ERASMUS UNIVERSITY ROTTERDAM

ECONOMETRICS AND OPERATIONS RESEARCH

BACHELOR THESIS

# To pool or not to pool

## Possible heterogeneous slopes in a functional coefficient model

MAREK OERLEMANS
453890

SUPERVISED BY DR. W. WANG

SECOND ASSESSOR: DR. J.W.N. REUVERS

This paper investigates the use of different estimators when dealing with possible heterogeneous slope coefficients among individuals in a panel data setting. We consider both parametric estimators and non-parametric time-variant estimators. The main estimator investigated is an averaging estimator proposed in Wang, Zhang, and Paap (2019). We review the performance of this estimator and extend it to a time-variant coefficient estimator. Results show the estimator performing particularly well with a group structure among individual slopes and performing quite robust when coefficients are completely homogeneous or heterogeneous.



July 7, 2019

# Contents

# 1. Introduction

When estimating coefficients in time series or in regressions we find that the variance of the used estimators might lead to misspecified parameters. A solution to this problem might lie in panel regression. Here we have not only a time dimension on the data, but we also have multiple individuals in the data. In contrast to regular regression and time series where only one individual is used. By combining the data on multiple individuals we are able to reduce the variance of the coefficient estimates. Although, this does require more restrictive assumptions which, when falsely assumed, can cause bias. We find ourselves with a very familiar problem: the bias-variance trade-off. At one of the ends of the spectrum with minimum variance, we find the pooled estimator, which assumes all slope coefficients are equal among individuals. On the other end of the spectrum we find the classical individual estimator, which is unbiased but has a higher variance. In literature this problem is dealt with repeatedly, but a fix-it-all solution has not been found. In Wang et al. (2019) an averaging estimator is proposed which takes a weighted average of the different models that are considered, according to a certain criterion. The models considered are the pooled and individual model, but also intermediate model where a group structure is assumed, such that individuals are divided into groups with equal coefficients within groups. The specific criterion investigated is the Mallows criterion, which is a consistent estimator for the squared (forecast) error. The research into such averaging estimators in a panel data setting is quite recent and current research on heterogeneous slopes in panel data is more focused on different kinds of models. In this paper we will study different estimators in different settings and determine which one is the best with a special focus on averagin estimators. Another assumption often taken for granted in panel data is the constantness of slopes over time. Here there are two main paths to choose: do we consider changes over time to be continuous or do we consider discontinuities. We investigate the former. We propose a new time-varying coefficient approach which is based on kernel smoothing estimation methods. We combine this kernel estimator with the Mallows criterion to investigate the behaviour with possible slope heterogeneity among individuals. Our main research question will be: does the effect of the Mallows pooling estimator differ between parametric and non-parametric settings? We test the models in a simulation setting. In reviewing the time-invariant coefficient methods we find similar results as Wang et al. (2019). The estimator based on Mallows criterion outperforms others when the data has a medium degree of heterogeneity such as a group structure. The estimator performs quite well in other cases, but the individual and pooled estimators have a better performance in the heterogeneous and homogeneous model respectively. We find that when the number of individuals changes the outcome does not change, but the difference in performance becomes greater. When looking at the time-variant coefficient model we find that the pooled estimator performs very well compared to the Mallows criterion and to estimators based on AIC and BIC. We find that this is the case especially when the number of individuals is low. When the number of individuals rises we find that Mallows pooling average outperforms the other models. Overall we find that the pooling

estimator albeit unbiased when the assumptions do not match performs quite well. We find that in the parametric models the Mallows pooling method has better estimates when heterogeneity arises. In the non-parametric model the Mallows pooling average does not drastically outperforms other estimators and only outperforms the pooled estimator when both the number of individuals and number of time observations is high. The methods we discuss can improve the decision-making process in multiple fields, because we are able to generate better estimates of effects in panel data models. Panel data models are found across multiple fields, examples are finance, economics, psychology and medicine. Alike normal panel data models we find that non-parametric models have uses in the same fields. Most literature on these models stems from the medical field, but recent literature in the area of finance and economics has shown it has diverse applications.
This paper has the following structure. In section 2, we will give a short summary of the literature on the topic until now. In section 3, the methods we use are considered. Already existing methods are reviewed and methods extending those are introduced. The setup of the simulation is also considered. In section 4, the results will be given and shortly examined. Section 5 will discuss the results and possibilities for future research on this topic.

## 2. Literature review

In literature we find that heterogeneous slopes in panel data are researched as early as 1970 in Swamy (1970), where a Random Coefficient Model (RCM) is proposed with a GLS type estimator as a solution for estimating the average effect (see also M. H. Pesaran and Yamagata (2008)). In M. Pesaran and Smith (1995) four methods for dealing with possible heterogeneous slopes are evaluated: individual estimation, pooled estimation, aggregated estimation and averaging estimation. It is found that pooling and aggregating, which need extra assumptions, can lead to misleading estimates and the assumption of homogeneity should be made carefully. However, Baltagi and Griffin (1997) finds that the pooling estimator performs better than the individual estimator in terms of root mean square error, even though the model is not correctly specified due to a false assumption of homogeneous slopes. It is thus not fair to rule out pooling the data although the assumptions might not hold, because the variance reducing properties of the pooling outweigh the bias due to violated assumptions.
In other literature the focus is not concentrated on average effect, but instead on the individual effect. As policy makers often have choices to make between individuals, these effects are valuable to estimate correctly. In G. S. Maddala, Trost, Li, and Joutz (1997) a shrinkage estimator is proposed as homogeneity does not hold and the individual estimator gives false results. In this method the pooled estimates and the individual estimates are combined to make a new estimate. In the previously mentioned study of Baltagi and Griffin (1997) the shrinkage estimator does however not outperform the pooling estimator. In a later study G. Maddala, Li, Srivastava, et al. (2001) is again favourable for the shrinkage estimator. Recently, Wang et al. (2019) uses another method to

combine models. Here not only the pooled and individual estimator are combined, but a weighted estimate is taken over a space of numerous models. Wang et al. (2019) builds upon the ideas proposed in Hansen (2007), where Mallows criterion is used to combine multiple models to form a new estimator. Mallows criterion, a consistent estimator for the squared error, is found to outperform other information criteria such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) when averaging weights are chosen by smoothing the values of these criteria. Mentioned should also be the study Wan, Zhang, and Zou (2010) where the optimality of Hansen's model is shown with even fewer assumptions. In Wang et al. (2019) this method is extended to the panel data case to make an averaging estimator with superior mean squared (forecast) error compared to multiple previously mentioned methods. We will revisit the results of Wang et al. (2019) and compare the Mallows averaging estimator to multiple estimators among which are the shrinkage estimator and the smoothed BIC and AIC estimators.

In different studies the problem of heterogeneous slopes is solved by defining which parameters are heterogeneous and estimating accordingly. A good example of this is M. H. Pesaran, Shin, and Smith (1999), where a distinction is made between homogeneous long-run parameters and heterogeneous short-run parameters. This method does however require a correct specification for the homogeneous and heterogeneous parameters. In more recent research a distinction is not made between parameters, but between individuals. For example in Ando and Bai (2015), where a model is considered with unobserved group factors. In literature different methods for estimating group structure is discussed, two major methods are a k-means approach which is discussed in Bonhomme and Manresa (2015) and Lin and Ng (2012) or a classifier-Lasso (c-lasso) approach as discussed in Su, Shi, and Phillips (2016). Where the latter is found to be less computationally demanding and have favourable inferential properties over the former. We use the c-lasso estimator for finding the correct groups in the data and compare the estimator to other estimators as well.

The methods for c-lasso are also extended to not only include time-invariant coefficient models, but to extend to the non-parametric time-variant case as well in Su, Wang, and Jin (2019). Here a sieve estimation method is proposed in combination with the c-lasso penalty, such that each individual coefficient function is shrunken towards a group coefficient function. Time-variant coefficient models, also known as functional coefficient models are a topic of research on their own. In Hastie and Tibshirani (1993) the methods of kernel estimation are discussed in detail and in Hoover, Rice, Wu, and Yang (1998) smoothing splines and polynomial methods are discussed. In Li, Chen, and Gao (2011) a kernel method approach based on Taylor expansion is extended to panel data. However, except for Su et al. every non-parametric model assumes parameter homogeneity and other literature does not extend the parametric models to non-parametric cases. In Bernoth and Erdogan (2012) the time-varying kernel model of Sun, Carroll, and Li (2009) is applied to model sovereign bond spread, inferential properties and a method to select the optimal bandwidth are also discussed. Again the c-lasso methods will be used to find the correct groups. Furthermore we will extend it with kernel methods and use it in comparison to other methods.

# 3. Methodology

## 3.1. General model

We consider the model

$$y_{it} = f(x_{it}) + \varepsilon_{it} \tag{1}$$

for $i = 1...N$ and $t = 1...T$. We assume $\varepsilon_{it}$ is a zero-mean, time-independent process with variance $\sigma_\varepsilon^2$. Furthermore we assume that $\varepsilon_{it}$ and $\varepsilon_{jt}$ are independent if $i \neq j$. The current form of $f(\cdot)$ is very unrestricted, thus hard to use for inference. In Hastie and Tibshirani (1993) a model is proposed such that $f(\cdot)$ is a function linear in some of the independent variable $x_{it}$. We consider two sets of independent variables $\mathcal{X} = \left\{x_{itj} : i = 1...N, t = 1...T, j = 1...k\right\}$ and $\mathcal{Z} = \{z_{it} : i = 1...N, t = 1...T\}$. The model we can consider then is

$$y_{it} = \sum_{j=1}^{k} x_{itj}\beta_{ij}(z_{it}) \tag{2}$$

for $i = 1...N$ and $t = 1...T$. We consider two specific models, the first being the standard linear model such that $\beta_{ij}(z_{it}) = \beta_{ij}$ for $i = 1...N$ and $j = 1...k$. Such that all the coefficients are constants, but are allowed to vary between individuals. This results in the following model.

$$y_{it} = \sum_{j=1}^{k} x_{itj}\beta_{ij} \tag{3}$$

The second model we consider assumes $z_{it} = t$, such that we can rewrite the model to be the following.

$$y_{it} = \sum_{j=1}^{k} x_{itj}\beta_{itj} \tag{4}$$

for $i = 1...N$ and $t = 1...T$ Here the coefficients are allowed to change between individuals and over time. We assume though that $\beta$ changes smoothly over time. In section 3.2 the model in equation (3) will be investigated and in section 3.3 the model in equation (4) will be looked at. In the rest of this text we will write $X_{it} = (x_{it1}, x_{it2}, ..., x_{itk})$ for $i = 1...N$ and $t = 1...T$

## 3.2. Time-invariant coefficients

If we consider the coefficients in the model to be time-invariant we can rewrite equation 3 to be

$$y_i = X_i\beta_i + \varepsilon_i, \quad i = 1...N \tag{5}$$

where $y_i = (y_{i1}, y_{i2}, ..., y_{iT})'$, $X_i = (X'_{i1}, X'_{i2}, ..., X'_{iT})$ and $\beta_i = (\beta_{i1}, \beta_{i2}, ..., \beta_{ik})$. For each individual $i$ we can consider the individual least squares estimator $\beta_i = (X_i'X_i)^{-1}X_i'y_i$. Under certain assumptions this estimator is unbiased, consistent and attains a variance of $\sigma_\varepsilon^2(X_i'X_i)^{-1}$. We will write $\beta_{ind} = (\beta_1', \beta_2', ..., \beta_N')'$ as the $Nk$ by 1 individual estimator. However this estimator does not take the panel structure of the data into account, meaning that only the individual data is used for each individual estimation. On the other end of the spectrum we consider the pooled estimator. This estimator assumes that the coefficients are equal over all individuals, that is $\beta_i = \beta_j, \forall i, j \in 1, 2, ...N$. The pooled estimator is defined as $b = \left(\sum_{i=1}^{N} X_i' X_i\right)\left(\sum_{i=1}^{N} X_i' y_i\right)$. We define $\beta_{pool} = (b', b', ..., b')'$ as our $Nk$ by 1 pooled estimator. Because this pooled estimator uses the cross-section variation of the data the variance is lower, however this comes at the cost that the estimator is biased if the, often not justifiable, assumption of equal coefficients is violated. This problem seems to be a classic bias-variance trade-off situation. If we do not wish to use the estimator at either side of the spectrum, we could choose to use the intermediate estimator. This intermediate estimator is formed by restricting the coefficients by some restriction matrix $R$. The restriction $R\beta = 0$ forces certain structure on the coefficients. We only consider restriction matrices which force a group structure on the parameters, such that within groups the coefficients are equal, but they differ over multiple groups. By using the projection matrix $P = I_{Nk} - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R$ we get the restricted estimator to be $\beta_{res} = P\beta_{ind}$. We denote $(R_m, P_m)$ for the m-th restriction and projection matrix, such that $\hat{\beta}_{(m)} = P_m\beta_{ind}$ is the m-th estimator. If we wish to combine multiple models we could do so by averaging. If we have $M$ different estimators and a set of weights $(w_1, w_2, ..., w_M) \in \left\{(w_m)_{i=1}^{M} : \sum_{i=1}^{M} = 1, w_m \in [0,1] \text{ for all } m\right\}$ we can make an averaging estimator $\beta(w) = \sum_{i=1}^{M} w_m \hat{\beta}_{(m)} = \sum_{i=1}^{M} w_m P_m \beta_{ind} = P(w)\beta_{ind}$. To find the optimal weights in literature certain information criteria are used (see Hansen (2007)). Examples of information criteria used are the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and Mallows information criterion (also known as Mallows $C_p$). The last one will be used extensively throughout this paper. The Mallows criterion for averaging is given by

$$C_A(w) = ||P(w)\beta_{ind} - \beta_{ind}||_A^2 + 2tr(P'(w)AV) - ||\beta_{ind} - \beta||_A^2 \tag{6}$$

where $V = var(\beta_{ind})$ is the variance of the estimator and $||x||_A^2 = x'Ax$ is a weighted matrix norm. When $A = I$ this criterion focuses on accuracy of the estimator, if $A = X'X$ this criterion focuses on in-sample forecasting. The use of Mallows criterion for averaging purposes was introduced by Hansen (2007) and more of its properties are investigated in Wang et al. (2019). If the variance matrix $V$ is unknown it can be replaced by an estimator $\hat{V}$ in equation (6), for homoscedastic error terms $\hat{V} = \tilde{\sigma}_\varepsilon^2(X'X)^{-1}$ where $\tilde{\sigma}_\varepsilon^2 = \frac{(y-X\beta)'(y-X\beta)}{NT-Nk}$. We find the optimal weights, and thus the optimal estimates, in our averaging model by $w_{opt} = \arg\min_{w\in\mathcal{W}} C_A(w)$.
However, in practice another problem is luring. In some situations there might be a certain set of restriction matrices that make sense. More often than not such a clear set of restriction matrices might not be clear. As the number of individuals grows the number of possible groups grows drastically, thus computation for all possible groups is not feasible. Therefore we must shrink the space of possible restrictions. In other words, we must find groupings that make sense according to the data and there must not be too many. However, we do want to find the minimum value for Mallows criteria $C_A(w)$. As is discussed in Wang et al. (2019) and Zhang, Yu, Zou, and Liang (2016) even though the space is shrunken, the asymptotic optimality is still achieved in terms of

squared loss under certain assumptions. That is,

$$\frac{L_A(\hat{w}^s)}{\inf_{w \in \mathcal{W}} L_A(\hat{w})} \longrightarrow 1 \tag{7}$$

where $L_A(w) = ||\hat{\beta}(w) - \beta||_A^2$ and $\hat{w}^s$ is the optimal weight vector from the set of shrunken model space. The denominator denotes the most optimal weights when the space is not shrunken. The assumptions this requires and the proof can be found in Zhang et al. (2016) and Wang et al. (2019). To put the above into practice we use the methods developed in Su et al. (2016) where group structure in data is found by shrinking the coefficients in a regression to group coefficient for a pre-specified number of $K0$ groups using a lasso penalty. That is, they define $Q_{1,NT}(\beta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\psi(w_{it};\beta_i)$ as the profile log-likelihood function. A criterion function is then defined as

$$Q_{1NT,\lambda}^{(K_0)}(\beta,\alpha) = Q_{1,NT}(\beta) + \frac{\lambda}{N}\sum_{i=1}^{N}\prod_{k=1}^{K_0}||\beta_i - \alpha_k|| \tag{8}$$

By minimizing this criterion function the individual coefficient for individual $i$: $\beta_i$ is shrunken towards the group coefficient $\alpha_k$ for some $k \in \{1,...,K_0\}$. Because of the product in equation (8) once an individual coefficient and a group coefficient are equal the penalty on that individual disappears to 0. After this function has been minimized and estimates $(\alpha_1, \alpha_2, ..., \alpha_{K0})$ and $(\beta_1, \beta_2, ..., \beta_N)$ the groups are formed such that $G_k = \left\{i \in \{1,2,...,N\} : \hat{\beta}_i = \hat{\alpha_k}\right\}$ for $k \in \{1,2,...,K_0\}$ In Su et al. (2016) an information criterion is also developed to find the best number of groups. We compare the results from the estimates of this information criterion to the estimates found using Mallows criterion. We furthermore compare models selected by the Akaike information criterion, Bayesian information criterion. We also use these criteria to find smoothed weights. Let $CR_i, \quad i = 1,...,M$ be the values of the information criteria for every model, then the smoothed weights are found using the formula $w_i = \frac{\exp(-\frac{1}{2}CR_i)}{\sum_{j=1}^{M}\exp(-\frac{1}{2}CR_j)}$ as mentioned in Hansen (2007).

Another method of averaging is Bayesian averaging, here the posterior probability of each model is used to create weights for the model, details for this can be found in Wang et al. (2019). We furthermore compare the Mallows pooling average estimator to the Shrinkage estimator, which weighs the pooled and the individual model according to the test statistic with a null-hypothesis of homogeneous slopes, detail for this can be found in G. S. Maddala et al. (1997).

### 3.3. Time-variant coefficients

Where in the previous section we restricted our model to the form in equation (4), where $\beta_i$ is considered constant, we now assume the model is of the form

$$y_{it} = x_{it}\beta_i(\frac{t}{T}) + \varepsilon_i t, \text{for } i = 1,...,N \text{ and } t = 1,...,T \tag{9}$$

where $x_i t = (1, x_{it1}, x_{it2}, ..., x_{itk})$ and $\beta_i(\frac{t}{T}) = (\beta_{i0}(\frac{t}{T}), \beta_{i1}(\frac{t}{T}), \beta_{i2}(\frac{t}{T}), ..., \beta_{ik}(\frac{t}{T}))'$. We assume $\beta_{ij}(\frac{t}{T})$ is a sufficiently smooth function, that is we assume $\beta_{ij}(\frac{t}{T}) \in \mathbf{C}^2[0,1]$ such that the second derivative is continuous in the time interval. Rewriting each $\beta_{ij}(\frac{t}{T})$ using Taylor series we find $\beta_{ij}(\frac{t}{T}) = \beta_{ij}(\tau) + \beta'_{ij}(\tau)(\frac{t}{T} - \tau) + \mathcal{O}\Big((\frac{t}{T} - \tau)^2\Big)$ for $0 < \tau < 1$ such that we can approximate each coefficient function by the first two terms of their Taylor series. Following the methods in Li et al. (2011) we define

$$M_i(\tau) = \begin{pmatrix} X_{i1} & \frac{1-\tau T}{T} X_{i1} \\ \vdots & \vdots \\ X_{iT} & \frac{T-\tau T}{T} X_{iT} \end{pmatrix} \quad \text{and} \quad W(\tau) = \text{diag}\left( \frac{K\left(\frac{1-\tau T}{T}\right)}{h}, \ldots, \frac{K\left(\frac{T-\tau T}{T}\right)}{h} \right) \tag{10}$$

Where $\frac{K(\frac{\cdot}{h})}{h}$ is a kernel function with bandwith $h$, which makes $W(\tau)$ a weighing matrix that weighs the distance from each observation $t$ to the current $\tau$. We are now able to find the Taylor approximation coefficients by the following minimization problem:

$$\min_{\mathbf{a},\mathbf{b} \in \mathbb{R}^k} [y_i - M_i(\tau)(\mathbf{a}', \mathbf{b}')']' \, W(\tau) \, [y_i - M_i(\tau)(\mathbf{a}', \mathbf{b}')'] \tag{11}$$

such that we can estimate $\hat{\beta}_i(\tau)$ as

$$\begin{pmatrix} \hat{\beta}_i(\tau) \\ \hat{\beta}'_i(\tau) \end{pmatrix} = \left[ M'_i(\tau) W(\tau) M_i(\tau) \right]^{-1} M'_i(\tau) W(\tau) y_i \tag{12}$$

for each individual $i$. As we did in the parametric time-invariant case we can also pool the data under the assumptions of homogeneous slopes among all individuals. The pooled estimator becomes:

$$\begin{pmatrix} \hat{\beta}_{pool}(\tau) \\ \hat{\beta}'_{pool}(\tau) \end{pmatrix} = \left[ \sum_{i=1}^{N} M'_i(\tau) W(\tau) M_i(\tau) \right]^{-1} \sum_{i=1}^{N} M'_i(\tau) W(\tau) y_i \tag{13}$$

We are also able to extend the restricted model to a time-variant model. We assume groupings do not vary over time. Then we can consider the restrictions as follows $R\beta(\frac{t}{T}) = 0$ for $t = 1, \ldots, T$, where $\beta(\frac{t}{T}) = (\beta_1(\frac{t}{T})', \beta_2(\frac{t}{T})', \ldots, \beta_N(\frac{t}{T})')'$ is defined as before by the stacked individual slopes. Let $\hat{\beta}(\tau)$ and $\hat{\beta}'(\tau)$ be defined in the same way. Then for restriction matrix $R_m$ we find

$$\begin{pmatrix} \hat{\beta}_{(m)}(\tau) \\ \hat{\beta}'_{(m)}(\tau) \end{pmatrix} = P_m \begin{pmatrix} \hat{\beta}(\tau) \\ \hat{\beta}'(\tau) \end{pmatrix} \tag{14}$$

where $P_m$ is the $2Nk$-by-$2Nk$ projection matrix defined by

$$P_m = I_{2Nk} - (M'(\tau)W(\tau)M'(\tau))^{-1} \begin{pmatrix} R'_m & R'_m \end{pmatrix} \left[ \begin{pmatrix} R_m \\ R_m \end{pmatrix} (M'(\tau)W(\tau)M'(\tau))^{-1} \begin{pmatrix} R'_m & R'_m \end{pmatrix} \right]^{-1} \begin{pmatrix} R_m \\ R_m \end{pmatrix} \tag{15}$$

Equation (15) is the extended equivalent of the time-invariant projection matrix. We stack two restriction matrices as the same restrictions hold for the coefficient functions as well as for their derivatives. We furthermore see that $(X'X)^{-1}$ is simply replaced by its weighted time-variant equivalent $(M'(\tau)W(\tau)M'(\tau))^{-1}$. We follow the time-invariant methods by defining $P(w) = \sum_{m \in \mathcal{M}} w_m P_m$, where $\mathcal{M}$ is the set of restrictions and $\sum_{m\in\mathcal{M}} w_m = 1$. Then $\beta_{(m)}(\tau) = P(w)\beta(\tau)$ as in the time-invariant case.

Next, we extend the Mallows criterion to the time-variant case. Where above we have seen that we can again define $P(w)$ and $\beta_{(m)}(\tau)$. We can redefine Mallows criterion by taking the average over every time observation. That is, the Mallows criterion becomes the following:

$$C_A(w) = \frac{1}{T}\sum_{t=1}^{T} \|P(w)\hat{\beta}\left(\frac{t}{T}\right) - \hat{\beta}\left(\frac{t}{T}\right)\|_A^2 + 2\mathrm{tr}\left(P'(w)A\hat{V}\right) - \|\hat{\beta}\left(\frac{t}{T}\right) - \beta\left(\frac{t}{T}\right)\|_A^2 \tag{16}$$

We redefine $\hat{V}$ to use the time-variant data: $\hat{V} = \hat{\sigma}(X'X)^{-1} = \frac{\sum_{t=1}^{T}\left(y_t - X_t\hat{\beta}\left(\frac{t}{T}\right)\right)'\left(y_t - X_t\hat{\beta}\left(\frac{t}{T}\right)\right)}{(NT-Nk)}(X'X)^{-1}$. We find the optimal weights by minimizing the extended Mallows criterion. Similar to the time-invariant case the model space $\mathcal{M}$ can consist of an immense amount of models, making the minimization problem for the optimal weight vector computationally very demanding. Alike the time-invariant case, we use a certain model classifier to find groups among individuals. We use the c-lasso type estimator discussed in Su et al. (2019). We alter their model by using the kernel type regression described above instead of the sieve type regression they propose. In the sieve regression approach the following objective function is considered:

$$Q_{NT,\lambda}^{(K)}(\pi,\omega) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[y_{it} - Z_{it}\mathrm{vec}(\pi_i)]^2 + \frac{\lambda}{N}\sum_{i=1}^{N}\hat{\sigma}_i^{2-K}\prod_{k=1}^{K}\|\mathrm{vec}(\pi_i - \omega_k)\| \tag{17}$$

where $\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}[y_{it} - Z_{it}\pi_i]^2$. Here $y_{it}$ and $Z_{it}$ have been normalized to filter out individual effects and for ease of computation, resulting in a slightly different but equivalent formulation as in the original text of Su et al. (2019). By minimizing the function $Q_{NT,\lambda}^{(K)}$ individual coefficients $\pi_i$ and group coefficients $\omega_k$ can be found and equivalently as in the time-invariant case we consider individual $i$ to be in group $k$ if $\|\pi_i - \omega_k\| = 0$.

We propose to extend this objective function to include kernel estimation techniques. We will replace the first term which concerns the data fit by $\frac{1}{NT}\sum_{i=1}^{N}\sum_{\tau=1}^{T}[y_i - M_i(\tau)\mathbf{b}_{i\tau}]'W(\tau)[y_i - M_i(\tau)\mathbf{b}_{i\tau}]$. If we set $\mathbf{B}_i = (\mathbf{b}_{i1}, \mathbf{b}_{i2}, ..., \mathbf{b}_{iT})$ and let $\|\cdot\|_F^2$ be the squared Frobenius norm, then we can replace the penalty term by $\frac{\lambda}{N}\sum_{i=1}^{N}\hat{\sigma}_i^{2-K}\prod_{k=1}^{K}\|\mathbf{B}_i - \Omega_k\|_F^2$. We get the new objective function:

$$Q_{NT,\lambda}^{(K)}(\mathbf{B},\Omega) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{\tau=1}^{T}[y_i - M_i(\tau)\mathbf{b}_{i\tau}]'W(\tau)[y_i - M_i(\tau)\mathbf{b}_{i\tau}] + \frac{\lambda}{N}\sum_{i=1}^{N}\hat{\sigma}_i^{2-K}\prod_{k=1}^{K}\|\mathbf{B}_i - \Omega_k\|_F^2 \tag{18}$$

Using the above objective function we are still able to apply the methods and results as presented in Su et al. (2019). Note that for clarity we have written $\Omega_k$ instead of $(\omega_k(1), \omega_k(2), ..., \omega_k(T))$. Unfortunately, the above minimization problem, even though there exist convex methods to solve it (see the work of Su et al. (2019)), is too computationally intensive and thus out of the scope of the current study. As we only study simulated data and thus are aware of the structure of the data, we form groups based on the order and the number of groups. If we make $K0$ groups then the first

group will contain all individuals $i$, for which $i < \frac{N}{K0}$. Such that group $j$ will be the following set $G_j^{(K0)} = \left\{ i : \frac{(j-1)N}{K0} < i \leq \frac{jN}{K0} \right\}$.
As the above latent structure models only work with a pre-specified number of groups an information criteria is given to find the number of groups. A BIC-type information criteria is proposed: let $\hat{\sigma}_K^2 = \frac{1}{NT} \sum_{j=1}^{K} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{1}_{i \in \hat{G}_j} \left[ y_{it} - X_{it} \left[ \mathbf{1}_k \quad \mathbf{0}_k \right] \omega_j(t) \right]$ then we define $\text{IC}(K) = \ln \hat{\sigma}_K^2 + \rho k K$. We then choose the number of groups with the lowest information criterion.

## 3.4. Simulation

As this paper only handles theoretical results we will describe in detail how the data was simulated. The section is divided in two sections, the first regarding the simulation of the time-invariant coefficient data, the second will be about the simulation of the time-variant coefficient data.

### 3.4.1. Time-invariant coefficients

To measure the performance of different estimators we will test them on data with different properties. We will vary the number of individuals in the data and the number of observations in the time dimension. We will keep the number of explanatory variables fixed at $k = 5$. We will vary the number of individuals to vary between 10 and 30. The number of time observations will be either 20 observations, 40 observations or 80 observations. The different data generating processes we use all have different degrees of heterogeneity. We start by the homogeneous data generating process here

$$\beta_{ij} = 1 \quad \text{for } i = 1, \ldots, N \text{ and } j = 1, \ldots, k \tag{19}$$

Next we will consider a data generating process where the individuals belong to different groups. We choose the number of groups $K0$ to be equal to 3.

$$\beta_{ij} = \begin{cases} 1 & \text{for } i = 1, \ldots, \lfloor N/3 - 1 \rfloor \\ 3 & \text{for } i = \lfloor N/3 \rfloor, \ldots, \lfloor 2N/3 - 1 \rfloor \\ 5 & \text{for } i = \lfloor 2N/3 \rfloor, \ldots, N \end{cases} \quad \text{and } j = 1, \ldots, k \tag{20}$$

As a last case, we will consider a completely heterogeneous process. For this process every individual has different coefficients.

$$\beta_{ij} = 0.1 \times i \times j \quad \text{for } i = 1, \ldots, N \text{ and } j = 1, \ldots, k \tag{21}$$

Throughout the simulation we use a low level of noise with a regression $R^2$ equal to 0.89. We generate the data for the explanatory variables by a zero mean normal distribution with a standard deviation of 10. The errors are generated using a standard normal distribution. When estimating we focus on the forecast abilities of the estimator, therefore we choose $A = X'X$ thus modelling in-sample forecasting. The model we compare, some of which have been previously discussed, are the following: MPA (Mallows pooling average), BPA (Bayesian pooling average), Pooled, Individual, SAIC (smoothed AIC), SBIC (smoothed BIC), AIC, BIC, SHK (Shrinkage estimator as suggested by G. S. Maddala et al. (1997)) and c-lasso (the best group structure according to the c-lasso information criteria).

#### 3.4.2. Time-variant coefficients

For the time-variant coefficient case we will again compare our estimates for different values of number of individuals $N$ and number of time observations $T$. We will fix the number of explanatory variables to $k = 3$. For the number of individuals we will again assume $N = \{10, 30\}$ and for the number of time observations we will have $T = \{20, 40, 80\}$. We will again consider three types of heterogeneity: homogeneous, group structure and completely heterogeneous. As the coefficients are now functions over time we will consider polynomials of degree two. Such that every coefficient has the following form: $\beta_{ij}(t) = \alpha_{0ij} + \alpha_{1ij}t + \alpha_{2ij}t^2$. We can use polynomials as coefficient functions as they are infinitely differentiable over the real numbers. As we are indifferent what $\alpha_.$ is as long as it does following the heterogeneity structure, we set $\alpha_{0ij} = \alpha_{1ij} = \alpha_{2ij}$ for the homogeneous and grouped structure processes. For the homogeneous we set

$$\beta_{ij}(t) = 0.5 + 0.5t + 0.5t^2 \quad \text{for } i = 1,\ldots N \text{ and } j = 1,\ldots,k \tag{22}$$

For the grouped structure process we choose the following:

$$\beta_{ij}(t) = \begin{cases} 0.5 + 0.5t + 0.5t^2 & \text{for } i = 1,\ldots,\lfloor N/3 - 1 \rfloor \\ 2 + 2t + 2t^2 & \text{for } i = \lfloor N/3 \rfloor,\ldots,\lfloor 2N/3 - 1 \rfloor \\ 5 + 5t + 5t^2 & \text{for } i = \lfloor 2N/3 \rfloor,\ldots,N \end{cases} \quad \text{and } j = 1,\ldots,k \tag{23}$$

When considering the heterogeneous process we use the following coefficients:

$$\beta_{ij}(t) = ij + 2ijt + 3ijt^2 \quad \text{for } i = 1,\ldots,N \text{ and } j = 1,\ldots,k \tag{24}$$

We again consider a high explanatory power for this simulation with the $R^2 = 0.89$. We generate the explanatory variables by a zero mean normal distribution with a standard deviation of 10. We estimate the error term with a standard normal distribution. We consider only a selection of the estimation methods used in the time-invariant simulation. We compare the (altered) MPA, SAIC, SBIC, AIC, BIC, c-lasso and the smoothed version of the c-lasso criterion.

## 4. Results

### 4.1. Time-invariant coefficients

#### 4.1.1. $N = 10$

Here we will consider the results for the time-invariant models when the number of individuals is relatively low with 10 individuals.

**Table 1:** The loss calculated for every different model divided by the loss of the individual estimator where the number of individuals is set to 10. Bold faces values is the best performing model. In the process where $T = 20$ and the data generating process in heterogeneous we see that the individual estimator performs better than all others, albeit marginally. The abbreviations for the models are given in the section regarding the simulation.

| | DGP | MPA | c-lasso | BPA | SAIC | SBIC | AIC | BIC | pooled | SHK |
|---|---|---|---|---|---|---|---|---|---|---|
| $T = 20$ | Homo | **0,03** | 0,03 | 0,58 | 0,07 | 0,11 | 0,03 | 0,03 | 0,03 | 0,88 |
| | Group | **0,75** | 33,40 | 0,77 | 14,32 | 14,00 | 33,40 | 33,40 | 33,40 | 1,00 |
| | Hetero | 1,01 | 142,89 | 1,00 | 44,31 | 44,00 | 142,89 | 142,89 | 142,89 | 1,00 |
| $T = 40$ | Homo | 0,16 | **0,01** | 0,89 | 0,86 | 0,23 | 0,01 | 0,01 | 0,01 | 0,96 |
| | Group | **0,76** | 33,15 | 0,80 | 16,88 | 16,69 | 33,15 | 33,15 | 33,15 | 1,00 |
| | Hetero | **1,00** | 38,19 | 1,00 | 8,12 | 8,07 | 38,19 | 38,19 | 38,19 | 1,00 |
| $T = 80$ | Homo | 0,10 | **0,02** | 0,75 | 0,59 | 0,07 | **0,02** | **0,02** | **0,02** | 0,98 |
| | Group | 0,33 | 297,09 | **0,31** | 168,44 | 167,64 | 297,09 | 297,09 | 297,09 | 1,00 |
| | Hetero | **1,00** | 84,22 | 1,11 | 23,21 | 23,11 | 84,22 | 84,22 | 84,22 | 1,00 |

We see in the table that Mallows pooling average performs very well among its peers. Across the different number of time observations it is apparent that the pooled estimator and the estimator that select a single model based on information criteria are all equal or slightly better in performance compared to MPA. In the case of a group structure we see that MPA and BPA are very closely matched up and other estimators are significantly outperformed by those two. In case of heterogeneous data it is very clear that the individual estimator performs well, although MPA, BPA and the SHK estimators are marginally better or worse. Overall we see that the SHK estimator closely matches the performance of the individual estimator. We see that the smoothed criterion estimators (SAIC and SBIC) are easily outperformed by all other models indifferent of number of observations or coefficient structure. We see no specific differences over time between estimators, only the relative differences become greater.

#### 4.1.2. $N = 30$

Next, we will consider the time-invariant models with the number of individuals tripled such that $N = 30$.

**Table 2:** The loss calculated for every different model divided by the loss of the individual estimator where the number of individuals is set to 30. Boldfaced values denote the best performing model. Where no value is boldfaced, the individual estimator is better than all the others. The abbreviations for the models are given in the section regarding the simulation.

| | DGP | MPA | c-lasso | BPA | SAIC | SBIC | AIC | BIC | pooled | SHK |
|---|---|---|---|---|---|---|---|---|---|---|
| $T = 40$ | Homo | 0,10 | **0,00** | 0,35 | 0,19 | 0,21 | **0,00** | **0,00** | **0,00** | 0,91 |
| | Group | **0,34** | 24,93 | 0,59 | 9,77 | 9,72 | 24,93 | 24,93 | 24,93 | 1,00 |
| | Hetero | 13,63 | 62,92 | 16,15 | 35,28 | 35,26 | 62,92 | 62,92 | 62,92 | 1,00 |
| | | | | | | | | | | |
| $T = 40$ | Homo | 0,21 | **0,01** | 0,62 | 0,39 | 0,40 | **0,01** | **0,01** | **0,01** | 0,94 |
| | Group | **0,31** | 58,55 | 0,55 | 11,92 | 11,87 | 58,55 | 58,55 | 58,55 | 1,00 |
| | Hetero | 6,66 | 780,97 | 8,44 | 127,20 | 127,10 | 780,97 | 780,97 | 780,97 | 1,00 |
| | | | | | | | | | | |
| $T = 80$ | Homo | 0,25 | **0,00** | 0,68 | 0,53 | 0,16 | **0,00** | **0,00** | **0,00** | 0,98 |
| | Group | **0,31** | 88,86 | 0,43 | 15,73 | 15,69 | 88,86 | 88,86 | 88,86 | 1,00 |
| | Hetero | 31,38 | 590,74 | 29,84 | 176,83 | 176,80 | 590,74 | 590,74 | 590,74 | 1,00 |

With a greater number of individuals we see some changes. As before for the homogeneous data generating process the pooled estimator and the model selecting criterion based estimates (which select the pooled model) substantially outperform the other models. We see that as the number of individuals has risen the relative difference between the MPA and the pooled estimators has become greater. When the data generating process has a group structure for the coefficients we see that MPA outperforms all other models, unexpectedly even the c-lasso significantly. We see that the relative difference between MPA and BPA is also bigger compared to the $N = 10$ case. In the heterogeneous case we see that the individual estimator and the shrinkage estimator perform the best.

## 4.2. Time-variant coefficients

### 4.2.1. $N = 10$

Thirdly, we will look at a low number of individuals of 10 when the coefficients are allowed to vary over time.

**Table 3:** The loss calculated for every different model divided by the loss of the individual estimator where the number of individuals is set to 10. Bold faces values is the best performing model. Where no value is boldfaces, the individual estimator is better than all the others. The abbreviations for the models are given in the section regarding the simulation.

| | DGP | MPA | Pooled | SAIC | SBIC | SCL | AIC | BIC | CL |
|---|---|---|---|---|---|---|---|---|---|
| $T = 20$ | Homo | 0,55 | **0,36** | 0,67 | 0,67 | 0,68 | 0,67 | 0,67 | **0,36** |
| | Group | 0,68 | **0,66** | 0,81 | 0,81 | 0,81 | 1,00 | 1,00 | **0,66** |
| | Hetero | 0,69 | **0,33** | 0,53 | 0,53 | 0,53 | 1,00 | 1,00 | **0,33** |
| $T = 40$ | Homo | 0,63 | **0,33** | 0,60 | 0,60 | 0,60 | 1,00 | 1,00 | **0,33** |
| | Group | 0,72 | **0,37** | 0,69 | 0,69 | 0,69 | 1,00 | 1,00 | **0,37** |
| | Hetero | **0,36** | 0,63 | 0,72 | 0,72 | 0,72 | 1,00 | 1,00 | 0,63 |
| $T = 80$ | Homo | 0,67 | **0,63** | 0,78 | 0,78 | 0,78 | 1,00 | 1,00 | **0,63** |
| | Group | 0,77 | **0,35** | 0,64 | 0,64 | 0,65 | 1,00 | 1,00 | **0,35** |
| | Hetero | **0,52** | 0,54 | 0,70 | 0,70 | 0,71 | 1,00 | 1,00 | 0,54 |

When looking at a low number of individuals we find that the pooled model performs exceptionally well when compared to other models. We find that the CL (c-lasso) criterion model performs just as well as the pooled model, but this can be explained by the fact that the CL criterion penalizes on number of groups and thus the pooled model with only 1 group is selected. We see that the Mallows pooling average does not outperform other models, except the marginal difference with a heterogeneous generating process and a high number of time observations equal to 80. We find that the AIC and BIC consistently pick the individual model. Furthermore we see that the smoothed information criteria perform similarly as the MPA.

#### 4.2.2. $N = 30$

Lastly, we will consider the time-variant model with a number of individuals equal to 30.

**Table 4:** The loss calculated for every different model divided by the loss of the individual estimator where the number of individuals is set to 30. Bold faces values is the best performing model. Where no value is boldfaces, the individual estimator is better than all the others. The abbreviations for the models are given in the section regarding the simulation.

| | DGP | MPA | Pooled | SAIC | SBIC | SCL | AIC | BIC | CL |
|---|---|---|---|---|---|---|---|---|---|
| $T = 20$ | Homo | 0,66 | **0,59** | 0,74 | 0,74 | 0,74 | 1,00 | 1,00 | **0,59** |
| | Group | 0,54 | **0,34** | 0,67 | 0,67 | 0,68 | 1,00 | 1,00 | **0,34** |
| | Hetero | 0,59 | **0,36** | 0,60 | 0,60 | 0,60 | 1,00 | 1,00 | **0,36** |
| $T = 40$ | Homo | 0,58 | **0,51** | 0,65 | 0,65 | 0,65 | 1,00 | 1,00 | **0,51** |
| | Group | **0,43** | 0,43 | 0,65 | 0,65 | 0,65 | 1,00 | 1,00 | 0,43 |
| | Hetero | **0,60** | 0,60 | 0,82 | 0,82 | 0,82 | 1,00 | 1,00 | 0,60 |
| $T = 80$ | Homo | 0,59 | **0,31** | 0,63 | 0,63 | 0,64 | 1,00 | 1,00 | **0,31** |
| | Group | **0,38** | 0,51 | 0,64 | 0,64 | 0,65 | 1,00 | 1,00 | 0,51 |
| | Hetero | **0,34** | 0,56 | 0,61 | 0,61 | 0,62 | 1,00 | 1,00 | 0,56 |

When looking at the results where the number of individuals has tripled, we directly notice that now the Mallows pooling average performs much better compared to other models. We find that for a low number of time observations the pooled model still outperforms all other models and for a homogeneous data generating process we find that the pooling model is the best, as we would expect. However, when the data has a medium heterogeneous structure or is completely heterogeneous we find that the MPA outperforms all other models, albeit in some cases only marginally. We find again that the AIC en BIC criteria select the individual model and that the CL criterion selects the pooled model. We find that the smoothed models do perform constantly better than the individual model, but they are outperformed by the pooled model.

## 5. Discussion

Concluding, we find that in the case of time-invariant coefficient the pooled estimator and the Mallows pooling average clearly both perform fairly well. Not unexpectedly we find that when data is homogeneous the pooled estimator performs very well. When there exists a heterogeneous structure we find however that the Mallows pooling average performs the best. We find that this estimator also perform quite well compared to other estimator when the data is homogeneous. We can thus follow the conclusions from Wang et al. (2019) that the MPA is the best estimator in non-extreme cases and robust in extreme cases. In the time-variant model our findings are different. We see here that the pooling estimator clearly outperforms all other models when the number of individuals is low. When the number of individuals is high we find that the pooled estimator is only superior when the structure is homogeneous. We find that the Mallows pooling average performs marginally better in this case. Summarizing we would choose the Mallows pooling average estimator in time-invariant models, except if the assumption of homogeneity is very likely. In the time-variant case we would let our decision be based on the number of individuals and observations available. We would only choose MPA above the pooling estimator if those dimensions are high.

However, we have to take into consideration that in the time-variant model the best groups for the data have not been made. We extended a method known literature, but were not able to use it for calculations due to time-constraints. The time-variant Mallows pooling average might suffer from this.

In the future the proposed method can be studied, but where we used kernel methods to estimate the non-parametric coefficients more methods must be investigated as literature on the combination of panel data with non-parametric models is very scarce.

## References

Ando, T., & Bai, J. (2015). A simple new test for slope homogeneity in panel data models with interactive effects. *Economics Letters*, *136*, 112 - 117. Retrieved from http://www.sciencedirect.com/science/article/pii/S0165176515003778 doi: https://doi.org/10.1016/j.econlet.2015.09.019

Baltagi, B. H., & Griffin, J. M. (1997). Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics*, *77*(2), 303 - 327. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304407696018027 doi: https://doi.org/10.1016/S0304-4076(96)01802-7

Bernoth, K., & Erdogan, B. (2012). Sovereign bond yield spreads: A time-varying coefficient approach. *Journal of International Money and Finance*, *31*(3), 639–656.

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

Hansen, B. (2007). Least squares model averaging. *Econometrica*, *75*(4), 1175-1189. Retrieved from https://EconPapers.repec.org/RePEc:ecm:emetrp:v:75:y:2007:i:4:p:1175-1189

Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, *55*(4), 757–779.

Hoover, D. R., Rice, J. A., Wu, C. O., & Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, *85*(4), 809–822.

Li, D., Chen, J., & Gao, J. (2011). Non-parametric time-varying coefficient panel data models with fixed effects. *The Econometrics Journal*, *14*(3), 387-408. Retrieved from https://

onlinelibrary.wiley.com/doi/abs/10.1111/j.1368-423X.2011.00350.x doi: 10.1111/j.1368-423X.2011.00350.x

Lin, C.-C., & Ng, S. (2012). Estimation of panel data models with parameter heterogeneity when group membership is unknown. *Journal of Econometric Methods*, *1*(1), 42–55.

Maddala, G., Li, H., Srivastava, V., et al. (2001). A comparative study of different shrinkage estimators for panel data models. *Annals of Economics and Finance*, *2*(1), 1–30.

Maddala, G. S., Trost, R. P., Li, H., & Joutz, F. (1997). Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business & Economic Statistics*, *15*(1), 90-100. Retrieved from https://amstat.tandfonline.com/doi/abs/10.1080/07350015.1997.10524691 doi: 10.1080/07350015.1997.10524691

Pesaran, M., & Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, *68*(1), 79 - 113. Retrieved from http://www.sciencedirect.com/science/article/pii/030440769401644F doi: https://doi.org/10.1016/0304-4076(94)01644-F

Pesaran, M. H., Shin, Y., & Smith, R. P. (1999). Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association*, *94*(446), 621–634.

Pesaran, M. H., & Yamagata, T. (2008). Testing slope homogeneity in large panels. *Journal of Econometrics*, *142*(1), 50–93.

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

Su, L., Wang, X., & Jin, S. (2019). Sieve estimation of time-varying panel data models with latent structures. *Journal of Business & Economic Statistics*, *37*(2), 334-349. Retrieved from https://doi.org/10.1080/07350015.2017.1340299 doi: 10.1080/07350015.2017.1340299

Sun, Y., Carroll, R. J., & Li, D. (2009). Semiparametric estimation of fixed-effects panel data varying coefficient models. In *Nonparametric econometric methods* (pp. 101–129). Emerald Group Publishing Limited.

Swamy, P. A. V. B. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, *38*(2), 311–323. Retrieved from http://www.jstor.org/stable/1913012

Wan, A. T., Zhang, X., & Zou, G. (2010). Least squares model averaging by mallows criterion. *Journal of Econometrics*, *156*(2), 277 - 283. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304407609002838 doi: https://doi.org/10.1016/j.jeconom.2009.10.030

Wang, W., Zhang, X., & Paap, R. (2019). To pool or not to pool: What is a good strategy for parameter estimation and forecasting in panel regressions? *Journal of Applied Econometrics*, *0*(0). Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2696 doi: 10.1002/jae.2696

Zhang, X., Yu, D., Zou, G., & Liang, H. (2016). Optimal model averaging estimation for generalized linear models and generalized linear mixed-effects models. *Journal of the American Statistical Association*, *111*(516), 1775-1790. Retrieved from https://doi.org/10.1080/01621459.2015.1115762 doi: 10.1080/01621459.2015.1115762

## Appendix

In this appendix we will describe the code used to come to our results. All code is programmed in matlab. Every subsection is a different folder.

### Matlab

- **betaCalc**: in this function the coefficient estimates are calculated, such that one function can return multiple estimates from multiple estimators
- **BPA**: in this function the weights belonging to the Bayesian pooled average estimator are calculated
- **calculateLoss**: this function calculates the loss when given the real coefficient and the estimates, the function can take multiple estimates as input
- **criterion2**: this function is developed by Su et al. (2016) and finds the optimal c-lasso model according to their proposed criterion
- **findP**: this function uses other functions to calculate the groups structure and then forms the restriction matrices and projection matrices with this group structure
- **generateData**: this function generates the data in the way as is discussed in the paper
- **indEst**: this function returns the individual estimates for every individual
- **main**: this program runs all other programs, puts the results in a table and saves them.
- **MakeTable**: this function is used to put all data in a table
- **MallowsCrit**: this function computes the weights and coefficient estimates for the Mallowscriterion
- **optCLasso**: This function returns the optimal c-lasso estimate
- **P**: this function takes weights and the projection matrices in as arguments and returns the weighted average

- **PLS**$_e st : thisfunctionisdevelopedbySu\ etal.$ (2016)$andfindsthecoefficientsforthegroupstructuremodeltheyhave$

### Time-invariant

- **datagenerator**: this function generates the data according to the process described in the paper

- **groupfinder**: this methods finds the group structure and makes restriction and projection matrices based on that
- **individualest**: This function calculates the individual estimates
- **main**: in this function the main results are gathered for different N,T or degree of heterogeneity
- **main2**: this program runs main for multiple values
- **tableMaker**: this function calculates the loss for each estimate
- **timevarIC**: This function calculates the estimates for the different information criterion estimates such as SAIC, SBIC, SCL, AIC, BIC and CL
- **timevarMallows**: This function calculates the estimates for the Mallows pooling average estimator
- **timeVarZ**: this function calculates the values for the M and W matrices for every time step.