ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Econometrics and Operational Research

Title of thesis: Optimizing Predictions in Heterogeneous Panels with Pooling Averaging

Name of student: Thijs Wijnberg

Student ID number: 385754

Name of Supervisor: W. Wang

Name of Second Assessor: R. L. Lumsdaine

Date final version: 28 June 2017

Abstract

This paper is an extension of the research done in Wang et al. (2017). For static linear panel models with heterogeneous coefficients across individuals, the predictive performance of pooling averaging methods is investigated. Latent group structure identification using Classifier-Lasso with Partial Profile Likelihood, as proposed by Su et al. (2016), is used to obtain pooling specifications to average over. In a Monte Carlo experiment, it is shown that Mallows pooling averaging combined with latent group structure identification using Classifier-Lasso has the best performance in terms of MSPE when a moderate or large degree heterogeneity is present across individuals. Especially when the amount of individuals in the panel is large, the performance of the newly proposed way to obtain pooling specification using Classifier-Lasso has clear advantages over the method proposed by Wang et al. (2017). When a low degree of heterogeneity is present, the Classifier-Lasso estimator as proposed by Su et al. (2016) has the best performance. For this estimator, its finite sample performance is also investigated. The newly proposed method is also used to predict changes in sovereign credit default swap spread for a cross-country panel as an illustration.

1 Problem Description

As early as in 1970, panel models have been used to successfully capture the effects of heterogeneity across individuals in a panel (Swamy, 1970). However, care should be taken when choosing between a panel model with or without heterogeneous coefficients across individuals, as the results of available methods might vary greatly.

One can assume that each individual behaves differently and separately estimate the parameters for each individual. In this case, the coefficients are allowed to be heterogeneous across individuals. This approach takes full account of the heterogeneity across individuals. Among others Durlauf, Kourtellos and Minkin (2001) have used this approach to account for heterogeneity across countries.

In contrast, one can also completely ignore the effects of heterogeneity by assuming all individuals in the panel behave in the same way. This corresponds with imposing homogeneity across individuals and estimate one common coefficient for each regressor for all individuals. In this case one ignores the heterogeneity that might be apparent across individuals. When coefficients actually do vary across individuals, this generally results in biased coefficient estimates. As was pointed out in Durlauf et al. (2001), pooling all individuals often leads to invalid conclusions. However, the increase in efficiency by pooling often leads to better predictions in terms of Mean Squared Error (MSE) or Mean Squared Prediction Error (MSPE) (Baltagi and Griffin, 1997).

Wang, Paap and Zhang (2017) (hereafter WPZ) analyze this econometric dilemma to choose whether to pool or not to pool in panel models. As they point out, it is unclear beforehand which method would work best. In various empirical applications both methods have been applied, with no found strict dominance of the one method over the other. The decision whether to pool, involves a bias-variance tradeoff. Individual estimates usually result in a consistent, but not very efficient estimate, whereas pooling increase the efficiency of the estimates, at the cost of introduced bias.

Many estimators exist which are between these two extremes. However, few estimators make this explicit tradeoff between bias and efficiency. WPZ use a new technique to explicitly make this bias-variance tradeoff to construct estimations which are directly based on the MSE. This seems valuable as the MSE is a popular criterion for evaluation and model comparison. In order to make this tradeoff, WPZ propose pooling average methods, which combine different pooling specifications with appropriate weights. These method do not try to find one single heterogeneity structure to group the individuals in the panel, but use different specifications to create the best estimator in terms of MSE. They conclude that their Mallows pooling averaging (MPA) method is preferred in non-extreme cases of static panel models with heterogeneous parameters across individuals.

In order to reduce the amount of possible pooling specification to combine, WPZ use a screening procedure that does not require the estimation of all candidate models. However, as all the pooling average estimators use these chosen pooling specifications to average over, the method to find the pooling specifications might have large impact. Apart from this screening procedure proposed in WPZ, other methods can be used to obtain pooling specifications. When using a different methods to obtain the pooling specifications, the performance of the screening procedure proposed by WPZ can be assessed.

In this paper, the screening procedure applied in WPZ are compared to the Classifier Lasso (C-Lasso) approach with Partial Profile Likelihood (PPL) estimation from Su, Shi and Phillips (2016) (hereafter SSP). This method is used to obtain the pooling specifications to average over, just like the screening procedure proposed by WPZ. While the research by WPZ was focused on the MSE, this paper focuses on the MSPE for evaluation, which is frequently used when evaluating predictions in practice. In a Monte Carlo experiment, all pooling averaging methods used in WPZ are used in combination with the two methods to obtain the pooling specifications. In addition, the infeasible Oracle estimator is added which uses the given group structure in the data generating processes. This allows for a more detailed analysis of the bias-variance tradeoff. It is shown that the MPA methods sometimes outperform this infeasible estimator in terms of MSPE, by making this explicit bias-variance tradeoff. Next, The (finite sample) performance of the the C-Lasso estimator as proposed by SSP is investigated. Finally, The newly proposed method to obtain pooling specifications is used to make predictions for changes in sovereign credit risk in a cross-country panel, which is also analyzed in WPZ.

Section 2 briefly reviews some of the methods that are available to estimate parameters in heterogeneous panel data models. Section 3 introduces the model setup used in this study. Section 4 describes the used existing estimators. Section 5 elaborates on the newly proposed pooling averaging method. Section 6 describes the Monte Carlo experiment set-up and discusses the results. Section 7 contains an empirical application of the pooling average method. Finally, Section 8 concludes and points out topics for further research.

2 Literature Review

In the literature on panel models, many possible approaches exist to estimate heterogeneous parameters. One can make use of an average effect estimator, which estimates common coefficients for all individuals. Swamy (1970) proposed one of the first methods to adjust for varying parameters with his generalized least squares (GLS) type estimator. A different average effect estimator is the mean group estimator proposed by Pesaran and Smith (1995). One could also use the FGLS estimator to incorporate the variation of coefficients (Hsiao, 2014). These methods are similar in the sense that they use information from individual regressions to construct an estimator for the average effect.

When it is of interest to obtain individual coefficients, one can use other techniques. The easiest way to obtain individual estimates is to use the OLS estimator for each individual separately, which is referred to as the individual estimator. More sophisticated methods try to incorporate information from the other individuals to improve the estimation of the individual parameters, such as the shrinkage estimator proposed by Maddala, Trost, Li and Joutz (1997) or the mixed estimator of Lee and Griffiths (1979).

More recent methods estimate coefficients for groups of individuals with similar characteristics, which results in a grouped estimator. This has both the favorable property of increasing the efficiency, while decreasing the introduced bias compared to imposing homogeneity across all individuals.

When the grouping of the individuals is known to the researcher, one can choose to ignore the estimation of group structures in the data. Among others, the grouped fixed effects estimator proposed by Bester and Hansen (2016) can then be used.

However, if the group structure is not known to the researcher, the latent group structure has to be estimated. Among others, Sun (2005), Lin and Ng (2012), Bonhomme and Manresa (2015), Ando and Bai (2016) and SSP propose methods to identify a latent group structure in the data, without requiring any a priori knowledge about the grouping. SSP introduced a new variant of the lasso estimator, which is applied in this paper. This method simultaneously retrieves a latent group structure and the corresponding slope coefficients by using Lasso technology, without having any a priori knowledge about the number of groups.

WPZ proposed to combine the estimators from different pooling specifications by using pooling averaging. They argue that this avoids certain problems with pretest estimators, discussed in Danilov and Magnus (2004). In addition, by using the Mallows criterion to estimate weights for the different pooling specifications, one can asymptotically achieve the lowest possible error (Hansen, 2007).

As this is this an extension of WPZ, The goal remains to investigate which methods lead to the most accurate estimator for different sample sizes and heterogeneity structures. WPZ focussed on parameter accuracy, whereas this paper focusses on predicition error. Specific to this paper is the comparison of two methods that identify the heterogeneity structure: the screening procedure as proposed in WPZ, and the latent group structures identification by using the C-Lasso with PPL estimation as proposed by SSP. How these methods are employed is described in section 5.3.

3 Model Specification

Following WPZ, the specification from (1) is used for the static linear panel data model with heterogeneous slopes.

$$y_i = X_i\beta_i + u_i \qquad i = 1, \cdots, N. \tag{1}$$

Here, $y_i = (y_{i1}, \cdots y_{iT})'$ and $X_i = (X'_{i1}, \cdots, X'_{iT})'$ is a $T \times K$ matrix of explanatory variables, including an intercept as the first explanatory variable. Thus $X_{it,1} = 1$ for $t = 1, \cdots, T$. The series $\{y_{it}, X_{it}\}$ is assumed to be stationary. The coefficient $\beta_i = (\beta_{i1}, \cdots, \beta_{iK})'$ is assumed to be fixed over time, but is allowed to vary across individuals to allow for heterogeneity. This means that elements from β_i can differ from β_j when $i \neq j$. The error terms of each individual u_i are assumed to be independently and identically distributed (IID) across time. Next to this, $u_1, \cdots, u_N$ are assumed to be uncorrelated conditional on X_i for all i. The error term is allowed to be heteroskedistic across individuals. That is, u_i has mean zero and variance $\sigma_i^2 I_T$.

The Monte Carlo simulations is focused on cases where only exogenous regressors are used. In this case, the model specification from (1) is correct.

4 Existing methods

Heterogeneity across individuals can be accounted for in various ways. This section elaborates on how some existing methods estimate coefficients in panel models which are possibly heterogeneous. For each of the methods used in this study, the way they can be estimated is described in the corresponding subsection. First, common estimators are described. Second, individual estimators are described.

4.1 Common Estimators

First, the methods are described which use the same coefficients across all individuals. Although these estimators do not produce individual specific estimators, they might still perform well due to their increased efficiency. If the introduced bias is offset by the increase in efficiency, a better bias-variance tradeoff is made.

The most simple common estimator is the pooled estimator as defined in (2). Coefficients are assumed to be homogeneous across individuals when using this estimator.

$$\hat{\beta}_{pool} = \left(b^{'}, \cdots, b^{'}\right)^{'}, \quad \text{where } b = \left(\sum_{i=1}^{N} X_i^{'} X_i\right)^{-1} \sum_{i=1}^{N} X_i^{'} y_i \tag{2}$$

This pooled estimator ignores all heterogeneity and thus can be severely biased. However, because all data is used for this one parameter estimate b, this estimator is more efficient.

The second average effect estimator that is being used, is the FGLS estimator of the average effect as defined in (3).

$$\begin{aligned} \hat{\bar{\beta}}_{FGLS} &= \left(\sum_{i=1}^{N} X_i' \hat{\Psi}_i^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i' \hat{\Psi}_i^{-1} y_i\right), \text{with } \hat{\Psi}_i = X_i' \hat{\Delta} X_i + \hat{\sigma}_i^2 I_T \\ \text{and } \hat{\Delta} &= \frac{1}{N-1} \sum_{i=1}^{N} \left(\hat{\beta}_{i,ind} - \frac{1}{N}\sum_{i=1}^{N} \hat{\beta}_{i,ind}\right) \left(\hat{\beta}_{i,ind} - \frac{1}{N}\sum_{i=1}^{N} \hat{\beta}_{i,ind}\right)' \end{aligned} \tag{3}$$

where the $\hat{\beta}_{i,ind}$ denotes the individual estimator which is described in the next section, and $\hat{\sigma}_i^2 = (y_i - X_i\hat{\beta}_{i,ind})'(y_i - X_i\hat{\beta}_{i,ind})/(T-K)$. This FGLS estimator incorporates variation in the coefficients to produce a common estimate for all individuals (Hsiao, 2014, Section 6.2.2b).

More average effect estimators are available such as the the mean group estimator of Pesaran and Smith (1995) or the average effect GLS type estimator of Swamy (1970). However, these estimators are consistently outperformed by other estimators in the Monte Carlo experiment. For the sake of brevity, the description of these estimators has been omitted.

4.2 Individual Estimators

Instead of using a common estimate for all individuals, it is also possible to estimate coefficients separately for each individual. When heterogeneous policy decisions need to be made, the use of individual estimators allows for differentiation across countries or companies which make part of the panel.

The most simple way to create individual estimators is by using the ordinary least square (OLS) estimator for each individual separately. This amounts to using the estimator as defined in (4).

$$\hat{\beta}_{i,ind} = (X_i'X_i)^{-1}X_i'y_i, \quad \text{where } i = 1, \cdots, N \quad \text{and } T > K \tag{4}$$

This individual estimator $\hat{\beta}_{i,ind}$ is consistent if the model from (1) is correctly specified for this individual. This is the case in this study, as a static linear panel model is used. However, because it only uses a low amount of observations, it is not as efficient as the common estimators from the previous section.

The second individual estimator that is used in this study, is the shrinkage estimator of Maddala et al. (1997) as denoted in (5).

$$\hat{\beta}_{i,shrinkage} = \left(1 - \frac{v}{F}\right)\hat{\beta}_{i,ind} + \frac{v}{F}\hat{\beta}_{pool} \tag{5}$$

where $v = [(N-1)K-2]/[NT-NK+2]$ and F the test statistic for the null hypothesis $H_0 : \beta_1 = \cdots = \beta_N$, which relates to full homogeneity in the parameters across individuals.

The third individual estimator is the mixed estimator from Lee and Griffiths (1979) as denoted in (6). This estimator uses the estimate from the FGLS average effect from the previous section for its individual parameter estimation.

$$\hat{\beta}_{i,mix} = \hat{\bar{\beta}}_{FGLS} + \hat{\Delta}X_i'\left(X_i\hat{\Delta}X_i' + \hat{\sigma}_i^2 I_T\right)^{-1}(y_i - X_i\hat{\bar{\beta}}_{FGLS}) \tag{6}$$

with $\hat{\Delta}$ as defined in (3) and $\hat{\sigma}_i^2 = (y_i - X_i\hat{\beta}_{i,ind})'(y_i - X_i\hat{\beta}_{i,ind})/(T-K)$.

5 Pooling Averaging methods

Pooling averaging is a method newly proposed by WPZ. By using pooling averaging, different pooling specifications can be combined and weighted. This section elaborates on which steps need to be taken to make use of pooling averaging methods. First, a method to impose restriction on coefficients in a panel model is described. Second, the specification and advantages of pooling averaging is described. Third, the methods to obtain candidate pooling specifications is elaborated upon. Finally, the methods to weight the candidate pooling specifications are described.

5.1 Parameter Restrictions

For every a pooling specification, some coefficients are restricted to have equal value. These equality restriction can be imposed by using a restriction matrix such that $R_m\beta = 0$, where $\beta = (\beta_1', \cdots, \beta_N')'$ is a $NK \times 1$ vector. When one wants to impose the restriction $\beta_i = \beta_j$ for $j > i$, then one can use the restriction matrix as defined in (7).

$$R_m = \left(0_{K\times(i-1)K}, I_K, 0_{K\times(j-i-1)K}, -I_K, 0_{K\times(N-j)K}\right) \tag{7}$$

This allows us to construct the projection matrix P_m as defined in (8).

$$P_m = I_{NK} - (X'X)^{-1}R_m'(R_m(X'X)^{-1}R_m')^{-1}R_m \tag{8}$$

where $X = diag(X_1, \cdots, X_N)$ is an $NT \times NK$ matrix. In this case, all parameters of an individual are imposed to be equal to those of another individual. It is also possible to only impose specific coefficients to be equal such that $\beta_{i,l} = \beta_{j,l}$ but not necessarily $\beta_{i,m} = \beta_{j,m}$ for $l \neq m$.

By using this projection matrix, one can construct the grouped OLS estimator by $\hat{\beta}_{(m)} = P_m\hat{\beta}_{ind}$ with $\hat{\beta}_{ind} = \left(\hat{\beta}_{1,ind}', \cdots, \hat{\beta}_{N,ind}'\right)'$ being the $NK \times 1$ vector of individual OLS estimators. Here the m denotes under which pooling strategy the estimators are grouped. A pooling strategy consist of a set of restrictions on which coefficients are imposed to be equal. Typically, Some individuals are imposed to have the same coefficients, while others are allowed to differ. Each pooling strategy is characterized by a different R_m matrix, which results in a grouped estimator, $\hat{\beta}_{(m)}$, with a different degree of bias and variance.

5.2 Pooling Averaging

The newly proposed pooling averaging method combines different pooling specification to obtain parameter estimates. The pooling averaging estimator $\hat{\beta}(w)$ is given in (9).

$$\hat{\beta}(w) = \sum_{m=1}^{M} w_m\hat{\beta}_{(m)} = \sum_{m=1}^{M} w_m P_m\hat{\beta}_{ind} = P(w)\hat{\beta}_{ind} \tag{9}$$

where M is the number of candidate pooling specifications, $P(w) = \sum_{m=1}^{M} w_m P_m$ is an $NK \times NK$ projection matrix which is used to generate the pooling average estimates for each individual, conditional on weighting w as in (8). Here $w = (w_1, \cdots, w_M)'$ belongs to the set $\mathcal{W} = \{w \in [0,1]^M : \sum_{m=1}^{M} w_m = 1\}$. After combining the different estimators from the proposed pooling specifications, individual specific parameters are obtained.

Pooling averaging has the advantage that a bias-variance tradeoff can be made, as every pooling specifications has a different degree of bias and variance. By attributing weights based on the bias and variance properties of the pooling specifications, this tradeoff can be made explicitly. Next to this, the problems caused by pretesting are avoided, as this estimator is continuous and has a bounded risk (Danilov and Magnus, 2004). However, conducting inference is challenging, as only a point estimate is provided. WPZ propose to calculate confidence intervals by bootstrap, using cross-sectional resampling following Kapetanios (2008). In the empirical application, this method is used.

5.3 Candidate Pooling Specifications

The pooling strategies to construct the restriction matrix R_m as defined in (7) are usually not known to the researcher. Except for when one wants to fix the group structure to some clustering apparent from the data (for example SIC codes or geographical location), Some method has to be used to select candidate pooling specifications among all possible pooling strategies. The amount of possible pooling strategies rises very fast as the amount of individuals (N) or the amount of regressors (K) increase. Therefore, checking all possible specifications quickly becomes computationally difficult. Instead, econometric techniques have to be used to retrieve the best pooling specifications.

In this study, two methods are used to retrieve grouping strategies. The first method is the classifier-Lasso (C-Lasso) method by using penalized profile likelihood (PPL) estimation for latent group structure identification (hereafter LGSI), proposed by Su et al. (2016). The second method is the screening procedure (hereafter SP) proposed by Wang et al. (2017). It is the goal of this paper to investigate the relative performance when using the SP. By comparing this method to the LGSI, the performance of the computationally efficient SP can be assessed. In addition to these two methods, the infeasible Oracle estimator which uses the true pooling specification is added.

5.3.1 Latent Group Structure Identification

In order to identify the latent group structures and estimate parameters in panel data, the C-Lasso method can be used. In the case of (non-)linear panel models without endogeneity and with or without dynamic structure, PPL estimation can be used. (Su et al., 2016) The PPL estimation uses a profile log-likelihood function in combination with a penalty term, which is described in two steps.

For a linear panel model with heterogeneity across individuals in both the intercept and regressors, this would result in a profile log-likelihood function as defined in (10).

$$Q_{NT}(\hat{\beta}_{PPL}) = \min_{\hat{\beta}_{PPL}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{2}(y_{it} - \hat{\beta}'_{i,PPL} x_{it})^2 \tag{10}$$

Here, x_{it} is a $K \times 1$ vector of exogenous variables, with $x_{it,1} = 1$ as the intercept and y_{it} as the corresponding outcome. $\hat{\beta}_{i,PPL}$ is a $K \times 1$ vector with time-invariant, individual specific parameters and $\hat{\beta}_{PPL} = (\hat{\beta}'_{1,PPL}, \cdots, \hat{\beta}'_{N,PPL})'$ a $NK \times 1$ vector containing the parameters for all individuals.

In order to make sure that the $\hat{\beta}_{i,PPL}$ parameters are restricted to be equal when two individuals are in the same cluster, a mixed additive-multiplicative penalty term is added to obtain the PPL criterion function as defined in (11). minimizing this criterion function produces the C-Lasso estimates $\hat{\beta}_{PPL}$ and $\hat{\alpha}_{PPL}$ for a given amount of clusters, S.

$$Q^{(S)}_{NT,\lambda}(\hat{\beta}_{PPL}, \hat{\alpha}_{PPL}) = Q_{NT}(\hat{\beta}_{PPL}) + \frac{\lambda}{N} \sum_{i=1}^{N} \prod_{s=1}^{S} ||\hat{\beta}_{i,PPL} - \hat{\alpha}_s|| \tag{11}$$

Here, λ is a tuning parameter, $\hat{\alpha}_{PPL} = (\hat{\alpha}_1, \cdots, \hat{\alpha}_S)$, with $\hat{\alpha}_s$ being the $K \times 1$ vector with parameters for cluster s.

This method imposes the $\hat{\beta}_{i,PPL}$ parameters to follow a group pattern such that $\hat{\beta}_{i,PPL} = \sum_{s=1}^{S} \hat{\alpha}_s \mathbf{1}\{i \in \hat{G}_s\}$, where $\hat{\alpha}_j \neq \hat{\alpha}_k$ for $j \neq k$, $\cup_{s=1}^{S} \hat{G}_s = \{1, 2, \cdots, N\}$ and $\hat{G}_k \cap \hat{G}_j = \emptyset$ for any $j \neq k$. This ensures that all individual belong to a certain cluster $\hat{G}_s$,which have the same parameters $\hat{\alpha}_s$, as desired for a grouped estimator. By using this criterion function, the C-Lasso achieves simultaneous classification and consistent estimation in a single step. For more details see Su et al. (2016, Section 2).

In order to obtain multiple pooling specifications to average over, we pick a range of values for the amount clusters, S. For each amount of clusters, $\hat{\beta}_{PPL}$ and $\hat{\alpha}_{PPL}$ get estimated using PPL C-Lasso. Here, it is possible to limit the maximum amount of clusters S_{max} to reduce the computation time, or limit the amount of clusters in the panel. In total, $M = S_{max}$ candidate pooling specifications $\hat{\beta}_{(m)}$ are formed as described in (9). Here, the $\hat{\beta}_{(m)}$ corresponds to the $\hat{\beta}_{PPL}$, which is obtained with the LGSI for one value of S.

5.3.2 Screening Procedure

WPZ propose a screening method to rule out the "poor" models that incorrectly impose equality restrictions on parameters that should not be imposed to be equal. This overcomes the difficulty of having to select or average over the full model space, which is computationally difficult.

In this screening method, the estimated individual coefficients are normalized and the Bhattacharyya distances between coefficients are being calculated (Wang et al., 2017, equation (36)). Afterwards, agglomerative hierarchical clustering (AHC) is employed to form a hierarchical tree which maps which coefficients that are close to each other. After the tree is formed, the tree can be cut to produce the pooling specifications. The amount of clusters that are created for each variable, denoted as C, can be selected by the researcher. For $C = 1, \cdots, N$, a clustering for each variable is formed. Here, $C = 1$ corresponds to using the pooled estimator, while $C = N$ corresponds to the individual estimator.

Note that this method differs from clustering as is done in the LGSI. Figure 1 shows how the LGSI and SP methods differ for the case where $S = C = 2$. For the LGSI, clustering is done per individual, as is visible in the left table. For the SP, two clusters are created for each variable, which is visible in the right table. This means that in this case, this could effectively result in four different clusters of individuals if the clustering per variable does not coincide. This more flexible specification of clusters allows to capture a larger extent of heterogeneity. As a result, this might cause for the pooling specifications of the SP to have a relatively low degree of bias, compared to the pooling specification of the LGSI. As minimizing MSPE is the goal, it is not clear which of the two methods is more suitable. In the Monte Carlo simulation, these techniques are compared.

(LGSI)

		i			
		1	**2**	**3**	**4**
	1	1	1	2	2
k	**2**	1	1	2	2
	3	1	1	2	2

(SP)

		i			
		1	**2**	**3**	**4**
	1	1	2	2	2
k	**2**	1	1	2	2
	3	1	1	1	2

Figure 1: Visualization of a possible clustering for $N = 4, K = 3$ with $C = S = 2$. Numbers in cells indicate to which cluster the parameter k for individual i belongs. Left is the clustering on individuals, which is applied in the LGSI. Right is the clustering on each parameter, which is applied in the SP.

An advantage of the SP compared to the LGSI, is that it is possible to create clustering specifications with amounts of clusters up to N, whereas this is computationally difficult for the LGSI. A disadvantage of the SP is that is uses no objective function that is directly minimized and there might be some tradeoff between efficiency loss and diversification gains from having many candidate models.

5.3.3 Infeasible Oracle Estimator

As the true group structure is known in a Monte Carlo experiment where the Data generating process (DGP) is defined by the researcher, one is able to construct the infeasible Oracle estimator which uses the optimal pooling specification. The pooling specification is optimal in the sense that there is minimal introduced bias from pooling individuals, because they get grouped according to the given grouping in the DGP. The Oracle estimator as defined in (12) uses the true grouping m_0 which is given in the DGPs of the Monte Carlo simulation.

$$\hat{\beta}_{oracle} = \hat{\beta}_{(m_0)} \tag{12}$$

Of course, this true grouping is not known in practice, but adding this estimator in the Monte Carlo simulation allows us to see the performance of the estimator with the correct grouping structure. In terms of MSPE, this estimator could still be outperformed if the gain in efficiency is larger than the introduced heterogeneity bias when pooling. This can especially be the case when the heterogeneity across individuals is small. In this case, The increase in efficiency by pooling might be larger than the introduced bias, which results in a better prediction in terms of MSPE.

The same applies when the pooling structure needs to be estimated. Even if the true grouping would be identified, this would still not guarantee optimal predictions in terms of MSPE. Therefore, one would like to apply a different way of choosing among the different pooling specifications, such as pooling averaging.

5.4 Pooling Averaging Weights

The last step before obtaining estimates with pooling averaging, is choosing a method for the attribution of weights for the candidate pooling specifications. This corresponds to choosing the weights w_m from (9) to weight the candidate pooling specifications $\hat{\beta}_{(m)}$ into the final pooling averaging estimate $\hat{\beta}(w)$. When LGSI is used, m ranges from $m = 1, \cdots, M = S_{max}$. When SP is used, m ranges from $m = 1, \cdots, M = N$. Four methods are considered to choose the pooling averaging weights.

5.4.1 Simple Average

The first and most simple method is using equal weight for all candidate models. Thus, $w_m = \frac{1}{M}$ for $m = 1, \cdots, M$. This method is favored in the forecast combination literature. The effectiveness of this method is analyzed in the theoretical framework provided by Claeskens, Magnus, Vasnev and Wang (2016). For pooling averaging, this simple average is also expected to have good finite sample performance because of possible inaccuracy in estimating the weights due to errors in coefficient estimates and strong correlation between the estimates. Here, the simple average trades off a small bias for a larger gain in efficiency. (Wang et al., 2017)

5.4.2 Information Criterion

The second method that is applied, is using an information criterion (IC) to either create weights, or select one candidate model. The two information criteria that are used are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), where $AIC = 2K_m - 2\ln \hat{L}_m$ and $BIC = \ln(NT)K_m - 2\ln \hat{L}_m$, where K_m refers to the amount of different regressors used in the pooling specification and $\hat{L}_m$ being the maximized loglikelihood candidate pooling specification m. Selecting the two model specifications with the lowest values for these information criteria produces two pretest estimators. In this case the weight for the selected pooling specification is set to one, while the other weights are set to zero.

In addition to choosing one pooling specification with the information criteria, one can also use them to create weights. Buckland, Burnham and Augustin (1997) introduced the smoothed information criteria as defined in (13).

$$w_m^{AIC} = \frac{\exp(-AIC_m/2)}{\sum_{m=1}^{M} \exp(-AIC_m/2)} \quad \text{and} \quad w_m^{BIC} = \frac{\exp(-BIC_m/2)}{\sum_{m=1}^{M} \exp(-BIC_m/2)} \tag{13}$$

Here AIC_m and BIC_m are adjusted by subtracting the minimum AIC and BIC values over the different pooling specifications to avoid numerical problems.

5.4.3 C-Lasso Information Criterion

The third method to attribute weights to the candidate models is only applied when LGSI is used. Here, one chooses the candidate model which minimizes the IC as defined in (14), proposed by SSP. The use of this IC in combination with LGSI results in the C-Lasso estimator, which is asymptotically equivalent to the Oracle estimator as defined in (12).

$$IC(S, \lambda) \equiv \ln\left(\frac{1}{NT}\sum_{s=1}^{S}\sum_{i\in\hat{G}_s(S,\lambda)}\sum_{t=1}^{T}\left(y_{it} - \hat{\beta}'_{i,PPL}x_{it}\right)^2\right) + \rho_{NT}KS \tag{14}$$

Here the second summation sums up over the individuals that belong to cluster s, and uses their corresponding parameter $\hat{\beta}_{i,PPL} = \hat{\alpha}_s$. For linear models, SSP propose to use the tuning parameter $\rho_{NT} = \frac{2}{3}(NT)^{-1/2}$.

This IC is used to select the optimal amount of clusters, such that $\hat{S} = \text{argmin}_{1\leq S\leq S_{max}} IC(S, \lambda)$. For this optimal amount of clusters, $\hat{S}$, the $\hat{\beta}_{C\text{-}Lasso}$ is selected by using $w_{\hat{S}} = 1$ while all the other weights are being set to zero.

5.4.4 Mallows Pooling Averaging

The fourth and last method that is used to construct weights is by using the Mallows criterion. Hansen (2007) proposed to use this criterion, which is asymptotically optimal, as it achieves the lowest possible squared error.

To derive the Mallows pooling averaging (MPA) criterion in estimating a heterogeneous panel, the approach from WPZ is followed. This corresponds to the case where one is interested in prediction and thus chooses $A = X'X$ in equation (26) from WPZ. The Mallows Criterion is an unbiased estimator of the squared risk, which makes it a good approximate for the MSPE (Wang et al., 2017, Theorem 5.1). In addition, WPZ derived the risk bounds of the MPA, which tells us how the MPA performs in the worst situation.

The feasible variant of the Mallows criterion is described in (15).

$$\mathcal{C}^*_{X'X}(w) = ||XP(w)\hat{\beta}_{ind} - X\hat{\beta}_{ind}||^2 + 2tr[P'(w)X'X\hat{V}] - ||X\hat{\beta}_{ind} - X\beta||^2 \quad (15)$$

where $||\theta||^2 = \theta'\theta$, $\hat{\beta}_{ind} = (\hat{\beta}'_{1,ind}, \cdots, \hat{\beta}'_{N,ind})'$ is an $NK \times 1$ vector with the individual estimates, $X = diag(X_1, \cdots, X_N)$ is an $NT \times NK$ matrix with observations and β the $NK \times 1$ vector which corresponds to the true parameters in the DGP.

$\hat{V} = diag(\hat{V}_1, \cdots, \hat{V}_N)$ correspond to the estimates of the variance of the individual estimates $\hat{\beta}_{i,ind}$ for $i = 1, \cdots, N$. For the estimation of $\hat{V}$ three approaches are proposed by WPZ. Each of the approaches uses different assumptions about the error structure.

1. Homoskedasticity: If one assumes equal variance for all individuals, such that $var(u_i) = \sigma^2 I_T$ for all i, V can be estimated by $\hat{V}_{homo} = \tilde{\sigma}^2(X'X)^{-1}$ where $\tilde{\sigma}^2 = (y - X\hat{\beta}_{ind})'(y - X\hat{\beta}_{ind})/(NT - NK)$, where $y = (y'_1, \cdots, y'_N)'$.

2. Between-individual heteroskedasticity: If one assumes that every individual is allowed to have a different variance, such that $var(u_i) = \sigma_i^2 I_T$ for all i, V can be estimated by $\hat{V}_{bh} = diag(\hat{\sigma}_1^2(X'_1X_1)^{-1}, \cdots, \hat{\sigma}_N^2(X'_NX_N)^{-1})$ where $\hat{\sigma}_i^2 = (y_i - X_i\hat{\beta}_{i,ind})'(y_i - X_i\hat{\beta}_{i,ind})/(T - K)$.

3. Completely (conditional) heteroskedasticity: If we assume that u_i is (conditionally) heteroskedastic for each $i = 1, \cdots, N$, one can use

$$\hat{V}_{ch} = \frac{T}{(T-K)} diag\Bigg((X'_1X_1)^{-1}\sum_{t=1}^{T}\hat{u}_{1t}^2 X'_{1t}X_{1t}(X'_1X_1)^{-1}, \cdots \\ , (X'_NX_N)^{-1}\sum_{t=1}^{T}\hat{u}_{Nt}^2 X'_{Nt}X_{Nt}(X'_NX_N)^{-1}\Bigg),$$

   where $\hat{u}_{it}^2 = (y_{it} - X_{it}\hat{\beta}_{i,indiv})^2$ for $i = 1, \cdots, N$ and $t = 1, \cdots, T$.

As the third term in (15) is not affected by the choice of w, this function can be rewritten into a quadratic function in w, which can be minimized to obtain the weights that are used for the pooling averaging by $\hat{w}^* = (\hat{w}_1, \cdots, \hat{w}_M)' = \text{argmin}_{w \in \mathcal{W}} \mathcal{C}^*_{X'X}(w)$, with $\mathcal{W}$ as defined in section 3.5. MPA is applied with all three error structures, which results in three different weight vectors $\hat{w}^*$ for one set of candidate pooling specifications $\hat{\beta}_{(m)}$ where $m = 1, \cdots M$.

WPZ show that the MPA estimator $\hat{\beta}(\hat{w}^*)$ is asymptotically optimal, as the squared loss is asymptotically identical to that by the infeasible best possible model averaging estimator, conditional on a given set of estimators as in Hansen (2007).

6 Monte Carlo Simulation

In order to evaluate the performance of the different pooling strategies. All described methods are tested in a Monte Carlo experiment. The DGPS used in

WPZ, and one additional DGP are used. In particular, this experiment compares the performance of the two methods used to obtain pooling specifications. The first method being the LGSI proposed by SSP and the second method being the SP proposed by WPZ.

First, the DGPs are be described. Second, the chosen parameters for the methods and evaluation are decribed. Finally, the results are presented.

6.1 Data Generation Process

Five different DGPs are used to evaluate the performance of the described methods.

The benchmark setup is the static panel data model with coefficients possibly varying over individuals, but constant over time as defined in (16).

$$y_{it} = \sum_{l=1}^{3} x_{it,l}\beta_{i,l} + \epsilon_{it}, \quad i = 1, \cdots, N; \quad t = 1, \cdots, T \tag{16}$$

where $x_{it,1} = 1$ and the remaining regressors are independently generated from the standard normal distributions. The idiosyncratic errors ϵ_{it} are independently, normally distributed with mean zero and variance $\sigma^2_{\epsilon_i}$. The errors are thus heteroskedastic across individuals, but homoskedastic over time for each individual. The variances are defined as

$$\sigma^2_{\epsilon_i} = \frac{\sum_{l=1}^{3} \beta^2_{i,l}}{R^2} - \sum_{l=1}^{3} \beta^2_{i,l}$$

such that a theoretical R^2 can be picked, which is fixed at $R^2 = 0.9$ in this simulation. The slope coefficients $\beta_{i,l}$ have a different grouping pattern for every DGP, which is described below.

DGP 1 (Homogeneous): $\beta_{i,l} = 1$ for all i and l.

DGP 2 (Weakly heterogeneous):

$$\beta_{i,1}, \beta_{i,2} = \begin{cases} q_1, i = 1, \cdots, [N/2] \\ q_3, i = [N/2] + 1, \cdots, N \end{cases} \qquad \beta_{i,3} = \begin{cases} q_1, i = 1, \cdots, [N/3] \\ q_3, i = [N/3] + 1, \cdots, N \end{cases}$$

where q_i is the i-th element of the q vector defined below and $[N/2]$ denotes the nearest integer to $N/2$, where .5 is rounded upwards.

DGP 3 (Strongly heterogeneous):

$$\beta_{i,1}, \beta_{i,2} = \begin{cases} q_1, i = 1, \cdots, [N/4] \\ q_2, i = [N/4] + 1, \cdots, [2N/4] \\ q_3, i = [2N/4] + 1, \cdots, [3N/4] \\ q_4, i = [3N/4] + 1, \cdots, N \end{cases} \qquad \beta_{i,3} = \begin{cases} q_1, i = 1, \cdots, [N/5] \\ q_2, i = [N/5] + 1, \cdots, [2N/5] \\ q_3, i = [2N/5] + 1, \cdots, [3N/5] \\ q_4, i = [3N/5] + 1, \cdots, N \end{cases}$$

DGP 4 (Completely heterogeneous): $\beta_{i,l} = 0.1 \times i \times l$ for all i and l.

DGP 5 (Varying over parameters):

$$\beta_{i,1} = \begin{cases} q_1, i = 1, \cdots, [N/4] \\ q_3, i = [N/4] + 1, \cdots, N \end{cases} \qquad \beta_{i,3} = \begin{cases} q_1, i = 1, \cdots, [3N/4] \\ q_3, i = [3N/4] + 1, \cdots, N \end{cases}$$

$$\beta_{i,2} = \begin{cases} q_1, i = 1, \cdots, [N/2] \\ q_3, i = [N/2] + 1, \cdots, N \end{cases}$$

Two sets of coefficients q are considered: The set $q_B = [1.0, 1.5, 3.3, 3.0]'$ and the set $q_S = [1.0, 1.2, 2.3, 2.0]'$ with coefficients which are less far from each other. The choice of q only affects DGP 2,3 and 5, as the other DGPs do not use these coefficients.

The amount of individuals vary from $N \in \{5, 10, 30\}$ and the amount of time periods per individual vary from $T \in \{15, 40\}$, which results in six different combinations of N and T. Because of limitations on computational power, this simulation is based on 100 replications.

6.2 Methods

All methods that are described in this paper are used in this Monte Carlo simulation. Thus as common effect estimators, the pooled estimator and the FGLS estimator are used. As individual estimators, the separate OLS estimator for each individual, the shrinkage estimator from Maddala et al. (1997) and the mixed estimator from Lee and Griffiths (1979) are used. For the pooling averaging estimators, the simple average weighting, AIB and BIC selection, Smoothed AIC and BIC weighting and Mallows pooling averaging with homoskedastic, between-individual heteroskedastic and completely heteroskedastic error structure is included. All methods that require estimated pooling specifications, are applied both on the pooling specifications estimated with the SP and the LGSI. For the LGSI, the C-lasso estimator with the C-Lasso IC is also included. Finally, the infeasible Oracle estimator is included, as if the group membership for each individual is known.

For the LGSI, a maximum amount of clusters $S_{max} = 5$ is used. Allowing more clusters to be formed increases computational complexity. The true amount of clusters of individuals in the DGPS are $1, 3, 7, N$ and 4 clusters for DGP 1 to 5 respectively. (with the exception for the case $N = 5$ for DGP 3, which has 5 true clusters). This means that for DGP 3 and 4, the real amount of clusters in the DGP is not included in the amount of clusters which are used to estimate the pooling specifications for LGSI. Thus, for these DGPs, the true pooling specification can not be found at all. Therefore, the C-Lasso estimator can not be equivalent to the infeasible Oracle estimator in these cases. Nevertheless, it is interesting to see how the performance varies when the real group structure can not be estimated due to a cap on the maximum amount of clusters.

In order to guarantee convergence for the LGSI, the parameter $\lambda = c\, s_y^2 T^{-1/3}$ has been fine-tuned for each DGP. The best results were obtained when using the values $2^{-2}, 2^{-4}, 2^{-4}, 2^{-6}, 2^{-4}$ for c in DGP 1 to 5 respectively, where s_y^2 corresponds with the sample variance of y. Under this choice of tuning parameters, the correct number of groups is chosen with probability approaching one by the C-Lasso IC criterion as N and T approach infinity (Su et al., 2016). This simulation shows the finite sample properties of this estimator.

The methods are evaluated on squared loss of predictions $L(w) = ||X\hat{\beta} - X\beta||^2$, where $\hat{\beta}$ corresponds to the $NK \times 1$ vector with the coefficients estimates using one of the described methods. In the tables with results, the relative MSPE is displayed. This relative MSPE is obtained by dividing the MSPE by the MSPE when using the individual estimator.

6.3 Results

The results are discussed separately for each DGP. This allows for easy comparison of the results when T, N and q are varied. The simulations for one combination of values for $\{N, T, q\}$ are referred to as a case.

6.3.1 Homogeneous

The results for the simulation of DGP 1 are displayed in table A.1 in the appendix. As expected, the best results are obtained when a pooling estimator is used. In this DGP, the pooling estimator is equivalent to the infeasible Oracle estimator. The performance of the C-Lasso estimator is nearly equivalent to the infeasible Oracle estimator, as one cluster is chosen by the C-Lasso in 98% of the cases. When using MPA, there is a drop in performance. When inspecting the chosen weights for MPA, the weights seem to be spread more equally across cluster sizes when the N increases, while T does not affect the weighting. Especially the completely heteroskedastic error structure seems to pick the smaller amounts of clusters, due to its larger increase in the second term of (15), which causes for larger weights for a specification with a lower amount of clusters.

When comparing the results from the two methods to obtain a pooling specification, the LGSI outperforms the SP in every case. As the amount of individuals increase, the relative performance of the SP seems to worsen, whereas the relative performance of the LGSI increases. One possible explanation for the drop in performance, is the large amount of pooling specifications available when using the SP. When inspecting the MPA weights for the different pooling specifications when using the SP, a large amount of weight is attributed to pooling specifications with multiple clusters. For the cases with $N = 30$, MPACH attributes 58% of the weight to pooling specifications using three or more clusters per variable, while this percentage is 85% for the MPAHomo and MPABH. As this DGP has homogeneous parameters, it is unlikely that this results in good performance.

6.3.2 Weakly Heterogeneous

The results for the simulation of DGP 2 are displayed in table A.2 in the appendix. The best results are obtained when MPA is used with the LGSI. For the cases with $N = 30$, $T = 40$ the C-Lasso estimator seems to have results nearly as good as the infeasible Oracle estimator. this is expected, as they are asymptotically equivalent, given the Oracle property of the C-Lasso estimator (Su et al., 2016). However, for the cases with a lower N, the MPA estimators seem to outperform the C-Lasso estimator. When looking at the amount of clusters chosen by the C-Lasso, it seems that the chosen amount of clusters is lower when (N, T) is lower. For $N = 5$ and $N = 10$ the amount of clusters is chosen to be 3 in 24% of the cases with $T = 15$. When $T = 40$ this amount increases to 82%, which shows how much influence an increase in T has for the C-Lasso estimator.

For the cases with $N = 5$ the difference between the MPA with SP or MPA with LGSI is small. Just as in the previous DGP, the relative performance of the methods with LGSI increase when N is increased. This shows that the LGSI is preferred to the SP for larger N. Once again, a possible explanation is the large weights for specifications with a large amount of clusters. In this DGP, the SP could account for all heterogeneity in parameters by using just 2 clusters per variable. However, for the cases with $N = 30$, an average of 87% of the weight for MPACH is used for pooling specifications with 3 or more clusters per variable. For the MPAHomo and MPABH, this is even 95%.

When inspecting the three different error structures used for MPA, the completely heteroskedastic variant gives best results when the number of individuals is large. For the cases with $N = 5$ the between-individual heteroskedastic variant gives best results. This is expected, as the errors are chosen to be between-individual heteroskedastic in the DGPs.

When looking at the difference between the cases with q_B and q_S, the simulations with q_S seem to have slightly lower relative performance, especially for the cases with $N = 30$. A possible explanation for this, is that more observations are required to estimate the parameters successfully, when the heterogeneity in parameters is smaller.

6.3.3 Strongly Heterogeneous

The results for the simulation of DGP 3 are displayed in table A.3 in the appendix. The best results are obtained when MPA is used with the LGSI. The mixed estimator also performs very well for this DGP, especially for the cases $N = 5$ and $N = 10$. For the cases with $N = 5$ and $N = 10$, the MPA even outperforms the Oracle estimator in 7 out of 8 cases. Here, the introduced bias when picking a lower amount of groups is smaller than the increase in efficiency. This shows that the correct identification of groups should not merely be the goal when one want to decrease prediction error. It is possible to obtain lower prediction errors when using MPA. As the number of individuals increase, the oracle estimator increases in performance. This is expected, as the variance of the Oracle estimator shrinks when the number of individuals per clusters grow.

One would expect that the C-Lasso estimator is equivalent to the the Oracle estimator for cases with larger N and T, but this is not the case for this DGP. For cases with $N = 5$ and $N = 10$, C-Lasso underestimates the amount of clusters by picking 2.34 clusters on average. For $N = 30$, an average of 2.92 clusters are chosen. In any case, it never chooses five clusters. This is unexpected, as one might expect that the specification allowing the most heterogeneity would have the best performance in this case, as the true grouping in the DGP uses seven clusters.

The performance of the MPA once again seems to be better when the LGSI is used, with the difference getting bigger when N increases. When inspecting the differences in performance between the cases with q_B and q_S, it is clear that the prediction errors are larger for q_S, which shows that more observations are required when the differences between parameters are smaller.

The three different error structures used for MPA have very similar results for the cases where $N = 10$ and $N = 30$. This is expected, as the performance with the three structures are asymptotically equivalent (Wang et al., 2017, Theorem 5.3). For the cases with $N = 5$ the completely heteroskedastic variant clearly performs worse.

6.3.4 Completely Heterogeneous

The results for the simulation of DGP 4 are displayed in table A.4 in the appendix. The best results are obtained when using the mixed estimator, while the MPA with LGSI also has comparable performance. In all cases, the mixed estimator outperforms the Oracle estimator, which is equivalent to the individual estimator in this DGP. Especially for the cases with higher N, the individual estimators are outperformed by a larger extent.

When inspecting the amount of clusters that have been selected with the C-Lasso, the selected maximum amount of clusters has large impact. For the cases with $N = 5$, the amount of clusters chosen is 4.96. However, this amount is 3.54 and 2.23 for the cases $N = 10$ and $N = 30$ respectively. Again, a low amount of clusters is chosen when the true amount of clusters is not available. In the cases of $N = 10$, this even results in very bad performance.

For MPA, the cases with $T = 15$ have larger prediction errors than the cases with $T = 40$. The mixed estimator does not seem to have this problem as T increases, which makes it a more suitable estimator for this completely heterogeneous DGP.

Just like with the previous DGP, the three error structures used for MPA seem to have very similar results for higher N. Again, for the cases with $N = 5$ and $N = 10$ the completely heteroskedastic variant has lower performance.

6.3.5 Varying over Parameters

The results for the simulation of DGP 5 are displayed in table A.5 in the appendix. For the cases with $N = 5$ and $N = 10$, the best results are obtained when using the MPA with the SP. This method outperforms the infeasible Oracle estimator in all these cases. For the cases with $N = 30$, MPA with the LGSI and the C-Lasso have the best performance.

This DGP differs only slightly from the weakly heterogeneous DGP. However, there is a clear difference in performance between the two methods to obtain pooling specifications. Due to the differences in constructing pooling specifications as explained in figure 1 in section 5.3.3, the SP is able to obtain the best results. The LGSI needs four clusters to capture all heterogeneity, whereas the SP is able to capture all heterogeneity with just two clusters and thus has more efficient estimates. Although the SP clearly outperforms the LGSI in this DGP, it remains unclear if this DGP is a better abstraction of reality.

The C-Lasso estimator is nearly equal to the performance of the Oracle estimator as four clusters are chosen in 96% of all cases. For the cases with $N = 30$ this leads to the lowest MSPE. The MPA methods give similar results as the Classo estimator in this case.

When inspecting the difference across the error structures of the MPA, it is clear that the completely heteroskedastic variant performs best for the SP, even for the cases with $N = 5$. For the LGSI, the completely heteroskedastic error structure only has decreased performance for the cases $N = 5$, $T = 15$.

7 Explain Sovereign Credit Risk

In addition to the Monte Carlo simulation, the LGSI is also used in the empirical application from WPZ. Here, the determinants of sovereign credit default swap (CDS) are investigated for a cross country panel. A CDS contract is an insurance which protects the buyer from a credit event. The spread, given in basis points, is the insurance premium that buyers have to pay, which reflects the credit risk. Among others, Longstaff, Pan, Pedersen and Singleton (2011) have associated change in CDS spread with changes in macroeconomic variables. WPZ used the pooling averaging method to investigate the CDS determinants of a cross-country panel. This paper investigates the predictive performance for CDS spreads, using different methods to account for heterogeneity in a panel model, including pooling averaging.

A set of variables which is similar to the set used in Longstaff et al. (2011) is used to predict CDS for a panel of 19 countries.[1] The data set contains monthly data of local and global financial indicators of macroeconomic fundamentals and the five-year sovereign CDS. The local variables include local stock market returns ($lstock$), changes in local exchange rates ($fxrate$) and changes in foreign currency reserves ($fxres$). The global variables include the U.S. stock market returns ($gstock$), treasury yields ($trsy$), investment-grade corporate bond spreads (ig), equity premium (eqp), volatility risk premium ($volp$), equity flows (ef) and bond flows (bf). In this analysis, the sample from July 2009 until March 2016 is used.[2] Preliminary unit root testing shows that the change in CDS spread is generally stable for all countries.

7.1 Estimation Results

All methods investigated in the Monte Carlo simulation are used to make predictions for the CDS. For evaluation both the in-sample MSPE and out-of-sample MSPE is used. Because of computational limitations, a static forecast is used. The in-sample period is July 2009 until July 2014 (61 observations for each country). The out-of-sample period is chosen to be from August 2014 until March 2016 (20 observations for each country).

Just as in the Monte Carlo simulation, a maximum amount of five clusters has been used for the estimation of the candidate pooling specifications using LGSI. For the tuning parameter λ a value of $\lambda = 2^6\, s_y^2 T^{-1/3}$ is used.

The relative MSPE of a selection of the methods is shown in table 1.[3] The relative MSPE is obtained by dividing the obtained MSPE by the MSPE obtained when using the individual estimator.

For the in-sample predictions, the shrinkage estimator has the best performance by a large margin. It seems that this estimator, which can be regarded as a pooling average estimator using only the individual and pooled estimator, outperforms the methods that use multiple candidate pooling specifications.

[1]The countries Brazil, Bulgaria, Chile, China, Columbia, Croatia, Hungary, Japan, Korea, Malaysia, Mexico, Philippines, Poland, Romania, Russia, Slovak, South Africa, Thailand and Turkey are included in the panel.

[2]Starting from the end of the great recession in the United States as reported by the National Bureau of Economic Research.

[3]The omitted pooling averaging methods has similar performance and are omitted for the sake of brevity.

Table 1: Relative MSPE for change in CDS spread. In-sample period contains 61 periods per country. Out-of-sample forecast contains 20 periods per country. The lowest MSPE for each test subsample is boldfaced.

	Pooling	**FGLS**	**SHK**	**Mixed**
IS	0.827	0.942	**0.799**	0.924
OOS	0.936	0.983	0.954	0.976

(SP)	**Equal**	**BIC**	**SBIC**	**MPABH**	**MPACH**
IS	0.893	0.827	0.827	0.841	0.827
OOS	0.970	0.936	0.936	0.950	0.936

(LGSI)	**Equal**	**BIC**	**SBIC**	**MPABH**	**MPACH**	**CLasso**
IS	0.829	0.827	0.827	0.845	0.827	0.827
OOS	**0.933**	0.936	0.936	0.933	0.936	0.936

For the out-of-sample predictions, LGSI combined with equal weights or MPABH weights give the best results. The shrinkage estimator decreases in performance compared to the in-sample prediction, as the individual estimates might be biased due to instability. In contrast to the Monte Carlo simulation, the MSPE of the different methods is quite similar. This makes it difficult to draw any strong conclusions from this application.

The coefficient estimates and their confidence intervals when using MPABH with the LGSI are displayed in table A.6 in the appendix. In order to construct the 95% confidence intervals for the coefficient estimates, a bootstrap using cross-sectional resampling as proposed by Kapetanios (2008) is used with 100 replications.

Compared to WPZ, a more recent subsample is used. Local stock market returns, changes in local exchange rates and global stock market returns remain significant determinants for changes in CDS spread in this new subsample, whereas volatility risk premium is no longer significant. Instead, investment-grade corporate bond spreads, equity premiums and equity flows have a significant positive effect for the majority of the countries. Investigating why this change occurred lies beyond the scope of this paper.

8 Conclusion

Wang et al. (2017) proposed a new, computationally efficient and optimal pooling averaging method for potential heterogeneous static panel regressions. By using the Mallows criterion, the MSE could be minimized by explicitly making the bias-variance tradeoff for different pooling specifications.

This paper is an extension to the paper by Wang et al. (2017) in order to investigate how the performance of the proposed pooling averaging methods changes when a different method is used to obtain pooling specifications to average over. Where Wang et al. (2017) uses its own novel screening procedure to overcome computational difficulties, this paper uses the C-Lasso technique with Partial Profile Likelihood estimation proposed by Su et al. (2016). By limiting the amount of possible clusters in the panel, this method is still computationally feasible. Thus, a comparison of the two methods to obtain pooling specifications can be made. This paper compares these two methods in a Monte Carlo simulation and an empirical application.

The simulation included static linear panel models with between-individual heteroskedastic errors with different degrees of heterogeneity in the parameters. Just like Wang et al. (2017) concluded, MPA produces results with the lowest MSPE in non-extreme cases. When comparing the two methods to obtain pooling specifications, the latent group structure estimation using C-Lasso outperforms the screening procedure proposed by Wang et al. (2017) when the amount of individuals in the panel is large. In more detail, one should opt for a variance structure with between-individual heterogeneity when the amount of individuals is small, but for a completely heteroskedastic variance structure when the amount of individuals is larger.

In the case of homogeneity or a low degree of heterogeneity, the C-Lasso method combined with using the Information Criterion proposed by Su et al. (2016) has the best performance. The true grouping structure is more likely to be selected when the amount of individuals and amount of observations per individual increase. This leads to achieving the lowest MSPE possible when only a low degree of heterogeneity is present.

However, this is not true for the cases with strong or complete heterogeneity. In these cases the Mallows pooling averaging combined with C-Lasso latent structure estimation or the mixed estimator proposed by Lee and Griffiths (1979) should be used. In these cases where no clear grouping structure can be identified, these methods often even outperform the infeasible oracle estimator which uses the unobservable group structure.

In summary, the Mallows Pooling Averaging with C-Lasso latent structure estimation has optimal performance in nearly all cases. Only when the degree of heterogeneity is very low, one should opt for the C-Lasso paired with the information criterion as proposed by Su et al. (2016).

These results are based on the analyses done in this paper. For a variety of DGPs, the difference in performance is studied. A valuable extension would be to study the performance of the screening procedure of Wang et al. (2017) when the amount of clusters is limited. This might be able to increase the performance of this method when there are more individuals in the panel. Alternatively, many additional topics for further research exist, such as the investigating the performance of pooling averaging methods for panel data models when the regressors are not exogenous. In empirical research, the performance of pooling averaging could be studied more elaborately by applying the pooling averaging methods on different data sets.

The practical procedure for choosing a suitable estimator for a possibly heterogeneous static panel model given in Wang et al. (2017, Section 9) remains valid. For the latent group structure identification using the C-Lasso, the effect of different data properties is not investigated, but the performance of the estimators other than the pooling average estimators still hold. This paper investigated if the performance of the pooling averaging estimator as proposed by Wang et al. (2017) could be increased. And indeed, it could be improved by adding a new way to identify group structures that are used for the pooling averaging methods. In addition good finite sample performance, the latent group structure identification method using C-Lasso paired with pooling averaging also has good performance when the amount of individuals and the observations per individual increase.

References

Ando, T. and Bai, J. (2016). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics*, 31(1):163–191.

Baltagi, B. H. and 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.

Bester, C. A. and Hansen, C. B. (2016). Grouped effects estimators in fixed effects models. *Journal of Econometrics*, 190(1):197 – 208.

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

Danilov, D. and Magnus, J. (2004). On the harm that ignoring pretesting can cause. *Journal of Econometrics*, 122(1):27–46.

Durlauf, S., Kourtellos, A., and Minkin, A. (2001). The local solow growth model. *European Economic Review*, 45(4-6):928–940.

Hansen, B. E. (2007). Least squares model averaging. *Econometrica*, 75(4):1175–1189.

Hsiao, C. (2014). *Analysis of panel data.* Number 54. Cambridge university press.

Kapetanios, G. (2008). A bootstrap procedure for panel data sets with many cross-sectional units. *Econometrics Journal*, 11(2):377–395.

Lee, L. and Griffiths, W. (1979). *The Prior Likelihood and Best Linear Unbiased Prediction in Stochastic Coefficient Linear Models.* Discussion paper series / Center for economic research, Department of economics, University of Minnesota. University of New England.

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

Longstaff, F., Pan, J., Pedersen, L., and Singleton, K. (2011). How sovereign is sovereign credit risk? *American Economic Journal: Macroeconomics*, 3(2):75–103.

Maddala, G. S., Trost, R. P., Li, H., and 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.

Pesaran, M. and Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68(1):79 – 113.

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

Sun, Y. (2005). Estimation and inference in panel structure models.

Swamy, P. A. V. B. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, 38(2):311–23.

Wang, W., Paap, R., and Zhang, X. (2017). "to pool or not to pool: What is a good strategy?".

Appendix

Tables follow on the next page.

Table A.1: Results of the Monte Carlo simulation with DGP 1. Performed for all variation of N, T and q. numbers in cells corresponds to the relative MSPE performance compared to individual estimators. The three best performing methods are shaded in gray, excluding the infeasible Oracle estimator. Results are based on 100 replications

		LGSI									common	
N	T	MPAHomo	MPABH	MPACH	equal weight	SAIC	SBIC	AIC	BIC	Classo	FGLS	Pool
5	15	0.38	0.38	0.19	0.55	0.51	0.23	0.61	0.23	0.19	0.21	0.18
5	40	0.39	0.39	0.21	0.56	0.50	0.21	0.56	0.21	0.20	0.20	0.20
10	15	0.38	0.38	0.15	0.38	0.52	0.19	0.61	0.21	0.12	0.11	0.10
10	40	0.37	0.37	0.13	0.38	0.51	0.12	0.58	0.12	0.10	0.10	0.10
30	15	0.31	0.31	0.16	0.20	0.40	0.25	0.45	0.29	0.03	0.04	0.03
30	40	0.35	0.35	0.18	0.22	0.43	0.24	0.49	0.28	0.03	0.03	0.03

		SP								individual		infeasible
N	T	MPAHomo	MPABH	MPACH	equal weight	SAIC	SBIC	AIC	BIC	mix	Shrinkage	Oracle
5	15	0.50	0.51	0.23	0.64	0.72	0.36	0.80	0.42	0.68	0.77	0.18
5	40	0.52	0.52	0.25	0.66	0.69	0.29	0.79	0.30	0.65	0.91	0.20
10	15	0.56	0.57	0.27	0.74	0.79	0.51	0.83	0.58	0.49	0.67	0.10
10	40	0.58	0.58	0.27	0.76	0.80	0.39	0.84	0.49	0.46	0.88	0.10
30	15	0.68	0.69	0.43	0.86	0.86	0.72	0.87	0.73	0.37	0.60	0.03
30	40	0.72	0.73	0.47	0.87	0.88	0.70	0.90	0.72	0.34	0.86	0.03

Table A.2: Results of the Monte Carlo simulation with DGP 2. Performed for all variation of N, T and q. numbers in cells corresponds to the relative MSPE performance compared to individual estimators. The three best performing methods are shaded in gray, excluding the infeasible Oracle estimator. Results are based on 100 replications

			LGSI									common	
N	T	q	MPAHomo	MPABH	MPACH	equal weight	SAIC	SBIC	AIC	BIC	Classo	FGLS	Pool
5	15	B	0.78	0.67	0.98	0.73	0.85	0.94	0.93	1.03	1.18	3.70	3.52
5	15	S	0.79	0.70	1.01	0.83	0.86	0.91	0.89	1.03	1.16	5.12	4.82
5	40	B	0.80	0.69	0.85	1.24	0.86	0.77	0.92	0.83	0.98	8.96	8.79
5	40	S	0.74	0.64	0.71	1.47	0.82	0.61	0.85	0.63	0.63	12.17	11.93
10	15	B	0.44	0.38	0.39	0.45	0.49	0.57	0.52	0.64	0.74	2.72	2.65
10	15	S	0.53	0.45	0.42	0.55	0.63	0.56	0.68	0.61	0.71	3.76	3.64
10	40	B	0.43	0.37	0.32	0.83	0.49	0.37	0.53	0.37	0.55	7.19	7.09
10	40	S	0.51	0.39	0.24	1.07	0.64	0.30	0.70	0.32	0.22	9.71	9.62
30	15	B	0.19	0.17	0.17	0.32	0.22	0.24	0.23	0.24	0.32	2.75	2.72
30	15	S	0.36	0.33	0.24	0.42	0.45	0.40	0.46	0.43	0.29	3.66	3.62
30	40	B	0.16	0.14	0.09	0.66	0.19	0.13	0.20	0.13	0.09	7.53	7.51
30	40	S	0.40	0.35	0.16	0.87	0.51	0.34	0.52	0.37	0.08	9.40	9.37

			SP								individual		infeasible
N	T	q	MPAHomo	MPABH	MPACH	equal weight	SAIC	SBIC	AIC	BIC	mix	Shrinkage	Oracle
5	15	B	0.82	0.80	0.98	0.75	0.86	0.95	0.92	1.01	0.82	0.96	0.57
5	15	S	0.75	0.75	0.83	0.76	0.81	0.79	0.88	0.81	0.85	0.96	0.55
5	40	B	0.74	0.73	0.80	1.01	0.77	0.74	0.80	0.76	0.88	1.00	0.56
5	40	S	0.62	0.62	0.57	1.03	0.68	0.57	0.71	0.61	0.88	1.00	0.56
10	15	B	0.63	0.62	0.54	0.62	0.74	0.66	0.79	0.71	0.63	0.92	0.22
10	15	S	0.63	0.64	0.53	0.66	0.74	0.60	0.79	0.62	0.69	0.93	0.24
10	40	B	0.57	0.59	0.42	0.66	0.70	0.46	0.77	0.46	0.68	0.99	0.21
10	40	S	0.52	0.55	0.34	0.68	0.69	0.33	0.75	0.35	0.72	0.99	0.21
30	15	B	0.60	0.62	0.46	0.72	0.78	0.53	0.80	0.55	0.51	0.90	0.07
30	15	S	0.61	0.64	0.43	0.74	0.80	0.54	0.83	0.56	0.55	0.92	0.08
30	40	B	0.60	0.62	0.37	0.74	0.78	0.41	0.81	0.43	0.58	0.98	0.07
30	40	S	0.60	0.63	0.35	0.76	0.81	0.42	0.83	0.46	0.62	0.99	0.08

Table A.3: Results of the Monte carlo simulation with DGP 3. Performed for all variation of N, T and q. numbers in cells corresponds to the relative MSPE performance compared to individual estimators. The three best performing methods are shaded in gray, excluding the infeasible Oracle estimator. Results are based on 100 replications

			LGSI									common	
N	T	q	MPAHomo	MPABH	MPACH	equal weight	SAIC	SBIC	AIC	BIC	Classo	FGLS	Pool
5	15	B	0.92	0.80	1.19	0.78	0.98	1.17	1.11	1.32	1.53	2.93	2.81
5	15	S	0.90	0.85	1.17	0.84	0.98	1.07	1.05	1.15	1.32	3.66	3.50
5	40	B	0.90	0.76	0.94	1.04	0.97	1.04	1.09	1.09	1.40	7.26	7.14
5	40	S	0.97	0.88	1.23	1.40	1.01	1.04	1.09	1.11	1.41	9.49	9.33
10	15	B	0.50	0.43	0.46	0.45	0.58	0.58	0.63	0.62	0.64	2.49	2.44
10	15	S	0.67	0.61	0.64	0.62	0.80	0.81	0.85	0.90	0.86	3.61	3.51
10	40	B	0.58	0.48	0.53	0.81	0.63	0.70	0.69	0.71	1.14	6.87	6.78
10	40	S	0.78	0.72	0.78	1.15	0.87	0.87	0.94	0.90	1.40	8.77	8.70
30	15	B	0.35	0.32	0.32	0.40	0.43	0.44	0.44	0.45	0.51	2.46	2.41
30	15	S	0.56	0.53	0.48	0.56	0.66	0.64	0.68	0.68	0.76	3.32	3.28
30	40	B	0.28	0.26	0.25	0.66	0.33	0.43	0.36	0.44	0.47	6.48	6.45
30	40	S	0.56	0.55	0.54	1.07	0.64	0.63	0.66	0.65	0.79	9.01	8.98

			SP								individual		infeasible
N	T	q	MPAHomo	MPABH	MPACH	equal weight	SAIC	SBIC	AIC	BIC	mix	Shrinkage	Oracle
5	15	B	0.92	0.87	1.13	0.83	0.94	1.20	1.01	1.31	0.85	0.96	1.00
5	15	S	1.01	0.98	1.17	0.89	1.05	1.21	1.12	1.25	0.86	0.96	1.00
5	40	B	0.90	0.85	0.99	1.13	0.91	1.11	0.98	1.20	0.87	0.99	1.00
5	40	S	1.05	1.01	1.24	1.22	1.05	1.36	1.13	1.41	0.91	1.00	1.00
10	15	B	0.67	0.67	0.63	0.63	0.78	0.79	0.82	0.84	0.60	0.91	0.65
10	15	S	0.78	0.76	0.77	0.71	0.86	0.90	0.91	0.95	0.65	0.93	0.62
10	40	B	0.64	0.64	0.54	0.67	0.73	0.62	0.80	0.64	0.68	0.99	0.64
10	40	S	0.80	0.80	0.78	0.77	0.92	0.95	0.98	0.98	0.72	0.99	0.64
30	15	B	0.64	0.66	0.53	0.72	0.78	0.63	0.81	0.65	0.51	0.89	0.22
30	15	S	0.75	0.76	0.67	0.80	0.88	0.81	0.89	0.83	0.56	0.91	0.23
30	40	B	0.63	0.66	0.48	0.72	0.80	0.55	0.82	0.57	0.55	0.98	0.19
30	40	S	0.75	0.76	0.62	0.77	0.89	0.82	0.91	0.86	0.62	0.99	0.23

Table A.4: Results of the Monte Carlo simulation with DGP 4. Performed for all variation of N, T and q. numbers in cells corresponds to the relative MSPE performance compared to individual estimators. The three best performing methods are shaded in gray, excluding the infeasible Oracle estimator. Results are based on 100 replications

		LGSI									common	
N	T	MPAHomo	MPABH	MPACH	equal weight	SAIC	SBIC	AIC	BIC	Classo	FGLS	Pool
5	15	0.94	0.95	1.38	2.61	1.01	1.14	1.02	1.20	1.08	27.63	26.06
5	40	0.99	1.00	1.58	7.33	1.00	1.00	1.00	1.00	1.00	83.30	81.73
10	15	0.80	0.76	0.90	1.29	1.13	1.16	1.19	1.34	1.60	9.99	9.73
10	40	1.18	1.17	1.31	3.08	1.92	1.94	1.97	2.15	2.79	26.70	26.40
30	15	0.31	0.27	0.26	0.33	0.36	0.42	0.37	0.44	0.46	1.31	1.27
30	40	0.49	0.47	0.46	0.72	0.64	0.64	0.64	0.65	0.87	3.29	3.26

		SP								individual		infeasible
N	T	MPAHomo	MPABH	MPACH	equal weight	SAIC	SBIC	AIC	BIC	mix	Shrinkage	Oracle
5	15	1.09	1.19	1.83	4.12	1.05	1.24	1.07	1.28	0.77	0.99	1.00
5	40	1.04	1.09	1.72	12.58	1.00	1.03	1.00	1.02	0.76	1.00	1.00
10	15	0.98	0.96	1.16	1.26	1.03	1.20	1.07	1.25	0.63	0.97	1.00
10	40	1.14	1.19	1.56	2.99	1.11	1.54	1.16	1.58	0.62	1.00	1.00
30	15	0.57	0.58	0.47	0.56	0.77	0.77	0.80	0.82	0.43	0.82	1.00
30	40	0.69	0.69	0.62	0.61	0.83	0.81	0.85	0.84	0.47	0.97	1.00

Table A.5: Results of the Monte Carlo simulation with DGP 5. Performed for all variation of N, T and q. numbers in cells corresponds to the relative MSPE performance compared to individual estimators. The three best performing methods are shaded in gray, excluding the infeasible Oracle estimator. Results are based on 100 replications

			LGSI									common	
N	T	q	MPAHomo	MPABH	MPACH	equal weight	SAIC	SBIC	AIC	BIC	Classo	FGLS	Pool
5	15	B	0.86	0.86	1.03	1.76	0.87	0.85	0.88	0.85	0.85	4.98	4.73
5	15	S	0.84	0.84	1.20	1.33	0.86	0.83	0.88	0.83	0.82	5.16	4.88
5	40	B	0.83	0.83	0.81	3.41	0.85	0.80	0.87	0.80	0.80	11.64	11.42
5	40	S	0.84	0.84	0.88	1.99	0.85	0.80	0.87	0.80	0.80	13.59	13.31
10	15	B	0.44	0.44	0.43	0.97	0.46	0.41	0.48	0.41	0.41	3.60	3.47
10	15	S	0.47	0.47	0.48	0.69	0.48	0.42	0.50	0.42	0.42	3.80	3.71
10	40	B	0.43	0.43	0.40	1.03	0.45	0.39	0.48	0.39	0.40	8.64	8.54
10	40	S	0.46	0.46	0.41	0.94	0.48	0.40	0.52	0.40	0.40	10.50	10.43
30	15	B	0.17	0.17	0.14	0.73	0.19	0.13	0.21	0.13	0.13	3.66	3.61
30	15	S	0.20	0.20	0.16	0.76	0.22	0.14	0.24	0.15	0.13	4.22	4.18
30	40	B	0.18	0.18	0.15	0.83	0.20	0.15	0.22	0.14	0.15	9.91	9.88
30	40	S	0.19	0.19	0.14	0.65	0.21	0.13	0.23	0.13	0.13	11.27	11.23

			SP								individual		infeasible
N	T	q	MPAHomo	MPABH	MPACH	equal weight	SAIC	SBIC	AIC	BIC	mix	Shrinkage	Oracle
5	15	B	0.56	0.57	0.43	2.81	0.67	0.46	0.75	0.48	1.01	1.00	0.82
5	15	S	0.58	0.59	0.44	1.22	0.70	0.49	0.78	0.51	1.01	0.99	0.79
5	40	B	0.59	0.59	0.39	6.84	0.71	0.40	0.78	0.42	1.00	1.00	0.80
5	40	S	0.59	0.59	0.40	2.48	0.71	0.39	0.80	0.39	1.00	1.00	0.80
10	15	B	0.51	0.52	0.26	1.23	0.72	0.37	0.78	0.44	1.00	1.00	0.39
10	15	S	0.52	0.54	0.28	0.80	0.73	0.37	0.79	0.40	0.98	0.99	0.40
10	40	B	0.53	0.54	0.28	2.42	0.71	0.28	0.77	0.31	1.00	1.00	0.40
10	40	S	0.56	0.56	0.29	1.16	0.75	0.33	0.80	0.37	0.99	1.00	0.40
30	15	B	0.55	0.56	0.27	0.81	0.78	0.43	0.80	0.47	0.98	0.99	0.12
30	15	S	0.59	0.61	0.31	0.78	0.82	0.53	0.84	0.56	0.94	0.98	0.13
30	40	B	0.62	0.62	0.33	0.96	0.82	0.45	0.85	0.51	0.99	1.00	0.14
30	40	S	0.64	0.64	0.35	0.84	0.83	0.51	0.86	0.55	0.97	1.00	0.13

Table A.6: Coefficient estimates of determinants of change in CDS spread by using MPA with between-individual heteroskedastic error structure using the LGSI. Constructued 95% confidence interval by bootstrapping in parentheses. Boldfaced numbers are significantly different from zero using the constructed 95% confidence interval.

variable	Brazil	Bulgaria	Chile	China	Columbia	Croatia
lstock	**−0.329**	**−0.127**	**−0.296**	**−0.336**	**−0.336**	**−0.127**
	(−0.368, −0.085)	(−0.374, −0.055)	(−0.377, −0.064)	(−0.38, −0.109)	(−0.369, −0.109)	(−0.444, −0.043)
fxrates	**0.184**	**0.284**	**0.165**	**0.155**	**0.155**	0.284
	(0.148, 0.317)	(0.115, 0.335)	(0.115, 0.343)	(0.099, 0.306)	(0.052, 0.306)	(−0.011, 0.315)
fxres	−0.039	−0.02	0.049	−0.019	−0.019	−0.02
	(−0.067, 0.015)	(−0.071, 0.04)	(−0.08, 0.044)	(−0.085, 0.046)	(−0.085, 0.094)	(−0.085, 0.155)
gstock	**−0.287**	**−0.335**	**−0.361**	**−0.267**	**−0.267**	**−0.335**
	(−0.378, −0.208)	(−0.378, −0.162)	(−0.426, −0.164)	(−0.423, −0.162)	(−0.488, −0.162)	(−0.622, −0.018)
trsy	0.027	−0.072	−0.003	−0.045	−0.045	−0.072
	(−0.08, 0.073)	(−0.105, 0.073)	(−0.105, 0.057)	(−0.099, 0.034)	(−0.099, 0.076)	(−0.104, 0.152)
ig	**0.075**	0.063	0.048	0.093	0.093	**0.063**
	(0.039, 0.098)	(−0.02, 0.107)	(−0.023, 0.106)	(−0.023, 0.11)	(−0.01, 0.11)	(0.035, 0.11)
eqp	**0.058**	**0.087**	**0.151**	**0.179**	**0.179**	**0.087**
	(0.046, 0.218)	(0.046, 0.214)	(0.052, 0.214)	(0.067, 0.214)	(0.066, 0.206)	(0.066, 0.214)
volp	0.02	0.183	−0.01	−0.003	−0.003	0.183
	(−0.042, 0.161)	(−0.053, 0.271)	(−0.053, 0.269)	(−0.052, 0.121)	(−0.043, 0.121)	(−0.049, 0.136)
ef	**0.093**	**0.111**	**0.155**	**0.076**	**0.076**	**0.111**
	(0.075, 0.123)	(0.067, 0.137)	(0.07, 0.124)	(0.065, 0.125)	(0.07, 0.13)	(0.028, 0.161)
bf	**−0.03**	−0.137	0.006	**−0.096**	**−0.096**	**−0.137**
	(−0.148, −0.013)	(−0.174, 0.009)	(−0.174, 0.013)	(−0.144, 0.017)	(−0.144, 0.019)	(−0.138, 0.058)

Table A.6: Coefficient estimates of determinants of change in CDS spread by using MPA with between-individual heteroskedastic error structure using the LGSI. Constructued 95% confidence interval by bootstrapping in parentheses. Boldfaced numbers are significantly different from zero using the constructed 95% confidence interval. (Continued)

variable	Hungary	Japan	Korea	Malaysia	Mexico	Philippines
lstock	**−0.336**	**−0.336**	**−0.336**	**−0.329**	**−0.231**	**−0.296**
	(−0.41, −0.043)	(−0.403, −0.043)	(−0.41, −0.08)	(−0.378, −0.139)	(−0.378, −0.176)	(−0.363, −0.109)
fxrates	0.155	0.155	0.155	**0.184**	**0.234**	**0.165**
	(−0.011, 0.306)	(−0.011, 0.381)	(−0.011, 0.368)	(0.159, 0.381)	(0.163, 0.287)	(0.123, 0.306)
fxres	−0.019	−0.019	−0.019	−0.039	−0.081	0.049
	(−0.091, 0.192)	(−0.091, 0.155)	(−0.075, 0.158)	(−0.085, 0.049)	(−0.085, 0.051)	(−0.085, 0.059)
gstock	**−0.267**	**−0.267**	**−0.267**	**−0.287**	**−0.331**	**−0.361**
	(−0.664, −0.164)	(−0.622, −0.164)	(−0.605, −0.164)	(−0.394, −0.18)	(−0.402, −0.159)	(−0.397, −0.093)
trsy	−0.045	−0.045	−0.045	0.027	0.022	−0.003
	(−0.099, 0.18)	(−0.11, 0.152)	(−0.11, 0.152)	(−0.103, 0.044)	(−0.093, 0.044)	(−0.137, 0.041)
ig	**0.093**	**0.093**	**0.093**	**0.075**	**0.092**	**0.048**
	(0.016, 0.117)	(0.035, 0.117)	(0.03, 0.117)	(0.065, 0.108)	(0.059, 0.11)	(0.054, 0.115)
eqp	**0.179**	**0.179**	**0.179**	**0.058**	**0.12**	**0.151**
	(0.066, 0.21)	(0.066, 0.205)	(0.083, 0.21)	(0.074, 0.191)	(0.043, 0.199)	(0.058, 0.253)
volp	−0.003	−0.003	−0.003	0.02	0.012	−0.01
	(−0.06, 0.121)	(−0.06, 0.121)	(−0.06, 0.076)	(−0.039, 0.121)	(−0.023, 0.121)	(−0.038, 0.121)
ef	**0.076**	**0.076**	**0.076**	**0.093**	**0.122**	**0.155**
	(0.065, 0.168)	(0.072, 0.161)	(0.065, 0.191)	(0.067, 0.172)	(0.067, 0.187)	(0.061, 0.181)
bf	−0.096	−0.096	−0.096	−0.03	−0.047	0.006
	(−0.138, 0.094)	(−0.134, 0.063)	(−0.114, 0.089)	(−0.144, 0.042)	(−0.144, 0.042)	(−0.162, 0.04)

Table A.6: Coefficient estimates of determinants of change in CDS spread by using MPA with between-individual heteroskedastic error structure using the LGSI. Constructued 95% confidence interval by bootstrapping in parentheses. Boldfaced numbers are significantly different from zero using the constructed 95% confidence interval. (Continued)

variable	Poland	Romania	Russia	Slovak	S. Africa	Thailand	Turkey
$lstock$	**−0.336**	**−0.127**	**−0.329**	**−0.336**	**−0.329**	**−0.296**	**−0.231**
	(−0.41, −0.172)	(−0.392, −0.109)	(−0.397, −0.109)	(−0.397, −0.236)	(−0.395, −0.264)	(−0.384, −0.217)	(−0.364, −0.15)
$fxrates$	**0.155**	**0.284**	**0.184**	**0.155**	**0.184**	**0.165**	**0.234**
	(0.084, 0.294)	(0.133, 0.306)	(0.133, 0.325)	(0.134, 0.282)	(0.124, 0.271)	(0.129, 0.286)	(0.138, 0.304)
$fxres$	−0.019	−0.02	−0.039	−0.019	−0.039	0.049	−0.081
	(−0.085, 0.044)	(−0.081, 0.032)	(−0.081, 0.044)	(−0.055, 0.044)	(−0.051, 0.053)	(−0.052, 0.054)	(−0.076, 0.038)
$gstock$	**−0.267**	**−0.335**	**−0.287**	**−0.267**	**−0.287**	**−0.361**	**−0.331**
	(−0.403, −0.093)	(−0.387, −0.162)	(−0.367, −0.093)	(−0.365, −0.157)	(−0.376, −0.182)	(−0.414, −0.173)	(−0.364, −0.185)
$trsy$	−0.045	−0.072	0.027	−0.045	0.027	−0.003	0.022
	(−0.137, 0.051)	(−0.099, 0.029)	(−0.137, 0.047)	(−0.098, 0.047)	(−0.092, 0.037)	(−0.082, 0.037)	(−0.082, 0.04)
ig	**0.093**	**0.063**	**0.075**	**0.093**	**0.075**	**0.048**	**0.092**
	(0.062, 0.115)	(0.069, 0.11)	(0.06, 0.115)	(0.059, 0.109)	(0.059, 0.102)	(0.056, 0.106)	(0.067, 0.106)
eqp	**0.179**	**0.087**	**0.058**	**0.179**	**0.058**	**0.151**	**0.12**
	(0.057, 0.253)	(0.077, 0.214)	(0.023, 0.253)	(0.01, 0.211)	(0.012, 0.211)	(0.034, 0.208)	(0.08, 0.208)
$volp$	−0.003	0.183	0.02	−0.003	0.02	−0.01	0.012
	(−0.038, 0.169)	(−0.03, 0.181)	(−0.037, 0.214)	(−0.026, 0.21)	(−0.017, 0.173)	(−0.013, 0.071)	(−0.02, 0.082)
ef	**0.076**	**0.111**	**0.093**	**0.076**	**0.093**	**0.155**	**0.122**
	(0.061, 0.184)	(0.065, 0.146)	(0.061, 0.12)	(0.064, 0.112)	(0.066, 0.192)	(0.07, 0.193)	(0.07, 0.189)
bf	−0.096	**−0.137**	**−0.03**	**−0.096**	−0.03	0.006	−0.047
	(−0.162, 0.002)	(−0.144, −0.014)	(−0.162, −0.019)	(−0.124, −0.019)	(−0.134, 0.012)	(−0.124, 0.029)	(−0.129, 0.004)