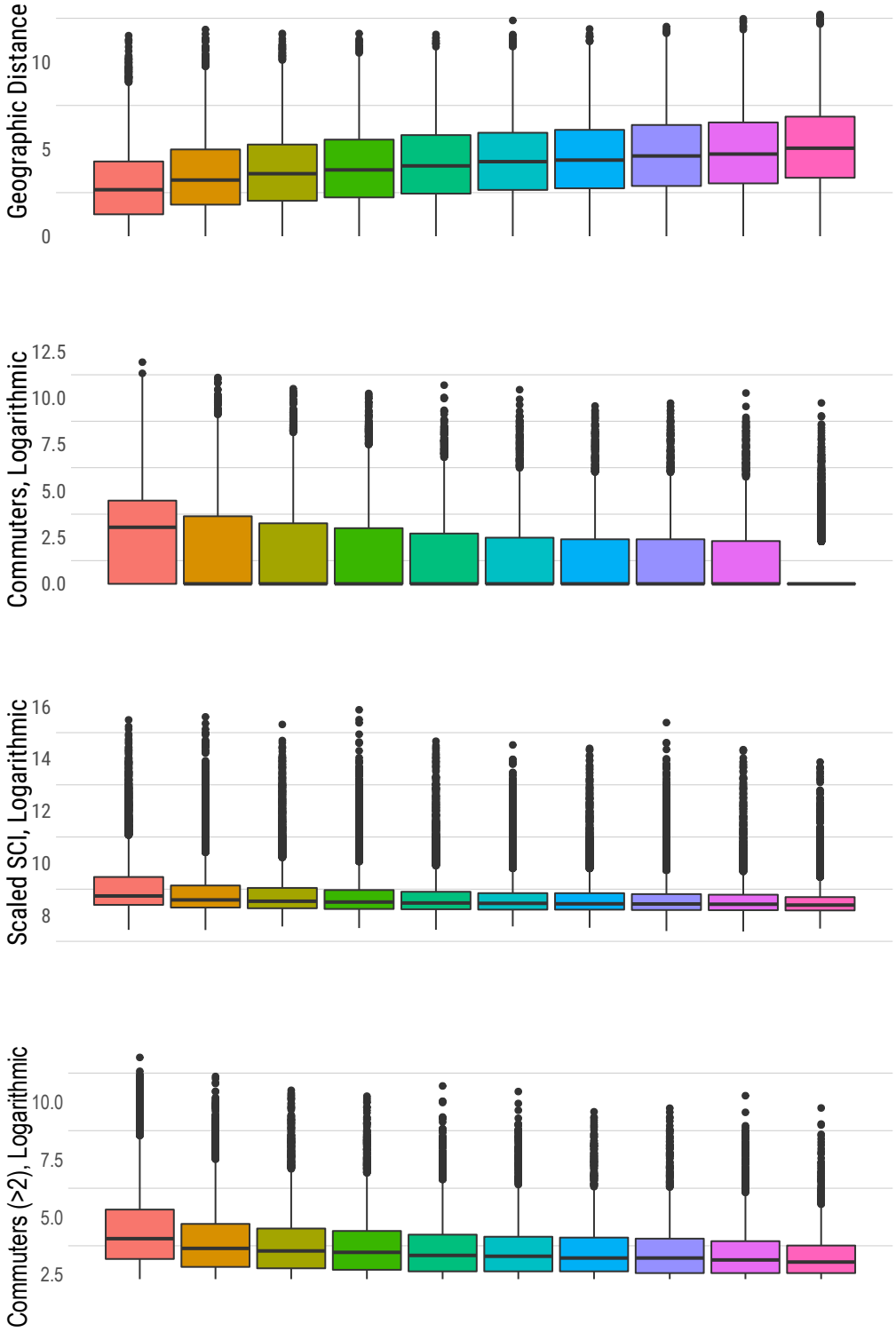
Note: Each observations is a pair of German districts, excluding Berlin. Nearest neighbours: pairs of districts that correspond to the minimum distance for at least one of the two districts, Complementary sample denotes the sample of district pairs that are not nearest neighbours. Variable Same State denotes location in the Same State, Region Type denotes that both districts are urban/rural. Geographic distance is measured in 10,000 metres. Data sources: see text.

Figure 2.5.1: Boxplots Nearest Neighbours



Note: Boxplots for geographic distance, log commuter flows and log scaled SCI. Nearest neighbours: pairs of districts that correspond to the minimum distance for at least one of the two districts, Complementary sample denotes the sample of district pairs that are not nearest neighbours. The green (right) boxplots correspond to the sample of Nearest Neighbours, the orange (left) boxplots to the Complementary Sample. Each observation is a pair of German districts. Scaled SCI is a measure for connectedness of districts via Facebook. Geographic distance is measured in 10,000 metres. Data sources: see text

Figure 2.5.2: Boxplots along the Distance Distribution



Note: Boxplots for geographic distance, logarithmic commuter flows and logarithmic scaled SCI; the bottom graph displays logarithmic commuter flows for all district pairs with at least 3 reported commuters. Geographic Distance is measured in 10,000 metres. Scaled SCI is a measure for connectedness of districts via Facebook. One boxplot corresponds to 10% of the sorted DTW distance distribution, sorted from left (smallest distances) to right. Each observation is a pair of German districts, excluding Berlin. Data sources: see text.

2.5.2 Regression Analysis

In this section I look at the relationships between the DTW distance and different economic networks in a more systematic way. In a first step using linear regressions. To account for the possible non-linearities that are suggested by the boxplots I will further apply regression forests in a second step.

I estimate the following regression specification:

$$d_{DTW}(i,j) = \alpha + \beta C(i,j) + \gamma I_{S(i)=S(j)} + \eta d_{geo}(i,j) + \delta |X_i - X_j| + \epsilon_{i,j}, \quad (2.10)$$

where the computed DTW distances $d_{DTW}(i,j)$ between each pair of districts $i, j, i \neq j$ are modelled as the sum of an intercept α, a vector $C(i,j)$ of economic connectedness measures such as log gross commuter flow between district i and j, an indicator $I_{S(i)=S(j)}$ denoting whether district i and j belong to the same state, the geographical distance $d_{geo}(i,j)$ and the absolute differences between regional characteristics X as well as an error term.

The DTW distance measures dynamic difference in a flexible way. The more similar two districts are in their dynamics the smaller is the DTW distance. However, it is not directly clear how to interpret specific numerical increases, such as an increase of 10 in the DTW distance. Therefore, I interpret the outcome on an ordinal scale. To facilitate interpretation of the regressions and estimated coefficients, I additionally compute which variables are connected to the DTW distance being small (<25th percentile), not large (<75th percentile) or below the median. Specifically, I compute the regression:

$$P(d_{DTW}(i,j) < q) = \alpha + \beta C(i,j) + \gamma I_{S(i)=S(j)} + \eta d_{geo}(i,j) + \delta |X_i - X_j| + \epsilon_{i,j}, \quad (2.11)$$

where I use the 25th, 50th and 75th percentile of all district-pair differences as q. As these regressions aim at facilitating interpretation of coefficients I use a linear probability model.

In all regressions I account for the data structure when computing standard errors by using multi-way clustering (Cameron et al., 2011) and cluster on both districts i and j. To facilitate comparisons I normalize all connectedness measures as well as the differences between regional variables to have mean zero and unit variance, except for differences in indicator variables (region type, high housing costs and same state) and geographic distance which is measured in 10,000 metres.

The measures of connectedness include log gross trade flow, log gross commuter flow, travel data and scaled SCI. Travel data is differentiated into five different categories: commuting to educational institutions, running errands such as shopping or doctor visits, business trips, holiday trips (5 days and more) and other private trips. The variable scaled SCI contains more recent information (2016) compared to the travel data (2010) but measures social contacts in presence less precisely. Further, social media connections conditional on travel data should not have an impact. Therefore, I either include the five measures of travel data or scaled SCI in the regressions. Regarding the differences in regional characteristics, I choose a set of 7 groups of characteristics: Urbanity and Density, Employment and Inequality, Sectoral Structure, Income, Age, Elderly Care and Health Provision, Education and Child Care and Geographical Connectedness. Table 2.C.1 reports all variables included in each of the groups. [9] In order to explore which of the variables are most relevant, I apply a LASSO estimator and compute a post LASSO estimation. For the model selection I choose the tuning parameter with 10-fold cross validation and select the sparsest model with a mean squared error within one standard deviation of the minimum mean squared error (compare Hastie et al., 2017). For the estimation of equation 2.11 I use a LASSO-type logistic regression to select variables (Friedman et al., 2010).

Results with DTW Distance as Outcome. Table 2.5.3 displays the results from the post LASSO regression of equation 2.10: a larger economic connectedness via trade flows implies a smaller DTW distance while the coefficient for business trips is not significant. Scaled SCI does even have a positive association with DTW distance, the same holds for the same state indicator. Both relevant in terms of coefficient size and significant is the coefficient on personal travel such that districts that are more connected via private travel have more similar dynamic patterns. The coefficient on log commuters is large and significant, but only when controlling for scaled SCI instead of travel patterns. A larger geographic distance corresponds to larger DTW distances (significant at least at the 10% level). The travel patterns errands and holiday stays are not selected by the LASSO estimator as well as only a few covariates measuring differences in regional characteristics: median income, the share of children in daycare, the share of employees working in the primary sector (specification (1)) and number of students in tertiary education (compare Table

[9] In all estimations with the DTW distance as outcome or indicators based on the DTW distance I leave out the regional characteristics "Hotel Stays/ Year", "Commuter Inflow", "Commuter Outflow" as these are very similar to included network characteristics.

2.5.3 and Table 2.C.1). This implies that a wide range of variables is correlated with dynamic patterns. The largest coefficients are estimated for differences in the share of low income households, distance to the next international airport, share of infants in daycare and physicians per inhabitants: larger differences in these variables are associated with larger dynamic differences. On the other hand especially larger differences in the share of high income households and population density are associated with smaller differences in dynamic patterns. Still a large fraction of the variance remains unexplained ($R^2 \approx 0.2$). To put the coefficients into perspective: The median DTW distance is 79.79 (Mean 83.83). Table 2.D.1 (Appendix 2.D) additionally presents results for regressions with overall DTW distance as dependent variable when including the network variables only, the full set of variables or a smaller set of differences in regional characteristics. All three regressions were computed without an additional Lasso Step. The overall pattern regarding the different network characteristics is the same.

When computing DTW distances within the three waves separately (compare Table 2.5.4)[10], the results show important differences regarding the different network measures. Personal Trips reduce the DTW distance significantly for wave 1 and 3. The same holds with a much smaller coefficient for Holiday Stays. The association with trade flows is mainly present in the third wave where it plays a large and relevant role. This suggests that during the first wave, in which many businesses were closed, similar dynamic patterns were driven by private networks, while in the third wave both private and economic networks were important. For the second wave the network measures do not significantly reduce the DTW distance, an exception is the commuter flow which is significant only at the 10% level. But both geographic distance and difference in the driving distance to international airports increase the DTW distance. This corresponds to the second wave starting after the holiday season.

[10]The 7-day incidence was also normalized for each wave separately before computing the DTW distance

Table 2.5.3: Regression Results: DTW Distance

	Dependent variable:	
	DTW Distance	
	(1)	(2)
Same State	2.156* (1.239)	
Commuters, Log		−3.175*** (0.588)
Trade Flow, Log	−1.030** (0.502)	−1.467*** (0.487)
Commuting to Education, Log	1.728*** (0.593)	
Business Trips, Log	−0.917 (1.543)	
Personal Trips, Log	−3.883** (1.750)	
Scaled SCI, Log		3.123*** (0.663)
Geographic Distance	0.851* (0.462)	1.830*** (0.371)
Region Type	1.496** (0.686)	1.382* (0.711)
Population Density	−3.665*** (0.642)	−3.961*** (0.556)
Average Living Space	1.350** (0.569)	1.512*** (0.533)
High Housing Costs	−1.257** (0.638)	−1.326** (0.662)
Recreational Areas	−1.877** (0.783)	−2.005*** (0.772)
Female Share	−0.571 (0.528)	−0.452 (0.535)
Broadband Availability	1.316** (0.547)	1.479*** (0.545)
Long Term Unemployment	−1.653*** (0.493)	−1.477*** (0.486)
Relative Female Wage	1.715 (1.120)	1.451 (1.125)
Turnout Federal Election	1.235** (0.617)	1.274** (0.620)
Industry Share	1.038 (0.669)	1.155* (0.656)
Share Services	0.976 (0.912)	0.843 (0.899)
Share Individualized Services	1.177** (0.540)	1.072* (0.552)
Share Primary Sector		0.274 (0.527)
GDP per Capita	0.454 (0.805)	0.591 (0.783)
Low Income Households (Share)	6.376*** (1.570)	6.857*** (1.550)
High Income Households (Share)	−6.466*** (1.413)	−6.986*** (1.397)
Inhabitants 65 to 85	1.709** (0.759)	1.717** (0.756)
Inhabitants 85 and older	0.304 (0.538)	0.331 (0.535)
Personell in Nursing Homes	0.825 (0.541)	0.854 (0.538)
Patients in Nursing Homes	0.954* (0.561)	0.963* (0.561)
Persons in Need of Care	1.324* (0.739)	1.299* (0.748)
Hospital Beds	−1.304* (0.723)	−1.291* (0.732)
Physicians per Inhabitants	2.165** (0.980)	1.840* (0.987)
Early School Leavers	0.924* (0.547)	0.894* (0.536)
Infants in Daycare (Share)	2.316** (0.946)	2.458*** (0.931)
Distance to Airport	2.464*** (0.610)	2.296*** (0.595)
Constant	79.844*** (2.077)	76.077*** (1.766)
R^2	0.22	0.21

Note: $N = 79800$. Significance levels: *p<0.1; **p<0.05; ***p<0.01. Linear regression results (post LASSO) with DTW distance of districts as dependent variable. Each observation is of a pair two districts. DTW distances computed with COVID-19 7 day incidence rates in 400 districts, excluding Berlin. Cluster Robust Standard errors in parentheses. Geographic distance is measured in 10,000 metres. Variables on regional characteristics (Region Type - Distance to Airport) measure the absolute difference of these characteristics for the district pairs. All input variables are normalized, except for Geographic Distance and indicator variables (Same State, Region Type, High Housing Costs). Trade and Travel data only included in column (1), Scaled SCI only included in column (2). Data sources: see text.

Table 2.5.4: Regression Results: DTW Distance - 3 Waves Separately

	Dependent variable:		
	DTW Distance		
	Wave 1	Wave 2	Wave 3
Commuters, Log		−0.727* (0.400)	−0.322 (0.294)
Same State	−1.948*** (0.587)	0.880 (0.783)	1.452* (0.773)
Trade Flow, Log	−0.085 (0.260)	−0.567 (0.350)	−0.956*** (0.306)
Errands, Log	0.865*** (0.286)		1.244*** (0.324)
Commuting to Education, Log		0.744** (0.362)	
Business Trips, Log		−0.203 (0.953)	−1.301 (1.073)
Holiday Stays, Log	−0.607*** (0.173)	−0.360 (0.253)	−0.563** (0.239)
Personal Trips, Log	−1.994*** (0.688)	−0.316 (1.507)	−2.043** (1.033)
Geographic Distance	−0.455* (0.254)	1.057** (0.470)	
Region Type	−0.441** (0.192)	0.984** (0.484)	
Population Density	0.664* (0.391)	−1.576*** (0.405)	−1.201*** (0.383)
Average Living Space	−0.408* (0.215)	1.536*** (0.389)	1.073*** (0.381)
High Housing Costs	−0.801*** (0.199)	−1.098*** (0.374)	−0.826*** (0.275)
Recreational Areas	0.034 (0.458)		
Female Share	−0.890*** (0.189)	−0.573** (0.267)	−0.311 (0.258)
Broadband Availability	0.420 (0.328)		
Unemployment	1.573*** (0.467)	0.959* (0.501)	
Long Term Unemployment	−0.616** (0.273)	−2.066*** (0.443)	−0.705** (0.298)
Relative Female Wage	−0.526 (0.404)	−0.669 (0.480)	
Turnout Federal Election	0.363 (0.320)	0.145 (0.381)	
Industry Share	0.206 (0.266)		
Share Services			0.784** (0.315)
Share Individualized Services		1.004** (0.436)	0.450* (0.270)
Share Primary Sector	0.256 (0.316)	0.476 (0.393)	
Median Income	0.052 (0.259)		
GDP per Capita		−0.028 (0.206)	
Low Income Households (Share)	−1.107*** (0.228)	2.557*** (0.951)	
High Income Households (Share)		−3.344*** (0.846)	−0.895*** (0.186)
Inhabitants 65 to 85	0.944** (0.392)	0.869** (0.431)	
Inhabitants 85 and older	−0.271 (0.277)	−0.411 (0.390)	
Personell in Nursing Homes	0.344 (0.256)	1.522*** (0.475)	0.222 (0.311)
Patients in Nursing Homes	−0.210 (0.202)	0.666** (0.329)	
Persons in Need of Care	0.080 (0.306)		0.332 (0.423)
Hospital Beds	−0.451** (0.217)		
Physicians per Inhabitants	−0.189 (0.308)		0.856 (0.567)
Students in Tertiary Education	0.290 (0.329)		−0.492 (0.351)
Early School Leavers	0.437 (0.506)	0.376 (0.282)	0.405 (0.270)
Infants in Daycare (Share)	1.041** (0.411)		−0.835** (0.397)
Children in Daycare (Share)	0.544* (0.291)	0.458 (0.412)	
Distance to Airport	0.164 (0.235)	2.324*** (0.601)	1.616*** (0.625)
Constant	22.598*** (1.195)	33.782*** (1.723)	26.447*** (0.635)
R^2	0.13	0.16	0.14

Note: $N = 79800$. Significance levels: *p<0.1; **p<0.05; ***p<0.01. Linear regression results (post LASSO). Dependent variable: DTW distance of district pairs in wave 1, wave 2, wave 3. Each observation is a pair of two districts. Cluster Robust Standard errors in parentheses. DTW distances computed with COVID-19 7 day incidence rates in 400 districts, excluding Berlin. Geographic distance measured in 10,000 metres. Regional characteristics (Region Type - Distance to Airport) measure the absolute difference of these characteristics. Input variables are normalized, except for Geographic Distance and indicator variables (Same State, Region Type, High Housing Costs). Data sources: see text.

Probabilities for small/large DTW distances. In the following I look at the results of regression 2.11 that investigates how the network measures correspond to changes in the probability for high and low DTW distances. Table 2.5.5 shows the results. Personal travel flows are correlated with increased probabilities for small and below median DTW distances, commuter flows increase the probability for small distances. However, both measures do not play a significant role in separating large distances from the first three quartiles. An increase of one standard deviation (sd) regarding personal trips increases the probability for a small distance by 7.5% (6.7% for below median). An increase of commuter flows by one sd increases the probability for a small distance by 1.8%. For geographic distance the relationship is reverse: geographically very distant districts have more frequently large distances but no significant association exists between small geographic and small DTW distances, conditional on all other covariates. A large geographic distance is correlated with a higher probability for a large distance (1.7% per 10 km) and below median distance but is not significant for the probability of small distances. Trade flows correspond to lower distances in all three specifications 1.3-1.8% per sd (below 25th and 75th percentile significant only at the 10% level). Commuting to education is even corresponding to a smaller probability for small and below median distances, errands to a smaller probability for a distance below the 75th percentile. Belonging to the same state, conditional on all covariates, is corresponding to a reduced probability for small (4.4%) and below median (4.8%) distances. Table 2.D.2 in Appendix 2.D shows the regression results without a prior LASSO step, that overall show the same pattern.

When computing DTW distances within the three waves separately (compare Table 2.5.6),the results show important differences regarding the different network measures. A one sd increase in personal trips increases the probability for a small DTW distance by 7% in wave 1, 4% in wave 2 and 10% in wave 3. District-pairs in the same state had a 10% higher probability for a small DTW distance in wave 1, while the coefficient is small and insignificant in wave 2 and 3. The association with trade flows is only present in the third wave (3.4%), with commuters only in wave 2 (2.6%) and 3 (3.2%). This suggests as before that during the first wave, in which many businesses were closed, similar dynamic patterns were driven by private networks and state differences, while in the third wave both private and economic networks were important. For the second wave, which started after the holiday season, difference in the driving distance to international airports lower the probability for a small DTW distance. But in this specification the geographic distance is not significant in wave 2. This is in line with the in Table 2.5.5 that

suggest that geographic distance is not correlated with the probabilities for small distances. The second wave has also the largest coefficients regarding differences in urbanity and density variables and income structure.

Regression Forests. The descriptive analysis using the box plots could suggest underlying non-linear relationships, as well as the different patterns for small and large distances in Table 2.5.5. To account for this, I further assess variable importance by explicitly allowing for potential non-linearities: I use both the DTW distances and regional overall registered cases as outcome in a random forest (Breiman, 2001). Random forests have several advantages in the given situation. As an ensemble of trees, forests can very flexibly detect the relationship of variables. They can handle a large number of input variables, are relatively robust to including irrelevant input variables (Breiman, 2001) and require little parameter tuning (Hastie et al., 2017, p. 590). Importantly, random forests have a naturally integrated variable importance measure that can be used as a variable selection tool. A disadvantage is, that random forests are not easily interpretable in the sense of an underlying model or coefficients. Therefore, the analysis is complementary to the linear regression analysis which gives a quantitative interpretation but less model flexibility. The forests are computed with 5000 regression trees. A description of the estimation is given in Appendix 2.A.

Figure 2.5.3 plots for both forests the ten most important variables, according to the permutation importance. Personal trips are the most important variable regarding the DTW distance. Business trips, geographic distance and commuters are also among the ten most important variables. The most important difference in regional characteristics is the difference in the share of infants in daycare, followed by the differences in population density, relative female wage and distance to airport. The same covariates were included as in the regressions (see Table 2.C.1 for a full list). Overall similar covariates are prominent as in the LASSO regressions and the network measures play an important role. There are however also differences in the interpretation such as the high importance of business trips and students in tertiary education.

Regarding cumulative cases several variables have a high importance measure that correspond to the sectoral composition and the income structure. Further, density related variables play a role as well as the turnout in federal elections. The share of infants in daycare is the only variable that has in both specifications a high importance measure.

Table 2.5.5: Regression Results: DTW Distance below Percentiles

	Dependent variable:		
	Small and Large DTW Distances		
	<25	<50	<75
Commuters, Log	0.018** (0.009)	0.005 (0.010)	
Same State	−0.044** (0.019)	−0.048** (0.023)	−0.025 (0.018)
Trade Flow, Log	0.013* (0.007)	0.018** (0.009)	0.013* (0.007)
Commuting to Education, Log	−0.016** (0.008)	−0.025*** (0.009)	
Holiday Stays, Log	−0.0004 (0.006)	−0.004 (0.007)	
Errands, Log			−0.025*** (0.008)
Business Trips, Log			0.013 (0.023)
Personal Trips, Log	0.075*** (0.016)	0.067*** (0.022)	0.021 (0.026)
Geographic Distance	−0.007 (0.005)	−0.017** (0.007)	−0.017** (0.007)
Region Type	−0.048*** (0.012)	−0.020* (0.012)	
Population Density	0.044*** (0.008)	0.064*** (0.010)	0.053*** (0.009)
Average Living Space	−0.021*** (0.006)	−0.026*** (0.009)	−0.017* (0.009)
High Housing Costs	0.015* (0.009)	0.032*** (0.010)	0.023** (0.009)
Recreational Areas	−0.002 (0.008)	0.012 (0.012)	0.035*** (0.012)
Female Share			0.010 (0.008)
Broadband Availability	−0.016** (0.007)	−0.018* (0.009)	−0.017** (0.009)
Long Term Unemployment	0.024*** (0.007)	0.034*** (0.008)	0.019*** (0.007)
Relative Female Wage	−0.001 (0.009)	−0.013 (0.014)	−0.024 (0.016)
Turnout Federal Election	−0.014** (0.007)	−0.022** (0.011)	−0.015 (0.009)
Industry Share	−0.015* (0.008)	−0.022** (0.011)	−0.013 (0.011)
Share Services			−0.014 (0.014)
Share Individualized Services	−0.009 (0.006)	−0.012 (0.009)	−0.019** (0.008)
Median Income	−0.0004 (0.009)	0.007 (0.012)	
Share Primary Sector			−0.005 (0.007)
GDP per Capita	−0.011 (0.008)	−0.017 (0.012)	−0.010 (0.013)
Low Income Households (Share)	−0.069*** (0.017)	−0.106*** (0.023)	−0.098*** (0.024)
High Income Households (Share)	0.067*** (0.017)	0.104*** (0.021)	0.096*** (0.021)
Inhabitants 65 to 85	−0.023*** (0.008)	−0.028** (0.012)	−0.021* (0.012)
Inhabitants 85 and older	−0.005 (0.006)	−0.006 (0.009)	−0.004 (0.009)
Personell in Nursing Homes	−0.019*** (0.006)	−0.018** (0.009)	−0.006 (0.008)
Patients in Nursing Homes	−0.017** (0.007)	−0.011 (0.009)	−0.007 (0.009)
Persons in Need of Care	−0.012 (0.008)	−0.032*** (0.011)	−0.023** (0.011)
Hospital Beds			0.023** (0.011)
Physicians per Inhabitants	−0.018** (0.008)	−0.028** (0.012)	−0.030* (0.015)
Students in Tertiary Education	0.005 (0.009)	0.006 (0.012)	
Early School Leavers	−0.008 (0.006)	−0.015* (0.009)	−0.014* (0.009)
Infants in Daycare (Share)	−0.018* (0.010)	−0.033** (0.015)	−0.032** (0.014)
Children in Daycare (Share)	−0.007 (0.006)	−0.005 (0.009)	−0.004 (0.008)
Distance to Airport	−0.028*** (0.006)	−0.043*** (0.010)	−0.039*** (0.010)
Constant	0.302*** (0.024)	0.573*** (0.033)	0.812*** (0.032)
R^2	0.14	0.15	0.12

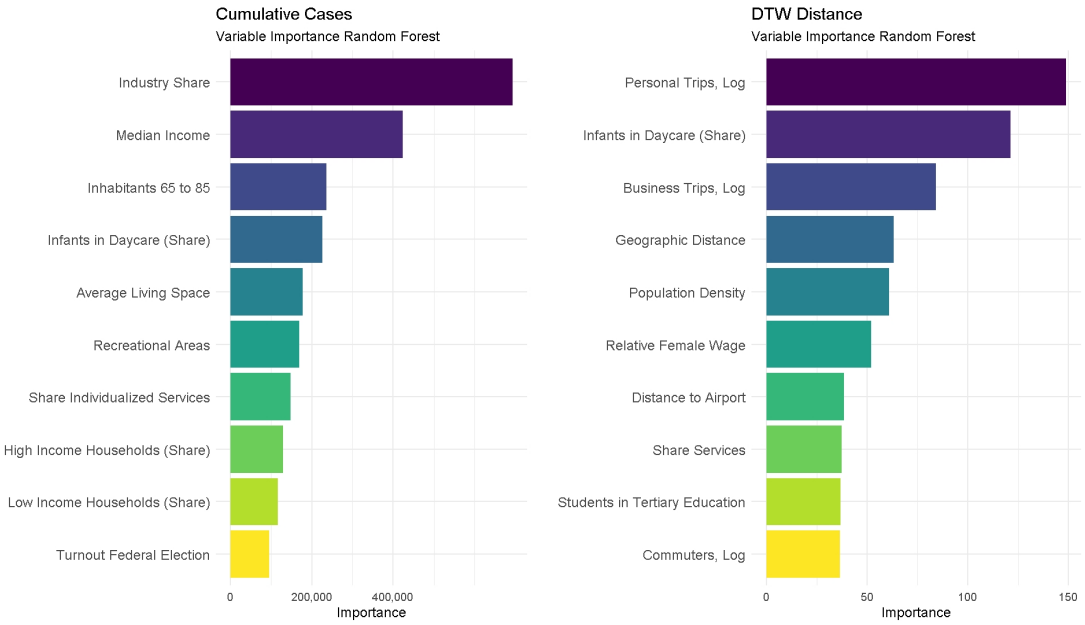
Note: $N = 79800$. Significance levels: *p<0.1; **p<0.05; ***p<0.01. Linear regression results (post LASSO). Dependent variable: DTW distance of the two districts below 25th, 50th and 75th percentile. COVID-19 7 day incidence rates in 400 German districts used as time series in DTW, excluding Berlin. Each observation is of a pair two districts. Cluster Robust Standard errors in parentheses. Geographic distance is measured in 10,000 metres. Regional characteristics (Region Type - Distance to Airport) measure the absolute difference of these characteristics for the district pairs. Input variables are normalized, except for Geographic Distance and indicator variables (Same State, Region Type, High Housing Costs). Data sources: see text.

Table 2.5.6: Small DTW Distances - 3 Waves Separately

	Dependent variable:		
	DTW Distance below 25th Percentile		
	Wave 1	Wave 2	Wave 3
Commuters, Log	0.014 (0.009)	0.026*** (0.009)	0.032*** (0.009)
Same State	0.098*** (0.026)	0.018 (0.022)	−0.015 (0.020)
Errands, Log	−0.033*** (0.008)		
Trade Flow, Log		−0.0004 (0.007)	0.034*** (0.007)
Commuting to Education, Log		−0.010 (0.009)	−0.046*** (0.008)
Holiday Stays, Log	0.034*** (0.007)	0.013** (0.006)	0.011* (0.006)
Geographic Distance	0.019*** (0.007)	−0.007 (0.006)	0.013** (0.006)
Personal Trips, Log	0.068*** (0.019)	0.040** (0.018)	0.100*** (0.017)
Region Type	0.011 (0.008)	−0.051*** (0.013)	−0.022** (0.011)
Population Density	−0.019* (0.011)	0.044*** (0.009)	0.046*** (0.009)
Average Living Space		−0.040*** (0.008)	−0.031*** (0.007)
High Housing Costs	0.013 (0.009)	0.003 (0.011)	0.013* (0.007)
Recreational Areas	−0.032*** (0.012)	−0.011 (0.011)	−0.014 (0.013)
Female Share	0.023*** (0.008)		
Broadband Availability	−0.018** (0.008)	−0.006 (0.008)	−0.006 (0.009)
Unemployment	−0.073*** (0.012)	−0.008 (0.010)	0.004 (0.009)
Long Term Unemployment	0.031*** (0.010)	0.044*** (0.008)	0.018** (0.008)
Relative Female Wage	0.010 (0.010)	0.012 (0.010)	−0.003 (0.009)
Turnout Federal Election	−0.004 (0.009)	−0.006 (0.008)	−0.007 (0.008)
Industry Share		−0.006 (0.009)	0.001 (0.007)
Share Services	0.007 (0.014)	0.009 (0.009)	−0.022** (0.009)
Median Income	−0.004 (0.012)		
Share Individualized Services		−0.005 (0.008)	−0.003 (0.006)
Share Primary Sector		−0.006 (0.009)	0.001 (0.007)
GDP per Capita	−0.009 (0.013)	−0.009 (0.009)	0.002 (0.008)
Low Income Households (Share)	0.011 (0.019)	−0.073*** (0.018)	−0.005 (0.017)
High Income Households (Share)	0.033* (0.019)	0.086*** (0.016)	0.029* (0.017)
Inhabitants 65 to 85	−0.018 (0.011)	−0.020** (0.009)	−0.004 (0.009)
Inhabitants 85 and older	0.015 (0.009)	0.018** (0.008)	0.008 (0.007)
Personell in Nursing Homes	−0.017** (0.008)	−0.028*** (0.007)	−0.021*** (0.007)
Patients in Nursing Homes		−0.013* (0.007)	−0.010 (0.006)
Persons in Need of Care	−0.005 (0.009)	−0.001 (0.008)	−0.001 (0.008)
Physicians per Inhabitants	0.011 (0.013)		
Students in Tertiary Education	−0.008 (0.011)		
Hospital Beds		−0.013* (0.008)	−0.006 (0.007)
Early School Leavers		−0.006 (0.006)	−0.007 (0.006)
Infants in Daycare (Share)	−0.021* (0.011)	−0.005 (0.011)	0.014 (0.010)
Children in Daycare (Share)	−0.008 (0.008)	−0.009 (0.007)	0.006 (0.006)
Distance to Airport		−0.025*** (0.007)	−0.013* (0.008)
Constant	0.148*** (0.030)	0.299*** (0.027)	0.201*** (0.025)
R^2	0.09	0.10	0.11

Note: $N = 79800$. Significane levels: *p<0.1; **p<0.05; ***p<0.01. Linear regression results (post LASSO). Dependent variable: DTW distance of the districts below 25th percentile - computed for wave 1,2,3 separately. COVID-19 7 day incidence rates in 400 German districts used as time series input, excluding Berlin. Each observation is of a pair two districts. Cluster Robust Standard errors in parentheses. Geographic distance is measured in 10,000 metres. Regional characteristics (Region Type - Distance to Airport) measure the absolute difference of these characteristics for the district pairs. All input variables are normalized, except for Geographic Distance and indicator variables (Same State, Region Type, High Housing Costs). Data sources: see text.

Figure 2.5.3: Variable Importance Plots



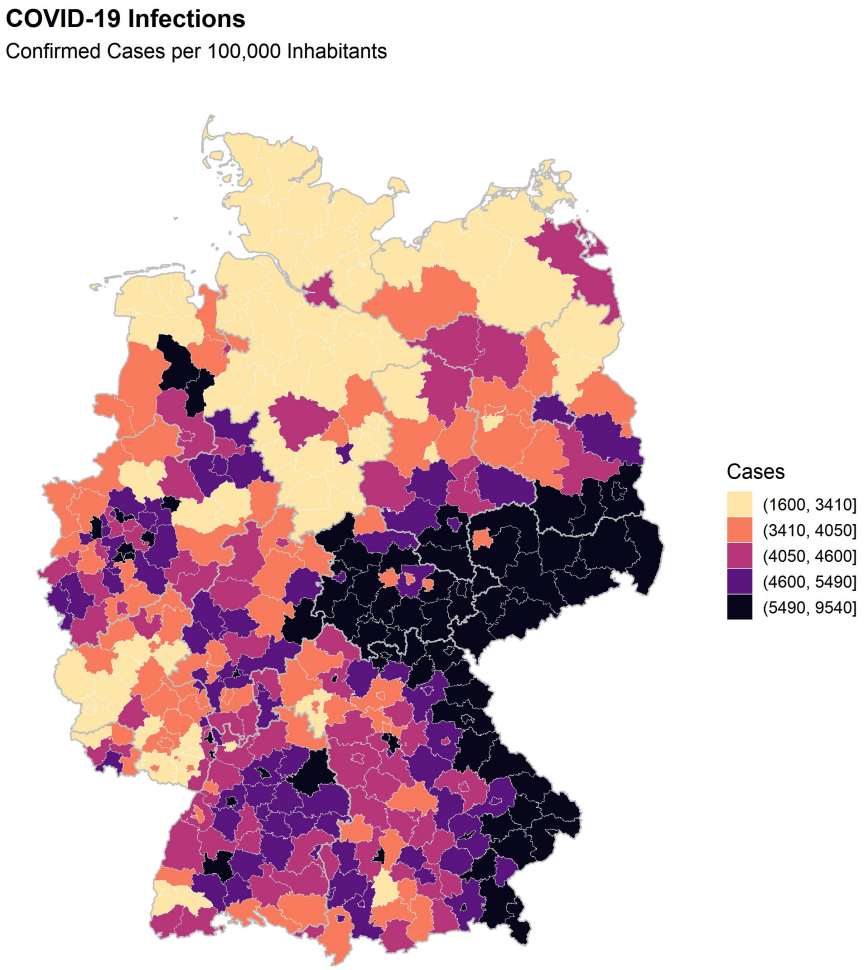
Note: Variable importance plots for random forests using linear regression trees. Variable Importance is measured as permutation importance, see for a description. The left figure corresponds to a forest with registered COVID-19 cases per 100,000 inhabitants, observations are 400 German districts, Berlin excluded. The right figure corresponds to a forest with the DTW distance as outcome variable, observations are 79800 district-pairs, district pairs with Berlin are excluded. Data sources: see text.

2.6 Geographic Pattern of Cumulative Cases

Geographic Pattern. As a comparison, I now look at the geographic distribution of cumulative cases by 100,000 inhabitants across German districts. A South-North difference is visible from Figure 2.6.1 as opposed to the East-West pattern present in the DTW map. The Northern districts have substantially lower cumulative cases per inhabitants. High numbers are present in different parts of Baden-Wurttemberg, Bavaria, Thuringia, Northrhine-Westphalia and Saxony, Saxony-Anhalt but also in parts of Lower-Saxony (North-West of Germany). There is no clear East-West pattern as opposed to the differences in the dynamic pattern illustrated in Figure 2.4.3. However, the districts with the highest intensities are mainly clustered in the Southern region of East Germany and in Bavaria. Figure 2.6.2 plots the distribution in urban and rural regions separately. The distribution in urban districts is more centred around medium values while the distribution in rural districts seems to be more evenly distributed pointing to a higher variance in rural districts.

Regional Characteristics. The two maps, Figure 2.6.1 and Figure 2.4.3, show similarities but also differences regarding the geographic structures of dynamic patterns and cumulative cases. Further, districts with small geographic distance can be very similar but also very distinct regarding the pandemic situation, which points towards the importance of regional characteristics. The previous section 2.5 points out which covariates are associated with smaller and larger dynamic distances and which network measures are associated with similar dynamic patterns. It is still not answered whether the same variables that correspond with similar dynamic patterns also correspond to higher overall infection rates. In the following, I analyze the relationship of pre-pandemic characteristics and cumulative cases in more detail. I choose the same set of 7 groups of characteristics as in section 2.5: Urbanity and Density, Employment and Inequality, Sectoral Structure, Income, Age and Health Care Provision, Education and Child Care and Geographical Connectedness. Table 2.C.1 reports all variables included in each of the groups. In addition to the analysis before, I also add the variables commuter inflow (share), commuter outflow (share) and hotel stays per year. These variables were left out in the analysis of DTW distances as they are closely related to flow variables in the analysis. While the regressions in section 2.5 included the absolute difference between two districts as independent variables I now include the realization of each district. Analogously to the previous regressions I select the most relevant

Figure 2.6.1: Cumulative Cases in German Districts: 5 Quantiles



Note: Map of COVID-19 cumulative cases per 100,000 inhabitants until June 30, 2021 across German districts. Thicker grey lines denote state borders. Data sources: RKI, Destatis, own computation, geographical data: BKG.

covariates with a Lasso estimator.[11] This estimation selects 19 out of 34 covariates. Table 2.6.1, column (1) presents the results of a post LASSO estimation.

All dependent variables are normalized such that the estimated coefficients correspond to an increase of one standard deviation of the non-normalized variable. Variables with the largest positive coefficients are population density, where a one sd increase corresponds to an increase by 625 infections, share of inhabitants 85 and older, commuter outflow and industry share. While large negative coefficients are estimated for median income, average living space per inhabitant, unemployment rates, broadband availability and recreational areas. This can suggest a high correlation with income and living factors as well as contact structures induced through work (out-commuting, unemployment and industry share). More densely populated areas with lower income levels experienced higher infection rates. As before, I note that the differences in variables do not reflect exogenous shocks and that the analysis is of descriptive nature. The LASSO selected variables explain nearly 60% of the regional variation of registered COVID-19 cases per 100,000 inhabitants.

I further integrate the group structure by estimating a group LASSO (Yuan and Lin, 2006) that selects which groups of variables have the strongest relationship with the outcome. Variable selection is on group level: All coefficients in each group are either all zero or all non-zero by combining a ridge (L2) penalty within groups and a LASSO (L1) penalty across groups. In the group LASSO specification the estimator deselects the variable group Education and Childcare. The share of infants in daycare is significantly negatively correlated with overall infection rates (standard post LASSO) or not selected as a relevant variables (group LASSO). This is in contrast to the positive association of differences in the share of infants in daycare with differences in dynamic patterns. This example illustrates that different variables are associated with the two different measures. Further, variables implying a similar dynamic pattern do not necessarily imply that the variable is correlated with higher cumulative infections.

To asses whether this is also the case for the inter-district network measures, that play an important role regarding the DTW distances, I aggregate these on district level: For commuters, trade flow and travel data I compute the sum of all flows to and from other districts.[12] Specifically, I compute the sum of the non-log

[11] For the model selection I choose the tuning parameter with 10-fold cross validation and select the sparsest model with a means squared error within one standard deviation of the minimum mean squared error (compare Hastie et al., 2017).

[12] I do not aggregate the measure scaled SCI. Given the definition of the variable, which is scaled by the product of Facebook users in both districts, an aggregation has no clear interpretation.

transformed variables and divide the sum by the number of inhabitants. The aggregated variables are measures of overall connectedness to all other German districts. I only include the flows to other German districts, this makes the estimations comparable to the analysis of dynamic patterns but abstracts from international connections.

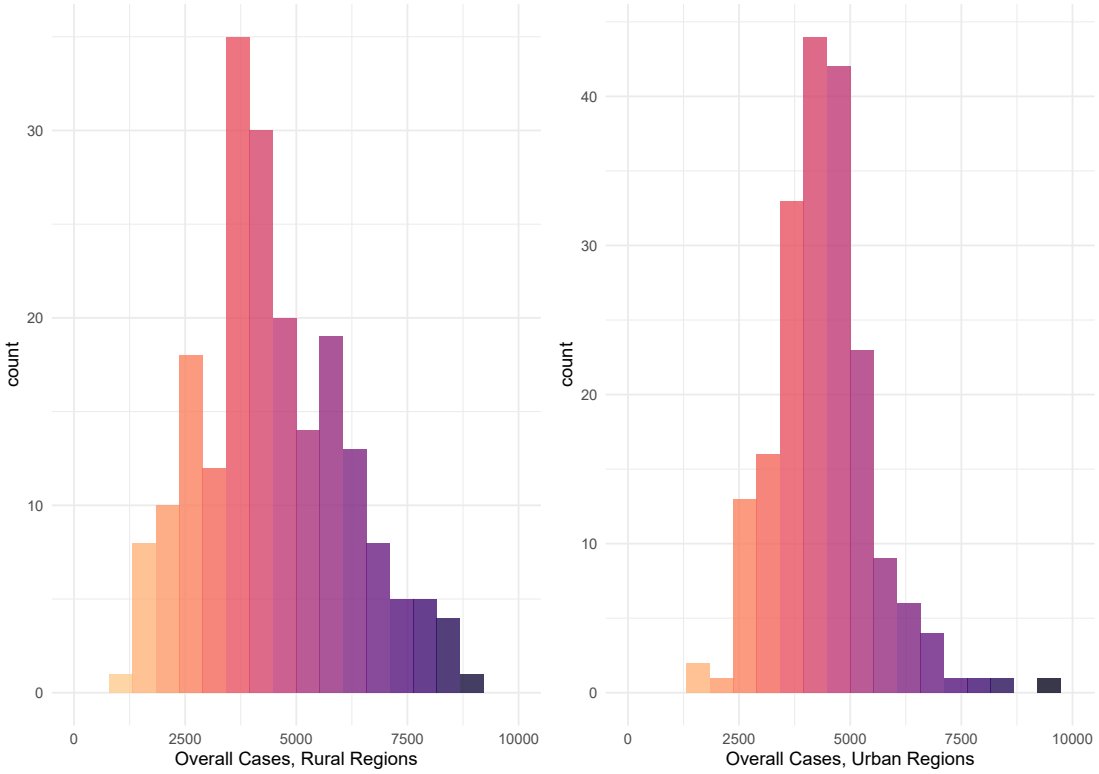
Table 2.6.1, column (2), shows the results. Out of the travel data only travel due to errands is selected but the coefficient is not significant in the post LASSO regression. In the case of highly correlated regressors the LASSO could be misleading as it may pick only one of the correlated variables. However, the variable selection patterns are identical when using an elastic net specification with parameter $\alpha = 0.8$. Further, computing the random forest variable importance with the aggregated flow variables included, those are not among the most relevant variables, compare Table 2.D.4. Although the results have to be interpreted with caution these could imply that the flow variables play a larger role regarding dynamic patterns than in the overall number of cases. The results of my estimations do not point towards higher overall infection rates of highly connected districts.

Table 2.6.1: Cumulative Cases, Regional Characteristics and Regional Network Measures

	Dependent variable:	
	Cases per 100,000 Inhabitants	
	(1)	(2)
Population Density	625.496*** (85.531)	597.080*** (85.111)
Average Living Space	−592.849*** (88.866)	−606.001*** (90.089)
Recreational Areas	−288.547*** (99.655)	−299.221*** (100.005)
Female Share	−112.665 (75.092)	−114.605 (74.717)
Broadband Availability	−306.755*** (92.882)	−332.943*** (93.864)
Unemployment	−344.209*** (107.389)	−348.992*** (106.815)
Turnout Federal Election	−153.346 (104.621)	−160.572 (105.890)
Industry Share	434.152*** (87.880)	446.542*** (89.957)
Employees Individual Related Services	−262.059*** (93.155)	−244.607** (94.643)
Share Primary Sector	−110.876 (75.566)	−112.048 (72.690)
Median Income	−792.836*** (103.350)	−812.937*** (104.077)
Inhabitants 85 and older	554.229*** (86.994)	523.225*** (90.581)
Personell in Nursing Homes	−148.192* (83.807)	−134.814 (85.760)
Patients in Nursing Homes	−101.315 (62.476)	−101.056 (62.018)
Hospital Beds	281.819*** (99.648)	259.142** (102.838)
Infants in Daycare (Share)	−247.884** (100.704)	−245.355** (99.945)
Incommuters	15.689 (128.737)	9.415 (127.554)
Outcommuters	487.350*** (166.212)	429.091** (176.976)
Hotel Stays/ Year	−89.508 (61.924)	−74.687 (62.972)
Errands		99.404 (71.461)
Constant	4,447.001*** (48.714)	4,447.001*** (48.625)
R^2	0.55	0.55

Note: $N = 400$. Significance levels: *p<0.1; **p<0.05; ***p<0.01. Results from post LASSO regressions. Dependent variable: Cumulative Covid 19 cases per 100,000 inhabitants until June 30 2021. Heteroscedasticity Robust Standard errors in parentheses. Covariates are normalized, except indicator variables (Region Type, High Housing Costs). Column (1) includes regional characteristics, column (2) additionally aggregated network measures. Data sources: see text.

Figure 2.6.2: Histogram of Cumulative Cases



Note: Histogram of COVID-19 cumulative cases per 100,000 inhabitants until June 30, 2021 across German districts. Plots are seperately for rural (left) and urban (right) districts. Data sources: see text.

2.7 Conclusion

This paper applies the method Dynamic Time Warping to regional COVID-19 incidence rates in Germany. The method first aligns the time series and afterwards computes the distances af the aligned series. Thereby it allows for differences in speed and timing. Combining the DTW measure with a hierarchical clustering algorithm I identify 5 clusters that mainly differ in the relative intensity of the three waves compared to each other. A geographic mapping of the clusters shows a clear East-West pattern. In Germany, the first wave was much less severe compared to the second and third wave. This difference was even more pronounced in East Germany. Further, in many cases several districts in the same clusters are geographical neighbouring districts. However, there are also many cases where neighbouring districts lie in different clusters. Geographic distance clearly seems to play a role but also leaves room for further underlying factors. In the second part of the analysis I look at the correlation of the DTW distance and different economic networks, such as commuter and trade flows, private and business travel but also political state borders. I find a large and influential role of private travel. Both when using a LASSO estimator as well as when applying random forests. A one standard deviation increase in (logarithmic) personal travel corresponds to an increase of 7.5% for a small DTW distance. The significance and relevance of the other network measures varies across analyses and waves but altogether the connectedness plays an important part. Higher trade and travel flows are associated with higher probabilities for small distances in the third wave, commuter flows in wave 2 and wave 3. On the other hand district-pairs that are both in the same state have a much higher probability for a small distance in wave 1. The patterns are not identical for all network measures. The role of business travel is ambiguous as it is in general not selected by the LASSO estimations but has a very high estimated variable importance in the random forests. Errands, commuting and belonging to the same state are in some specifications associated with larger DTW distances. The clear correlation of infectious spread with Facebook friendships (measure scaled SCI) that has been established by Kuchler et al. (2021) is not present in my analysis. The role of networks such as travel and transport was already pointed out by other literature such as Harris (2020) and Jo et al. (2021) for the cases of COVID-19 as well as Oster (2012) and Adda (2016) for other infectious diseases. The number of cumulative cases per 100,000 inhabitants is on the other hand not clearly correlated with the network variables. This results underlines that different factors can play a role regarding the timing and the overall intensity.

One reason for this could be a change of influence factors over time. This is in line with Doblhammer et al. (2022, 2021) who find that connectedness measures such as airports and borders play a role only in the beginning of a wave. Also Berkessel et al. (2021) find that the role of income regarding infections changes over time. One possible explanation for the non-present correlation of cumulative cases and network measures, which I cannot proof, could be the dominance of intra-district networks and contact structure regarding cumulative cases such that inter-district networks would mainly influence the timing and starting of waves. I note that I treat containment measures as given and do not study their role. Potentially, these can mitigate the effect that the connectedness variables have on overall cases. Regarding the underlying data used more recent data on travel and trade flows as well as network variables that take into account international connections could improve the analysis. A further improvement of my analysis step would be the possibility of a clear quantitative interpretation. A possibility could be to combine the analysis with a spatial panel model that builds on the insights of the DTW analysis. Potentially this could involve a prior determination of an adequate lag structure of the spatial model. Mastroeni et al. (2021) and Stübinger and Schneider (2020) use DTW to quantify the time distortion between time series. In the presence of exogenous shocks or adequate instrumental variables a spatial framework or a panel analysis could also imply a causal analysis. This causal perspective and clear quantification is a clear advantage of the papers by Oster (2012) and Adda (2016). With the analysis at hand I can on the other hand take a broader perspective. In a first step I can identify different patterns that would not become obvious in a panel specification. In a second step I can differentiate and compare different network measures. Compared to spatial panel models however this analysis does not require pre-specified weighting matrices, based on geographic distance or other connectivity measures such as commuter flows (compare Adda, 2016) nor a pre-specified time lag structure.

Appendix

2.A Estimators

The **Random Forest** (see Breiman, 2001, for a discussion) is computed as follows: For each forest I compute 5000 linear regression trees where each tree is computed on a bootstrap sample using N observations, where N is the sample size. The minimum node size is 5. At each split a random subset of all covariates is considered. This reduces the correlation between the trees. The recommended value for regression trees is $p/3$, where p denotes the number of all covariates. Prediction and prediction error in a random forest are computed with out-of-bag (OOB) prediction: the outcome prediction of a certain observation is computed by using only the trees with the bootstrap samples that do not contain the specific observation. This is similar to N-fold cross-validation (Hastie et al., 2017)[p. 593]. The permutation importance measure estimates how much the OOB error changes when a specific variable is added to the model. Therefore the values for this specific variable are permuted among observations and for each tree the OOB prediction error with and without permutation are compared and the difference is the variable importance measure. I use the implementation by Wright and Ziegler (2017) which computes the original random forest by Breiman (2001).

The **LASSO** estimator $\hat{\boldsymbol{\beta}}^{Lasso}$ Tibshirani (1996) is characterized by the following optimization problem:

$$\min_{(\beta_0,\boldsymbol{\beta})\in\mathbb{R}^{p+1}} \left\{ \frac{1}{2N}\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j| \right\}$$

Where y_i denotes the outcome y of observation i and x_{ij} the value of regressor j for observation i, $\boldsymbol{\beta}$ is the coefficient vector with the j-th component β_j corresponding to the j-th regressor, β_0 denotes the intercept. p is the number of regressors and N the number of observations.

The **elastic net** estimator $\hat{\boldsymbol{\beta}}^{Elnet}$ (Zou and Hastie, 2005) is a combination of

LASSO and Ridge estimators, that contains a convex combination of a L1 penalty (LASSO) and a L2 penalty (Ridge). The estimator is determined by the following optimization problem:

$$\min_{(\beta_0,\boldsymbol{\beta})\in\mathbb{R}^{p+1}} \left\{ \frac{1}{2N}\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda(\alpha\sum_{j=1}^{p}|\beta_j| + (1-\alpha)\frac{1}{2}\sum_{j=1}^{p}\beta_j^2) \right\}$$

Where y_i denotes the outcome y of observation i and x_{ij} the value of regressor j for observation i, $\boldsymbol{\beta}$ is the coefficient vector with the j-th component β_j corresponding to the j-th regressor, β_0 denotes the intercept. p is the number of regressors and N the number of observations and $0 \leq \alpha \leq 1$. For $\alpha = 1$ the elastic net estimator equals the LASSO estimator, for $\alpha = 0$ the Ridge estimator. I use the implementation by Friedman et al. (2010). When not stated otherwise I choose the regularization parameter λ by using 10 fold cross validation and use the sparsest model within one standard deviation of the minimum mean squared error (compare Hastie et al., 2017).

Further, I compute the **Group LASSO** estimator (Yuan and Lin, 2006) with the implementation in the R package gglasso (Yang et al., 2020). The following definition is from Hastie et al. (2017). Let $\boldsymbol{Y}$ denote the outcome vector for all observations $i = 1, \ldots, N$, with regressor matrix $\boldsymbol{X} \in \mathbb{R}^{N*p}$. The p regressors are organized into L different groups such that group l contains p_l regressors. Let $\boldsymbol{X}_l$ be the submatrix with N rows and p_l columns that contains the regressors belonging to group l and $\boldsymbol{\beta}_l \in \mathbb{R}^{p_l}$ the corresponding subvector of $\boldsymbol{\beta}$, β_0 the intercept, then the group LASSO is determined by the following minimization:

$$\min_{(\beta_0,\boldsymbol{\beta})\in\mathbb{R}^{p+1}} \left\{ \|\boldsymbol{Y} - \beta_0\mathbf{1} - \sum_{l=1}^{L}\boldsymbol{X}_l\boldsymbol{\beta}_l\|_2^2 + \lambda\sum_{l=1}^{L}\sqrt{p_l}\|\beta_l\|_2 \right\},$$

where $\mathbf{1}$ is an appropriate vector of ones.

2.B Computation

Computations and figures are realized using R version 4.1.2 (R Core Team, 2021). Further, I use the following R packages and resources: tidyverse (Wickham et al., 2019), readxl (Wickham and Bryan, 2019), sf (Pebesma, 2018), dtw (Giorgino, 2009), cluster (Maechler et al., 2021), glmnet (Friedman et al., 2010), gglasso (Yang et al., 2020), ranger (Wright and Ziegler, 2017), lmtest (Zeileis and Hothorn, 2002),

sandwich (Zeileis, 2004; Zeileis et al., 2020), cowplot (Wilke, 2020), hrbrthemes (Rudis, 2020), stargazer (Hlavac, 2018), dendextend (Galili, 2015).

2.C Data Appendix

RKI Data. COVID-19 data is reported by RKI and was made available for download on a daily basis by Geo Esri. The dataset contains all confirmed cases as a time series and additionally corrections of the file issued one day in advance. I have downloaded the dataset on September 17, 2021. The corrections are reported by the variable "NeuerFall" ("new case"). I discard all data rows coded by "NeuerFall=-1" as these do only denote corrections compared to the earlier file. I confirm that this data procession leads to the same cumulative cases in each of the 16 German states as reported by the RKI.

Log Transformation. I log-transform the network measures between districts when indicated in the main text and tables. Before log-transformation I change all values of zero to ones. Values between 0 and 1 are not reported in the data.

Regional Characteristics: Definitions. Table 2.C.1 reports all regional characteristics that are used in the analyses. All variables are downloaded from the INKAR database (BBSR Bonn, 2021), as at 17th and 19th of August 2021. Data is collected by the Federal Institute for Building, Urban Affairs and Spatial Research (BBSR) from different original sources such as the federal statistical office and the institute for employment research (IAB). In the following I provide additional information on the definition of variables, the database website provides definitions for all variables as well as information on the original data sources.

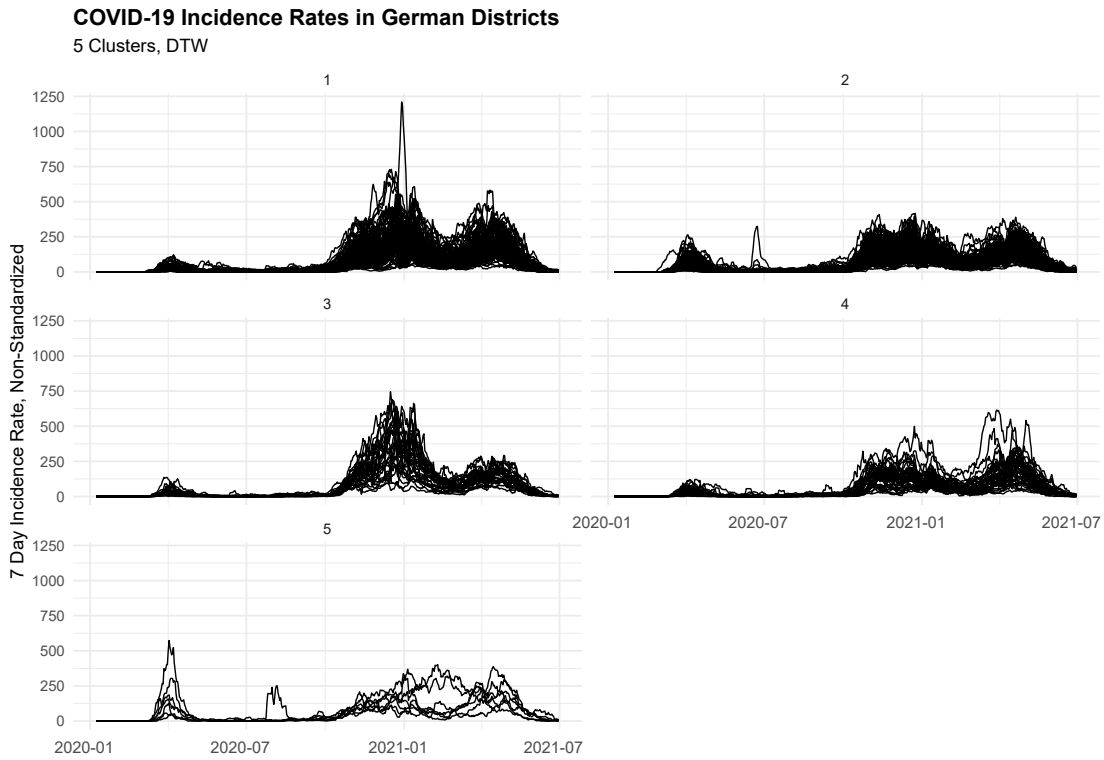
Table 2.C.1: List of Regional Variables

Group Name	Variable	Year	Definition
Urbanity and Density	Urban/ Rural Region	2017	Indicator variable coded as 1 for urban region.
	Population Density	2017	Number of inhabitants per square kilometre.
	Average Living Space	2017	Average square metres per person.
	High Housing Costs (> Median)	2017	Own computation: Indicator variable coded as 1 for average housing costs per square metre above the median over all districts.
	Recreational Areas	2017	Measured in square kilometre per 10,000 inhabitants.
	Female Share	2017	Percentag share of female inhabitants among all inhabitants.
	Broadband Availability	2017	percentage of households with access to at least 50 Mbit/s.
Employment and Inequality	Unemployment Rate	2017	Number of persons unemployed divided by sum of persons unemployed, employed, self-employed and civil servants (excluding soldiers).
	Long Term Unemployment Rate	2017	Percentage share of all persons unemployed that are unemployed for at least 1 year.
	Relative Female Wage	2017	Median gross monthly wage of female employees (full-time) in percentage of the median gross monthly wage of male employees (full-time).
	Turnout Federal Election	2017	Number of second votes (valid and invalid) in percentage of all inhabitants entitled to vote.
Sectoral Structure	Industry Share	2017	Number of employees in the industry sector, within the mandatory social security scheme, per 100 inhabitants between 15 and 65.
	Share Services	2017	Number of employees in the service sector, within the mandatory social security scheme, per 100 inhabitants between 15 and 65.
	Share Individualized Services	2017	Percentage share of employees within the mandatory social security scheme that are employed in individualized services. Individualized services are services where the customer plays an integral part during the service, such as teaching.
	Share Primary Sector	2016	Employees in the primary sector per 100 employees.
Income	Median Income	2017	Median gross monthly income (in Euros) of employees (full-time) within the mandatory social security scheme.
	GDP per Capita	2017	Measured in 1,000 Euros per inhabitant.
	Low Income Households (Share)	2016	Percentage of households with a household income below 1,500 Euros.
	High Income Households (Share)	2016	Percentage of households with a household income above 3,600 Euros. Share of medium income households is left out as a reference level for share of low income households and share of high income households.
Age and Elderly Care	Inhabitants, Age 65-85	2017	Own computation: sum of the original variables inhabitants 65-75 and inhabitants 75-85. Persons in need of care are measured as the total number times 100 divided by number of inhabitants.
	Inhabitants, 85 and older	2017	Percentage share among all inhabitants.
	Personnel in Nursing Homes	2017	Number of employees in nursing homes times 100 divided by number of patients in nursing homes.
	Patients in Nursing Homes	2017	Number of patients in nursing homes in percentage of all persons in need of care.
	Persons in Need of Care	2017	Number of persons in need of care times 100 divided by number of inhabitants.
	Physicians per Inhabitants	2017	Registered doctors (excluding psychotherapists) per 10,000 inhabitants.
	Hospital Beds	2016	Hospital beds per 1,000 inhabitants.
Education and Child Care	Students in Tertiary Education	2017	Students at universities and universities of applied sciences per 1,000 inhabitants.
	Early School Leavers	2017	Persons that leave school without a qualification in percentage of all persons leaving school (with and without qualification).
	Infants in Daycare (Share)	2017	Percentage share of children under the age of 3 in daycare.
	Children in Daycare (Share)	2017	Percentage share of children between 3 and 6 in daycare.
Geographical Connectedness	Commuter Inflow	2017	Persons commuting into the district as percentage share of all employees in the mandatory social security scheme.
	Commuter Outflow	2017	Persons commuting out of the district as percentage share of all employees in the mandatory social security scheme.
	Distance to Airport	2017	Driving distance in minutes from the district centre to the next international airport.
	Hotel Stays/Year	2017	Guest-nights in touristic accommodation with at least 10 available beds.

Note: List of regional characteristics from the INKAR database used in the analyses. In all estimations with the DTW distance as outcome or indicators based on the DTW distance I leave out the regional characteristics "Hotel Stays/ Year", "Commuter Inflow", "Commuter Outflow" as these are very similar to included network characteristics.

2.D Further Results

Figure 2.D.1: Clustered Time Series, Non-Standardized



Note: Time Series of non-standardized COVID-19 7 day incidence rates in 400 German districts and 12 Berlin regions. Observations are assigned into 5 clusters. The clustering is computed with standardized COVID-19 7 day incidence rates, hierarchical clustering based on DTW distance. Data sources: RKI, Destatis, Federal Statistical Office Berlin Brandenburg, own computations.

Table 2.D.1: DTW Distance - Different Covariate Sets

	Dependent variable:		
	DTW Distance		
	(1)	(2)	(3)
Commuters, Log	−1.153* (0.683)	−0.849 (0.651)	−0.385 (0.584)
Same State	1.520 (1.363)	2.889** (1.344)	2.215* (1.255)
Trade Flow, Log	−1.748*** (0.574)	−1.213** (0.549)	−1.016** (0.500)
Commuting to Education, Log	0.996 (3.001)	1.920 (2.619)	1.301 (2.466)
Errands, Log	1.149 (2.883)	0.229 (2.564)	0.751 (2.404)
Business Trips, Log	−0.106 (1.713)	1.025 (1.861)	−1.058 (1.541)
Holiday Stays, Log	0.096 (0.508)	0.257 (0.480)	0.325 (0.460)
Personal Trips, Log	−6.269*** (2.193)	−6.781*** (2.164)	−3.594* (1.835)
Geographic Distance	0.871* (0.516)	0.704 (0.515)	0.967** (0.481)
Region Type			1.444** (0.715)
Population Density		−2.902*** (0.705)	−3.565*** (0.683)
Average Living Space		1.543** (0.635)	1.301** (0.560)
High Housing Costs			−1.319** (0.645)
Recreational Areas		0.933 (0.725)	−1.880** (0.786)
Female Share			−0.594 (0.524)
Broadband Availability			1.330** (0.558)
Unemployment		−0.734 (0.632)	−0.411 (0.627)
Long Term Unemployment			−1.511*** (0.537)
Relative Female Wage			1.821 (1.136)
Turnout Federal Election		1.218* (0.678)	1.366** (0.654)
Industry Share		1.442** (0.626)	1.067 (0.665)
Share Services			0.822 (0.894)
Share Individualized Services			1.158** (0.540)
Share Primary Sector			0.082 (0.530)
Median Income		2.529*** (0.777)	−0.204 (0.768)
GDP per Capita			0.549 (0.908)
Low Income Households (Share)		1.431** (0.686)	6.630*** (1.614)
High Income Households (Share)			−6.591*** (1.447)
Inhabitants 85 and older		1.902*** (0.533)	0.357 (0.532)
Inhabitants 65 to 85			1.689** (0.775)
Personnel in Nursing Homes			0.820 (0.538)
Patients in Nursing Homes		1.240** (0.589)	0.937* (0.563)
Persons in Need of Care			1.326* (0.738)
Hospital Beds			−1.347* (0.735)
Physicians per Inhabitants			2.141** (0.974)
Students in Tertiary Education			0.293 (0.786)
Early School Leavers			0.955* (0.549)
Infants in Daycare (Share)			2.264** (0.974)
Children in Daycare (Share)		0.425 (0.547)	0.234 (0.529)
Distance to Airport		2.295*** (0.634)	2.451*** (0.603)
Constant	79.981*** (2.296)	80.521*** (2.331)	79.407*** (2.152)
R^2	0.11	0.12	0.22

Note: $N = 79800$. Results from linear regressions. Each observation is a pair of two districts. Dependent variables: DTW distance of the two districts. Cluster Robust Standard errors in parentheses. Significance levels: *p<0.1; **p<0.05; ***p<0.01. Geographic distance measured in 10,000 metres. Dependent variables are normalized, except for Geographic Distance and indicator variables (Same State, Region Type, High Housing Costs). Data sources: see text.

Table 2.D.2: DTW Distance: Quartiles - Full Covariate Set

	Dependent variable:		
	DTW Distance below percentiles		
	<25	<50	<75
Commuters, Log	0.019** (0.009)	0.006 (0.010)	−0.005 (0.008)
Same State	−0.042** (0.019)	−0.049** (0.023)	−0.026 (0.018)
Trade Flow, Log	0.013* (0.007)	0.017* (0.009)	0.014* (0.007)
Commuting to Education, Log	−0.075 (0.046)	−0.015 (0.045)	0.019 (0.032)
Errands, Log	0.057 (0.045)	−0.011 (0.043)	−0.047 (0.031)
Business Trips, Log	0.002 (0.019)	0.006 (0.025)	0.015 (0.023)
Holiday Stays, Log	−0.001 (0.006)	−0.003 (0.008)	−0.006 (0.007)
Geographic Distance	−0.007 (0.005)	−0.017** (0.007)	−0.017** (0.007)
Personal Trips, Log	0.075*** (0.022)	0.063** (0.030)	0.026 (0.028)
Region Type	−0.047*** (0.012)	−0.021* (0.012)	0.001 (0.010)
Population Density	0.040*** (0.009)	0.062*** (0.011)	0.053*** (0.010)
Average Living Space	−0.020*** (0.006)	−0.025*** (0.009)	−0.017* (0.009)
High Housing Costs	0.016* (0.009)	0.034*** (0.010)	0.023** (0.009)
Recreational Areas	−0.002 (0.008)	0.012 (0.012)	0.035*** (0.012)
Female Share	0.003 (0.006)	0.012 (0.008)	0.010 (0.008)
Broadband Availability	−0.016** (0.007)	−0.018* (0.009)	−0.018** (0.009)
Unemployment	0.009 (0.008)	0.010 (0.011)	0.003 (0.009)
Long Term Unemployment	0.021*** (0.007)	0.029*** (0.009)	0.018** (0.008)
Relative Female Wage	−0.003 (0.009)	−0.015 (0.014)	−0.025 (0.016)
Turnout Federal Election	−0.017** (0.007)	−0.025** (0.011)	−0.017* (0.010)
Industry Share	−0.016** (0.008)	−0.025** (0.011)	−0.014 (0.010)
Share Services	−0.002 (0.010)	−0.014 (0.014)	−0.011 (0.014)
Share Individualized Services	−0.010 (0.006)	−0.014 (0.009)	−0.018** (0.008)
Share Primary Sector	0.003 (0.008)	0.003 (0.009)	−0.005 (0.007)
Median Income	0.00003 (0.009)	0.007 (0.013)	0.002 (0.012)
GDP per Capita	−0.009 (0.009)	−0.010 (0.014)	−0.010 (0.014)
Low Income Households (Share)	−0.073*** (0.017)	−0.111*** (0.023)	−0.099*** (0.024)
High Income Households (Share)	0.070*** (0.017)	0.107*** (0.022)	0.097*** (0.021)
Inhabitants 65 to 85	−0.023*** (0.008)	−0.029** (0.012)	−0.020 (0.013)
Inhabitants 85 and older	−0.006 (0.006)	−0.007 (0.009)	−0.005 (0.009)
Personell in Nursing Homes	−0.019*** (0.006)	−0.018** (0.009)	−0.007 (0.008)
Patients in Nursing Homes	−0.017** (0.007)	−0.012 (0.009)	−0.008 (0.009)
Persons in Need of Care	−0.012 (0.008)	−0.032*** (0.011)	−0.023** (0.011)
Hospital Beds	0.003 (0.009)	0.011 (0.014)	0.023** (0.011)
Physicians per Inhabitants	−0.019* (0.010)	−0.030** (0.015)	−0.030** (0.015)
Students in Tertiary Education	0.004 (0.009)	0.004 (0.013)	−0.005 (0.011)
Early School Leavers	−0.008 (0.006)	−0.014* (0.009)	−0.015* (0.009)
Infants in Daycare (Share)	−0.016 (0.010)	−0.030** (0.015)	−0.032** (0.015)
Children in Daycare (Share)	−0.008 (0.006)	−0.006 (0.009)	−0.004 (0.008)
Distance to Airport	−0.028*** (0.006)	−0.043*** (0.010)	−0.038*** (0.010)
Constant	0.301*** (0.024)	0.574*** (0.033)	0.813*** (0.033)
R^2	0.14	0.16	0.12

Note: *p<0.1; **p<0.05; ***p<0.01. $N = 79800$. Results from linear regressions. Each observation is of a pair of two German districts, excluding Berlin. Dependent variables: DTW distance of the two districts below 25th, 50th, 75th percentile. Cluster Robust Standard errors in parentheses. Geographic distance measured in 10,000 metres. Covariates are normalized, except for Geographic Distance and indicator variables (Same State, Region Type, High Housing Costs). Data sources: see text.

Table 2.D.3: Cumulative Cases: Variable Selection with Group Lasso

	Dependent variable:
	Cases per 100,000 Inhabitants
Region Type	128.379 (151.476)
Population Density	604.236*** (93.938)
Average Living Space	−455.858*** (86.911)
High Housing Costs	50.495 (160.318)
Recreational Areas	−386.163*** (94.787)
Female Share	−158.916* (88.034)
Broadband Availability	−289.102*** (99.084)
Unemployment	−283.969** (142.585)
Long Term Unemployment	16.616 (97.858)
Relative Female Wage	−46.757 (115.033)
Turnout Federal Election	−238.192** (112.383)
Industry Share	468.916*** (110.944)
Share Services	98.442 (203.181)
Share Individualized Services	−301.258*** (109.119)
Share Primary Sector	−82.859 (76.288)
Median Income	−828.271*** (143.375)
GDP per Capita	44.214 (170.108)
Low Income Households (Share)	53.750 (309.636)
High Income Households (Share)	36.324 (272.320)
Inhabitants 65 to 85	−129.181 (135.161)
Inhabitants 85 and older	611.353*** (113.152)
Personell in Nursing Homes	−45.285 (125.875)
Patients in Nursing Homes	−162.593 (111.899)
Persons in Need of Care	−147.465 (155.985)
Hospital Beds	256.241** (104.304)
Physicians per Inhabitants	137.186 (147.072)
Commuter Inflow	−194.665 (169.138)
Commuter Outflow	868.671*** (255.696)
Distance to Airport	−66.724 (75.283)
Hotel Stays/ Year	−34.317 (81.279)
Constant	4,363.576*** (109.335)
R^2	0.55

Note: $N = 400$. Results from OLS regression after variable selection with group LASSO. Dependent variable: Cumulative COVID-19 cases per 100,000 inhabitants until June 30 2021. Heteroscedasticity Robust Standard errors in parentheses. Significance levels: *p<0.1; **p<0.05; ***p<0.01. Covariates are normalized, except indicator variables (Region Type, High Housing Costs). Data sources: see text.

Table 2.D.4: Cumulative Cases: Variable Importance with Aggregated Network Measures

	Variable Importance
Industry Share	738100.071
Median Income	406693.705
Inhabitants 65 to 85	212536.666
Infants in Daycare (Share)	180195.686
Average Living Space	165683.965
Recreational Areas	149046.514
Share Individualized Services	143175.855
High Income Households (Share)	116044.350
Low Income Households (Share)	108151.893
Turnout Federal Election	101225.809
Relative Female Wage	64216.371
Distance to Airport	62058.186
Errands	54534.074
Inhabitants 85 and older	51042.696
Unemployment	47127.340
Population Density	44571.099
Commuter Inflow	43894.529
Share Primary Sector	40126.800
Personal Trips	33917.288
Commuter Outflow	31953.607
Children in Daycare (Share)	28701.912
Persons in Need of Care	28439.306
Business Trips	28263.853
Long Term Unemployment	26496.146
Hospital Beds	23530.932
GDP per Capita	21111.125
Broadband Availability	20134.695
Patients in Nursing Homes	19155.462
Personell in Nursing Homes	16235.508
Holiday Stays	13389.007
Hotel Stays/ Year	12672.317
Early School Leavers	12453.750
Commuting to Education	10724.944
Physicians per Inhabitants	9795.936
Share Services	8646.013
Female Share	7872.130
Students in Tertiary Education	6173.045
Commuters	5312.356
Region Type	4792.391
Trade Flow	4298.839
High Housing Costs	1694.083

Note: Variable Importance Measures from Random Forest. Dependent Variable is cumulative cases per 100,000 Inhabitants. Data sources: see text.

Chapter 3

Do Minimum Wages Encourage Capital Deepening?

Joint with Christina Gathmann and Terry Gregory.

This chapter is based on but substantially exceeds my dissertation proposal and master thesis (Zapp, 2017).

3.1 Introduction

How minimum wages affect workers and firms has been one of the most hotly debated questions in labor economics. A long line of research has explored the link between minimum wage hikes, wages and employment (see Kennan, 1995; Flinn, 2011; Neumark and Wascher, 2008; Schmitt, 2015; Manning, 2021, for comprehensive surveys). While minimum wages increase the wage of affected workers, the employment effects turn out to be modest or close to zero (see e.g. Card and Krueger, 1994; Dube et al., 2010; Neumark et al., 2014; Meer and West, 2016; Dube and Reich, 2016; Cengiz et al., 2019).

The absence of sizable employment effects has pushed the research to consider other margins of adjustments (see Clemens, 2021; Manning, 2021, for recent surveys).[1] The type of adjustments firms undertake in response to higher labor costs for low-skilled workers induced by minimum wages is likely to vary across industries and the underlying production technology. If low- and high-skilled workers are good substitutes, firms could respond to higher Wages for low-wage

[1]They discuss various potential adjustment margins including changes to the non-wage benefit package and job amenities; but also costs borne by the firm through reduced profits (e.g. Draca et al., 2011) and firm value (Bell and Machin, 2018) or more firms exiting the industry (Luca and Luca, 2019)

workers through labor-labor substitution: firms substitute from low productivity workers to more productive workers (Gregory and Zierahn, 2022), invest more in the human capital of their workforce (e.g. Neumark and Wascher, 2001; Sutch, 2011) or reduce turnover, which brings down hiring costs and increases the value of a match to the firm (e.g. Brochu and Green, 2013; Dube and Reich, 2016; Portugal and Cardoso, 2006).

Alternatively, firms might adjust to higher labor costs for low-skilled workers by substituting labor with capital and adjusting their production technology. Globalization and the digital revolution open up many new opportunities to automate certain jobs and replace them by machines or robots (Brynjolfsson and McAfee, 2014). While the scale for automation is still debated, it is undisputed that many jobs held by low-skilled workers, esp. those involving routine tasks that can be easily codified and performed by machines, are likely to disappear in the near future (Autor, 2015; Arntz et al., 2017). High minimum wages could speed up these technological adjustments. As minimum wages raise the labor costs for firms employing low-skilled workers, employers have an incentive to automate, invest in labor-saving technology like machines or software and replace less-skilled workers. Such capital-labor substitution is more likely for jobs with a high share of routine tasks, which can be more easily performed by a machine, for instance (Aaronson and Phelan, 2019; Lordan and Neumark, 2018). To what extent minimum wages encourage capital-labor substitution is an open question empirically: some find evidence for capital deepening (see Harasztosi and Lindner, 2019; Dai and Qiu, 2022), others report a decline in capital investments (Gustafson and Kotter, 2022).[2] Further, such changes in labor demand do not only affect incumbent firms but also have consequences for firm dynamics and the characteristics of firms entering a market subject to a minimum wage (e.g. Aaronson et al., 2018).

Firms might also react to higher labor costs by outsourcing some tasks that are performed by less-skilled workers to firms that are not covered by a minimum wage. Overall employment effects could then be small though the activity has been shifted away from the industry or location subject to the minimum wage regulation.[3]

Finally, firms might also pass some of the additional costs on to consumers through raising prices for their products and services (e.g. Aaronson, 2001; Aaronson

[2]For China, there is some evidence that its minimum wage policy has encouraged capital-labor substitution (Hau et al., 2020; Mayneris et al., 2018).

[3]Offshoring in order to reduce the labor cost of production, in contrast, would not fall in this category. With offshoring, the activities are shifted abroad and we would observe negative employment effects in the affected industries or locations.

and French, 2007; Aaronson et al., 2008; Harasztosi and Lindner, 2019). A firm's ability to adjust prices depends on the elasticity of demand for its goods and/or services. This depends, in turn, on the scope of the market. Firms that produce widely traded goods or services may face large demand elasticities and thus have little capacity to raise prices. By contrast, firms that produce "nontradable" goods and services may face smaller demand elasticities and have more substantial scope for passing cost increases to consumers.

In this paper, we provide novel evidence how firms operating in industries with very different production technologies adjust to the adoption of minimum wages in their industry. The variation we use comes from industry-specific minimum wages that were introduced in Germany between 1997 and 2014. Our analysis has a number of unique features. First, we track how firms respond differently to minimum wages depending on the importance of physical capital, the importance of routine tasks and thus the possibilities for capital-labor substitution. As Germany is one of the world leaders in robot use in manufacturing, we might expect that capital deepening and investments in automation might play a more important role than in other countries with lower labor costs.

Second, we analyze not only the adjustment behavior among incumbent firms but also shed light on the characteristics of firms entering the industries with a minimum wage. Depending on the the production technology, capital deepening might occur mostly within incumbent firms or, in case of a putty-clay technology, mostly through firms entering an industry. Furthermore, unlike most of the literature, we analyze the first-time adoption of a minimum wage rather than incremental changes to an existing minimum wage. Firm-level responses to an incremental change are likely to differ from responses to the adoption of a minimum wage. Under rational expectations, firms should have priced in future incremental changes to an existing minimum wage in their long-run factor demands. Firms that get exposed to a minimum wage for the first time, in contrast, might change firms' cost minimization in response to a relative price increase for low-skilled labor.

Finally, the minimum wages adopted at the industry level are high. The Kaitz indices vary between 50% and 90% of the median wage – and are thus substantially higher than most minimum wages analyzed in the literature. In East Germany, around one in three male full-time workers earning below the 15th percentile would be affected by the minimum wage. In West Germany, the share is just one out of five male workers earning below the 15th percentile (Brüll and Gathmann, 2020). As minimum wages bite more deeply into the East German wage distribution, we focus our analysis on firms operating in East Germany. The existing literature has

focused on the introduction of the national minimum wage in Germany in 2015, which is substantially lower than the industry-specific wages we analyze.[4]

For the empirical analysis, we use rich balance sheet data of firms before and after the adoption of minimum wages from the Dafne database provided by Bureau van Dijk (BvD). Crucial for our purpose is that we observe a detailed industry indicator at the 5-digit level (equivalent to NACE rev. 2) in order to identify the firms in industries that adopted a minimum wage. We also observe the exact geographic location of each firm, which we need to assign the relevant, region-specific minimum wage to each firm. Furthermore, the data is a rare case where we have detailed financial information on the capital stock – like the capital assets, technical equipment and machinery, factory and office equipment – as well as several proxies for outsourcing like the amount of prefabrication.

We first analyze the responses of incumbent firms. Because firms in industries adopting a minimum wage might differ from the average firm, we use a matching approach to find suitable controls in closely related industries.[5] The matching performs well in eliminating observable differences between firms in covered industries and suitable control firms in uncovered industries. We then use an event study approach to flexibly compare the evolution of firm-level outcomes of firms in treated sectors to their matched controls. Because we match firms within the same broad sector, shocks to labor demand or supply in the same broad industry does not affect our results. Our estimation further controls for time-invariant unobservable differences between treated firms and their matched controls through firm fixed effects.

We have four main findings. First, incumbent firms increase their capital intensity relative to matched control firms in industries with a high capital-labor ratio and a large share of routine tasks. We see no effect in sectors relying heavily on low-skilled non-routine labor. We find evidence for a catch-up effect within the treated industries: it is mostly firms with prior lower capital intensity that invest in capital deepening relative to control firms. Second, we find little or no negative effect on employment; capital deepening occurs through additional investments in capital. Third, firms entering the industries with a minimum wage are more capital intensive than entrants before the minimum wage was introduced; the higher capital intensity of entrants after the adoption of a minimum wage also

[4]These recent studies suggest no (see Dustmann et al., 2022) or small displacement effects (e.g. Bossler and Gerner, 2020; Caliendo et al., 2018).

[5]More specifically, to increase the homogeneity between treated and control sample and reduce the computational burden, we only select control firms in East Germany, located in either urban or rural area and in the same broad sector than the treated firms.

holds relative to entrants in control industries. Finally, we find little evidence for an increase in outsourcing activities of firms. Yet, firms raise their revenues, which suggests that some of the additional labor costs are passed through to consumers. As the affected industries mostly provide local services, firms seem to have some room to raise consumer prices.

The paper is structured as follows. Section 3.2.1 introduces Germany's minimum wage policy and the sectors covered. Section 3.3 introduces our main data sources and provides descriptive evidence of the key variables. We describe our empirical estimation strategy to identify the minimum wage effects among incumbent firms in Section 3.4. Section 3.5 reports the results for incumbents, while Section 3.6.1 analyzes firm entry. Finally, Section 3.7 discusses the implications of our findings and concludes.

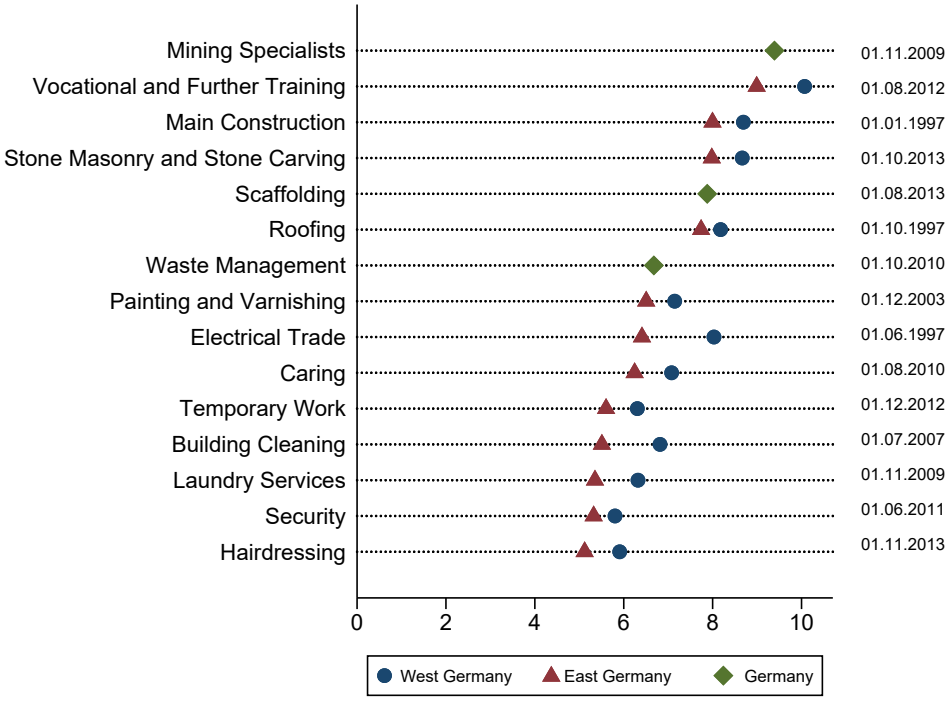
3.2 Institutional Background

3.2.1 The Introduction of Industry-Specific Minimum Wages

Minimum wages in Germany were first introduced at the industry level starting in the late 1990s – many years before a national minimum wage came into effect on January 1, 2015. With the free movement of workers in the EU, German firms increasingly faced competition from Eastern Europe where labor costs were relatively low. The competition was strongest in traditional craft industries like construction as many tasks can be performed by posted workers from abroad. Several of these industries also saw the wage inequality between workers in the industry rising. To counter the rising competition from cheap Eastern European labor and the challenge of rising wage disparity, minimum wages were adopted in 1997 in the main construction sector, roofing and electric trade. Several other industries followed over the next fifteen years.

Figure 3.2.1 gives an overview of industry-specific minimum wage regulations that were introduced in Germany until 2014, together with the corresponding minimum wage levels at the time of introduction. The industries cover traditional crafts and trades with a mostly medium-skilled workforce subject to strict occupational entry regulations (e.g. the requirement to hold a master craftsman's degree to open a company). Examples for such industries are construction, roofing, painting, varnishing and electrical trades. Yet, minimum wages were also adopted in highly labor-intensive industries that involve less complex tasks of low- or medium-skilled workers like laundry services, building cleaning, caring, hairdressing or security

Figure 3.2.1: Industry-Specific Real Minimum Wages in Germany



Notes: The figure shows the industry-specific minimum wages per hour that were adopted in Germany between 1997 and 2013. Minimum wages deflated to prices in 1997. The industries are indicated on the left-hand side, the date of introduction at the right-hand side. Green diamonds show the minimum wage for industries that adopted one for Germany as a whole. The other industries introduced lower minimum wages in East Germany (shown as red triangles) than in West Germany (shown as blue circle).

services.[6] In some industries like waste management or caring, the public sector is an important provider; others offer highly specialized products and services like mining, stone masonry or scaffolding.

It is important to stress that a minimum wage is binding for all employees working in a certain industry in Germany irrespective of where the employee received their salary (domestically or abroad).[7] That implies that all craftsmen hired for a construction site in Germany, for instance, have to be paid at least the

[6]Some special regulations and exceptions apply. In laundry services, workers are only covered if firms supply their services to commercial clients like large firms, hotels or restaurants. In building cleaning, workers are only covered if they are predominantly engaged in cleaning activities. In caring, hospitals and facilities are exempted from the regulations if they mainly provide ambulatory nursing services, medical prevention, rehabilitation, services to promote participation in community life or educating sick or disabled persons.

[7]See the Posted Workers Act ("Arbeitnehmer-Entsendegesetz") that was passed in 1996.

minimum wage agreed in the German collective bargaining agreement, which are traditionally negotiated at the industry level. Hence, a foreign company (e.g. a Polish construction firm) that is hired for constructing buildings in Germany has to pay its workers at least the minimum wage, similar to domestic firms.

Figure 3.2.1 shows the range of minimum wages implemented in Germany before 2014 – just before the national minimum wage of 8.50 Euros was implemented in 2015. Levels vary mostly between East and West Germany and between groups of workers (e.g. skilled vs. unskilled workers).[8] Most industries raised the minimum wages over time. We focus in our analysis on the effects of the adoption of a minimum wage and its initial level, not on subsequent marginal changes in the level of the minimum wage. It is important to stress that industry-level minimum wages remain in place even after 2015 if they lie above the statutory minimum wage. Figure 3.2.1 shows that most industries implemented lower minimum wages in East than in West Germany to reflect the lower wage levels in East Germany. Yet, the lower minimum wage does not fully compensate the wage differentials between East and West Germany. As a result, minimum wages are much higher relative to the median wage in East Germany than in West Germany. Our main analyses therefore focus on East Germany where we expect the effects to be stronger than in West Germany.

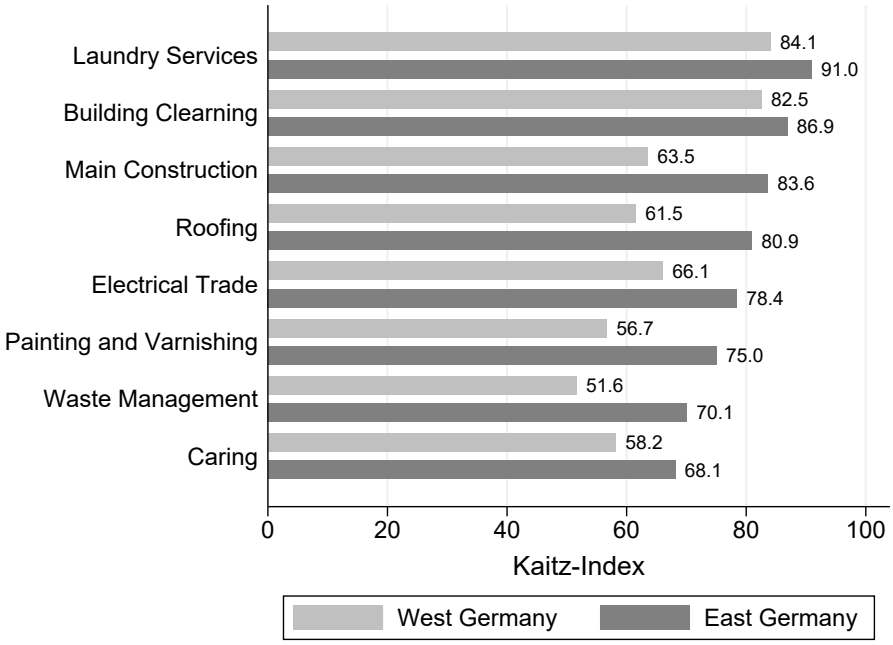
3.2.2 Bite of the Minimum Wage

To see where the minimum wage is located in the wage distribution, Figure 3.2.2 shows the Kaitz-Index, measured as the hourly minimum wage divided by the median hourly wage using the latest year available before the minimum wage introduction, by industry and East/West.[9] The figure is currently restricted to industries that adopted a minimum wage until 2010, excluding mining. A value of 100% means that the minimum wage equals the median wage. Figure 3.2.2 documents that industries vary a lot in the bite of the minimum wage. In caring, the Kaitz-Index varies between 58-68%. In laundry services, in stark contrast, the Kaitz-Index is 91% one year before its adoption. Note that the minimum wages relative to median wages are much higher than the national minimum wage Germany introduced in 2015 where the Kaitz-Index is only 46% (see Table 3 in Manning, 2021). These industry-specific minimum wages are also high in international comparison: the average Kaitz-Index in OECD countries was around

[8] In rare cases such as security services, the minimum wage varies even between federal states.
[9] Section 3.3 provides a detailed discussion of the data source used.

50% in 2018 with a range between 30% and 70% (see Gregory and Zierahn, 2022, for a more detailed discussion).[10]

Figure 3.2.2: Kaitz Index for the Industry-Specific Minimum Wages



Notes: The figure shows the Kaitz-Index measured as the hourly minimum wage level divided by the median hourly wage one year before the policy introduction. The figure shows all industries that introduced a minimum wage until 2010; we also exclude mining. See Section 3.3 for details on the data.

3.3 Data Sources

3.3.1 Firm-Level Data

Our empirical analyses uses firm-level information from the Dafne database provided by Bureau van Dijk.[11] Dafne contains data from balance sheets and income statements. The data are originally collected and published by credit rating agencies, official company registers and company reports. It is considered the most comprehensive balance sheet dataset on German firms. BvD's firm data has been

[10] Gregory and Zierahn (2022) note that Kaitz indices are typically larger in low-wage industries, which makes them an attractive subject to study as we might expect to find larger effects of a high minimum wage in low-wage industries.

[11] For Germany, Bureau van Dijk relies on data from Creditreform, the largest credit rating agency in Germany.

extensively used in the economic literature (see Jäger et al., 2020; Autor et al., 2020, for examples).

The data include the five-digit industry of a firm, which is crucial to identify firms in which an industry-level minimum wage was introduced from firms in closely related industries where no minimum wage was adopted.[12] In addition, the data report the location, founding year, date of closure, legal form, shareholder structure and sales of the firm, which will be used below to identify control firms that resemble firms in industries adopting a minimum wage along observable characteristics. Most importantly for our purpose, the data contain information on the capital stock distinguishing between its fixed capital like buildings, machinery or assets; and current assets like inventories. Information on the workforce is limited and most comprehensive for the number of employees.

For our analyses, we drop all firms operating in financial services, real estate, agriculture and forestry, public administration, social security, defense, membership organisations (e.g. religious organisations) as well as arts and entertainment. These sectors have not introduced a minimum wage and differ substantially in their capital-labor structure and the production process from sectors that introduced a minimum wage. We further restrict our sample to limited liability firms two years prior to the adoption of a minimum wage.[13] To fill gaps in our data, we exploit the panel dimension of the data. We impute missing values of a variable for a firm if the same firm has valid entries for the same variable in the year before and after. We do not extrapolate outside the observed time frame of the variable for the given firm. Specifically we impute the following variables: Log employees, log capital (tangible assets), log capital per employee, log liabilities, log balance sheet total, equity share, the ratio of fixed assets per balance sheet total, investment intensity (fixed assets divided by current assets), log fixed assets, log inventory and log current assets. Log transformed variables are imputed after the log transformation.

Our main independent variable is the firm's capital intensity. We calculate a firm's capital intensity using tangible assets, which measures the value of all physical goods purchased and used in production process long-term, divided by the number of employees in the firm.[14]

In addition to the capital structure of the firm, we also investigate whether

[12] We use a unique walkover to bridge changes in the industry classification over our observation period (Bersch et al., 2014).

[13] The sample also includes limited partnerships with a limited liability company as general partner.

[14] Figure 3.A.1 in Appendix 3.A shows the different concepts of capital in the Dafne database. We focus on the broad category of tangible assets because more detailed categories contain many missing values.

the firm adapts by outsourcing or offshoring tasks in the production process. Outsourcing and offshoring are difficult to measure at the firm level with balance sheet data. As a proxy, we use the inventories held by the firm, which combines advanced payments, raw materials, intermediate inputs, goods in process and finished goods. Inventories thus include any pre-fabricated materials or goods that the firm uses as inputs in the production process or sells in combination with its products and services. Yet, the measure also includes finished goods, which could be influenced by business cycle effects or industry-specific demand changes.[15] Hence, our measure should be considered a rough proxy for outsourcing and offshoring that also contains other categories unrelated to the concept we want to capture. We use the number of employees to scale capital and inventories. Finally, we use information on revenues to study potential adjustments through consumer prices.

3.3.2 Employment and Wage Data

We further use data on employment and wages from a different data source to illustrate the bite and characteristics of the industries adopting a minimum wage.[16] The data are taken from Ganserer et al. (2022) based on a 2-percent random sample of workers subject to social security contributions, thus excluding civil servants and self-employed individuals by the Institute for Employment Research (IAB). The data include individual employment histories together with detailed worker characteristics including age, education, gender, daily wage, workplace location and the occupation of a worker. Most importantly, the data include an industry code at the detailed 5-digit level, which is crucial to identify the industries that adopted a minimum wage.

Social security data do not include information on hours worked, which is imputed from information on working hours in the Microcensus. The Microcensus is an annual survey of one percent of all households in Germany and includes information on the number of hours worked per week. Average weekly working hours are calculated in 5376 cells along the following individual characteristics: industry, year, region, age, education, gender, employment status and type of work (for a more detailed discussion of the procedure see Ganserer et al., 2022). After

[15] Inventories are part of the current assets that a firm will use or sell within a short time period. As such, it differs from our measure for capital intensity, which covers all fixed tangible assets like equipment and machines, for instance.

[16] Because of data protection constraints, we cannot currently link the Dafne data to the employment and wage records at the establishment level.

merging the information on hours worked by cell to the social security data, the merged data is aggregated to the industry-year level. We then calculate average hourly wages in each industry and year and several other statistics describing the employment and wage structure of the treatment industries (see Table 3.3.1 below).

We further calculate the task structure within minimum wage industries using information on tasks performed in occupations from Berufenet, a database that classifies all occupations according to the five standard task groups: analytic non-routine, interactive non-routine, manual non-routine, manual-routine and cognitive routine (see Dengler et al., 2014, for details). The task data we use cover 334 occupations (3-digit level), which we weigh by occupational employment shares in each industry (taken from the Microcensus in the year prior to the adoption of a minimum wage) and aggregate it to the industry level.

3.3.3 Selection of Industries

We focus in our analysis on industries with a sizable number of affected employees (more than 30,000) and for which the minimum wage was adopted during our sample period from 2005 to 2014. For incumbent firms, we require in addition non-missing values for our key variables and a sufficient number of firms (at least 100) in each industry to implement our empirical strategy.[17] When analyzing the effects of minimum wages on incumbent firms, we therefore focus on three industries: waste management, caring and scaffolding. To study firm entry, we rely on a broader sample of industries including waste management, caring, scaffolding, hairdressing, security services, stone masonry and stone carving, building cleaning and laundry service.[18]

Table 3.3.1 characterizes the workforce and firm structure for the three industries. As our employment and wage data currently stops in 2010, workforce characteristics for scaffolding are missing. The figures suggest that industries differ in terms of employment and firm structures. In particular, the caring sector is large in size, has a high share of women and its workforce is well educated. Wage management, in contrast, employs a higher share of men, mostly in blue-collar jobs.

Firm size also varies a lot: firms in the caring sector are larger than the average firm in waste management and scaffolding. Interestingly, firms' average log tangible

[17] Appendix Table 3.D.1 shows the number of firms in our balance shett data for the other industries, which introduced a minimum wage between 2005 and 2014.

[18] We do not analyze temporary work because that industry sends workers temporarily to work in other industries but does not produce anything itself. We exclude vocational and further training because the minimum wage only applies to specific firms that we cannot clearly identify in our data. In the mining industry we do not observe entering firms.

Table 3.3.1: Worker and Firm Characteristics in Minimum Wage Industries

	Waste Management	Caring	Scaffolding
West:			
Workforce characteristics:			
Median hourly wage (in euros)	15.5	14.6	N.A.
Share of whilte-collar workers	21.7	39.1	N.A.
Share of workers with vocational degree	75.8	82	N.A.
Share of workers with university degree	5.6	7.2	N.A.
Share of female workers	14.3	81.1	N.A.
Share of workers with age above 40	34	34.3	N.A.
Number of workers	2091	12848	N.A.
Firm characteristics:			
Log tangible assets	11.639	11.408	11.493
Log employees	2.284	3.197	2.349
Log capital-labor ratio	9.837	8.443	9.464
East:			
Workforce characteristics:			
Median hourly wage (in euros)	11.4	11	N.A.
Share of white-collar workers	20	37.6	N.A.
Share of workers with vocational degree	87.2	86.8	N.A.
Share of workers with university degree	6.4	8.1	N.A.
Share of female workers	19.8	82.2	N.A.
Share of workers with age above 40	26.3	37.7	N.A.
Number of workers	796	3244	N.A.
Firm characteristics:			
Log tangible assets	11.860	11.780	11.749
Log employees	2.420	3.250	2.377
Log capital-labor ratio	9.834	8.692	9.531

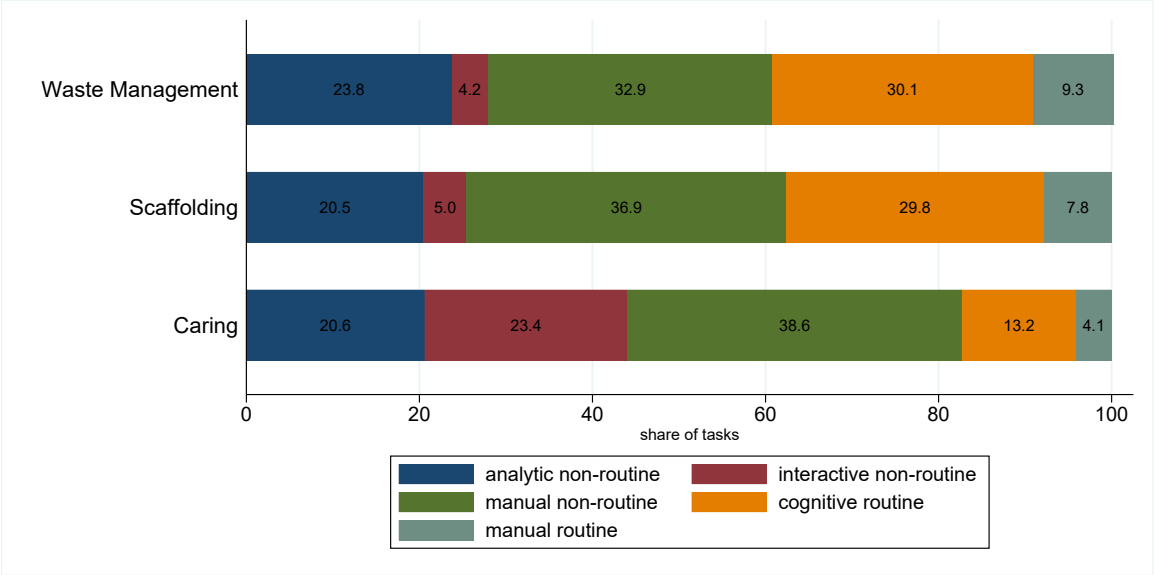
Notes: Source: SIAB and Microcensus data for workforce characteristics (see Section 3.3), Dafne for firm characteristics. Numbers of workers are calculated from the 2 percent random sample of the workforce subject to social security contributions. The numbers refer to 2009, the year prior to the adoption of a minimum wage, for caring and waste management.

assets are similar across the three industries. As a consequence, the log capital-labor ratio is higher in scaffolding and waste management than in the caring sector. These industry differences are similar between East and West Germany.

Given the industry structure just documented, how would firms respond to the introduction of a minimum wage with a sizable bite? As discussed in the introduction, firms have many margins to adjust to a minimum wage. One adjustment margin is that firms might substitute away from low-skilled workers

toward capital. According to the Routine-Replacing Technological Change (RRTC) hypothesis, we can expect more automation in industries with more routine tasks. Such routine tasks follow a set protocol, making them codifiable in software, and are hence more likely to be substituted by machines and algorithms (see e.g. Autor et al., 2003). To provide more insights into such potential for automation, Figure 3.3.1 shows the work task shares for our selected industries (see Section 3.3 for details on the data). Roughly 39 percent of tasks performed in waste management are "routine" (represented by the orange and anthracite bars in the figure). The corresponding share is only 15 percent in the caring sector. Actual decisions to automate will certainly depend on many more factors beyond the degree of routine work. Yet, judged by the amount of routine work, the potential for automation is relatively higher in waste management, somewhat less in scaffolding and relatively small in caring.

Figure 3.3.1: Occupational Task Structures in Minimum Wage Industries



Notes: The figure shows the average task shares in the occupations of the minimum wage industries waste management, scaffolding and caring calculated from the Microcensus one year before the policy introduction. Information on occupational tasks are taken from the Berufenet data base. Industries are aggregated at the 3-digit industry level.

3.4 Empirical Approach

Firms in industries covered by a minimum wage might differ systematically from firms in industries without a minimum wage in terms of the underlying production function, the capital intensity and demand shocks. To adjust for such differences in

levels and trends, we combine a matching strategy with an event study approach. We describe each of them in turn. Here, we focus on incumbent firms that are observed throughout three years prior treatment to four years after treatment and have valid information on key variables such as number of employees, tangible assets, liabilities, equity ratio, founding year and total balance sheet. Below, we also investigate whether minimum wages affect firm entry.

3.4.1 Matching Step

To identify suitable controls, we match treated firms in industries identified at a very detailed 5-digit level to firms that belong to the same broad sector (we distinguish 21 broad sectors) but whose main activity is in a 5-digit industry without a minimum wage.[19] More specifically, we use Mahalanobis distance matching with replacement to select suitable control firms that operate in closely related industries but are not covered by a minimum wage. We match on the following firm characteristics or a subset thereof: the total balance sheet (in logs), firm age (in years) or log firm age, liabilities (in logs) and the ratio of total fixed assets per total balance sheet. All matching variables are measured two years prior to the adoption of a minimum wage.

We impose that firms located in urban (rural) areas can only be matched to firms that are located in urban (rural) areas. We focus in our main results on East Germany where the bite of the minimum wage and hence, the total number of affected workers in the treatment industries is much higher than in West Germany (see Figure 3.2.2). As such, we require that treated and control firms operate in East Germany. We then compute the Mahalanobis distance for all firms in a treated industry and all potential control firms within the same broad industry separately for each minimum wage industry.[20] We then select for each treated firm the nearest neighbor, the firm with the minimum Mahalanobis distance within the same broad

[19]The broad industry "Water supply; sewerage, waste management and remediation activities" contains a single treated industry (waste management) but very few potential control industries. To implement matching, we additionally allow firms in the broad sector "Construction" that are not subject to a minimum wage introduction as potential controls.

[20]Specifically, we calculate the following metric:

$$\text{Matched Control } x_{nn} := \arg\min_{y \in C_x} \sqrt{(y-x)'\Sigma^{-1}(y-x)},$$

where x denotes the covariate vector of the treated firm, x_{nn} its selected nearest neighbor, C_x denotes the set of covariate vector representing all potential controls and Σ the sample variance-covariance matrix of the covariates used calculated with the whole set of treated and potential controls. By scaling the Euclidean distance with the inverse sample covariance matrix, the Mahalanobis distance is scale invariant and automatically corrects for correlations of covariates.

industry, the same region and the same settlement pattern. The matching step relies on two main assumptions. The conditional independence assumption is satisfied when we include all covariates in the matching procedure that influence both the treatment status and the outcome in the broad industry. The second assumption is that we have enough overlap in the joint covariate space to ensure common support in treatment states. This second assumption implies that the treatment state is non-deterministic for all joint realizations of the covariates.

3.4.2 Event Study Approach

We then use an event study approach to flexibly trace how outcomes change with the adoption of a minimum wage affects over time. We first pool our sample of matched treated and control firms over all industries. Specifically, we estimate variants of the following model:

$$Y_{fi\tau} = \sum_{\tau=-3}^{-2} \alpha_\tau MW_{i\tau} + \sum_{\tau=0}^{+3} \beta_\tau MW_{i\tau} + \theta_\tau + \gamma_f + \varepsilon_{fi\tau} \tag{3.1}$$

where τ denotes the time (in years) relative to the minimum wage introduction. The minimum wage is introduced between $\tau = -1$ and $\tau = 0$ and our reference period is $\tau = -1$. $Y_{is\tau}$ is the outcome of firm f operating in industry i in time-period τ.

$MW_{i\tau}$ is an indicator equal to one if the firm operates in year τ in an industry i that adopts a minimum wage; and zero otherwise. θ_τ are indicators for the time relative to the minimum wage introduction; these parameters adjust for general time trends that affect treated and control firms in a similar fashion.[21] The specification further includes firm fixed effects (γ_f) to control for firm-specific unobserved differences in production, market structure and demand conditions.

The coefficients α_τ identify whether there are any differential changes in outcome variables between treated and untreated firms prior to the adoption of a minimum wage. The primary parameters of interest are β_τ, which represent the effect of a minimum wage policy on firms in covered industries relative to similar firms in uncovered industries. Each coefficient measures the cumulative effect τ years after adoption relative to the pre-reform year. As the coefficients in an event study may not identify the treatment effect on the treated in the presence of heterogeneity across treated units or over time, we report below an imputation

[21] As waste management and caring introduce their minimum wage in the same year and we only have one post-adoption year for scaffolding, we do not include calendar fixed effects in addition to the period fixed effects.

estimator that is robust to the presence of heterogeneous treatment effects. We report this imputation estimator after our main results.

The tables report standard errors clustered at the firm level. To check the sensitivity of our approach, we also estimate standard errors clustered at the (5-digit) industry level. Each table reports at the bottom the 95% confidence intervals for the coefficient of the last post-event period estimated from a wild bootstrap. We now turn to our main empirical results for incumbent firms.

3.5 Results for Incumbent Firms

3.5.1 Matching Step

Table 3.5.1 provides evidence on the matching step. In each of the three industries, we have at least 100 treated incumbents subject to a minimum wage and their respective control firms, which we observe three years before and four years after the minimum wage was adopted: waste management ($N_f = 478$), the nursing and caring sector ($N_f = 593$) and the scaffolding industry ($N_f = 152$).

A comparison of treated and control firms (column (1)) and treated and all untreated in the matching set (in column (3)) shows that matching does work well in eliminating the differences in observable firm characteristics. The top panel shows no differences in the firm variables we match on; the bottom panel of Table 3.5.1 shows that firms in treated and control industries look very similar even along the dimensions we do not match on. Matching eliminates all differences in firm inputs and structure in the caring industry. In scaffolding, matching eliminates most differences with the exception of the firm's equity ratio and revenues. In waste management, matched control firms have fewer employees and and lower revenues, but look very similar to treated firms otherwise.

Table 3.5.1: Covariate Balance before and after Matching

	Waste Management				Caring				Scaffolding			
	All Firms		Control-Treated Firms		All Firms		Control - Treated Firms		All Firms		Control - Treated Firms	
	Diff	Std	Diff	Std	Diff	Std	Diff	Std	Diff	Std	Diff	Std
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log Balance Sheet Total	-1.194***	0.068	-0.105	0.094	0.417***	0.113	0.065	0.100				
Log Liabilities	-0.988***	0.074	-0.064	0.098	0.467***	0.111	0.104	0.100				
Fixed Assets\	-0.266***	0.011	-0.015	0.017	-0.001	0.017	0.004	0.019	-0.161***	0.017	-0.001	0.026
Balance Sheet Total												
Firm Age	0.410	0.566	-0.067	0.370								
Log Firm Age					-0.002	0.036	0.014	0.036	0.042	0.052	0.004	0.063
Investment Intensity									0.361	1.139	-0.024	0.135
Log Current Assets									0.076	0.107	-0.022	0.114
Log Tangible Assets pEmp	-1.685***	0.090	-0.039	0.122	0.375***	0.123	0.163	0.130	-0.763***	0.147	-0.079	0.171
Log Employees	-0.560***	0.046	-0.146**	0.070	0.142	0.093	-0.092	0.085	-0.221***	0.074	-0.034	0.095
Log Tangible Assets	-2.294***	0.103	-0.187	0.135	0.505***	0.162	0.048	0.161	-1.028***	0.165	-0.132	0.167
Log Balance Sheet Total									-0.155	0.110	-0.041	0.104
Log Liabilities									-0.114	0.122	0.130	0.119
Equity Ratio	-0.042***	0.011	-0.007	0.015	-0.023	0.015	-0.016	0.016	-0.048**	0.020	-0.076***	0.026
Firm Age					0.513	0.557	-0.008	0.450	1.581	0.978	0.066	0.738
Log Firm Age	-0.064**	0.029	0.007	0.031								
Investment Intensity	-0.749	0.849	0.346	0.332	-0.746	0.500	0.344	0.558				
Log Revenues	-0.776***	0.070	-0.236**	0.095	0.835***	0.125	0.166	0.108	0.157	0.108	0.186*	0.098
Log Current Assets	-0.712***	0.066	-0.097	0.089	0.487***	0.096	0.049	0.079				
Log Wage Sum	-0.821***	0.124	0.003	0.155	0.727***	0.116	0.083	0.101	0.003	0.260	-0.068	0.268
Log Wage	0.011	0.053	0.131*	0.067	0.267***	0.048	0.202	0.051	-0.007	0.114	-0.295	0.190

Notes: Dafne Dataset, Incumbent sample for East Germany, all values computed for the matching year, two years prior treatment. The upper rows display differences in the matching variables. The bottom rows differences in variables not matched on. Results from two sample t-tests with the null hypotheses of zero differences. Differences: Control - Treated. All firms: set of all firms that can be selected as nearest neighbours in the matching step and treated firms; Matched sample: selected nearest neighbour control firms and treated firms.

3.5.2 Rising Capital Intensity

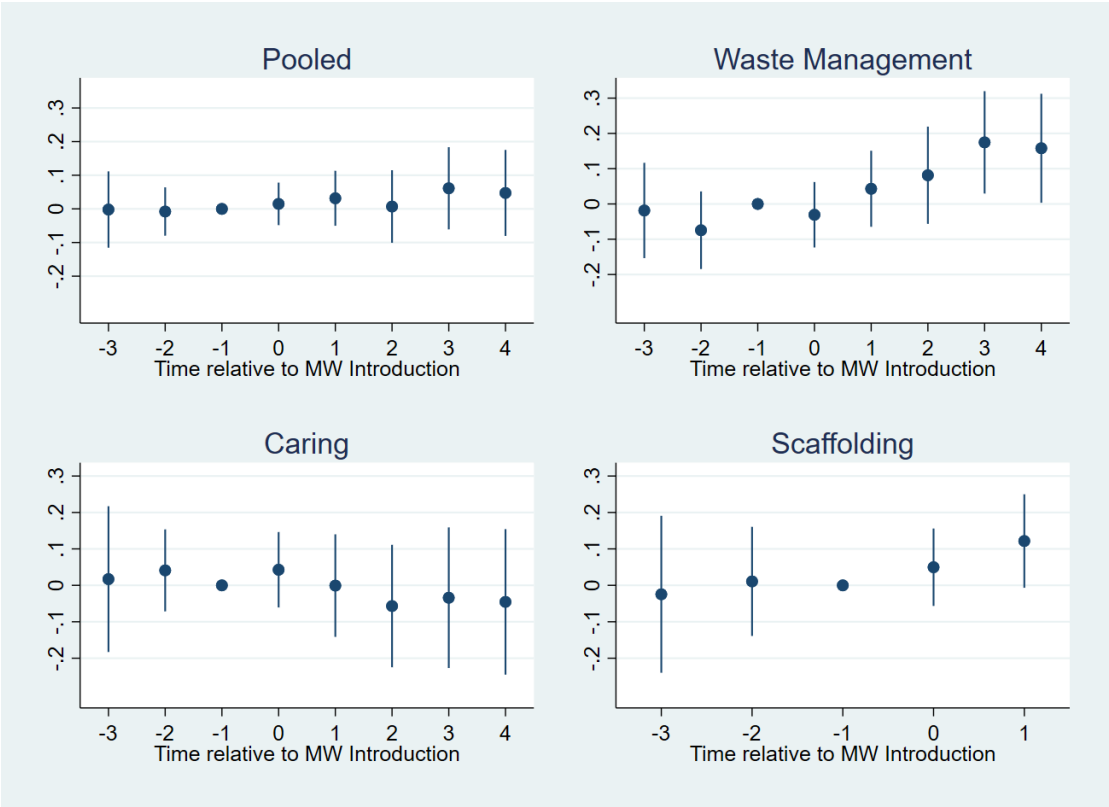
Figure 3.5.1 plots the coefficients from estimating equation (3.1) where the dependent variable is the capital-labor ratio. The time period spans from three years before and up to four years after the adoption of the minimum wage in treated firms and matched control firms. Year zero corresponds to the year when the industry first adopts a minimum wage, while the Whiskers represent the 95% confidence intervals.

Firms subject to a minimum wage experienced very similar trends in their capital intensity than their respective control firms in the period before the adoption. After the adoption, we observe a rise in the capital-labor ratio in waste management and scaffolding compared to control firms with the gap widening with time since adoption. In caring, in contrast, we see only a muted response in treated firms. Figure 3.5.1 thus provides first support for the hypothesis that minimum wages encourage capital-labor substitution in minimum wage industries where a sizable share of tasks performed fall into the routine category.

We next quantify the effect of a minimum wage on firm's capital intensity in Table 3.5.2. Pooling across the three industries, we find only a small, insignificant increase in the capital intensity of incumbent firms. Estimating the model in equation (3.1) separately for each industry reveals some interesting patterns: we find a very strong response in waste management and scaffolding in line with the visual findings in Figure 3.5.1. In waste management, the capital-labor ratio increases by about 16 percentage points four years after the adoption of a minimum wage. There is a similar strong response in the scaffolding industry after the minimum wage was introduced. We find no effect of the minimum wage on the capital intensity in caring.

The results in Table 3.5.2 raise the question whether the higher capital intensity comes from additional investments in physical capital or rather from reducing the number of employed workers. A decline in the number of workers employed would speak to employment adjustments as the main mechanism to adapt to the minimum wage. Actual investments in physical capital would rather support the view of capital deepening and capital-labor substitution. Firms might also combine the two processes by substituting away from labor toward physical capital through automation, for instance. To shed light on this question, we re-estimate our event study in equation (3.1) where the dependent variables are now either log capital or log employees. Table 3.5.3 shows no evidence for a decline in employment within four years of the adoption of a minimum wage compared to control firms. On

Figure 3.5.1: Results: Capital-Labor Ratio



Notes: The figure shows coefficient plots from estimating equation (3.1) where the dependent variable is log capital intensity (log capital per employee). The coefficients show the effect relative to the year prior to the minimum wage adoption (in $\tau = -1$). Sample includes incumbent firms in the treated industries waste management, caring and scaffolding and matched control firms, all located in East Germany. The top left panel shows pooled estimates, the other three panels estimates for each industry separately. The estimation controls for period and firm fixed effects. Standard errors are clustered at the firm level. Data source: Dafne.

Table 3.5.2: Minimum Wages and Capital Intensity

	Pooled (1)	Waste Industry (2)	Caring (3)	Scaffolding (4)
3 years before	-0.002	-0.019	0.017	-0.025
	(0.058)	(0.069)	(0.102)	(0.109)
2 years before	-0.008	-0.075	0.041	0.011
	(0.037)	(0.056)	(0.057)	(0.076)
1 year before				
0 years after	0.015	-0.031	0.043	0.050
	(0.032)	(0.047)	(0.053)	(0.054)
1 year after	0.032	0.043	-0.001	0.122*
	(0.042)	(0.055)	(0.072)	(0.065)
2 years after	0.007	0.081	-0.057	
	(0.055)	(0.070)	(0.086)	
3 years after	0.061	0.175**	-0.034	
	(0.062)	(0.074)	(0.098)	
4 years after	0.047	0.158**	-0.045	
	(0.065)	(0.079)	(0.102)	
Wild Bootstrap CI	[-0.107, 0.240]	[-0.035, 0.394]	[-0.216, 0.106]	[-0.052, 0.291]
R^2	0.01	0.01	0.01	0.04
N	18,656	7,648	9,488	1,520

Notes: Results from equation 3.1 where the dependent variable is log capital intensity (log capital per employee). Sample includes incumbent firms in the treated industries waste management, caring and scaffolding and matched control firms, all located in East Germany. The specification includes period and firm fixed effects. Standard errors clustered at the firm level are shown in parentheses. Wild bootstrap CI denotes 95% confidence intervals for the coefficient four years after treatment where standard errors are clustered at the 5-digit industry level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

the contrary, the pooled estimates in column (1) suggest even some employment growth though none of the coefficients reaches statistical significance. Looking at the individual industries, we see no employment adjustments in waste management (see column (3)) and even employment growth in scaffolding and the caring sector (see columns (5) and (7)). Turning to capital, we find that there is much stronger growth in physical capital (see column (2) for the pooled estimates) than in employment. The strongest response occurs in waste management and scaffolding, both traditionally capital-intensive industries. In waste management, the capital stock grows by 14.7 percentage points within four years after the adoption of the minimum wage (see column (4)). In scaffolding, capital increases by 15 percentage points (see column (8)). Both set of estimates are statistically significant at the 5% resp. 10% level. In caring, there is little additional investment in capital; here, the main response seems to be an increase in employment. The latter response would be consistent with firms in the caring sector having some monopsony power in their local labor markets relative to control firms (Manning, 2013).

Table 3.5.3: Estimates for Log Employees and Log Capital

	Pooled		Waste Management		Caring		Scaffolding	
	Log Employees (1)	Log Capital (2)	Log Employees (3)	Log Capital (4)	Log Employees (5)	Log Capital (6)	Log Employees (7)	Log Capital (8)
3 years before	-0.045	-0.054	-0.026	-0.057	-0.056	-0.042	-0.061	-0.087
	(0.034)	(0.054)	(0.032)	(0.066)	(0.061)	(0.095)	(0.067)	(0.085)
2 years before	-0.014	-0.014	0.012	-0.080	-0.026	0.050	-0.053	-0.056
	(0.023)	(0.034)	(0.026)	(0.052)	(0.040)	(0.055)	(0.045)	(0.062)
1 year before								
0 years after	-0.009	-0.003	0.003	-0.037	-0.028	0.004	0.031	0.078*
	(0.016)	(0.026)	(0.024)	(0.038)	(0.027)	(0.044)	(0.031)	(0.042)
1 year after	-0.002	0.032	-0.005	0.043	-0.011	-0.008	0.040	0.151**
	(0.023)	(0.038)	(0.025)	(0.051)	(0.041)	(0.064)	(0.036)	(0.059)
2 years after	0.020	0.027	-0.021	0.047	0.048	0.002		
	(0.028)	(0.050)	(0.032)	(0.064)	(0.045)	(0.078)		
3 years after	0.018	0.086	-0.016	0.148**	0.039	0.026		
	(0.030)	(0.059)	(0.035)	(0.069)	(0.048)	(0.092)		
4 years after	0.039	0.089	-0.003	0.147*	0.067	0.033		
	(0.033)	(0.063)	(0.037)	(0.077)	(0.052)	(0.096)		
Wild B CI	[-0.121, 0.164]	[-0.026, 0.222]	[-0.070,0.071]	[-0.038,0.373]	[-0.018,0.169]	[-0.132,0.234]	[-0.019,0.165]	[0.042,0.342]
R^2	0.07	0.04	0.02	0.02	0.11	0.06	0.04	0.16
N	18,656	18,656	7,648	7,648	9,488	9,488	1,520	1,520

Notes: Results from equation 3.1 where the dependent variables are log employees (odd columns) and log capital (even columns). The sample includes incumbent firms in the minimum wage industries (waste management, caring and scaffolding) and matched control firms. Treated and matched control firms are located in East Germany. All specifications include period indicators relative to treatment and firm fixed effects. Standard errors clustered at the firm level are shown in parentheses. Wild Bootstrap CI: Wild bootstrap 95% confidence intervals with clustering at the 5-digit industry level for the coefficient four years after treatment. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

The evidence thus far reveals that in capital-intensive industries like waste management or scaffolding, firms in industries that adopt a minimum wage mainly adjust by investing more in physical capital thus raising their capital intensity. In contrast, there is no evidence of capital deepening in the caring sector, which is traditionally labor-intensive but with a low routine task intensity. Yet, do we see the adjustments in all firms or are the adjustments concentrated in a subset of firms in the treated industries? To shed more light on this question, we split the sample by capital intensity two years prior to the adoption of the minimum wage. We thus define a treated firm as capital intensive (high) if it has a capital-labor ratio above the median in its industry and as not capital-intensive (low) if it is below the median. Control firms are then assigned to the same group as their corresponding match.[22]

We re-estimate equation (3.1) separately for capital-intensive and less capital-intensive firms. As before, we include period fixed effects and firm fixed effects and cluster the standard errors at the firm level. Table 3.5.4 reveals that across the three treated industries, it is the less-capital intensive firms (see column (1)) that invest heavily in capital deepening. After the minimum wage adoption, the capital-labor ratio increases by a stunning 20 percentage points in firms with low capital intensity. In contrast, we see no effect in firms that had a high capital intensity prior to the adoption of the minimum wage (see column (2) of Table 3.5.4). Looking at the individual industries, we see that it is in the capital-intensive industries (waste management and scaffolding) that incumbents with a low capital-labor ratio try to catch up by investing in capital deepening (see column (3) for waste management and column (7) for scaffolding). We see no change in the capital-labor ratio for capital-intensive firms in the two industries.

3.5.3 Alternative Adjustment Margins

Outsourcing Instead of capital deepening, firms in industries that adopt a minimum wage could also adjust by outsourcing some of their activities to other firms that operate in industries that are not subject to a minimum wage. While it is difficult to track outsourcing in balance sheet data, one proxy is the amount of inventory a firm holds. Inventory contains all materials and pre-fabricated products that firms use as inputs in their production or service. As not all firms report their inventories so the sample of firms is somewhat smaller than for capital

[22]Results look very similar if we split the sample at the mean capital intensity before the adoption instead.

Table 3.5.4: Who Invests in Capital Deepening in Response to a Minimum Wage?

	Pooled		Waste Industry		Caring		Scaffolding	
	Low	High	Low	High	Low	High	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
3 years before	0.003	0.002	0.017	-0.041	0.020	0.015	-0.092	0.107
	(0.103)	(0.060)	(0.103)	(0.089)	(0.192)	(0.092)	(0.173)	(0.135)
2 years before	-0.071	0.036	-0.157**	-0.022	-0.033	0.090*	0.034	0.009
	(0.061)	(0.040)	(0.080)	(0.073)	(0.101)	(0.054)	(0.146)	(0.060)
1 year before								
0 years after	0.084	-0.029	0.053	-0.084	0.097	0.007	0.123	0.015
	(0.057)	(0.034)	(0.066)	(0.060)	(0.101)	(0.046)	(0.099)	(0.058)
1 years after	0.175**	-0.055	0.236***	-0.080	0.095	-0.064	0.277**	0.095
	(0.079)	(0.041)	(0.086)	(0.067)	(0.145)	(0.060)	(0.109)	(0.078)
2 years after	0.182*	-0.104**	0.279***	-0.044	0.098	-0.158**		
	(0.104)	(0.052)	(0.105)	(0.085)	(0.172)	(0.070)		
3 years after	0.255**	-0.061	0.387***	0.040	0.141	-0.149*		
	(0.118)	(0.056)	(0.117)	(0.084)	(0.196)	(0.077)		
4 years after	0.224*	-0.064	0.393***	0.008	0.082	-0.129		
	(0.122)	(0.059)	(0.122)	(0.090)	(0.201)	(0.081)		
R^2	0.03	0.02	0.07	0.01	0.02	0.04	0.10	0.02
N	7,478	11,178	2,976	4,672	3,760	5,728	742	778
Capital-Labor Ratio	8.183	10.606	9.029	11.085	7.357	10.275	8.710	10.344

Notes: Results from equation 3.1 where the dependent variable is log capital per employee. Sample includes all firms in the treated industries waste management, caring and scaffolding and their matched controls, all located in East Germany. Regressions include period effects and firm fixed effects. The sample of treated firms is split according to log capital labour ratio (above/below 5-digit industry median) two years prior treatment into capital-intensive (high) and not capital-intensive (low) firms. The bottom row shows the average log capital-labor ratio in the respective sample in the year before the adoption of the minimum wage ($\tau = -1$). Standard errors clustered at the firm level are reported in parentheses. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Data source: Dafne.

intensity. We then re-estimate equation (3.1) where the dependent variable is now log inventory per employee. The results are shown in the left-hand side of Table 3.5.5. While the pooled estimates indicate that the inventory of firms increases after the adoption of a minimum wage by almost 12 percentage points relative to control firms, the coefficients are not statistically significant at conventional levels. The effect is mainly driven by the waste management industry where four years after the introduction of the minimum wage, the inventory per employee is 13 percentage points higher. We see no adjustment in the caring industry where capital plays a less important role compared to labor.

Revenues and Consumer Prices Rather than adjusting their capital stock or rely more on pre-fabricated inputs and outsource some services, firms could roll over some of the additional labor costs to consumers by raising prices. While prices are difficult to measure at the establishment or detailed industry level, the revenues changes in the treated firms should mostly reflect price changes. Using revenues as dependent variable, odd columns in Table 3.5.5 show that the revenues of firms indeed increased after the adoption of a minimum wage relative to the matched control group. In the pooled sample, revenues increase by 9.4 percentage points four years after the introduction of the minimum wage. Interestingly, revenues increase the most (by 16.1 percentage points) in waste management (see column (3)), which is also the industry where capital deepening and outsourcing plays a prominent role relative to the control group. In scaffolding, the coefficients are positive in both post-adoption years, but do not reach statistical significance (see column (7) of Table 3.5.5). In contrast, there is no visible effect on revenues in the caring industry. Overall then, incumbent firms did not react to the introduction of the minimum wage at the employment margin; instead, they respond with increasing consumer prices and capital deepening. The exception is the caring industry where we see no adjustments along these margins.

Table 3.5.5: Outsourcing and Revenues of Firms

	Pooled		Waste Management		Caring		Scaffolding	
	Log Revenues	Log Inventory per Employee	Log Revenues	Log Inventory per Employee	Log Revenues	Log Inventory per Employee	Log Revenues	Log Inventory per Employee
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
3 years before	-0.013	-0.033	0.026	-0.060	-0.054	-0.159	-0.069	0.385*
	(0.034)	(0.096)	(0.052)	(0.144)	(0.044)	(0.153)	(0.089)	(0.207)
2 years before	0.025	-0.005	0.048	-0.067	0.015	0.005	-0.038	0.207
	(0.021)	(0.073)	(0.032)	(0.116)	(0.028)	(0.094)	(0.049)	(0.186)
1 year before								
0 years after	0.006	0.067	0.024	0.037	-0.026	0.040	0.068	0.245
	(0.024)	(0.061)	(0.038)	(0.092)	(0.022)	(0.081)	(0.070)	(0.191)
1 year after	0.031	0.064	0.079**	0.129	-0.030	-0.000	0.051	-0.022
	(0.026)	(0.073)	(0.038)	(0.113)	(0.031)	(0.102)	(0.083)	(0.183)
2 years after	0.030	0.018	0.055	0.010	-0.003	-0.024		
	(0.031)	(0.075)	(0.046)	(0.111)	(0.032)	(0.101)		
3 years after	0.040	0.042	0.059	-0.000	0.015	0.050		
	(0.037)	(0.083)	(0.055)	(0.117)	(0.037)	(0.118)		
4 years after	0.084*	0.117	0.161**	0.129	-0.000	0.042		
	(0.044)	(0.097)	(0.073)	(0.145)	(0.040)	(0.116)		
Wild Bootstrap CI	[0.006,0.162]	[-0.046, 0.293]	[0.045, 0.288]	[-0.0946, 0.311]	[-0.048, 0.088]	[-0.191, 0.334]	[-0.423, 0.339]	[-1.296, 0.475]
R^2	0.07	0.00	0.05	0.01	0.22	0.01	0.07	0.02
N	5,974	9,686	3,166	5,234	2,246	3,558	562	894

Notes: Results from estimating equation (3.1) where the dependent variables are log sales in odd columns and log inventories per employee as a proxy for outsourcing in even columns. The sample includes incumbent firms in three industries (waste management, caring and scaffolding) and their matched control firms, all located in East Germany. We only include observations where the outcome is non-missing for both treated and control firm. Regressions include period and firm fixed effects. Standard errors are clustered at the firm level. Alternatively, we report 95% confidence intervals based on clustering at the 5-digit industry level based on a wild bootstrap procedure for the coefficient four years after treatment. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

3.5.4 Specification Checks

A primary concern with our estimation strategy is that the event study in equation (3.1) might not identify the treatment effect on the treated in the presence of heterogeneity across treated units or over time since adoption (Borusyak et al., 2021; de Chaisemartin and D'Haultfœuille, 2020; Sun and Abraham, 2021). To check the sensitivity of our results to that assumption, we implement the imputation estimator by Borusyak et al. (2021).[23] Table 3.5.6 reports the results for the pooled sample as well as the individual industries. For the post-treatment period, estimates of the treatment effect are reported. For the pre-event years, the table reports a test for parallel trends.[24] The results are similar to the main results reported in Table 3.A.1. There is clear evidence for capital deepening in the waste management industry; furthermore, the coefficient for scaffolding is positive and similar to that in Table 3.A.1 but just about misses statistical significance at conventional levels.

Our results for incumbent firms indicate that capital deepening is an important adjustment mechanism in waste management. Our matching procedure uses control firms selected from the same broad industry group. As the broad industry group is not very large, the matching procedure might not select the best potential control groups. To check the sensitivity of our results to this restriction, we implement another matching estimator for firms in the waste management industry where we allow control firms from all industry sectors; the matching variables are the same as in the main analysis. The left-hand side of Appendix Table 3.D.2 shows that the matching approach again balances the observable characteristics of treated firms in waste management and their alternative control firms. We then re-estimate our baseline model in equation (3.1) using the alternative control firms. Appendix Table 3.D.3 shows that even with this alternative control group, firms in waste management invest in more capital after the adoption of a minimum wage. The coefficient four years after the adoption indicate that capital intensity has increased by 19.1 percentage points, which is even slightly larger than in our main results (see column (1)). Column (2) further shows that firms in waste management reduce

[23] Rather than using post-treatment observations of earlier treated units as controls for units treated later, the estimator corrects for the potential bias in the case of heterogeneous treatment effects using a three-step approach: first, non-treated units and pre-treatment observations of treated units are used to estimate the basic model (excluding the treatment effect). In the second step, the model from step one is extrapolated to treated units by imputing non-treated potential outcomes. Finally, average the estimated treatment effects to obtain the treatment effect on the treated.

[24] The tests for parallel pre-trends are constructed using a separate regression of the outcome on the pre-event period indicators and the firm fixed effects for non-treated observations only.

Table 3.5.6: Minimum Wages and Capital Intensity: Imputation Estimator

	Pooled	Waste Management	Caring	Scaffolding
	(1)	(2)	(3)	(4)
3 years before				
2 years before	0.017	-0.056	0.024	0.035
	[0.042]	[0.053]	[0.079]	[0.085]
1 year before	0.050	0.019	-0.017	0.025
	[0.053]	[0.069]	[0.102]	[0.109]
0 years after	0.077	0.000	0.023	0.054
	(0.043)	(0.047)	(0.074)	(0.075)
1 year after	0.065	0.074	-0.020	0.126
	(0.049)	(0.055)	(0.087)	(0.084)
2 years after	-0.013	0.112	-0.076	
	(0.061)	(0.070)	(0.097)	
3 years after	0.047	0.206**	-0.053	
	(0.069)	(0.086)	(0.107)	
4 years after	0.037	0.189**	-0.065	
	(0.071)	(0.090)	(0.110)	
N	18,656	7,648	9,488	1,520

Notes: Results from implementing the imputation estimator by Borusyak et al. (2021), which identifies the treatment effect on the treated in the presence of heterogeneous treatment effects. The sample includes incumbent firms in one of the treated industries and their matched control firms, all located in East Germany and active during the 2005-2014 period. Dependent variable: log capital per employee. Standard errors clustered at the firm level are reported in parentheses. Row 2 years before and 1 year before report test statistics and standard errors for a test on pre-trends. The tests for parallel pre-trends are constructed using a separate regression of the outcome on the pre-event period indicators and the firm fixed effects for non-treated observations only. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

their workforce in response to the minimum wage by about 5.1 percentage points. This employment adjustment occurs within two years after the minimum wage was adopted. Column (3) further confirms that firms in waste management invest in more capital in the longer-run: four years after the minimum wage was adopted, physical capital has increased by 15.6 percentage points, similar to our baseline results (see Table 3.A.1).

Our results have thus far focused on East Germany where the minimum wage is higher relative to the median wage than in West Germany. Yet, do we see similar effects for West German firms where the minimum wage is still substantial with the Kaitz-index ranging from 58.2 to 84.1 (see Figure 3.2.2). To shed some light on potential adjustments in West Germany, we implement a matching estimator for the waste management industry where we have seen the strongest effects in East Germany. The matching variables are here the log of balance sheet, fixed assets divided by balance sheet total, firm age, investment intensity and the equity ratio. The right-hand side of Appendix Table 3.D.2 (columns (5)-(8)) shows that the matching procedure balances the matching variables and most observable characteristics. For a few characteristics, the matching does not achieve full balance but substantially reduces the difference between treated and control firms relative to a non-matched sample including all incumbent firms (employees, tangible assets, revenues, the wage sum and average wage). We then proceed by estimating the event study in equation (3.1). The results are reported in Appendix Table 3.D.4. We find that incumbent firms in West Germany invest in capital deepening just like their East German counterparts. The effect on the capital-labor ratio four years after the adoption of a minimum wage is with 11.8 percentage points somewhat lower than in East Germany (see column (1)). Like East German firms, West German incumbents invest heavily in physical capital by almost 18 percentage points in the longer-run (see column (3)). Unlike East German firms, West German firms in waste management actually increase the size of their workforce after the introduction of the minimum wage though the effect is with 4.4 percentage points (see column (2)) much smaller than the additional investments in physical capital.

3.6 Capital Intensity among Entering Firms

So far, we have focused on incumbent firms that are exposed to a minimum wage in their industry and compared their capital structure to those of suitable control firms in related industries. One explanation we did not find effects on capital intensity in all industries is that the effect of a minimum wage could be concentrated among

new entrants into the industry. The total effect of a minimum wage policy on the capital-labor ratio is an effect on incumbent firms plus a composition effect through firm entry and exit.[25]

We would expect that most of the adjustments occur by entrants if firms have a putty-clay technology in which the capital structure is largely fixed after entry. Putty-Clay models suggest that a firm's capital intensity is a long-term (and often irreversible) decision taken at firm entry (compare Aaronson et al., 2018). After a firm has entered an industry and chose its capital intensity, the capital structure of incumbent firms is largely fixed. And even in industries where we see adjustments in the capital structure among incumbents (waste management and scaffolding), it could be that firm entry plays an important role in the adjustment to a minimum wage.

We now investigate whether firms entering the industries that adopted a minimum wage are more capital intensive after the adoption than firms entering prior to the adoption of a minimum wage. Our analysis relies on data for all eight minimum wage industries, for which we observe entering firms in our data (building cleaning, laundry services, waste management, caring, security services, scaffolding, hairdressing, stone masonry and carving). To identify firm entry, we take the year a firm was founded from the data.[26] We then compare firms entering an industry after the adoption of a minimum wage to firms entering the same industry before the adoption.[27] As before, we restrict the analysis to the 2005-2014 period and calculate capital intensity just like for incumbent firms.

Specifically, we now estimate variants of the following model:

$$\log(K\backslash L)_{fit} = \beta Entry\ \ after\ \ MW_{ft} + \alpha Just\ \ Founded_{ft} + \theta_t + \gamma_i + \varepsilon_{fit}, \quad (3.2)$$

where $\log(K\backslash L)_{fit}$ denotes the capital intensity in entering firm f, industry i and calendar year t. $Entry\ \ after\ \ MW_{ft}$ is an indicator that takes the value of one if a firm f was founded in the year the minimum wage was adopted or later. We also include year fixed effects to control for business cycle dynamics and longer-run trends in the capital structure of entering firms (θ_t). We further control for detailed industry fixed effects (5-digits) in order to adjust for differences in the production process and capital intensity across industries. Because we only have few entrants

[25] We might also expect a decline in entry as the entry barriers have increased after the introduction of a minimum wage; we will analyze this margin in the next version of the draft.

[26] In contrast, firm exit is not well defined in the Dafne database as we cannot distinguish it clearly from missing data. We therefore focus on firm entry here.

[27] Because incumbent firms in each of the treated industry have a much higher capital intensity than entrants, we do not compare entrants to incumbent firms.

with multiple observations in the Dafne database, we cannot include firm fixed effects here.[28]

Because not all entrants might have fully built up their capital stock or hired all of their workers in the year of entry, we add in subsequent specification an indicator for the year a company was founded ($Just\ \ Founded_{ft}$). Alternatively, we control below for the age of a firm. As for incumbent firms, we cluster standard errors at the firm level but also report confidence intervals from a wild bootstrap procedure.

The coefficient of interest is β that measures whether firms that where founded after the minimum wage introduction are more capital intensive than firms founded before the minimum wage introduction. Table 3.6.1 shows the results. The pooled specification for all eight industries in columns (1)-(4) of Table 3.6.1 shows that firms entering an industry after the adoption of a minimum wage indeed start off with a higher capital intensity than firms entering the industry before the minimum wage was introduced. The coefficient is sizable (22 percentage points, see column (1)) but does typically not reach statistical significance.

We then analyze the capital intensity of entrants separately for the industries for which we did the incumbent analysis and have enough entering firms over the period: waste management ($N_{entry} = 349$) and the caring sector ($N_{entry} = 878$).[29] Columns (5) and (6) in Table 3.6.1 show that entrants have a substantially higher capital intensity in the waste management industry after the minimum wage is introduced though the coefficient does not reach statistical significance. As we saw above that the adoption of a minimum wage increased the capital intensity of those incumbents with an initially low capital intensity, these results point to sizable adjustments in the capital structure in response to the adoption of a minimum wage. For the caring sector, columns (7) and (8) show no effect on the capital intensity of entrants, which is in line with the evidence for incumbent firms. As such, our evidence indicates that in capital-intensive industries, a minimum wage triggers additional capital investments both among incumbent firms, esp. those with a low capital intensity initially, and among firm entrants. In labor-intensive industries, in turn, additional capital investments do not play a role in the adjustment to a minimum wage policy.[30]

[28] Appendix Table 3.D.5 shows the number of entrants observed in our balance sheet data for the industries that adopted a minimum wage between 2005 and 2014.

[29] Appendix Table 3.D.5 shows that only three establishment in our database enter the scaffolding industry after the introduction of a minimum wage.

[30] Appendix Table 3.D.6 suggests similar results even if we include entrants of all legal forms in the estimation.

Table 3.6.1: Capital Intensity among Entering Firms

	Pooled	Pooled	Pooled	Pooled	Waste	Waste	Caring	Caring
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Entry after MW	0.220	0.211	0.306**	0.223	0.261	0.429	-0.078	-0.249
	(0.139)	(0.139)	(0.141)	(0.156)	(0.351)	(0.386)	(0.178)	(0.208)
Just Founded		0.103			-0.498	-0.447	0.493**	0.365*
		(0.168)			(0.418)	(0.395)	(0.244)	(0.217)
Firm Age			0.012	-0.007		0.024		-0.047
			(0.017)	(0.024)		(0.060)		(0.037)
Post MW				0.183		0.500*		0.274
				(0.115)		(0.271)		(0.188)
WB CI	[-0.763,0.222]	[-0.753,0.235]	[-0.582,0.304]	[-0.597,0.046]				
R^2	0.09	0.09	0.09	0.09	0.03	0.02	0.01	0.00
N	6,540	6,540	6,540	6,540	1,124	1,124	3,021	3,021
Industry FE	yes	yes	yes	yes	no	no	no	no
Year FE	yes	yes	no	no	yes	no	yes	no

Notes: Results from estimating equation (3.2). Entry after MW is an indicator equal to one if the founding year was in year the minimum wage was adopted or later; Just Founded is an indicator equal to one if the firm was founded in the current year. Post MW is an indicator equal to one if the year of observation is in the year of adopting the minimum wage or after; and zero otherwise. The sample includes all firms entering the treated industries in East Germany over the 2005-2014 period in columns (1)-(4). The sample is restricted to firms entering the waste management industry (in columns (5)-(6)) or caring sector (in columns (7)-(8)). Standard errors are clustered at the firm level. We also report confidence intervals from a wild bootstrap procedure with clustering at the 5-digit industry level. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

One disadvantage of our analysis of entering firms according to equation (3.2) is that we only include entrants in industries that introduced a minimum wage. Hence, any industry-specific shocks that encourage or reduce entry and coincide with the adoption of a minimum wage, e.g. through changes in industry-specific regulation, for instance, cannot be separated from the minimum wage effect on entrants. As we cannot perform matching on characteristics prior to the minimum wage adoption for entering firms, we instead use the industry composition of control firms in the incumbent sample to generate a control group of entrants. The basic idea is that incumbent firms reflect the technology used in a particular industry well. As such, we would expect that firms entering an industry will mimic incumbent firms along important characteristics. Even if they do not closely mirror incumbents, we could expect entering firms in minimum wage industries to be close along observable characteristics and technology to entrants in the control industries selected in the sample of incumbent firms.

We thus weigh entering firms in control industries in order to replicate the industry composition of control firms in the incumbent sample. Industries that were not in the control sample are assigned a weight of zero. We then estimate variants of the following model:

$$\log(K\backslash L)_{fit} = \beta Entry\ \ after\ \ MW_{fit} + \alpha Just\ \ Founded_{ft} + \theta_t + \gamma MW_i + \varepsilon_{fit} \tag{3.3}$$

where $\log(K\backslash L)_{fit}$ characterizes the capital intensity among entering firm f in industry i and year t. $Entry\ \ after\ \ MW_{fit}$ an indicator equal to one if firm f entered an industry after the minimum wage was adopted. For entrants in control industries, we assign the year of adoption of the respective treated industry to the firms in the weighted control industries. MW_i is an indicator equal to one if industry i adopts a minimum wage; and zero otherwise. We further include year fixed effects (θ_t) and an indicator if a firm was founded in current year ($Just\ \ Founded_{ft}$). In alternative specifications, we control for firm age instead. As before, standard errors are clustered at the firm level.

Our coefficient of interest β reveals whether entrants in treated firms have become more capital intensive after the introduction of a minimum wage relative to entrants in the control firms. Table 3.6.2 shows a similar pattern than we saw for entering firms in treated industries in Table 3.2. Firms entering the waste management industry are more capital intensive after the industry adopted a minimum wage relative to entrants in control firms, though the coefficients do not

Table 3.6.2: Capital Intensity among Entrants in Treated and Control Industries

	Waste (1)	Waste (2)	Waste (3)	Caring (4)	Caring (5)	Caring (6)
Entry after MW * MW	0.237	0.237	0.235	0.034	0.031	0.029
	(0.411)	(0.409)	(0.410)	(0.260)	(0.260)	(0.260)
Entry after MW	0.041	0.041	0.233	-0.102	-0.123	-0.211*
	(0.230)	(0.222)	(0.235)	(0.199)	(0.200)	(0.210)
MW	0.824***	0.824***	0.837***	-0.650***	-0.651***	-0.645***
	(0.178)	(0.178)	(0.180)	(0.160)	(0.160)	(0.160)
Just Founded		0.004			0.255	
		(0.242)			(0.180)	
Firm Age			0.039			-0.038
			(0.035)			(0.032)
Post			0.437***			0.185
			(0.165)			(0.143)
Year FE	yes	yes	no	yes	yes	no
R^2	0.06	0.06	0.05	0.02	0.02	0.02
N	9,568	9,568	9,568	6,587	6,587	6,587

Notes: Results from estimating equation (3.3) where the dependent variable is log capital per employee. Entry after MW is an indicator equal to one if a firm enters after the minimum wage was adopted where the adoption year of the treated industries are assigned to the firms in the control sample of that industry. MW is an indicator equal to one if the firm enters a treated industry; and zero if the firm enters a control industry. Just Found is an indicator equal to one if the firm was founded in the current calendar year; and zero if the firm has been founded in an earlier or later calendar year. Post is equal to one for the year when the minimum wage was adopted and later years. The sample includes firms entering a treated industry and those entering control industries, which are weighted according to the incumbent sample, in East Germany in the 2005-2014 period. Standard errors clustered at the firm level are reported in parentheses. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Data Source: Dafen.

reach statistical significance (see columns (1)-(3) of Table 3.2). In contrast, we do not find any effect of the minimum wage on the capital intensity of entering firms in the labor-intensive caring sector (see columns (4)-(6) of Table 3.2).

3.7 Conclusion

We provide novel evidence on how minimum wages affect firm-level investments in capital and work organization. In contrast to most of the literature, we study the impact of the first-time adoption of a minimum wage rather than changes under an existing minimum wage policy. We exploit the German setting, which is uniquely suited to analyze the labor demand responses of adopting a minimum wage policy. Based on a large dataset on firms including detailed financial information on the capital stock (capital assets, technical equipment and machinery, factory and

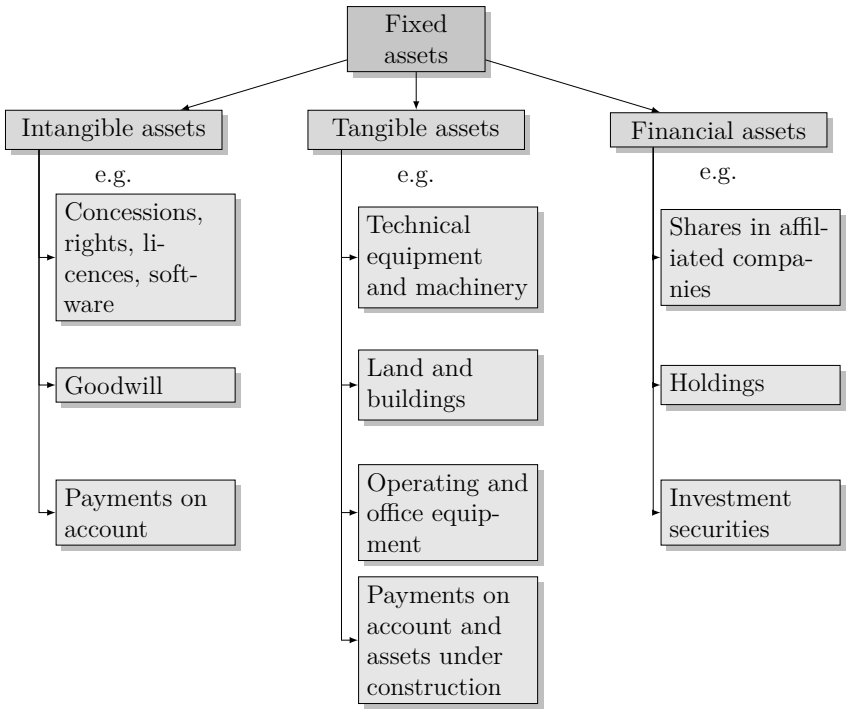
office equipment), we combine matching with an event study approach to flexibly compare measures of the capital structure and technology in incumbent firms covered by a minimum wage and suitable control firms. Our results indicate sizable investments in capital deepening among treated incumbent firms in industries with a high routine task share. We see a similar development among entrants in those industries. In contrast, we see no additional capital investments in labor-intensive industries with a low routine task share. In addition, we observe suggestive evidence that incumbent firms rely more on pre-fabricated inputs, but also a sizable pass through of the additional costs to consumers.

Appendix

3.A Capital Assets in German Balance Sheets

Figure 3.A.1 gives an overview of the structure of capital assets in German balance sheets. Accordingly, fixed assets is the highest and most broadest category regarding firms' capital assets and includes intangible assets, tangible assets and financial assets. Tangible assets include land and buildings, technical equipment and machinery, operating and office equipment as well as payments on account and assets under construction. As the number of missings increases, the more specific the balance sheet items are, we focus on tangible assets in our analyses.

Figure 3.A.1: Structure of Capital Assets in German Balance Sheets



3.B Data cleaning

The raw Dafne database consists of a base file that is continuously updated which contains for each firm up to ten years history on financial information and historical updates. We use an excerpt from Dafne 2018 and merge all available historical updates (2004-2017) using an internal firm identifier using observations from 1994 onwards. We further use information on firm ID changes provided by Bureau van Dijk. In cases where we have several variables containing information for a given firm and year, we use the most recent information. If the most recent variable turns out to be missing, we fill the variable with the most recent information available. We create an annual panel for the time period 2005-2014. We only use information on unconsolidated accounts. Balance sheet information is reported for a specific date. We therefore assign accounting information which is reported by the firm before (after) March 31 to the previous (current) calendar year. Finally, the raw data require a fair amount of data cleaning. In particular, we set an observation to missing if the values lie outside plausible ranges such as impossible negative values or implausible jumps. For the capital variables fixed assets and tangible assets we put a value to missing if in a specific year a firm changes its relative place in the year specific distribution of this variable by more than 50 percentiles compared to the mean. We regard implausible negative entries from the following variables as missing information: fixed assets, number of employees. balance sheet total, total wage bill, tangible assets, total assets, total liabilities, current assets, total revenues, inventory. We set the founding year of the firm to missing if it is prior to 1800. We compute a proxy for average wage by dividing the total wage bill by the number of employees. We observe the industry on 5-digit level. We transform the industry variables for all firms to the coding WZ 2008, this means that we transfer for data from the historical updates 2004-2007 the coding WZ 2003 to WZ 2008. For the recoding we use a unique walkover that recodes the WZ 2003 code into the corresponding WZ 2008 mode (Bersch et al., 2014). To avoid wrong industry coding we merge the list of 5-digit WZ 2008 codes to the dataset. Some observations report a 5-digit industry code that is not present in the official WZ 2008 5-digit coding but represents a 3 or 4 digit coding followed by zeros. We interpret these as correct 3 or 4 digit codes. If within this 3 or 4 digit sector all 5-digit industries belong to a minimum wage industry we assign the firms to this minimum wage industry. If the 3 or 4 digit sector contains both minimum wage and non minimum wage industries we exclude these firms as controls in the matching step. As described in the main text we drop some industries from our

sample. These correspond to the WZ 2008 codes 01-03, 64-68, 84-86 and 90-94. We differentiate between firms in East Germany (including Berlin) and West Germany (excluding Berlin) by using the municipality codes provided by Dafne. As limited liability companies and limited partnerships we define firm-year observations with legal status limited liability firms ('Gesellschaft mit beschränkter Haftung', "GmbH" or 'GmbH & Co. KG'). To distinguish firms in urban/rural districts we use the classification from the Federal Office for Building and Regional Planning. Dafne does not contain information on firm exit/closure. We drop firm-year observations were both the reported year is after the last year with time-varying information and the observed year is after the estimated firm exit as estimated by the MUP (Bersch et al., 2014). Before log transforming variables we add +1 to the original variable. When defining the incumbent samples we use firms (treated and controls) that have non-missing information for all years in the following key characteristics: log balance sheet total, log tangible assets, log liabilities, equity ratio, firm age (founding year), number of employees.

3.C Computation

Computations are realized in Stata including the user written commands did_imputation (Borusyak et al., 2021), reghdfe (Correia, 2017) and coefplot (Jann, 2014).

3.D Additional Results

Table 3.D.1: Small Industries in East Germany

Industry	MW Year (1)	Firms (per year) (2)	Years (3)
Building Cleaning	2007	82	2005-2011
Hairdressing	2013	80	2010-2014
Security Services	2011	88	2008-2014
Stone Masonry and Stone Carving	2013	89	2010-2014
Laundry Services	2009	33	2006-2013

Notes: The table shows the year of the adoption of a minimum wage (in column (1)), the number of incumbent firms in East Germany with non-missing information on the key variables in the Dafne Dataset (in column (2)) and the time period for which we have information on the firms (in column (3)). All industries are analysed for a period three years prior to the minimum wage until 4 years after the minimum wage. If this lies outside the sample period 2005-2014, a smaller time window is used. Data source: Dafne.

Table 3.D.2: Covariate Balance for Robustness Checks

	Waste: All Industries as controls				Waste: West Germany			
	Control -Treated		All Firms		Control - Treated		All Firms	
Variables	Diff	Std	Diff	Std	Diff	Std	Diff	Std
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Balance Sheet Total	-0.004	0.096	-0.518***	0.073	-0.054	0.059	-1.060***	0.037
Log Liabilities	-0.001	0.098	-0.424***	0.078				
Fixed Assets\	-0.011	0.017	-0.169***	0.038	-0.012	0.011	-0.245***	0.007
Balance Sheet Total								
Firm Age	0.029	0.384	0.422	0.556				
Log Firm Age					-0.001	0.030	0.293***	0.025
Investment Intensity					-0.109	0.426	-1.318***	0.224
Equity Ratio					-0.005	0.009	-0.049***	0.006
Log Tangible Assets pEmp	-0.118	0.112	-1.052***	0.089	0.015	0.085	-1.752***	0.052
Log Employees	0.048	0.071	-0.184***	0.055	-0.257***	0.043	-0.399***	0.026
Log Tangible Assets	-0.062	0.135	-1.266***	0.110	-0.264***	0.091	-2.177***	0.056
Equity Ratio	0.022	0.015	-0.002	0.011				
Log Firm Age	0.002	0.032	-0.076***	0.030				
Firm Age					-0.465	0.626	8.112***	0.702
Log Liabilities					-0.076	0.064	-0.941***	0.042
Log Revenues	0.042	0.105	-0.130	0.082	-0.388***	0.059	-0.859***	0.036
Log Current Assets	-0.008	0.094	-0.194***	0.071	-0.003	0.056	-0.610***	0.038
Investment Intensity	1.703	1.304	-0.003	2.940				
Log Wage Sum	0.246	0.158	0.042	0.122	-0.961***	0.162	-0.879***	0.082
Log Wage	0.091	0.061	0.067	0.045	-0.228**	0.096	-0.046	0.035

Notes: Dafne Dataset, Incumbent Samples, all values computed for the matching year 2008, two years prior treatment; The upper rows display the matching variables. The bottom rows non-matched variables. Results from two sample t-test with the null hypotheses of zero differences. Differences: Control - Treated. All Firms: all controls that can be selected as nearest neighbours in the matching step and treated; Control - Treated: selected nearest neighbour control firms and treated firms.

Table 3.D.3: Capital Intensity (Waste Management) with Alternative Set of Control Firms

	Log Capital per Employee (1)	Log Employees (2)	Log Capital (3)
3 years before	-0.046	0.047*	-0.007
	(0.057)	(0.027)	(0.056)
2 years before	-0.028	0.031*	-0.023
	(0.040)	(0.019)	(0.037)
0 years after	0.019	-0.012	-0.010
	(0.037)	(0.018)	(0.034)
1 year after	0.074	-0.038	0.021
	(0.046)	(0.026)	(0.039)
2 years after	0.143***	-0.051*	0.064
	(0.054)	(0.030)	(0.046)
3 years after	0.187***	-0.029	0.135**
	(0.060)	(0.034)	(0.055)
4 years after	0.191***	-0.015	0.156***
	(0.061)	(0.037)	(0.058)
Wild Bootstrap CI	[0.067, 0.045]	[-0.086, 0.065]	[0.025, 0.283]
R^2	0.01	0.04	0.03
N	7,648	7,648	7,648

Notes: Results from equation 3.1 where the dependent variable is log capital intensity (log capital per employee). Sample includes incumbent firms in the treated industry waste management in East Germany and matched control firms. All industries are allowed as controls, except for industries that introduce a minimum wage before 2015. The specification includes period indicators relative to the treatment year and firm fixed effects. Standard errors clustered at the firm level are shown in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Wild bootstrap CI denotes 95% confidence intervals for the coefficient on 4 years after treatment for standard errors clustered at the 5-digit industry level. Data source: Dafne.

Table 3.D.4: Minimum Wages and Capital Intensity among Incumbent Firms in West Germany

	Log Capital per Employee (1)	Log Employees (2)	Log Capital (3)
3 years before	0.039	-0.021	0.036
	(0.046)	(0.018)	(0.044)
2 years before	0.017	-0.009	0.015
	(0.032)	(0.016)	(0.025)
0 years after	0.017	0.033**	0.075***
	(0.029)	(0.013)	(0.027)
1 year after	0.041	0.044***	0.102***
	(0.041)	(0.017)	(0.038)
2 years after	0.069	0.042**	0.128***
	(0.047)	(0.020)	(0.044)
3 years after	0.113**	0.049**	0.179***
	(0.051)	(0.021)	(0.048)
4 years after	0.118**	0.044*	0.179***
	(0.054)	(0.024)	(0.051)
Wild Bootstrap CI	[-0.028, 0.236]	[-0.007, 0.100]	[0.059, 0.291]
R^2	0.01	0.06	0.04
N	20,496	20,496	20,496

Notes: Results from equation 3.1 where the dependent variable is log capital intensity (log capital per employee). Sample includes incumbent firms in the treated industry waste management in West Germany and matched control firms. The specification includes period indicators relative to the treatment year and firm fixed effects. Standard errors clustered at the firm level are shown in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Wild bootstrap CI denotes 95% confidence intervals for the coefficient on 4 years after treatment for standard errors clustered at the 5-digit industry level. Data source: Dafne.

Table 3.D.5: Descriptives on Firm Entry in East German Sample

	No. of Entrants before Treatment (1)	No. of Entrants after Treatment (2)
Building Cleaning	80	247
Hairdressing	111	3
Laundry Services	28	20
Caring	549	329
Scaffolding	83	3
Security Services	99	34
Stone Masonry	53	2
Waste Management	270	79

Notes: Number of firms in East Germany with founding years before MW introduction (column (1)) and in the year of or after MW introduction (column (2)) in the specific industry. Sample restriction: limited liability firms and limited partnerships. Temporary work and vocational and further training excluded. Zero observations in the mining industry. Data source: Dafne.

Table 3.D.6: Capital Intensity among Firm Entrants (All Legal Forms)

	Pooled	Pooled	Pooled	Pooled	Waste	Waste	Caring	Caring
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post Found	0.172	0.162	0.246*	0.179	0.069	0.255	-0.041	-0.229
	(0.129)	(0.130)	(0.132)	(0.147)	(0.326)	(0.359)	(0.171)	(0.199)
New Found		0.107			-0.355	-0.281	0.530**	0.409**
		(0.153)			(0.382)	(0.363)	(0.228)	(0.202)
Firm Age			0.011	-0.005		0.033		-0.048
			(0.016)	(0.022)		(0.057)		(0.036)
Post				0.148		0.503*		0.227
				(0.111)		(0.265)		(0.182)
WB CI	[-0.770,0.191]	[-0.767,0.178]	[-0.592,0.235]	[-0.596,0.036]				
R^2	0.09	0.09	0.09	0.09	0.03	0.02	0.01	0.00
N	7,038	7,038	7,038	7,038	1,217	1,217	3,178	3,178
Industry FE	yes	yes	yes	yes	no	no	no	no
Year FE	yes	yes	no	no	yes	no	yes	no

Notes: Results from estimating equation (3.6.1) where we include entrants of all legal forms. New Found equals one in the founding year, and zero otherwise. Post Found equals one if the founding year was in the same year or after the minimum wage introduction, and zero for the pre-adoption years. Post equals one if the year of observation is equal to or after the adoption of a minimum wage. The sample includes firms of all legal forms from the Dafne database entering one of the treated industries in East Germany between 2005 and 2014. Standard errors are clustered at the firm level. We also report confidence intervals (0.95) from a wild bootstrap procedure clustered at the 5-digit industry level. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Bibliography

Aaronson, D. (2001). Price pass-through and the minimum wage. *Review of Economics and Statistics*, 83(1):158–169.

Aaronson, D. and French, E. (2007). Product market evidence on the employment effects of the minimum wage. *Journal of Labor Economics*, 25(1):167–200.

Aaronson, D., French, E., and MacDonald, J. (2008). The minimum wage, restaurant prices, and labor market structure. *Journal of Human Resources*, 43(3):688–720.

Aaronson, D., French, E., Sorkin, I., and To, T. (2018). Industry dynamics and the minimum wage: A putty-clay approach. *International Economic Review*, 59(1):51–84.

Aaronson, D. and Phelan, B. J. (2019). Wage shocks and the technological substitution of low-wage jobs. *Economic Journal*, 129(617):1–34.

Adda, J. (2016). Economic activity and the spread of viral diseases: Evidence from high frequency data. *Quarterly Journal of Economics*, 131(2):891–941.

Aggarwal, C. C. and Reddy, C. K., editors (2014). *Data clustering. algorithms and applications*. CRC Press.

Aghabozorgi, S., Shirkhorshidi, A. S., and Wah, T. Y. (2015). Time-series clustering – a decade review. *Information Systems*, 53:16 – 38.

Amt für Statistik Berlin Brandenburg (2020). Statistischer Bericht A I 5 - hj 2 / 19 Einwohnerinnen und Einwohner im Land Berlin am 31. Dezember 2019; Grunddaten; 3. korrigierte Ausgabe.

Arntz, M., Gregory, T., and Zierahn, U. (2017). Revisiting the risk of automation. *Economics Letters*, 159:157–160.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 135(2):645–709.

Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, 29(3):3–30.

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, 118(4):1279–1333.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., and Wong, A. (2018). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3):259–280.

Bartscher, A., Seitz, S., Siegloch, S., Slotwinski, M., and Wehrhöfer, N. (2021). Social capital and the spread of Covid-19: Insights from European countries. *Journal of Health Economics*, 80:102531.

Bauer, A. and Weber, E. (2021). COVID-19: how much unemployment was caused by the shutdown in Germany? *Applied Economics Letters*, 28(12):1053–1058.

BBSR Bonn (2021). INKAR Datenbank. www.inkar.de.

Bell, B. and Machin, S. (2018). Minimum wages and firm value. *Journal of Labor Economics*, 36(1):159–195.

Berger, M. and Tutz, G. (2018). Tree-structured clustering in fixed effects models. *Journal of Computational and Graphical Statistics*, 27(2):380–392.

Berkessel, J., Ebert, T., Gebauer, J., Johnsson, T., and Oishi, S. (2021). Pandemics initially spread among people of higher (not lower) social status: Evidence from COVID-19 and the Spanish flu. *Social Psychological and Personality Science*, Advance online publication.

Bersch, J., Gottschalk, S., Müller, B., and Niefert, M. (2014). The Mannheim Enterprise Panel (MUP) and firm statistics for Germany. *ZEW Discussion Paper*, 14-104.

Bluhm, R. and Pinkovskiy, M. (2021). The spread of COVID-19 and the BCG vaccine: A natural experiment in reunified Germany. *The Econometrics Journal*, 24(3):353–376.

Bondell, H. D., Krishna, A., and Ghosh, S. K. (2010). Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics*, 66(4):1069–1077.

Bonhomme, S., Lamadon, T., and Manresa, E. (2022). Discretizing unobserved heterogeneity. *Econometrica*, 90(2):625–643.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. arXiv preprint arXiv:2108.12419.

Bossler, M. and Gerner, H.-D. (2020). Employment effects of the new german minimum wage: Evidence from establishment-level microdata. *ILR Review*, 73(5):1070–1094.

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.

Brochu, P. and Green, D. (2013). The impact of minimum wages on labor market transitions. *Economic Journal*, 123(573):1203–1235.

Brodeur, A., Gray, D., Islam, A., and Bhuiyan, S. (2021). A literature review of the economics of COVID-19. *Journal of Economic Surveys*, 35(4):1007–1044.

Brüll, E. and Gathmann, C. (2020). Evolution of the East German wage structure. ZEW Discussion Paper 20-081, ZEW – Leibniz Centre for European Economic Research.

Brynjolfsson, E. and McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. *WW Norton & Company*.

Bundesministeriums für Verkehr und digitale Infrastruktur (BMVI). Berlin (2014). Prognose der deutschlandweiten Verkehrsverflechtungen 2030.

Caliendo, M., Fedorets, A., Preuss, M., Schröder, C., and Wittbrodt, L. (2018). The short-run employment effects of the German minimum wage reform. *Labour Economics*, 53:46–62.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249.

Campello, R. J. G. B., Kröger, P., Sander, J., and Zimek, A. (2020). Density-based clustering. *WIREs Data Mining and Knowledge Discovery*, 10(2):e1343.

Campello, R. J. G. B., Moulavi, D., and Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In Pei, J., Tseng, V., Cao, L., Motoda, H., and Xu, G., editors, *Advances in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science*, volume 7819, pages 160–172. Springer.

Campello, R. J. G. B., Moulavi, D., Zimek, A., and Sander, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 10(1):1–51.

Card, D. and Krueger, A. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4):772–793.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The Effect of Minimum Wages on Low-Wage Jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454.

Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., and Leskovec, J. (2021). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 589:82–87.

Chatterjee, A. and Lahiri, S. N. (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association*, 106(494):608–625.

Chatterjee, A. and Lahiri, S. N. (2013). Rates of convergence of the adaptive lasso estimators to the oracle distribution and higher order refinements by the bootstrap. *The Annals of Statistics*, 41(3):1232–1259.

Chiquet, J., Rigaill, G., Sundqvist, M., and Dervieux, V. (2020). *aricode: Efficient computations of standard clustering comparison measures*. R package.

Clemens, J. (2021). How do firms respond to minimum wage increases? understanding the relevance of non-employment margins. *Journal of Economic Perspectives*, 35(1):51–72.

Correia, S. (2017). reghdfe: Stata module for linear and instrumental-variable/GMM regression absorbing multiple levels of fixed effects. Statistical software components s457874.

Croissant, Y. and Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, 27(2):1–43.

Cuadrado, J. L. (2020). *VeryLargeIntegers: Store and operate with arbitrarily large integers*. R package.

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., and Swinton, J. (2019). *xtable: Export tables to LaTeX or HTML*. R package.

Dai, X. and Qiu, Y. (2022). Minimum wage hikes and capital deepening: Evidence from U.S. establishments. Working Paper.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

De Ridder, D., Sandoval, J., Vuilleumier, N., Azman, A. S., Stringhini, S., Kaiser, L., Joost, S., and Guessous, I. (2021). Socioeconomically disadvantaged neighborhoods face increased persistence of SARS-CoV-2 clusters. *Frontiers in Public Health*, 8.

Decoster, A., Minten, T., and Spinnewijn, J. (2021). The income gradient in mortality during the Covid-19 crisis: Evidence from Belgium. *The Journal of Economic Inequality*, 19(3):551–570.

Dengler, K., Matthes, B., and Paulus, W. (2014). Berufliche tasks auf dem deutschen arbeitsmarkt. *Eine alternative Messung auf Basis einer Expertendatenbank. FDZ-Methodenreport*, 12:2014.

Deriso, D. and Boyd, S. (2019). A general optimization framework for dynamic time warping. Working Paper arXiv:1905.12893, arXiv.

Doblhammer, G., Kreft, D., and Reinke, C. (2021). Regional characteristics of the second wave of SARS-CoV-2 infections and COVID-19 deaths in Germany. *International Journal of Environmental Research and Public Health*, 18(20):10663.

Doblhammer, G., Reinke, C., and Kreft, D. (2022). Social disparities in the first wave of COVID-19 infections in Germany: A county-scale explainable machine learning approach. *BMJ Open*, 12:e049852.

Draca, M., Machin, S., and van Reenen, J. (2011). Minimum wages and firm profitability. *American Economic Journal: Applied Economics*, 3(1):129–151.

Dragano, N., Hoebel, J., Wachtler, B., Diercke, M., Lunau, T., and Wahrendorf, M. (2021). Soziale Ungleichheit in der regionalen Ausbreitung von SARS-CoV-2. *Bundesgesundheitsblatt*, 64:1116–1124.

Dube, A., Lester, T. W., and Reich, M. (2010). Minimum wage effects across state borders: Estimates using contiguous counties. *Review of Economics and Statistics*, 92(4):945–964.

Dube, Arindrajit, T. W. L. and Reich, M. (2016). Minimum wage shocks, employment flows and labor market frictions. *Journal of Labor Economics*, 34(3):663–704.

Dustmann, C., Lindner, A., Schönberg, U., Umkehrer, M., and vom Berge, P. (2022). Reallocation effects of the minimum wage. *Quarterly Journal of Economics*, 137(1):267–328.

Ehlert, A. (2021). The socio-economic determinants of covid-19: A spatial analysis of German county level data. *Socio Economic Planning Sciences*, 78:101083.

Ester, M. (2014). Density-based clustering. In Aggarwal, C. C. and Reddy, C. K., editors, *Data Clustering. Algorithms and Applications*, pages 111–124. CRC Press.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, page 226–231. AAAI Press.

Facebook Data for Good Program (2021). Social Connectedness Index (SCI). Dataset. https://dataforgood.facebook.com.

Fan, Y. and Li, R. (2012). Variable selection in linear mixed effects models. *The Annals of Statistics*, 40(4):2043–2068.

Feller, W. (1971). *An introduction to probability theory and its applications*. Wiley.

Flinn, C. J. (2011). *The minimum wage and labor market outcomes*. MIT Press.

Fox, J. and Weisberg, S. (2019). *An R companion to applied regression*. Sage, third edition.

Franses, P. H. and Wiemann, T. (2020). Intertemporal similarity of economic time series: An application of dynamic time warping. *Computational Economics*, 56:59–75.

Frühwirth-Schnatter, S., Pittner, S., Weber, A., and Winter-Ebmer, R. (2018). Analysing plant closure effects using time-varying mixture-of-experts Markov chain clustering. *Annals of Applied Statistics*, 12(3):1796–1830.

Frühwirth-Schnatter, S., Pamminger, C., Weber, A., and Winter-Ebmer, R. (2016). Mothers' long-run career patterns after first birth. *Royal Statistical Society Series A*, 179(3):707–725.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.

Galili, T. (2015). dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics*, 31(22):3718–3720.

Ganserer, A., Gregory, T., Murmann, S., and Zierahn, U. (2022). Minimum wages and solo self-employment. Working Paper.

Ganzer, A., Schmucker, A., vom Berge, P., and Wurdack, A. (2017). Sample of integrated labour market biographies - regional file 1975-2014 : (siab-r 7514). *FDZ Datenreport. Documentation on Labour Market Data 201701 en.*

GeoBasis-DE / BKG (2021a). Bundesländergrenzen 2019.

GeoBasis-DE / BKG (2021b). Kreisgrenzen 2019.

Giorgino, T. (2009). Computing and visualizing dynamic time warping alignments in R: The dtw package. *Journal of Statistical Software*, 31(7):1–24.

Goujon, A., Natale, F., Ghio, D., and Conte, A. (2021). Demographic and territorial characteristics of COVID-19 cases and excess mortality in the European Union during the first wave. *Journal of Population Research*. Advance online publication.

Gregory, T. and Zierahn, U. (2022). When the minimum wage really bites hard: The negative spillover effect on high-skilled workers. *Journal of Public Economics*, 206:104582.

Gustafson, M. and Kotter, J. D. (2022). Higher minimum wages reduce capital expenditures. Working Paper.

Hahn, J. and Moon, H. R. (2010). Panel data models with finite number of multiple equilibria. *Econometric Theory*, 26(3):863–881.

Hahsler, M., Piekenbrock, M., and Doran, D. (2019). dbscan: Fast density-based clustering with R. *Journal of Statistical Software*, 91(1):1–30.

Harasztosi, P. and Lindner, A. (2019). Who pays for the minimum wage? *American Economic Review*, 109(8):2693–2727.

Harris, J. E. (2020). The subways seeded the massive coronavirus epidemic in New York City. NBER Working Paper 27021, National Bureau of Economic Research.

Hastie, T., Tibshirani, R., and Friedman, J. (2017). *The elements of statistical learning : data mining, inference, and prediction*. Springer, second edition.

Hau, H., Huang, Y., and Wang, G. (2020). Firm response to competitive shocks: Evidence from China's minimum wage policy. *The Review of Economic Studies*, 87(6):2639–2671.

Heiler, P. and Mareckova, J. (2021). Shrinkage for categorical regressors. *Journal of Econometrics*, 223(1):161–189.

Heinzl, F. and Tutz, G. (2014). Clustering in linear-mixed models with a group fused lasso penalty. *Biometrical Journal*, 56(1):44–68.

Hlavac, M. (2018). *stargazer: Well-formatted regression and summary statistics tables*. R package.

Ibragimov, I. (1956). On the composition of unimodal distributions. *Theory of Probability and its Applications*, 1(2):255–260.

Itakura, F. (1975). Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 23(1):67–72.

Jann, B. (2014). Plotting regression coefficients and other estimates. *The Stata Journal*, 14(4):708–737.

Jäger, S., Schoefer, B., and Heining, J. (2020). Labor in the Boardroom. *The Quarterly Journal of Economics*, 136(2):669–725.

Jo, Y., Hong, A., and Sung, H. (2021). Density or connectivity: What are the main causes of the spatial proliferation of COVID-19 in Korea? *International Journal of Environmental Research and Public Health*, 18(10):5084.

Kennan, J. (1995). The elusive effects of minimum wages. *Journal of Economic Literature*, 33(4):1950–1965.

Keogh, E. and Pazzani, M. (2001). Derivative dynamic time warping. In *Proceedings of the 2001 SIAM International Conference on Data Mining (SDM)*, pages 1–11, Chicago, USA.

Keogh, E. and Ratanamahatana, C. (2005). Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7(3):358–386.

Knittel, C. R. and Ozaltun, B. (2020). What does and does not correlate with COVID-19 death rates. NBER Working Paper 27391, National Bureau of Economic Research.

Kotsakos, D., Trajcevski, G., Gunopulos, D., and Aggarwal, A. G. (2014). Time-series data clustering. In Aggarwal, C. C. and Reddy, C. K., editors, *Data Clustering. Algorithms and Applications*, pages 357–379. CRC Press.

Kuchler, T., Russel, D., and Stroebel, J. (2021). JUE Insight: The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook. *Journal of Urban Economics*, 127:103314.

Lai, R. (2020). *arrangements: Fast generators and iterators for permutations, combinations, integer partitions and compositions*. R package.

Li, Y., Wang, S., Song, P. X.-K., Wang, N., Zhou, L., and Zhu, J. (2018). Doubly regularized estimation and selection in linear mixed-effects models for high-dimensional longitudinal data. *Statistics and Its Interface*, 11(4):721–737.

Liao, T. W. (2005). Clustering of time series data – a survey. *Pattern Recognition*, 38(11):1857–1874.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137.

Lordan, G. and Neumark, D. (2018). People versus machines: The impact of minimum wages on automatable jobs. *Labour Economics*, 52:40–53.

Luca, D. L. and Luca, M. (2019). Survival of the fittest: The impact of the minimum wage on firm exit. Working Paper 25806, National Bureau of Economic Research.

Lumley, T. (2020). *biglm: Bounded memory linear and generalized linear models*. R package.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Le Cam, L. M. and Neyman, J., editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 5, pages 281–297. University of California Press, Berkeley.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2021). *cluster: Cluster analysis basics and extensions*. R package.

Manning, A. (2013). *Monopsony in motion: Imperfect competition in labor markets*. Princeton University Press.

Manning, A. (2021). The elusive employment effect of the minimum wage. *Journal of Economic Perspectives*, 35(1):3–26.

Mastroeni, L., Mazzoccoli, A., Quaresima, G., and Vellucci, P. (2021). Decoupling and recoupling in the crude oil price benchmarks: An investigation of similarity patterns. *Energy Economics*, 94:105036.

Mayneris, F., Poncet, S., and Zhang, T. (2018). Improving or disappearing: Firm-level adjustments to minimum wages in China. *Journal of Development Economics*, 135:20–42.

Meer, J. and West, J. (2016). Effects of the minimum wage on employment dynamics. *Journal of Human Resources*, 51(2):500–522.

Müller, M. (2007). *Information Retrieval for Music and Motion*, chapter Dynamic Time Warping, pages 69–84. Springer.

Mogi, R. and Spijker, J. (2021). The influence of social and economic ties to the spread of COVID-19 in Europe. *Journal of Population Research*, Advance online publication.

Mueen, A. and Keogh, E. J. (2016). Extracting optimal performance from dynamic time warping. In *KDD'16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2129–2130, San Francisco, USA.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1):69–85.

Neumark, D., Salas, J. M. I., and Wascher, W. (2014). Revisiting the minimum wage—employment debate: Throwing out the baby with the bathwater?. *Industrial and Labor Relations Review*, 67(s3):608–648.

Neumark, D. and Wascher, W. (2001). Minimum wages and training revisited. *Journal of Labor Economics*, 19(3):563–595.

Neumark, D. and Wascher, W. (2008). *Minimum wages*. MIT Press.

Oster, E. (2012). Routes of infection: Exports and HIV incidence in Sub-Saharan Africa. *Journal of the European Economic Association*, 10(5):1025–1058.

Papageorge, N. W., Zahn, M. V., Belot, M., van den Broek-Atlenburg, E., Choi, S., Jamison, J. C., and Tripodi, E. (2021). Socio-demographic factors associated with self-protecting behavior during the covid-19 pandemic. *Journal of Population Economics*, 34(2):691–738.

Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. *The R Journal*, 10(1):439–446.

Portugal, P. and Cardoso, A. R. (2006). Disentangling the minimum wage puzzle: An analysis of worker accessions and separations. *Journal of the European Economic Association*, 4(5):988–1013.

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raihan, T. (2017). Predicting US recessions: A dynamic time warping exercise in economics. Working Paper.

Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J., and Keogh, E. (2012). Searching and mining trillions of time series subsequences under dynamic time warping. In *18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 12*, Beijing, China.

Reddy, C. K. and Vinzamuri, B. (2014). A survey of partitional and hierarchical clustering algorithms. In Aggarwal, C. C. and Reddy, C. K., editors, *Data Clustering. Algorithms and Applications*, pages 87–107. CRC Press.

Robert Koch Institut (2021a). COVID-19 Fallzahlen in Deutschland. `https://www.arcgis.com/home/item.html?id=dd4580c810204019a7b8eb3e0b329dd6`, retrieved September 17, 2021.

Robert Koch Institut (2021b). Täglicher Lagebericht des RKI zur Coronavirus-Krankheit-2019 (COVID-19) 30.06.2021 – Aktualisierter Stand für Deutschland.

Robitzsch, A., Grund, S., and Henke, T. (2020). *miceadds: Some additional multiple imputation functions, especially for 'mice'*. R package.

Rohart, F., San Cristobal, M., and Laurent, B. (2014). Selection of fixed effects in high dimensional linear mixed models using a multicycle ECM algorithm. *Computational Statistics and Data Analysis*, 80(C):209–222.

Rojas-Valenzuela, I., Valenzuela, O., Delgado-Marquez, E., and Rojas, F. (2021). Estimation of COVID-19 dynamics in the different states of the United States during the first months of the pandemic. *Engineering Proceedings*, 5(53):1–9.

Rudis, B. (2020). *hrbrthemes: Additional themes, theme components and utilities for 'ggplot2'*. R package.

Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49.

Schelldorfer, J., Bühlmann, P., and De Geer, S. V. (2011). Estimation for high-dimensional linear mixed-effects models using ℓ_1-penalization. *Scandinavian Journal of Statistics*, 38(2):197–214.

Schilling, J., Buda, S., Fischer, M., Goerlitz, L., Grote, U., Haas, W., Hamouda, O., Prahm, K., and Tolksdorf, K. (2021a). Retrospektive Phaseneinteilung der COVID-19-Pandemie in Deutschland bis Februar 2021. *Epidemiologisches Bulletin*, 15:8–17.

Schilling, J., Tolksdorf, K., Marquis, A., Faber, M., Pfoch, T., Buda, S., Haas, W., Schuler, E., Altmann, D., Grote, U., Diercke, M., and RKI COVID-19 Study Group (2021b). Die verschiedenen Phasen der COVID-19-Pandemie

in Deutschland: Eine deskriptive Analyse von Januar 2020 bis Februar 2021. *Bundesgesundheitsblatt*, 64:1093–1106.

Schmitt, J. (2015). Explaining the small employment effects of the minimum wage in the united states. *Industrial Relations*, 54(4):547–581.

Schuppert, A., Polotzek, K., Schmitt, J., Busse, R., Karschau, J., and Karagiannidis, C. (2021). Different spreading dynamics throughout Germany during the second wave of the COVID-19 pandemic: a time series study based on national surveillance data. *The Lancet Regional Health Europe*, 6:100151.

Statistik der Bundesagentur für Arbeit (2020). Tabellen, Pendlerverflechtungen der sozialversicherungspflichtig Beschäftigten nach Kreisen, Stichtag 30. Juni 2019.

Statistisches Bundesamt (2016). 4 Millionen Jobs vom Mindestlohn betroffen. Press Release, no. 121, April 6, 2016.

Statistisches Bundesamt (2020). Bevölkerung: Kreise, Stichtag 31.12.2019.

Stübinger, J. and Schneider, L. (2020). Epidemiology of coronavirus COVID-19: Forecasting the future incidence in different countries. *Healthcare*, 8(2):99.

Su, L., Shi, Z., and Phillips, P. C. B. (2016). Identifying latent structures in panel data. *Econometrica*, 84(6):2215–2264.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199. Themed Issue: Treatment Effect 1.

Sutch, R. (2011). The unexpected long-run impact of the minimum wage: an educational cascade. In Rhode, P. W., Rosenbloom, J. L., and Weiman, D. F., editors, *Economic Evolution and Revolution in Historical Time*, pages 387–418. Stanford University Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Tibshirani, R. J. and Taylor, J. (2011). The solution path of the generalized lasso. *Annals of Statistics*, 39(3):1335–1371.

Tutz, G. and Oelker, M. (2017). Modelling clustered heterogeneity: Fixed effects, random effects and mixtures. *International Statistical Review*, 85(2):204–227.

Tutz, G. and Schauberger, G. (2015). Extended ordered paired comparison models with application to football data from German Bundesliga. *AStA Advances in Statistical Analysis*, 99(2):209–227.

von Bismarck-Osten, C., Borusyak, K., and Schönberg, U. (2022). The role of schools in the transmission of SARS-CoV2 virus: Quasi-experimental evidence from Germany. *Economic Policy*, eiac001.

Wang, G.-J., Xie, C., Han, F., and Sun, B. (2012). Similarity measure and topology evolution of foreign exchange markets using dynamic time warping method: Evidence from minimal spanning tree. *Physica A: Statistical Mechanics and its Applications*, 391(16):4136–4146.

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1):1–29.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

Wickham, H. and Bryan, J. (2019). *readxl: Read Excel files*. R package.

Wickham, H., François, R., Henry, L., and Müller, K. (2021). *dplyr: A grammar of data manipulation*. R package.

Wickham, H. and Miller, E. (2021). *haven: Import and export 'SPSS', 'Stata' and 'SAS' files*. R package.

Wilke, C. O. (2020). *cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*. R package.

Wooldridge, J. M., editor (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, second edition.

Wooldridge, J. M. (2019). Correlated random effects models with unbalanced panels. *Journal of Econometrics*, 211(1):137–150.

Wright, M. N. and Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77(1):1–17.

Yang, Y., Zou, H., and Bhatnagar, S. (2020). *gglasso: Group Lasso penalized learning using a unified BMD algorithm*. R package.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B - Statistical Methodology*, 68(1):49–67.

Zapp, K. (2017). Minimum wages and automation: An empirical analysis for Germany. Dissertation Proposal, Universität Mannheim. Recognized as Master Thesis.

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, 11(10):1–17.

Zeileis, A. and Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, 2(3):7–10.

Zeileis, A., Köll, S., and Graham, N. (2020). Various versatile variances: An object-oriented implementation of clustered covariances in R. *Journal of Statistical Software*, 95(1):1–36.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B - Statistical Methodology)*, 67(2):301–320.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die Arbeit selbständig angefertigt und die benutzten Hilfsmittel vollständig und deutlich angegeben habe.

Mannheim, 04.04.2022

Kristina Zapp

Curriculum Vitae

Education

2015-2022 (expected)	**Ph.D. in Economics** University of Mannheim Graduate School for Economics and Social Sciences Parental leave 09/2020-05/2021
05/2019-06/2019	**Research Stay** Copenhagen Business School Department of Economics
2018	**M.Sc. Economics** University of Mannheim
2012-2015	**B.Sc. Mathematics** Heidelberg University
2009-2012	**B.Sc. Economics** University of Mannheim
08/2011-01/2012	Exchange Semester, Tilburg University

Work Experience

Since 2017	**Researcher** ZEW – Leibniz Centre for European Economic Research Department Labour Markets and Human Resources Parental leave 09/2020-05/2021