Erasmus University Rotterdam

Erasmus School of Economics

Bachelor Thesis Econometrics and Operations Research

Improving Performance for Pooling Averaging Strategies in Forecasting Sovereign CDS Spreads

Name student: B.J.W. Walpot
Student ID number: 452721

Supervisor: W. Wang
Second assessor: J.W.N. Reuvers

Abstract

This paper examines the estimation and forecasting of heterogeneous panel data models by various estimation techniques, of which a particular pooling averaging approach is of central interest. Furthermore, the research addresses the potential time-varying nature of financial products. This is done by incorporating optimal window size selection in the estimation of regression parameters. These methods are applied to forecast sovereign CDS spreads. By considering two different angles of approach for window determination, it is found that these methods can help improving forecasting performance significantly.

Contents

1 Introduction **2**

2 Literature Review **4**

3 Data **6**

4 Methodology **6**

4.1 Model Specification 6

4.2 Estimation of Regression Parameters 7

4.3 Shrinking the Number of Pooling Strategies 8

4.4 Choosing Mallows Pooling Averaging Weights 9

4.5 Alternative Pooling Averaging Strategies 10

5 Application to CDS Swaps **11**

5.1 Determining Local and Global Effects 12

5.2 Forecasting Sovereign Credit Risk 13

6 Optimal window selection with structural breaks **14**

6.1 Individual Optimal Windows 16

6.2 Group Optimal Windows 18

7 Conclusion and Discussion **20**

A Appendix **25**

A.1 Data Description 25

A.2 Expanding Window Forecasting Results 26

A.3 Code Explanation 27

A.3.1 Estimation 27

A.3.2 Forecasting 27

1 Introduction

Sovereign credit risk and credit debt of countries have obtained particular attention since the outbreak of the worldwide financial crisis and the euro crisis. In order to investigate the concept of credit risk of various countries, sovereign credit default swaps (CDS) need to be thoroughly studied. For the sake of policy making, it is important to know whether these financial products are mainly driven by local variables or global variables. Next to that, it is also of great interest to forecast the values of sovereign CDS swaps, especially in these times when some European countries suffer from political and financial stability, such as the United Kingdom and Italy.

Much research that has been performed on sovereign CDS swaps only examines the effect of variables by individual regressions or a pooled regression model, such as respectively Longstaff et al. (2011) and Remolona et al. (2008). Individual ordinary least squares (OLS) regression has the beneficial characteristic that there is no bias in estimation. However, since this estimation technique does not exploit any cross-sectional information, it usually has a larger variance than necessary. Pooled estimation maximally uses this cross-country information, resulting in efficient but biased estimation results. This estimation method namely does not consider individual effects, while existing literature implicates that heterogeneity is present in the effect of determinants across countries.

This choiche on whether to estimate series individually or pooled, is a dilemma that is commonly known in econometrics as "to pool or not to pool". However, the choiche does not have to be between these two extremes. In order to obtain estimates for these individual effects in a heteroskedastic panel, an optimal trade-off between zero bias and efficiency can be obtained by choosing a weighted combination between individual OLS and pooled estimation. In the research of Maddala et al. (1997), one possible method for choosing weights between these two estimators is proposed.

In this paper, the Mallows Pooling Averaging (MPA) estimation technique, proposed by Wang et al. (2018), is considered. This method attempts to find an optimal bias-variance trade-off by combining multiple pooling specifications and choosing weights that minimize the well-known Mallows criterion. Since the number of various pooling specifications can already become very large for a reasonable number of individuals, I use a pre-screening method that makes all pooling averaging techniques computationally feasible. In order to do this, individuals will be structured into groups. Regression coefficients are homogeneous within groups but heterogeneous across different groups. Recently, new methods have been developed for specifying group membership and group estimators, for example Ando and Bai (2016) proposed grouping from the starting-point of unobserved group factors. The preliminary process that I will use concerns the Classifier-Lasso (C-Lasso) approach, proposed by Su et al. (2016) and also implemented by Wang et al. (2018). In addition to this, to account for possible heterogeneity within groups, the sample will be conditioned on two structural break points that are detected by Wang et al. (2018).

In this paper I will re-estimate the effects of local and global variables on sovereign CDS estimates using the proposed methods of Wang et al. (2018). Also, forecasts will be constructed for the various estimation techniques and their associated forecasting performances are examined and compared.

This research adds to the literature with respect to modelling heterogeneity for the effects of the determinants to obtain the best performing forecasts possible. In order to improve the forecasting performance of the MPA technique, the estimation window size will be determined explicitly for each forecast by choosing it such that an MSFE expression is minimized. The proposed approach of Inoue et al. (2017) for selecting the optimal window size for a univariate time series will be used and adjusted for the multivariate panel model that is under consideration. This adjustment will be done in two ways; namely by determining the optimal window size for each country individually, and alternatively by selecting it for groups of countries, over the various grouping structures the C-Lasso procedure establishes. Individual window selection intuitively provides accurate estimations for the individual window sizes, but ignores potentially valuable cross-sectional information. Grouped window selection uses this information, but may yield inaccurate estimated windows, since the groups from the C-Lasso procedure remove valuable individual country information. This can particularly be the case when all countries are shrinked in a very small number of groups. Results on the forecasting performance suggest that in most cases, both approaches provide better forecasts than the original MPA technique.

The research is structured in the following way. In Section 2, a literature overview on heterogeneous panel estimation is provided. A description of the used data can be found in Section 3. The methodology is described in Section 4 and its empirical application to sovereign CDS spreads in Section 5, which together represents the part of my research that replicates the methods of Wang et al. (2018). In Section 6, I specify my extension on optimal estimation window selection and provide results on forecasting performance of these proposed techniques. Section 7 concludes the research and offers recommendations for further research.

2 Literature Review

For approximately five decades, extensive research has been performed on estimating the slope parameters in heterogeneous panel data models. It commenced with Swamy (1970), who proposed a mean estimator that is asymptotically efficient and consistent and an unbiased variance estimator for one form of the random coefficient model (RCM). This estimator is now commonly known in the field of econometrics as the generalised least squares (GLS) estimator. Hsiao (1975), Hildreth and Houck (1968), Rosenberg (1972) and others have proposed other closely related types of the stationary RCM, each with a particular set of assumptions. A consistent estimator for heterogeneous panels can also be found under more general assumptions than made by Swamy (1970). Pesaran and Smith (1995) suggested a mean group estimator for dynamic random coefficient models, which is simply an average of individual OLS estimators. This estimator is algebraically equivalent to the GLS estimator in case of a large number of observations. In this context, an estimator can also be obtained from a Bayesian perspective, which assumes all variables and parameters are random variables. It is shown in Hsiao and Pesaran (2008) that this is a weighted combination of the individual OLS estimator and Swamy's mean estimator. Mean and mean grouped estimators work very well when the average causality is most important. However, in this paper I want to forecast sovereign CDS swaps for fourteen individual countries as best as possible. This means the interest is drawn to more sophisticated estimators that are also able to take other effects than only the mean into account. Maddala et al. (1997) estimates the heterogeneous individual effects in a cross-sectional way. This is done by means of taking a weighted combination of the pooled estimator and the individual estimator. In this way, all heterogeneous estimators are shrinked towards a pooled estimator. Baltagi and Griffin (1997) proves the necessity of such a method, since it shows that purely pooled models cannot capture heterogeneous individual effects and that purely individual models do not account for general changes or underlying shocks. Methods and techniques that attempt to account for individual as well as common effects, have substantially improved over the last years, mainly because of a wide range of research that is performed on the specification of groups of individuals. Ando and Bai (2016) developes a new estimation technique for heterogeneous panel data models with unobserved factor structures. The regression coefficients in this type of model can be homogeneous or group specific. The estimation of the number of groups and group membership of all individuals require new estimation techniques, such as the k-means approach or the Classifier-Lasso procedure. These methods are proposed and proven to be asymptotically efficient in several papers, namely in Su et al. (2016), Lin and Ng (2012) and Bonhomme and

Manresa (2015). A clear distinction need to be made between estimation when estimation accuracy is of central interest and estimation when forecasting performance is most important. As forecasting is of central interest for the application to sovereign CDS swaps, the estimated group structure, number of groups and estimated group coefficients do not need to be correct. They need to perform best from a forecasting perspective, that is by giving us the lowest forecasting error possible.

The k-means and C-Lasso procedures require knowledge of the correct number of groups. The fact that this is usually not known and more importantly that it is not necessarily relevant for the forecasting performance of the estimates, drives Wang et al. (2018) to propose a new estimate for heterogeneous panel models. This estimation technique makes a trade-off between bias and variance by minimizing the MPA criterion. In this way the grouping structure and parameters can be estimated in such a way that the MSFE is minimised, resulting in an optimal forecasting performance.

Pesaran et al. (2013) argues that parameter instability is a common cause of bad forecasting performance in financial products. Much research has been done on testing for structural breaks in panel data models. Works to consider are for example those of Evans and Kim (2011), Qian and Su (2016), Li et al. (2016). In order to account for parameter instability, Wang et al. (2018) applies the methods of Baltagi et al. (2016) and estimates the parameters heterogeneous for these different regimes. The paper of Wang et al. (2018), on which this research is based, investigates whether ignoring these structural breaks or taking them into account in the estimation window, is best for the forecasting performance.

However, like most described research on estimating parameters in heterogeneous panels, the window size is ignored as a potential source of forecasting improvement. Pesaran and Timmermann (2007) develop multiple techniques for explicitly choosing the estimation window optimal with regard to minimizing forecasting errors, when a forecasting model is conditioned on structural breaks. An alternative perspective on this is comprehended by Pesaran et al. (2013), which derives expressions for optimal weights under a conditioned forecasting model. More recently, Inoue et al. (2017) introduces a new approach for optimal rolling window size selection in forecasting models with potential breaks, in which these breaks are generalised to being smooth parameter changes over all time points. By combining this approach for optimal window selection with the proposed estimation technique of Wang et al. (2018), I hope to find improved forecasting results for the concerned empirical application of sovereign CDS spreads.

3 Data

In order to examine the sovereign credit risk of a country, the sovereign credit default swap spreads are used as a proxy. By means of a sovereign CDS spread, an investor can offset his credit risk, for example the risk of a default, to another investor. In exchange, a lasting premium has to be payed for swapping this risk, just as is done with an insurance contract. The 5-year sovereign CDS spreads for 14 different countries are used, which are Brazil, Bulgaria, Chile, China, Hungary, Japan, South-Korea, Malaysia, Philippines, Poland, Romania, Slovak, South Africa and Thailand. These are expressed as monthly basis points running from January 2003 to January 2016, in total 156 observations over time. The main summary statistics of these sovereign CDS spreads for the different countries are denoted in Table A.1, which can be found in Section A.1 of the Appendix.

Because I also want to explore the determinants of sovereign credit risk, I use the monthly data on local stock market return, foreign currency reserves and local exchange rates for each of the 14 countries for local examination. Regarding global determinants, the effects of the U.S. stock market returns, treasury yields, high-yield corporate bonds spreads, equity premium, volatility risk premium, equity flows and bond flows are obtained. The data on these local and global variables have the same time span as the data on sovereign CDS spreads. A comprehensive explanation on the definition of these variables can be found in Longstaff et al. (2011).

The data on CDS spreads, local and global variables is the same as in Wang et al. (2018) and conveniently is obtained upon request of one of the authors. In this research on which this paper is based, the data of Longstaff et al. (2011) for the concerned 14 countries is used and extended to 2016. Regarding the sovereign CDS spreads, first differencing is necessary to obtain the alteration data, since the CDS spreads are expressed as level data.

4 Methodology

4.1 Model Specification

The setup model is specified in the following way.

$$y_i = X_i\beta_i + u_i, \quad i = 1,\ \dots,\ N. \tag{4.1.1}$$

In this specification, y_i is the vector of dependent variables for country i and time points $t = 1,\ \dots,\ T$. X_i is a matrix of explanatory variables; $X_i = [X'_{i1},\ \dots,\ X'_{iT}]'$, in which $X_{it} = [x_{it,1},\ \dots\ , x_{it,K}]$ with K the number of explanatory variables. It is assumed that the system is stationary and exogenous, since there are no lagged y_i present in the set of explanatory variables.

There are three possibilities for the heterogeneity structure of the error terms. First of all, they can be homoskedastic within and across individuals i, which means that the variance of u_i is constant over time and the same for all countries. A more comprehensive option concerns error terms that are heteroskedasticity between countries, but homoskedastic within countries over time. In the most general case, errors are conditionally heteroskedastic, which means they are varying across individuals and within the time span of each individual country. Each of these options yields a different estimated variance matrix, as we will see in computations later on in this paper. In the first case, the regression parameter $\beta_i = [\beta_{i1},\ ...,\ \beta_{ik}]'$ is assumed to be time-invariant and heteroskedastic across individuals, as is assumed in the setup model of Wang et al. (2018).

4.2 Estimation of Regression Parameters

Since I have assumed that the system 4.1.1 is stationary and exogenous, Ordinary Least Squares (OLS) can be used in order to estimate the regression coefficients β_i individual for each country i. In case of correct specification, which I will assume for the moment, the vector of OLS estimators $\hat{\beta} = (\hat{\beta}_1',\ ...,\ \hat{\beta}_N')'$ is unbiased. However, it is probably not efficient, as it does not make use of the information that lies in the cross-sectional variation of the different countries. In order to make optimal use of this information, one can consider the pooled estimator $\hat{\beta}_{pool} = (b',\ ...\ , b')'$, which can be computed by the following expression:

$$b = (\sum_{i=1}^{N} X_i'X_i)^{-1} \sum_{i=1}^{N} X_i'X_i\hat{\beta}_i, \tag{4.2.1}$$

in which $\hat{\beta}_i$ represents the individual OLS estimator of country i. The OLS estimator on the one hand and the pooled estimator on the other hand are two extremes in possible estimation methods. While the individual OLS estimator is unbiased but inefficient, the pooled estimator has minimal variance but a relatively high level of bias. The essence of this paper is to combine these two extreme estimation methods, such that an optimal trade-off between unbiasedness and efficiency will be achieved, from a perspective of forecasting.

A range of possible combinations between the individual and pooled estimator is represented by a general pooling strategy, in which regression coefficients are estimated through OLS with some of these coefficients restricted to be equal. This restricted OLS estimator can generally be expressed by the following:

$$R_m\beta = 0, \tag{4.2.2}$$

in which R_m is the restriction matrix that corresponds to the mth pooling strategy. The restriction matrix is directly derived from the imposed restrictions on the coefficients from the different countries. If, for example, countries i and j are restricted to have the same regression coefficients for the mth pooling strategy, then $R_m = [0_{k\times(i-1)k}, I_k, 0_{k\times(j-i-1)k}, -I_k, 0_{k\times(N-j)k}]$. Every restriction matrix corresponds to a certain projection matrix; the projection matrix for the mth pooling strategy can be constructed as follows:

$$P_m = I_{Nk} - (X'X)^{-1}R'_m(R_m(X'X)^{-1}R'_m)^{-1}R_m. \tag{4.2.3}$$

Ideally, we would like to directly evaluate or compute which pooling strategy has the optimal trade-off between bias and variance. This can be done by computing the Mean Squared Error (MSE) for estimation accuracy or the Mean Squared Forecasting Error (MSFE) for forecasting performance. However, selecting the actual true restrictions by this approach is not feasible, since the (forecasting) error cannot distinguish whether the possible cause concerns a bias in estimation or loss in efficiency. Therefore, the focus in this research is on minimizing this error, instead of uncovering the true relations between regression parameters of different countries.

4.3 Shrinking the Number of Pooling Strategies

In general, for the different pooling averaging strategies that I will consider hereafter, there will not be chosen one pooling strategy based on some criterion. Instead, multiple pooling strategies will be considered and a weighted combination of the estimates that results from these approaches will be chosen in an optimal way. This means that for a pooling averaging strategy, the weights will be chosen based on some criterion. A general expression for a pooling averaging strategy is as follows:

$$\hat{\beta}(w) = \sum_{m=1}^{M} w_m \hat{\beta}_{(m)} = \sum_{m=1}^{M} w_m P_m \hat{\beta} = P(w)\hat{\beta}. \tag{4.3.1}$$

Since the number of pooling strategies can become very large, first shrinking the number of potential pooling strategies will be implemented, in order to allow for computational feasibility. Pooling strategies that restrict parameters to be the same while actually these are significantly different from each other, are not desired to be considered. A procedure that I will use as a preliminary step before considering optimal weights for the different pooling strategies is the classifier-Lasso (C-Lasso) procedure, according to Su et al. (2016).

This is a recently developed estimation method for panel data that can be used in case of heterogeneous panel data with an unknown group classification. Su et al. (2016) proposes two different approaches for this sort of panel estimation. Since there are no lagged versions of the dependent variable or endogenous variables in the model, I use the Penalized Profile Likelihood (PPL) method for model shrinkage. In this approach, it is assumed that the regression parameters β_i can be grouped by the following general structure:

$$\beta_i = \sum_{k=1}^{K} \alpha_k \mathbf{1}\{i \ \epsilon \ G_k\}, \tag{4.3.2}$$

where $\cup_{k=1}^{K} G_k = \{1, 2, ..., G_{max}\}$ denotes the set of different true groups. Each country will be grouped in one of these G_k groups and all countries in one group share the same regression parameter α_k. The estimates of α and β from the C-Lasso procedure are obtained by minimizing the so-called PPL criterion function:

$$Q_{1NT,\lambda_1}^{K_0} = Q_{1,NT}(\beta) + \frac{\lambda_1}{N} \sum_{i=1}^{N} \prod_{k=1}^{K_0} \|\beta_i - \alpha_k\|, \tag{4.3.3}$$

where K_0 denotes the true number of groups and $Q_{1,NT}(\beta)$ the profile log-likelihood function that is proposed by Hahn and Newey (2004) and Hahn and Kuersteiner (2011). λ_1 represents the tuning parameter, a certain chosen constant on which I will elaborate later in this paper.

For the pooling averaging estimator, optimal weights will be chosen over the C-Lasso estimates, obtained for several different numbers of groups $G = 1, \ ... \ , G_{max}$. Each estimated vector of regression parameters $\hat{\beta}_{(m)}$ of the m-th pooling strategy corresponds to a grouping structure with an associated number of groups in the C-Lasso procedure. Since the estimates for the various group structures are combined in an optimal way for a pooling averaging approach, the information criterion that is proposed by Su et al. (2016) will not be used in the pooling averaging procedure.

4.4 Choosing Mallows Pooling Averaging Weights

The procedure for choosing optimal weights in pooling averaging estimation, which of central interest in this research, is the Mallows pooling averaging approach. The feasible Mallows criterion is specified as follows:

$$C_A^*(w) = \|P(w)\hat{\beta} - \hat{\beta}\|_A^2 + 2tr[P'(w)A\hat{V}] - \|\hat{\beta} - \beta\|_A^2 \tag{4.4.1}$$

The distance $\|\gamma\|_A$ is defined as $\gamma' A \gamma$ for any vector or matrix γ. In case forecasting performance is of central interest, A is set equal to $X'X$, and when obtaining accurate estimation results is the ultimate aim, $A = I_{NK}$. As is already specified in equation 4.3.1, $P(w)$ is the projection matrix and $P(w)\hat{\beta}$ is the eventual coefficient estimate that results from Mallows pooling averaging, for which the weights w are to be find. This criterion is a generalization of the criterion that was first proposed by Hansen (2007), for which A should be equal to $X'X$ and the variance of the errors is assumed to be homoskedastic. The optimal weights can naturally be obtained by minimizing the criterion with respect to vector of weights w:

$$\hat{w}* = \arg\min_{w \epsilon W} C_A^*(w) \tag{4.4.2}$$

$\hat{V}$ is the estimated variance-covariance matrix that depends on the assumptions for the errors u_i in model setup 4.1.1. The three different cases for the error structure are already described in Section 4.1. In case of homoskedasticity, V is estimated by:

$$\hat{V}_{homo} = \tilde{\sigma}(X'X)^{-1},$$

in which $\tilde{\sigma}^2$ denotes the variance of the OLS residuals, that is $\tilde{\sigma}^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta})/(NT - NK)$. When the variance is heteroskedastic between individuals, this expression is:

$$\hat{V}_{bh} = \text{diag}(\hat{\sigma}_1^2 Q_1^{-1}, ..., \hat{\sigma}_N Q_N^{-1}),$$

where $\hat{\sigma}_i$ is the estimated variance of individual i that is obtained from the OLS residuals u_i; $\hat{\sigma}_i'\hat{\sigma}_i/(T-k)$. In addition to that, Q_i is computed as $Q_i = X_i'X_i/T$ for $i = 1, ..., N$. In the most general case, the variance is assumed to be different across groups and conditionally heteroskedastic within individuals. The estimated variance then can be obtained by the formula:

$$\hat{V}_{ch} = \frac{1}{T(T-k)}\text{diag}(Q_1^{-1}\sum_{t=1}^{T}\hat{u}_{1t}^2 X_{1t}' X_{1t} Q_1^{-1}, ..., \sum_{t=1}^{T}\hat{u}_{Nt}^2 X_{Nt}' X_{Nt} Q_N^{-1}).$$

Now that the procedure of choosing weights through minimizing Mallows Criterion is clear, let us have a closer look on the theoretical justification of this approach. Under four conditions that are specified by Wang et al. (2018), it can be proved that the MPA estimator in combination with the proposed C-Lasso approach as a preliminary model shrinkage procedure, exhibits asymptotic optimality:

$$\frac{L_A(\hat{w}^S)}{inf_{w\epsilon W} L_A(w)} \to 1.$$

In this expression, $L_A(\hat{w}^S)$ is the squared loss function $L_A(w) = \|\hat{\beta}(w) - \beta\|_A^2$ for the weights $\hat{w}^S$ associated with the pooling averaging estimator $\hat{w}^S = \arg\min_{w\epsilon W^S} C_A^*(w)$. In this pooling averaging expression, W^S is the restricted space of all possible weight vectors; $W^S = \{w \;\epsilon\; [0,1]^M : \sum_{m\in M^S} w_m = 1 \text{ and} \sum_{m\notin M^S} = 0\}$. The corresponding model space can be mathematically denoted by M^S and is a subset of $M = \{1, \; ..., \; M\}$. Conveniently, all weights that are restricted to sum up to one for all pooling techniques in the model space are included in the restricted model space, for which zero weight should be assigned to all pooling strategies that are shrinked out by, in our case, the C-Lasso procedure. The proof of this optimality theorem is done in the research of Wang et al. (2018), which is based on an asymptotic optimality proof that is given by Zhang et al. (2016).

4.5 Alternative Pooling Averaging Strategies

In order to compare the estimation and forecasting performance of the MPA estimator with other estimation techniques, I will estimate the regression coefficients through various alternative techniques. Note that the associated forecasts for the estimated parameter $\hat{\beta}$ can be computed as $\hat{y} = X\hat{\beta}$, for each estimation technique. First of all, the forecasts from the MPA estimator can be directly compared to the individual and pooled forecasts.

Another estimator that immediately weights the individual and pooled estimators is the shrinkage estimator, which is proposed by Maddala et al. (1997). It is defined as:

$$\hat{\beta}_{shrinkage} = (1 - \frac{\nu}{F})\hat{\beta}_{ind} + \frac{\nu}{F}\hat{\beta}_{pool}, \tag{4.5.1}$$

where F is the test statistic for equal coefficients across individuals $i = 1, ..., N$, thus the F statistic for the pooled model restrictions. The degrees of freedom is given by $\nu = \frac{(N-1)K-2}{NT-NK+2}$.

Another class of estimators selects the group estimate that results from the C-Lasso procedure, based on some data driven criterion. From this class, I will use two approaches for selecting the number of estimated groups, namely the commonly known AIC-criterion and BIC-criterion. The grouped estimates of the C-Lasso procedure with the smallest associated AIC and BIC will be chosen as estimators for these two estimation methods, respectively. Another possibility is to minimize these data driven criteria over the various estimates that result from choosing different numbers of groups. This results in a AIC and BIC optimal choice of the weight vector, the so-called Smoothed AIC (SAIC) and Smoothed BIC (SBIC) estimators. The weights will be chosen by minimizing the SAIC and SBIC criterion, which are specified by Burnham and Anderson (2004) in the following way, respectively:

$$w_{i,SAIC} = \frac{\exp(-\Delta AIC_i/2)}{\sum_{m=1}^{M}\exp(-\Delta AIC_m/2)}, \tag{4.5.2}$$

where $\Delta AIC_i = AIC_i - \min_{m=1,...,M} AIC_m$, for all groups $i = 1, ..., M$.

$$w_{i,SBIC} = \frac{\exp(-\Delta BIC_i/2)}{\sum_{m=1}^{M}\exp(-\Delta BIC_m/2)}, \tag{4.5.3}$$

where $\Delta BIC_i = BIC_i - \min_{m=1,...,M} BIC_m$, for all groups $i = 1, ..., M$.

5 Application to CDS Swaps

Now that I have specified the assumed model setup and explained several pooling averaging strategies, of which MPA estimation is of most interest, the theory can be applied to various financial products. Pooling averaging estimation will be used to determine the effect of local and global variables on sovereign CDS spreads for fourteen different countries. The countries concerned in this paper and the local and global determinants used in the regression are already specified in Section 3.

Sovereign CDS swaps are nowadays particularly interesting, because since the worldwide financial crisis in 2008, credit risk of countries has become more important. In addition to that, investigating these financial products by a pooling averaging approach as MPA estimation, is very relevant. Current research has namely focused on estimating and predicting sovereign CDS swaps by individual regressions. In the research of Aizenman et al. (2013) for example, a common pattern is observed for five countries in the South-West Eurozone. Also, the results of Dieckmann and Plank (2011) on sovereign CDS swaps for mostly European countries suggest that a fiscal insurance mechanism drives the value of these financial products. The observed co-movement for sovereign CDS swaps has not been used in these works of literature for improving estimation and forecasting. Thus applying MPA estimation and forecasting for this kind of financial product can be seen as a useful continuation and extension on the already existing research. The model that will be estimated in this empirical application is given by:

$$\Delta CDS_{it} = \alpha_i + X'_{i,t-1}\beta_i + \epsilon_{it}, \quad i = 1, ..., N \quad t = 1, ..., T. \tag{5.0.1}$$

As becomes clear from this expression, the prices of CDS swaps are first differenced for each country individually, which yields $\Delta CDS_{it} = CDS_{it} - CDS_{i,t-1}$. In this way, all dependent variable series are stable, as can be proved by performing unit root tests. $X_{i,t-1}$ is the lagged vector of local and global determinants, for which exogeneity will be assumed. All variables in this model are normalized, as is done in Wang et al. (2018), in order to make the individual coefficients comparable to each other. Note that this means that the constant term α_i is automatically removed from the regression for each country i. For now, the vector of regression coefficients β_i is assumed to be homogeneous within countries and heterogeneous across countries.

The defined pooling average technique requires that the data series is stable over time. Considering the fact that a financial crisis is present in the time period, this can still be a questionable assumption. There are multiple ways to capture the potential heterogeneity for the regression coefficient in (5.0.1). One method is to allow for structural breaks. In the literature of Wang et al. (2018), two structural breaks are found through the structural break detection method of Baltagi et al. (2016). The time points of these discrete breaks are January 2009 and December 2009, which consequently results in three different regimes. In my research I will directly implement these break points for estimation and forecasting.

5.1 Determining Local and Global Effects

For determination of the local and global determinants of the fourteen concerned countries, I will apply MPA estimation to the model 5.0.1. As already mentioned in this paper, this approach combines individual characteristics and co-movement of countries in an optimal way. Since it requires stability of the estimated model, I will implement this pooling averaging estimation to the three regimes separately. To facilitate computational feasibility, a preliminary step of model shrinkage is required. The C-Lasso procedure will be implemented for this, with tuning parameter $\lambda = c_\lambda s_Y^2 T^{-1/3}$, as is used by Su et al. (2016). In this expression, s_Y represents the sample variance of ΔCDS_{it} for $i = 1, ..., N$ and $t = 1, ..., T$. Also, for the constant c, a geometrically increasing sequence $\{0.0625, 0.125, 0.25, 0.5, 1\}$ will be used. The preliminary C-Lasso will be performed for the possible groups $G = 1, ..., G_{max}$, in which the maximum number of groups G_{max} is set to $N/2$. Using the C-Lasso estimated for the various groups, I will compute optimal weights for the MPA estimator by minimizing Mallows Criterion, specified in 4.4.1. Since I want to make statistical inference and the mentioned C-Lasso procedure and MPA criterion only calculate point estimates, a bootstrap procedure is implemented in both pre-screening and estimation. Using this bootstrap procedure, the empirical variance and confidence intervals can be calculated using B cross-sectional resamples, as is proposed by Kapetanios (2008).

5.2 Forecasting Sovereign Credit Risk

The introduced pooling averaging approach by Wang et al. (2018) cannot only help improving estimation accuracy, but can also achieve better forecasting performance. In order to evaluate the MPA strategy on forecasting performance, it will be used for out-of-sample forecasting sovereign CDS swaps.

This will be done for the whole sample and for post-samples after each of the two structural breaks. For the purpose of comparison, other estimation methods will be utilised by forecasting the same data set of sovereign CDS swaps. As explained in Section 4.5, these methods are; C-Lasso, SAIC, SBIC, AIC, BIC, pooled estimation, FGLS, SHK and individual estimation. A proportion τ will be used to estimate the model, and consequently $1 - \tau$ % will be used for out-of-sample forecasting. I will evaluate the forecasts based on three out-of-sample proportions; namely 0.1, 0.05 and 0.01. The Root Mean Squared Forecasting Error (RMSFE) will be the evaluation criterion for the forecasts of each estimation method, averaged over the fourteen different countries. For the purpose of comparison, all RMSFEs of the different forecasting techniques will be divided by the full sample individual time series RMSFE.

The results on the out-of-sample forecasting performance for the rolling window based on a fixed window size 1 - τ % are represented in Table 1. A rolling estimation window is implemented instead of a fixed window, because of the upcoming extension of optimal window size, which is only relevant in rolling or expanding window models. In this way, forecasting performance for

Table 1: Out-of-sample forecasting comparison with fixed window size for three sample choices and out-of-sample proportions 0.1, 0.05 and 0.01

	Full sample			**Post-first-break sample**			**Post-second-break sample**		
1 - τ %	**0.1**	**0.05**	**0.01**	**0.1**	**0.05**	**0.01**	**0.1**	**0.05**	**0.01**
MPA	0.951	0.953	0.902	0.947	0.809	0.651	1.065	0.870	0.619
C-Lasso	0.952	0.953	0.902	0.947	0.809	0.651	1.065	0.870	0.619
SAIC	0.951	0.953	0.902	0.947	0.809	0.651	1.065	0.870	0.619
SBIC	0.952	0.953	0.902	0.947	0.809	0.651	1.065	0.870	0.619
AIC	0.952	0.953	0.902	0.947	0.809	0.651	1.065	0.870	0.619
BIC	0.952	0.953	0.902	0.947	0.809	0.651	1.065	0.870	0.619
Pool	0.952	0.953	0.902	0.947	0.809	0.651	1.065	0.870	0.619
FGLS	1.002	1.006	0.972	0.983	0.990	1.061	1.005	0.984	1.000
SHK	0.991	0.993	0.986	0.986	0.886	0.793	1.067	0.894	0.630
Indiv	1.000	1.000	1.000	0.993	0.928	0.841	1.090	0.931	0.663

Note: All RMSFEs are divided by the individual RMSFEs that are estimated with the full sample for the out-of-sample proportions 0.1, 0.05 and 0.01

this extension can conveniently be compared to original results. Forecasts are also performed for an expanding window technique, these results are shown in Table 5 in the Appendix. It can be clearly seen that the MPA estimator gives the best estimation results from a forecasting perspective, although several estimation techniques yield almost equally good forecasts. Apparently, the (smoothed) criterion estimation approaches are very good at selecting the best pooling strategies, which usually concerns the strategy in which all individual countries are contained in one group. This means a pooled model provides the best forecasts in most cases, as can also be justified by the RMSFE from the pooled model that is almost equal to that of MPA for all subsamples and out-of-sample forecasting fractions. Thus the MPA estimators, with the matrix A chosen in such a way that it provides optimal forecasts, is very different from those that are obtained in Section 5.1, when A is set such that estimation accuracy is the goal.

6 Optimal window selection with structural breaks

Estimating and forecasting financial products correctly regarding time-varying effects of the local and global parameters is a topic that needs extensive research. As Pesaran and Timmermann (2004) mention, not covering time-varying breaks in regression coefficients correctly, can become very costly. In the paper of Wang et al. (2018), the heterogeneity of the regression parameters is captured through selecting separate estimation windows with regard to the detected discrete time breaks. In this way, forecasts for the various estimation techniques are constructed by using an estimation window based on the whole sample period, the sample period after the first structural break and the sample period after the second one. In Section 5.2 however, it can be observed that this does not necessary result in improved forecasts. Selecting a fixed window size, based on the possibility of two discrete breaks, is a very convenient but also rather restrictive way of accounting for time-varying regression parameters in estimation. For this reason, it would be interesting to incorporate a technique for optimally selecting the window, at each point in time a forecast is executed. Note that this is something that is explicitly left as a topic that deserves further research in the paper of Wang et al. (2018).

For this purpose, I will use the recently proposed method of Inoue et al. (2017) to address the potential time-varying nature of the effects of the concerned local and global variables on the sovereign CDS spreads. This approach chooses an optimal rolling window size for each time point at which a forecast is to be made for the next unknown observation. This is done by minimizing the MSFE at the end of the estimation sample, which makes this approach very suitable for the objective of providing optimal forecasts. Note that this approach is originally specified for a univariate data generating process (DGP). This requires me to extend the technique, in order to make it compatible with the heteroskedastic panel model that is under consideration.

The methodology of Inoue et al. (2017) is actually specified for a data generating process (DGP) with a vector of time-varying parameters, in other words a model that allows for smooth breaks over time. However, simulation results in this research show that the MSFE criterion for choosing the optimal window size performs also well for a process with structural, discrete breaks. This is even the case for DGPs with only one time break. This makes the method very well appropriate for system 5.0.1 that is conditioned on two already obtained structural breaks, also when there would be no actual smooth change of the regression parameters over time.

Before explaining the methodology of this extension, I first need to clearly specify several important considerations. Firstly, when generalizing the procedure of Inoue et al. (2017) for a multivariate panel model, the optimal rolling window size can be determined for each country individually, or for each group in each of the $N/2$ grouping structures, resulting from the C-Lasso shrinkage procedure. Since it is my goal to develop a window selection technique that optimizes forecasting performance, I will implement both methods in order to possibly induce a preferred multivariate method for window selection.

Secondly, for both techniques, the resulting rolling window is used for estimating all parameters in the MPA criterion 4.4.1, determining the projection matrices and for choosing the matrix $A = X'X$. The estimation of the conditional heteroskedastic variance structure, as is explained in Section 4.4, is calculated from the whole rolling window using the fraction 1 - τ % of the concerned sample. In this way, I still get an estimated variance matrix that is as accurate as possible and does not use different samples for the explanatory variables that define the error terms over time. This means the original individual OLS estimates are used for estimating the matrix that captures covariance.

6.1 Individual Optimal Windows

In the individual window selection approach, the optimal window size R_i is chosen such that the MSFE at the last observation of the estimation sample, $t = T$, is minimized. Thus, for each forecast in the forecasting sample, an optimal window is selected from the prior rolling window of 1 - τ percentage of concerned sample. This will be done in such a way that the MSFE of the concerned forecast will be minimised. Following the derivation of Inoue et al. (2017), this is equivalent to minimizing the following expression:

$$(\hat{\beta}_{i,R} - \beta_i)' X_{i,T} X'_{i,T} (\hat{\beta}_{i,R} - \beta_i). \tag{6.1.1}$$

It is infeasible to minimize this expression, since the true value of the regression parameter is unknown. The actual parameter β_i has to be estimated through the most accurate estimation technique possible. In Inoue et al. (2017), $\hat{\beta}_{i,R}$ is calculated as the rolling OLS estimate and β_i is replaced with the local linear regression estimator $\tilde{\beta}_i$:

$$(\hat{\beta}_{i,R} - \tilde{\beta}_i)' X_{i,T} X'_{i,T} (\hat{\beta}_{i,R} - \tilde{\beta}_i). \tag{6.1.2}$$

In this paper, I take $\hat{\beta}_{i,R}$ to be the rolling OLS estimate that is calculated from the estimation window, with the R most recent observations of country i at the last observation $t = T$ of the estimation sample:

$$\hat{\beta}_{i,R} = (\sum_{t=T-R+1}^{T} X_{i,t-1} X'_{i,t-1})^{-1} (\sum_{t=T-R+1}^{T} X_{i,t-1} \Delta CDS_{i,t}) \tag{6.1.3}$$

$\tilde{\beta}_i$ represents the local linear regression estimator. Using this estimation technique for obtaining the actual regression parameter can be justified by the research of Cai (2007), which shows that this estimator exhibits the same asymptotic behaviour at interior points as at boundaries. The vector of the local linear regression estimate $\tilde{\beta}_i$ and its first derivative $\tilde{\beta}_i^{(1)}$ in the associated Taylor expansion for country i, is computed in the following ay:

$$\begin{aligned}\begin{pmatrix} \tilde{\beta}_{i,T} \\ \tilde{\beta}_{i,T}^{(1)} \end{pmatrix} = &\begin{pmatrix} \sum_{t=T-R_0+1}^{T} X_{i,t-1} X'_{i,t-1} & \sum_{t=T-R_0+1}^{T} X_{i,t-1} X'_{i,t-1} (\frac{t-T}{T}) \\ \sum_{t=T-R_0+1}^{T} X_{i,t-1} X'_{i,t-1} (\frac{t-T}{T}) & \sum_{t=T-R_0+1}^{T} X_{i,t-1} X'_{i,t-1} (\frac{t-T}{T})^2 \end{pmatrix}^{-1} \\ &\times \begin{pmatrix} \sum_{t=T-R_0+1}^{T} X_{i,t-1} \Delta CDS_{it} \\ \sum_{t=T-R_0+1}^{T} X_{i,t-1} \Delta CDS_{it} (\frac{t-T}{T}). \end{pmatrix}\end{aligned} \tag{6.1.4}$$

Using these two estimation techniques results in a feasible approach for selecting the optimal individual window at $t = T$:

$$\hat{R} = \arg\min_{R \epsilon \odot_R} (\hat{\beta}_{i,R} - \tilde{\beta}_i)' X_{i,T} X'_{i,T} (\hat{\beta}_{i,R} - \tilde{\beta}_i). \qquad (6.1.5)$$

Note that the initial optimal window size R_0 for this minimization procedure is set to be the one computed from the proposed technique of Pesaran and Timmermann (2007).

Table 2: Out-of-sample forecasting performance with optimal individual window size selection for three sample choices and out-of-sample proportions 0.1, 0.05 and 0.01

	Full sample			**Post-first-break sample**			**Post-second-break sample**		
1 - τ %	**0.1**	**0.05**	**0.01**	**0.1**	**0.05**	**0.01**	**0.1**	**0.05**	**0.01**
MPA	0.669	0.612	0.366	0.616	0.552	0.311	0.637	0.561	0.314
C-Lasso	0.669	0.612	0.366	0.616	0.552	0.311	0.637	0.561	0.314
SAIC	0.669	0.612	0.366	0.616	0.552	0.311	0.637	0.561	0.314
SBIC	0.669	0.612	0.366	0.616	0.552	0.311	0.637	0.561	0.314
AIC	0.669	0.612	0.366	0.616	0.552	0.311	0.637	0.561	0.314
BIC	0.669	0.612	0.366	0.616	0.552	0.311	0.637	0.561	0.314
Pool	0.669	0.612	0.366	0.616	0.552	0.311	0.637	0.561	0.314
FGLS	0.651	0.694	0.457	0.634	0.645	0.421	0.710	0.638	0.310
SHK	0.756	1.359	0.427	0.657	1.341	0.319	2.058	2.179	0.334
Indiv	0.673	0.627	0.399	0.603	0.617	0.309	0.626	0.614	0.321

Table 3: Out-of-sample forecasting performance with fixed window size for three sample choices and out-of-sample proportions 0.1, 0.05 and 0.01

	Full sample			**Post-first-break sample**			**Post-second-break sample**		
1 - τ %	**0.1**	**0.05**	**0.01**	**0.1**	**0.05**	**0.01**	**0.1**	**0.05**	**0.01**
MPA	0.621	0.657	0.426	0.618	0.558	0.307	0.694	0.600	0.292
C-Lasso	0.622	0.657	0.426	0.618	0.558	0.307	0.694	0.600	0.292
SAIC	0.621	0.657	0.426	0.618	0.549	0.307	0.694	0.600	0.292
SBIC	0.622	0.657	0.426	0.618	0.558	0.307	0.694	0.600	0.292
AIC	0.622	0.657	0.426	0.618	0.558	0.307	0.694	0.600	0.292
BIC	0.622	0.657	0.426	0.618	0.558	0.307	0.694	0.600	0.292
Pool	0.621	0.657	0.426	0.618	0.558	0.307	0.694	0.600	0.292
FGLS	0.654	0.694	0.459	0.637	0.634	0.421	0.714	0.632	0.313
SHK	0.647	0.684	0.465	0.632	0.611	0.374	0.696	0.616	0.297
Indiv	0.652	0.690	0.472	0.648	0.640	0.397	0.711	0.642	0.313

The resulting RMSFE for the different estimation techniques and out-of-sample proportions are reported in Table 2. To facilitate comparison between the forecasting performance of this extended method with the original forecasting results of Section 5.2, the results on forecasting performance for the initial fixed window size model are displayed in Table 3 again, now with RMSFEs that are not normalized by the RMSFE of individual regressions using the full sample period.

When comparing the results of Table 2 to those of Table 3, it becomes clear that extending MPA estimation with the optimal rolling window selection approach of Inoue et al. (2017) does not yield better forecasts in all cases. However, the constructed forecasts are better for the out-of-sample proportions 0.05 and 0.01 for the full sample and for the fractions 0.1 and 0.05, if only post-first-break or post-second-break data points are used. From this, it can be induced that selecting optimal windows in MPA estimation can be useful for rather small forecasting proportions, when all discrete breaks are ignored. It is also profitable for large proportions, when parameter stability is already incorporated in the selection of the estimation window through discrete breaks. The latter also suggests that the adjusted approach of Inoue et al. (2017) captures the smooth change of the regression parameters well within regimes. This also establishes the suspicion of smoothly time-varying regression parameters for the data set under consideration.

Regarding other estimation techniques, a significant improve for the individual forecasts can be observed for almost all cases. This improvement is of a certain level that it causes the individual estimation with optimal window selection to defeat the MPA estimation, in which optimal window sizes are not taken into consideration. This is particularly the case in situations in which the generalised MPA technique also beats the one from the original replication approach.

6.2 Group Optimal Windows

Adjusting optimal window selection for the concerned multivariate panel model can also be done by a pooled estimation approach for the optimal window size within each group. This means countries that are captured in the same group share the same optimal window size for all forecasts, which naturally is allowed to be time-varying. Countries that are grouped in different entities are allowed to have different optimal window sizes. While seven different grouping structures are constructed for the MPA estimation, this approach is implemented for each of the seven grouping structures, after which weights will be chosen to be optimal according to the usual MPA criterion.

One difficulty for this method concerns the fact that there are seven possible individual OLS estimates for β in this criterion 4.4.1, each corresponding to a window size selection for one grouping structure. I will choose β to be the vector of individual OLS estimates from optimal window selection in which all countries are shrinked into one group. It is chosen like this, since the pooled model proves itself to provide good forecasts in Section 5.2. This could namely suggest that a pooled window size vector is also the best choiche from a forecasting perspective. As this still concerns a suspicion and not something that is explicitly proved, choosing alternative window sizes for β regarding the grouped window selection remains a topic that deserves further research. Consequently, the pooled, shrinkage and FGLS estimation techniques are also based on the individual OLS estimates that select a pooled window size for one group.

Most methodological expressions for individual window selection can be utilised in this grouped procedure. For each time point $t = T$, at which a forecast should be constructed for the next observation, the feasible MSFE of each individual country is denoted by 6.1.2. In this expression, the rolling OLS estimate $\hat{\beta}_{i,R}$ and the local linear regression estimate $\tilde{\beta}_i$ are computed by the formulas 6.1.3 and 6.1.4, respectively. The crucial difference in computation now concerns the fact that the optimal window size $\hat{R}$ at each time point T is obtained through the following expression:

$$\hat{R}_i = \underset{R \epsilon \odot_R}{\arg\min} \sum_{i \epsilon G_k} (\hat{\beta}_{i,R} - \tilde{\beta}_i)' X_{i,T} X_{i,T}' (\hat{\beta}_{i,R} - \tilde{\beta}_i), \tag{6.2.1}$$

where G_k is the k-th group in a particular grouping structure. In this way, the average MSFE within each group is minimized with respect to the window size. This results in seven individual OLS-estimators, in each of which the window size is selected in an optimal way regarding the associated group structure. For every individual vector of estimates, the projection matrix is constructed with the same estimation window for each group of individuals. Minimizing the MPA criterion for each of these quantities results in RMSFEs that are shown in Table 4.

These results show that the MPA estimator for which the grouped procedure of choosing optimal window sizes is incorporated, yields forecasts that are better than those associated with the original MPA estimates for most cases. Nevertheless, the better forecasting performance is not consistent in any of the samples, since for all three sample cases, there is one out-of-sample proportion for which MPA with selected grouped window sizes performs worse than the previous MPA estimates. For the full sample and the post-first-break sample, it performs better with relatively small fractions 0.05 and 0.01, while for the post-second-break case, it performs best for the two biggest proportions 0.10 and 0.05. These results, together with the results in Table 2, imply that the window selection method of Inoue et al. (2017) can be useful for MPA estimation in a significant amount of cases, as a way of capturing the smooth change of regression parameters to improve forecasting performance.

Table 4: Out-of-sample forecasting performance with optimal group window size selection for three sample choices and out-of-sample proportions 0.1, 0.05 and 0.01

	Full sample			**Post-first-break sample**			**Post-second-break sample**		
1 - τ %	**0.1**	**0.05**	**0.01**	**0.1**	**0.05**	**0.01**	**0.1**	**0.05**	**0.01**
MPA	0.691	0.638	0.358	0.631	0.554	0.290	0.638	0.585	0.295
C-Lasso	0.691	0.638	0.358	0.631	0.554	0.290	0.638	0.585	0.295
SAIC	0.682	0.638	0.358	0.631	0.554	0.290	0.638	0.585	0.295
SBIC	0.691	0.638	0.358	0.631	0.554	0.290	0.638	0.585	0.295
AIC	0.682	0.638	0.358	0.631	0.554	0.290	0.638	0.585	0.295
BIC	0.691	0.638	0.358	0.631	0.554	0.290	0.638	0.585	0.295
Pool	0.691	0.638	0.358	0.631	0.554	0.290	0.638	0.585	0.295
FGLS	0.656	0.698	0.460	0.646	0.644	0.404	0.721	0.664	0.316
SHK	1.223	0.697	0.781	0.736	1.983	0.317	1.080	1.600	1.065
Indiv	0.682	0.669	0.389	0.639	0.621	0.281	0.644	0.620	0.355

When the two optimal window selection methods are particularly compared to each other, it can be said that the individual window method performs better for the rather large out-of-sample proportions 0.1 and 0.05, and that the grouped window method gives better forecasts for a small out-of-sample proportion of 0.01. The fact that this is the case for all three sample cases makes this notion quite credible, although it should be mentioned that it cannot be generalised directly for several other financial products or different data sets.

7 Conclusion and Discussion

In this research, I have replicated the paper of Wang et al. (2018) for estimating and forecasting sovereign CDS spreads through regression on several local and global determinants for 14 countries. This has been done by a proposed pooling average technique that minimizes the MPA criterion. I have extended the research by incorporating a time-varying estimation window that aims to further minimize the MSFE of the forecasts. This has been done by selecting an optimal time-varying window size, firstly for individual countries and after that for groups of countries.

For the replication part, MPA usually provides optimal forecasts when compared to alternative estimation techniques. However, the choiche of optimal weights for the considered fixed rolling window size does not improve forecasting performance significantly if MPA is compared to approaches that choose only one pooling technique based on some (smoothed) information criterion, such as C-Lasso IC, (S)AIC and (S)BIC. The reason for this concerns the notion that when forecasting is of central interest in the MPA procedure, mostly full weight is assigned to the pooled model. This is the reason why the RMSFE of MPA forecasts usually is quite similar to that of the purely pooled forecasts.

This paper shows that MPA can be further improved with respect to forecasting performance by addressing potential smooth change of regression parameters. Individual selection of optimal windows improves the forecasting errors for most out-of-sample proportions for different subsamples. Selecting optimal windows for estimated group structures of countries instead of individual countries also gives better forecasts for most cases. Since one of the two window selection approaches does not consistently yields better forecasts than the other, a general rule for window selection cannot be induced from the obtained results. Nevertheless, since the individual window technique commonly provides best forecasts for rather large out-of-sample proportions, and the grouped window technique does so for the smallest out-of-sample proportion, it can be suggested that the two approaches can be used in conjunction, dependent on the fraction and number of forecasting observations.

Further research can be performed on using the proposed methods for determining optimal window sizes in a broader range of financial products, such that general rules can be established for the forecasting performance of these techniques. Furthermore, incorporating optimal weights within estimation windows may also improve forecasting performance for the window optimization techniques.

References

Aizenman, J., Hutchison, M., and Jinjarak, Y. (2013). What is the risk of european sovereign debt defaults? fiscal space, cds spreads and market pricing of risk. *Journal of International Money and Finance*, 34:37–59.

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., Feng, Q., and Kao, C. (2016). Estimation of heterogeneous panels with structural breaks. *Journal of Econometrics*, 191(1):176–195.

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.

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

Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188.

Dieckmann, S. and Plank, T. (2011). Default risk of advanced economies: An empirical analysis of credit default swaps during the financial crisis. *Review of Finance*, 16(4):903–934.

Evans, P. and Kim, J. U. (2011). Stochastic convergence of the catch-up rate and multiple structural breaks in asian countries. *Economics Letters*, 111(3):260–263.

Hahn, J. and Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory*, 27(6):1152–1191.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319.

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

Hildreth, C. and Houck, J. P. (1968). Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association*, 63(322):584–595.

Hsiao, C. (1975). Some estimation methods for a random coefficient model. *Econometrica: Journal of the Econometric Society*, pages 305–325.

Hsiao, C. and Pesaran, M. H. (2008). Random coefficient models. In *The econometrics of panel data*, pages 185–213. Springer.

Inoue, A., Jin, L., and Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, 196(1):55–67.

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

Li, D., Qian, J., and Su, L. (2016). Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association*, 111(516):1804–1819.

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. A., Pan, J., Pedersen, L. H., and Singleton, K. J. (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. H., Pick, A., and Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, 177(2):134–152.

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

Pesaran, M. H. and Timmermann, A. (2004). How costly is it to ignore breaks when forecasting the direction of a time series? *International Journal of Forecasting*, 20(3):411–425.

Pesaran, M. H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1):134–161.

Qian, J. and Su, L. (2016). Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. *Journal of Econometrics*, 191(1):86–109.

Remolona, E. M., Scatigna, M., and Wu, E. (2008). The dynamic pricing of sovereign risk in emerging markets: Fundamentals and risk aversion. *Journal of Fixed Income, Vol. Spring.*

Rosenberg, B. (1972). The estimation of stationary stochastic regression parameters reexamined. *Journal of the American Statistical Association*, 67(339):650–654.

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

Swamy, P. A. (1970). Efficient inference in a random coefficient regression model. *Econometrica: Journal of the Econometric Society*, pages 311–323.

Wang, W., Zhang, X., and Paap, R. (2018). To pool or not to pool: What is a good strategy for parameter estimation and forecasting in panel regressions? *Journal of Applied Econometrics*.

Zhang, X., Yu, D., Zou, G., and 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.

A Appendix

A.1 Data Description

As is mentioned in Section 3, the summary statistics of the sovereign CDS spreads for all countries are shown below. These are from left to right the mean, standard deviation (SD), minimum value (min), median (med) and maximum value (max) for each country.

	Mean	**SD**	**Min**	**Med**	**Max**
Brazil	269.08	263.76	62.16	161.59	1733.33
Bulgaria	166.41	118.83	13.42	150.16	597.39
Chile	75.82	47.42	13.16	74.67	266.56
China	69.48	44.52	10.00	71.79	248.34
Hungary	188.24	164.00	11.00	163.00	623.20
Japan	41.82	35.69	2.17	39.50	146.47
Korea	84.89	2.57	14.30	67.34	432.48
Malaysia	88.36	55.74	12.86	85.32	296.39
Philippines	227.32	141.26	82.16	171.16	617.50
Poland	86.52	73.93	8.13	67.67	366.00
Romania	200.02	140.34	17.75	182.49	723.56
South Africa	150.52	82.22	25.06	150.25	459.93
Slovakia	66.97	65.21	6.00	50.33	298.29
Thailand	97.15	55.49	26.94	100.13	298.34

When these are compared to the summary statistics of Longstaff et al. (2011) for the same countries, I observe that the mean values of the various countries are closer to each other. Also, the maximum values are less extreme. This can be a sign of globalisation of the past several years, since the data set of Longstaff et al. (2011) terminates in 2010.

A.2 Expanding Window Forecasting Results

Table 5: Out-of-sample forecasting comparison with expanding window for three sample choices and out-of-sample proportions 0.1, 0.05 and 0.01

	Full sample			**Post-first-break sample**			**Post-second-break sample**		
1 - τ %	**0.1**	**0.05**	**0.01**	**0.1**	**0.05**	**0.01**	**0.1**	**0.05**	**0.01**
MPA	0.956	0.960	0.903	1.019	0.795	0.652	1.021	0.867	0.619
C-Lasso	0.956	0.960	0.903	1.019	0.795	0.652	1.021	0.867	0.619
SAIC	0.956	0.960	0.913	1.048	0.799	0.686	0.997	0.883	0.616
SBIC	0.956	0.960	0.913	1.048	0.799	0.686	0.997	0.883	0.616
AIC	0.968	0.960	0.913	1.048	0.799	0.686	0.997	0.883	0.616
BIC	0.968	0.960	0.913	1.048	0.799	0.686	0.997	0.883	0.616
Pool	0.956	0.960	0.903	1.019	0.795	0.652	1.021	0.867	0.619
FGLS	1.016	1.005	0.978	1.114	0.967	0.894	1.055	0.965	0.665
SHK	0.992	0.993	0.986	1.076	0.894	0.793	1.023	0.899	0.631
Indiv	1.000	1.000	1.000	1.108	0.936	0.842	1.046	0.939	0.663

A.3 Code Explanation

A.3.1 Estimation

1. main
   - The main program, which calculates the MPA estimates without bootstrap for the three different regimes
2. main_boot
   - The main program, which calculates the standard errors for the MPA estimates through the bootstrap technique proposed by Kapetanios (2008), for the three different regimes

Note: The other programs in the folder Estimation are also present in the folder Forecasting, although slightly adapted in order to be used by a program that provides and evaluates forecasts. These still have the exact same functions as the corresponding functions in the folder Estimation, which is the reason these are not explained in this subsection, but only once in subsection A.3.2.

A.3.2 Forecasting

1. mainrolling
   - Main program for calculating the RMSFE for all estimation techniques, subsamples and out-of-sample proportions
2. dataTransformation
   - Function that transforms the given data set by first differencing the sovereign CDS swaps, ensuring the explanatory variables are incorporated as one-period lagged variables
3. individualEstimation
   - Estimates individual OLS estimates for a given sample
4. CLasso
   - Shrinks the set of possible pooling strategies through grouping with the C-Lasso procedure. This code is literally obtained from 'https://github.com/zhentaoshi/C-Lasso/blob/master/ generic_functions/SSP_PLS_est.m'

5. numberOfRestrictions
   - Function that calculates the number of equality restrictions for a given grouping structure that is obtained from the C-Lasso procedure
6. CLassoIC
   - Function that calculates the performance of a given grouping structure by determining the sigma part of the information criterion that is specified by Su et al. (2016). This code is literally obtained from 'https://github.com/zhentaoshi/C-Lasso/blob/master/app_saving_PLS/hat_IC.m'
7. pooledEstimation
   - Function that calculates the pooled regression estimates for a given sample
8. shrinkageEstimation
   - Function that calculates the shrinked estimates for a given sample
9. FGLSEstimation
   - Function that calculates the FGLS estimates for a given sample
10. computeHOMOVariance
    - Function that estimates the variance-covariance structure for the estimation window data, by assuming homoskedastic errors
11. computeBHVariance
    - Function that estimates the variance-covariance structure for the estimation window data, by assuming a between-heteroskedastic error structure
12. computeCHVariance
    - Function that estimates the variance-covariance structure for the estimation window data, by assuming a conditionally heteroskedastic errors
13. calculateProjection
    - Function that calculates the projection matrix and the pooled regression estimator for a given sample and grouping structure
14. computeSSR

- Function that calculates the sum of squared residuals for a given sample and estimator, which is needed for several estimation techniques, such as the (S)AIC estimation and (S)BIC estimation approach

15. calculateWeightsMPA

    - Optimal MPA weights estimator that calculates the weights by minimizing Mallows criterion for given estimations of the variance matrix, individual parameters, projection matrix and matrix A

16. criterionEstimation

    - Program that picks the pooled estimator from the pooling strategy that minimizes a certain criterion, in our case the AIC and BIC, respectively

17. SAICEstimation

    - Function that returns the SAIC estimator by choosing weights over given pooling strategy estimators which minimizes the smoothed AIC criterion

18. SBICEstimation

    - Function that returns the SBIC estimator by choosing weights over given pooling strategy estimators which minimizes the smoothed BIC criterion

19. mainexpanding

    - Main program that does the same as mainrolling, but now for a window that expands as forecasts are iteratively obtained

20. calculateProjectionExpanding

    - Program that calculates the projection matrix and the associated pooled beta for a given grouping structure for an expanding window size

21. mainIndVarWinSize

    - Main program for calculating the RMSFE for all estimation techniques, subsamples and out-of-sample proportions, in which the optimal window selection approach of Inoue et al. (2017) is individually incorporated for all countries

22. calculatingVaryingWindowSize

- Function that calculates the optimal estimation window size for each country individually, at each time point a forecast is to be constructed

23. individualOptEstimation
    - Function that computes the individual OLS estimates, by using the individual optimal window sizes from the function calculatingVaryingWindowSize
24. pooledOptEstimation
    - Function that computes the pooled estimate, by using the individual optimal window sizes from the function calculatingVaryingWindowSize
25. shrinkageOptEstimation
    - Function that computes the shrinkage estimator, by using the individual optimal window sizes from the function calculatingVaryingWindowSize
26. calculateOptProjection
    - Program that calculates the projection matrix and associated pooled strategy estimator, like the function calculateProjection also does, but now by taking the optimal individual window sizes into account
27. calculateInitialWindowSize
    - Function that calculates the initial optimal window size, as proposed by Pesaran et al. (2013), which will be used for the minimization procedure of calculatingVaryingWindowSize
28. localLinearEstimation
    - Function that estimated the regression parameters of the individual countries via local linear regression, which is needed for estimating the MSFE in the function calculatingVaryingWindowSize
29. mainGrVarWinSize
    - Main program for calculating the RMSFE for all estimation techniques, subsamples and out-of-sample proportions, in which the optimal window selection approach of Inoue et al. (2017) is incorporated for each group over the different grouping structures, which result from the model shrinkage procedure

30. getOptimalGroupWindows
    - Function that computes the vector of optimal window size estimates for given estimation data and a particular grouping structure
31. calculateGroupWindow
    - Function that computes the optimal group window size for a given group from the function getOptimalGroupWindows
32. normalize
    - Normalizes input data, for our data this naturally concerns the sovereign CDS spreads and all explanatory local and global panel data