Supplementary file to

"To pool or not to pool: What is a good strategy for parameter estimation and forecasting in panel regressions?"

This supplementary file contains further theoretical properties of pooling averaging approaches. We also provide additional simulation studies and empirical results here that are not reported in the paper.

S.1 Unbiasedness of Mallows criterion

This section provides the theorem of the unbiasedness of Mallows criterion as an estimator or the squared risk. Consider the model

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

The pooling averaging estimator can be obtained by

$$\widehat{\beta}(w) = \sum_{m=1}^{M} w_m \widehat{\beta}_{(m)} = \sum_{m=1}^{M} w_m P_m \widehat{\beta} = P(w)\widehat{\beta}, \tag{S.2}$$

We propose to obtain the weights of pooling averaging estimator by minimizing the Mallows criterion as

$$\mathcal{C}_A(w) = \|P(w)\widehat{\beta} - \widehat{\beta}\|_A^2 + 2\mathrm{tr}[P'(w)AV] - \|\widehat{\beta} - \beta\|_A^2, \tag{S.3}$$

Theorem S.1. *Under model (S.1), the Mallows criterion defined in Equation (S.3) is an unbiased estimator of the squared risk $R_A(w)$.*

Proof. From (S.2), it is straightforward to show that

$$\begin{aligned} R_A(w) &= \mathrm{E}\{L_A(w)\} = \mathrm{E}\|\widehat{\beta}(w) - \beta\|_A^2 = \mathrm{E}\|P(w)\widehat{\beta} - \beta\|_A^2 \\ &= \mathrm{E}\|P(w)\widehat{\beta}\|_A^2 + \|\beta\|_A^2 - 2\beta' P'(w)A\beta \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}\{\mathcal{C}_A(w)\} &= \mathrm{E}\|P(w)\widehat{\beta} - \widehat{\beta}\|^2 + 2\mathrm{tr}\left\{P'(w)AV\right\} - \mathrm{tr}(AV) \\ &= \mathrm{E}\|P(w)\widehat{\beta}\|_A^2 + \mathrm{E}\|\widehat{\beta}\|_A^2 - \mathrm{tr}(AV) \end{aligned}$$

$$
\begin{aligned}
& -2\mathrm{E}\left\{\widehat{\beta}' P'(w) A \widehat{\beta}\right\} + 2\mathrm{tr}\{P'(w)AV\} \\
= \; & \mathrm{E}\|P(w)\widehat{\beta}\|_A^2 + \|\beta\|_A^2 - 2\beta' P'(w) A\beta. \qquad \text{(S.4)}
\end{aligned}
$$

So $\mathcal{C}_A(w)$ is an unbiased estimator of $R_A(w)$.

□

S.2 Equivalence of MPA and Stein-rule estimators

This sections provides the details on the relation between the MPA and the Stein-rule shrinkage estimator. Our pooling averaging estimator includes the shrinkage estimator of Maddala *et al.* (1997) as a special case. The shrinkage estimator is defined as

$$
\widehat{\beta}_{\text{shrinkage}} = \left(1 - \frac{\nu}{F}\right)\widehat{\beta} + \frac{\nu}{F}\widehat{\beta}_{\text{pool}}, \qquad \text{(S.5)}
$$

where $\nu = [(N-1)k - 2]/[NT - Nk + 2]$ and F is the test statistic for null hypothesis $H_0 : \beta_1 = \ldots = \beta_N$. More specifically, if we denote $\widetilde{R}$ as the restriction matrix associated with H_0 and

$$
\widetilde{\sigma}^2 = (Y - X\widehat{\beta})'(Y - X\widehat{\beta})/(NT - Nk), \qquad \text{(S.6)}
$$

the rank of $\widetilde{R}$ is $k(N-1)$ and the F statistic is

$$
F = (\widetilde{R}\widehat{\beta})'(\widetilde{R}(X'X)^{-1}\widetilde{R}')^{-1}(\widetilde{R}\widehat{\beta})/[(N-1)k\widetilde{\sigma}^2]. \qquad \text{(S.7)}
$$

The shrinkage estimator can be regarded as the pooling average of only the pooled and individual estimators.

Now we shall show how the Mallows pooling averaging estimator is associated with the shrinkage estimator of Maddala *et al.* (1997) in the context of combining only two estimators, $\widehat{\beta}$ and $\widehat{\beta}_{\text{pool}}$. In this case, the averaged estimator is $\widehat{\beta}(w) = w_1\widehat{\beta} + w_2\widehat{\beta}_{\text{pool}}$. Following Maddala *et al.* (1997), we assume $\sigma_1^2 = \cdots = \sigma_N^2 = \sigma^2$, and σ^2 can be estimated by $\widetilde{\sigma}^2$ as in (S.6). We consider the case with $A = X'X$. If the F statistic given by (S.7) is larger than 1, such that $1/F \in [0,1]$, then by minimizing $\mathcal{C}_A^*(w)$ we can obtain

$$
\widehat{w}_2 = \frac{1}{F}. \qquad \text{(S.8)}
$$

This result suggests that if we only average the pooled and individual estimators and $1/F \in [0,1]$, then the Mallows pooling averaging estimator is essentially a Stein-rule estimator (see Equation (2) of Maddala *et al.* (1997)). The weights of the Mallows pooling

averaging estimator and the shrinkage estimator defined by (S.5) are proportional to each other (see also Hansen, 2014). We provide the proof of (S.8) below.

Proof. Let $P_{\text{pool}} = I_{Nk} - (X'X)^{-1}\widetilde{R}'(\widetilde{R}(X'X)^{-1}\widetilde{R}')^{-1}\widetilde{R}$, where $\widetilde{R}$ is defined below (S.5). When $A = X'X$, we have

$$
\begin{aligned}
& \mathcal{C}_A^*(w) \\
= \; & \|w_1\widehat{\beta} + w_2\widehat{\beta}_{\text{pool}} - \widehat{\beta}\|_A^2 + 2\text{tr}[(w_1 I_{Nk} + w_2 P'_{\text{pool}})A\widehat{V}_{\text{homo}}] - \|\widehat{\beta} - \beta\|_A^2 \\
= \; & \|(1-w_2)\widehat{\beta} + w_2\widehat{\beta}_{\text{pool}} - \widehat{\beta}\|_A^2 + 2\text{tr}[((1-w_2)I_{Nk} + w_2 P'_{\text{pool}})A\widehat{V}_{\text{homo}}] - \|\widehat{\beta} - \beta\|_A^2 \\
= \; & w_2^2\|\widehat{\beta} - \widehat{\beta}_{\text{pool}}\|_A^2 + 2w_2\text{tr}(P'_{\text{pool}}A\widehat{V}_{\text{homo}}) + 2(1-w_2)\text{tr}(A\widehat{V}_{\text{homo}}) - \|\widehat{\beta} - \beta\|_A^2 \\
= \; & w_2^2\|\widehat{\beta} - \widehat{\beta}_{\text{pool}}\|_A^2 + 2w_2\widetilde{\sigma}^2\left\{\text{tr}(P_{\text{pool}}) - Nk\right\} + 2\widetilde{\sigma}^2 Nk - \|\widehat{\beta} - \beta\|_A^2, \qquad \text{(S.9)}
\end{aligned}
$$

where the last two terms have nothing to do with w. From (S.7) and

$$
\begin{aligned}
& \text{tr}((X'X)^{-1}\widetilde{R}'(\widetilde{R}(X'X)^{-1}\widetilde{R}')^{-1}\widetilde{R}) \\
= \; & \text{tr}((X'X)^{-1/2}\widetilde{R}'(\widetilde{R}(X'X)^{-1}\widetilde{R}')^{-1}\widetilde{R}(X'X)^{-1/2}) \\
= \; & \text{rank}(\widetilde{R}) \\
= \; & (N-1)k,
\end{aligned}
$$

we have

$$
\frac{\widetilde{\sigma}^2\left\{Nk - \text{tr}(P_{\text{pool}})\right\}}{\|\widehat{\beta} - \widehat{\beta}_{\text{pool}}\|_A^2} = \frac{\widetilde{\sigma}^2(N-1)k}{\widehat{\beta}'\widetilde{R}'(\widetilde{R}(X'X)^{-1}\widetilde{R}')^{-1}\widetilde{R}\widehat{\beta}} = \frac{1}{F}. \qquad \text{(S.10)}
$$

So when $1/F \in [0,1]$, we can obtain (S.8). □

S.3 Additional Monte Carlo Simulation

This section provides various extensions of simulation, including extensions in the data generation process and methods.

Comparison with more average-effect estimators

In this section, we compare our pooling averaging forecast with various average-effect-based forecasts. In particular, we consider Swamy's FGLS estimator, mean-group estimator,

and optimal average-effect estimator that chooses the weights by minimizing the risk. One of the main differences of our estimator from the average-effect estimator is that the latter averages over individual estimator $\hat{\beta}_i$ across *individual units*, whereas our estimator averages restricted estimators $\hat{\beta}_{(m)}$ over different *pooling strategies*. Thus the model space of our estimator is much larger and the resulting estimator is of distinct dimension. In particular, Swamy's estimator is essentially a generalized least square estimator (GLS), and its feasible version can be obtained by

$$\widehat{\bar{\beta}}_{\text{FGLS}} = \left(\sum_{i=1}^{N} X_i' \widehat{\Psi}_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' \widehat{\Psi}_i^{-1} y_i \right),$$

where $\widehat{\Psi}_i = X_i \widehat{\Delta} X_i' + \hat{\sigma}_i^2 I_T$, $\hat{\sigma}_i^2$ is the estimated variance of residuals for individual i, and

$$\widehat{\Delta} = \frac{1}{N-1} \sum_{i=1}^{N} \left(\widehat{\beta}_i - \frac{1}{N} \sum_{i=1}^{N} \widehat{\beta}_i \right) \left(\widehat{\beta}_i - \frac{1}{N} \sum_{i=1}^{N} \widehat{\beta}_i \right)'.$$

Equivalently, it can also be written as

$$\widehat{\bar{\beta}}_{\text{FGLS}} = \sum_{i=1}^{N} W_i \widehat{\beta}_i, \tag{S.11}$$

where $W_i = \left[\sum_{j=1}^{N} \{ \widehat{\Delta} + \hat{\sigma}_j^2 (X_j' X_j)^{-1} \}^{-1} \right] \left[\widehat{\Delta} + \hat{\sigma}_i^2 (X_i' X_i)^{-1} \right]^{-1}$. Another closely related estimator is the mean group estimator (Pesaran and Smith, 1995) that uses equal weights in (S.11), namely

$$\widehat{\bar{\beta}}_{\text{MG}} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\beta}_i.$$

Finally, we consider choosing the weights in the mean-group estimator optimally by minimizing the risk (rather than using equal weights $1/N$). In particular, we consider $\widehat{\bar{\beta}}_{\text{OPT}} = \sum_{i=1}^{N} \omega_i^* \widehat{\beta}_i$, where the weights ω_i^* for $i = 1, \ldots, N$ are chosen to minimize the risk. To this end, we first need to derive the MSE of $\hat{\beta}_{\text{AP}}$. We can show that

$$\begin{aligned}
\text{MSE}(\widehat{\bar{\beta}}_{\text{OPT}}) &= \sum_{i=1}^{N} \text{E} || \sum_{j=1}^{N} \omega_j^* \widehat{\beta}_j - \beta_i ||^2 \\
&= \sum_{i=1}^{N} \text{E} || \sum_{j=1}^{N} \omega_j^* \widehat{\beta}_j - \widehat{\beta}_i + (X_i' X_i)^{-1} X_i' u_i ||^2 \\
&= \sum_{i=1}^{N} \left[\text{E} || \sum_{j=1}^{N} \omega_j^* \widehat{\beta}_j - \widehat{\beta}_i ||^2 - \sigma_i^2 (1 - 2\omega_i^*) (X_i' X_i)^{-1} \right].
\end{aligned}$$

The risk of $\widehat{\bar{\beta}}_{\mathrm{OPT}}$ under the squared error loss can be written as the trace of $\mathrm{MSE}(\widehat{\bar{\beta}}_{\mathrm{OPT}})$, and it can be obtained by

$$R(\widehat{\bar{\beta}}_{\mathrm{OPT}}) = \sum_{i=1}^{N} \left[\| \sum_{j=1}^{N} w_j \widehat{\beta}_j - \widehat{\beta}_i \|^2 - \widehat{\sigma}_i^2 (1 - 2w_i) \mathrm{tr}(Z_i' Z_i)^{-1} \right].$$

If we denote $\omega^* := (\omega_1^*, \ldots, \omega_N^*)'$ and $\widehat{D}_i := (\widehat{\beta}_1 - \widehat{\beta}_i, \ldots, \widehat{\beta}_N - \widehat{\beta}_i)$ for $i = 1, 2, \ldots, N$, then the optimal weights that minimize the risk can be obtained by minimizing the following quadratic function

$$\omega^* = \arg\min_{\omega^*} \omega^{*'} \sum_{i=1}^{N} D_i' D_i \omega^* + 2v'w - \iota' v,$$

where $v := (v_1, \ldots, v_N)'$ with $v_i := \widehat{\sigma}_i^2 \mathrm{tr}(Z_i' Z_i)^{-1}$.

The results are given in Table S.1 and S.2. We find that our pooling averaging forecast outperforms the forecast based on any average-effect estimator in the heterogeneous panels, while the latter is more preferable in the homogeneous panels. Note that the MSFEs produced by FGLS, OPT, and MG are close to that of the pooled estimator, because they all estimate a common average-effect for all individual units.[1] Thus the similar comparison between the pooled and pooling averaging estimator also applies when we compare pooling averaging with Swamy's type and other average-effect estimator. Among the three average-effect estimator considered here, FGLS and OPT produce similar MSFE, marginally smaller than that of MG.

[1]Interestingly, the forecast using FGLS is not necessarily better than the pooled forecast, although FGLS is the best linear unbiased estimator in heterogeneous panels (the MSE of FGLS coefficient estimator is indeed smaller than that of the pooled estimator). Nevertheless, as shown in the response to Comment 5, forecasts using FGLS outperforms the pooled forecast when R^2 is decreased or when T is increased.

Table S.1: Risk comparison: Independent regressors with $R^2 = 0.9$

	DGP	MPA	C-Lasso	SAIC	SBIC	AIC	BIC	Pool	SHK	FGLS	OPT	MG
	1	0.130	0.454	0.487	0.144	0.591	0.160	**0.111**	0.742	0.112	0.125	0.126
$N = 10$	2	**0.341**	0.604	0.444	0.429	0.477	0.474	4.342	0.955	4.429	4.436	4.483
$T = 20$	3	**0.536**	0.829	0.724	0.786	0.801	0.830	2.640	0.934	2.695	2.697	3.099
	4	**0.918**	1.905	1.192	1.386	1.265	1.521	14.53	0.984	14.82	14.83	14.84
	1	0.175	0.314	0.364	0.265	0.408	0.301	**0.034**	0.701	0.037	0.041	0.051
$N = 30$	2	**0.158**	0.194	0.247	0.204	0.259	0.214	4.148	0.943	4.192	4.178	4.198
$T = 20$	3	**0.376**	0.655	0.509	0.536	0.524	0.567	2.317	0.911	2.353	2.337	2.417
	4	**0.396**	0.649	0.625	0.649	0.644	0.649	1.733	0.891	1.776	1.750	1.751
	1	0.131	0.466	0.488	0.127	0.579	0.133	**0.094**	0.876	0.098	0.103	0.102
$N = 10$	2	**0.268**	0.300	0.458	0.284	0.501	0.297	8.852	0.988	8.932	8.938	8.986
$T = 40$	3	**0.657**	1.264	0.897	1.072	1.000	1.174	5.153	0.981	5.209	5.208	5.302
	4	1.264	2.679	1.757	1.884	1.789	2.062	29.94	**0.996**	30.23	30.24	30.27
	1	0.177	0.319	0.373	0.248	0.420	0.288	**0.033**	0.854	0.034	0.036	0.039
$N = 30$	2	0.105	**0.089**	0.273	0.164	0.282	0.181	8.257	0.985	8.290	8.284	8.334
$T = 40$	3	**0.490**	1.013	0.673	0.754	0.721	0.779	4.596	0.975	4.623	4.612	4.702
	4	**0.651**	1.070	0.882	1.006	0.923	1.024	3.503	0.969	3.538	3.515	3.588

Notes:

1. Forecasts constructed using: MPA: Mallows pooling averaging estimator; C-Lasso: C-Lasso estimator with the number of groups determined by BIC; SAIC/SBIC: pooling averaging estimator based on relative values of AIC/BIC; AIC/BIC: estimator selected based on minimum value information criterion; Pool: pooled estimator; SHK: shrinkage estimator; FGLS: Swamy's estimator; OPT: Optimal average-effect estimator; MG: Mean group estimator.
2. All numbers are divided by the risk of the individual time series forecast.

Table S.2: Risk comparison: Autoregressive regressors with $R^2 = 0.9$

	DGP	MPA	C-Lasso	SAIC	SBIC	AIC	BIC	Pool	SHK	FGLS	OPT	MG
	1	0.114	0.421	0.472	0.132	0.571	0.138	**0.096**	0.749	0.122	0.157	0.157
$N = 10$	2	**0.345**	0.659	0.426	0.437	0.462	0.474	4.304	0.955	4.534	4.561	4.570
$T = 20$	3	**0.511**	0.809	0.680	0.761	0.749	0.803	2.520	0.930	2.653	2.674	2.686
	4	**0.920**	1.818	1.177	1.354	1.240	1.518	14.62	0.984	15.37	15.39	15.48
	1	0.141	0.283	0.333	0.218	0.376	0.254	**0.034**	0.704	0.041	0.054	0.056
$N = 30$	2	**0.172**	0.229	0.231	0.214	0.240	0.219	4.131	0.942	4.213	4.205	4.235
$T = 20$	3	**0.368**	0.661	0.476	0.522	0.494	0.555	2.309	0.910	2.369	2.359	2.381
	4	**0.408**	0.630	0.601	0.628	0.616	0.629	1.791	0.891	1.854	1.838	1.861
	1	0.124	0.450	0.485	0.121	0.572	0.127	**0.098**	0.878	0.106	0.120	0.127
$N = 10$	2	**0.295**	0.367	0.457	0.298	0.504	0.307	8.734	0.988	8.935	8.948	8.995
$T = 40$	3	**0.645**	1.257	0.877	1.077	0.976	1.169	5.122	0.981	5.259	5.269	5.298
	4	1.209	2.616	1.672	1.803	1.721	1.954	29.25	**0.997**	29.99	29.99	30.01
	1	0.156	0.299	0.353	0.209	0.399	0.249	**0.034**	0.855	0.036	0.043	0.045
$N = 30$	2	0.104	**0.096**	0.256	0.155	0.266	0.172	8.417	0.985	8.500	8.493	8.544
$T = 40$	3	**0.496**	1.018	0.645	0.731	0.682	0.751	4.625	0.975	4.679	4.670	4.691
	4	**0.654**	1.062	0.873	0.987	0.902	1.010	3.516	0.968	3.570	3.553	3.571

Notes:

1. Forecasts constructed using: MPA: Mallows pooling averaging estimator; C-Lasso: C-Lasso estimator with the number of groups determined by BIC; SAIC/SBIC: pooling averaging estimator based on relative values of AIC/BIC; AIC/BIC: estimator selected based on minimum value information criterion; Pool: pooled estimator; SHK: shrinkage estimator; FGLS: Swamy's estimator. OPT: Optimal average-effect estimator; MG: Mean group estimator.
2. All numbers are divided by the risk of the individual time series forecast.

Alternative pre-screening methods

To examine whether the performance of competing methods is sensitive to the choice of pre-screening method, here we consider a different type of clustering method from C-Lasso. One interesting alternative is the mixture-like iterative (M-Estimation) method proposed by Liu et al. (2018). Like C-Lasso, the M-estimation method also provides a consistent estimate of group membership if there is a group pattern of slope heterogeneity and the number of groups is not under-specified. This method has both pros and cons. A potential advantage of M-estimation over C-Lasso is that it does not involve a tuning parameter. Nevertheless, its optimization is an NP-hard problem. The iterative algorithm relies on the initial values, and the global optimum is not theoretically guaranteed. With sufficiently large number of trials of initial values, we expect that M-Estimation and C-Lasso produce similar forecasts. To use M-estimation for shrinking the model space, we employ Liu et al.'s (2018) procedure to estimate the latent group structure given the number of groups ranging from 1 to $G_{\max}$, and then average over forecasts obtained from different choices of this number. Thus the M-estimation method is used to shrink the model space, as the C-Lasso is used in the paper. Based on the pre-screened model space, all the methods (including MPA, IC-based methods, BPA, etc.) are implemented and compared. The results are summarized in Table S.3 and S.4. We find that using M-estimation and C-Lasso as a pre-screen method of model space leads to highly robust results. MPA remains the best method in most cases of heterogeneous panels, while the pooled forecast is the best choice in homogeneous panels. We also compare our pooling averaging forecast with the forecast produced by M-estimation method using the optimal number of groups (M-Est) proposed by Liu et al. (2018). The comparison results are also summarized in Table S.3 and S.4. Compared with the M-Est, MPA again demonstrates its advantages since it directly aims at minimizing the MSFE and trades off the efficiency and bias in a sensible way.

Another alternative pre-screening procedure can be based on agglomerative hierarchical clustering. The procedure starts with normalizing the estimated coefficients

$$\widehat{\beta}_{il} = \widehat{\beta}_{il} / \max\{|\widehat{\beta}_{1l}|, \ldots, |\widehat{\beta}_{Nl}|\}$$

for each $l = 1, \ldots, k$, so that the coefficients of the regressors have the same scale between $[-1, 1]$. Normalization avoids the numerical problems caused by extremely large numbers, and also allows us to group coefficients of different regressors using the same criteria. In

Table S.3: Risk comparison based on M-estimation: Independent regressors with $R^2 = 0.9$

	DGP	MPA	BPA	M-Est	SAIC	SBIC	AIC	BIC	Pool	SHK
	1	0.130	0.366	0.633	0.484	0.143	0.583	0.163	**0.101**	0.742
$N = 10$	2	**0.527**	0.584	0.917	0.716	0.689	0.769	0.797	4.342	0.955
$T = 20$	3	**0.716**	0.800	1.042	0.938	0.935	1.039	1.018	2.640	0.934
	4	1.062	1.356	2.269	1.362	1.493	1.447	1.668	14.53	**0.984**
	1	0.223	0.246	0.479	0.576	0.307	0.640	0.354	**0.034**	0.701
$N = 30$	2	**0.330**	0.411	0.698	0.582	0.501	0.597	0.533	4.148	0.943
$T = 20$	3	**0.540**	0.717	0.807	0.827	0.759	0.862	0.824	2.317	0.911
	4	**0.547**	0.724	0.803	0.820	0.747	0.847	0.810	1.733	0.891
	1	0.130	0.370	0.649	0.488	0.125	0.585	0.131	**0.094**	0.876
$N = 10$	2	**0.392**	0.483	0.848	0.645	0.444	0.701	0.469	8.852	0.988
$T = 40$	3	**0.798**	0.861	1.274	1.044	1.188	1.141	1.308	5.153	0.981
	4	1.436	2.061	4.092	1.944	2.025	1.998	2.192	29.94	**0.996**
	1	0.220	0.249	0.487	0.570	0.257	0.645	0.308	**0.033**	0.854
$N = 30$	2	**0.230**	0.288	0.769	0.506	0.374	0.525	0.401	8.257	0.985
$T = 40$	3	**0.598**	0.766	1.013	0.890	0.834	0.910	0.911	4.596	0.975
	4	**0.583**	0.775	0.935	0.858	0.818	0.884	0.885	3.503	0.969

Notes:

1. Forecasts constructed using: MPA: Mallows pooling averaging estimator; BPA: Bayesian model averaging; M-Est: M-estimation with the optimal number of groups; SAIC/SBIC: pooling averaging estimator based on relative values of AIC/BIC; AIC/BIC: estimator selected based on minimum value information criterion; Pool: pooled estimator; SHK: shrinkage estimator.
2. All numbers are divided by the risk of the individual time series forecast.

Table S.4: Risk comparison based on M-estimation: Autoregressive regressors with $R^2 = 0.9$

	DGP	MPA	BPA	M-Est	SAIC	SBIC	AIC	BIC	Pool	SHK
	1	0.114	0.353	0.621	0.478	0.130	0.569	0.142	**0.096**	0.745
$N = 10$	2	**0.547**	0.604	0.960	0.729	0.726	0.782	0.811	4.384	0.955
$T = 20$	3	**0.695**	0.789	1.002	0.910	0.901	1.021	0.972	2.598	0.929
	4	1.043	1.348	2.259	1.330	1.484	1.411	1.646	14.67	**0.984**
	1	0.189	0.238	0.458	0.547	0.281	0.613	0.323	**0.032**	0.701
$N = 30$	2	**0.352**	0.435	0.756	0.584	0.498	0.599	0.521	4.171	0.942
$T = 20$	3	**0.553**	0.698	0.804	0.816	0.738	0.851	0.802	2.326	0.909
	4	**0.572**	0.729	0.805	0.813	0.734	0.845	0.790	1.728	0.891
	1	0.121	0.364	0.628	0.470	0.113	0.564	0.113	**0.101**	0.875
$N = 10$	2	**0.433**	0.490	0.921	0.658	0.479	0.715	0.516	8.548	0.988
$T = 40$	3	**0.784**	0.851	1.233	1.037	1.186	1.150	1.305	5.185	0.980
	4	1.384	2.011	3.812	1.845	1.988	1.919	2.147	29.30	**0.996**
	1	0.205	0.244	0.473	0.559	0.246	0.628	0.292	**0.033**	0.855
$N = 30$	2	**0.235**	0.298	0.745	0.511	0.376	0.526	0.406	8.215	0.985
$T = 40$	3	**0.616**	0.782	1.002	0.907	0.854	0.938	0.921	4.535	0.975
	4	**0.597**	0.779	0.945	0.864	0.830	0.892	0.892	3.532	0.968

Notes:

1. Forecasts constructed using: MPA: Mallows pooling averaging estimator; BPA: Bayesian model averaging; M-Est: M-estimation with the optimal number of groups; SAIC/SBIC: pooling averaging estimator based on relative values of AIC/BIC; AIC/BIC: estimator selected based on minimum value information criterion; Pool: pooled estimator; SHK: shrinkage estimator.
2. All numbers are divided by the risk of the individual time series forecast.

the second step, we group the normalized coefficient estimates based on their differences. To incorporate estimation uncertainty in the coefficient estimates, we employ the Bhattacharyya distance. If we assume that the individual estimators are normally distributed, then the Bhattacharyya distance between two coefficient estimates can be obtained by

$$DB_{ij,l} = \frac{1}{4}\frac{(\widehat{\beta}_{i,l} - \widehat{\beta}_{j,l})^2}{\widehat{\sigma}_{i,l}^2 + \widehat{\sigma}_{j,l}^2} + \frac{1}{2}\ln\left(\frac{\widehat{\sigma}_{i,l}^2 + \widehat{\sigma}_{j,l}^2}{2\widehat{\sigma}_{i,l}\widehat{\sigma}_{j,l}}\right), \tag{S.12}$$

where $\widehat{\sigma}_{i,l}^2$ is the estimated variance of $\widehat{\beta}_{i,l}$, and $\widehat{\sigma}_l^2 = (\widehat{\sigma}_{i,l}^2 + \widehat{\sigma}_{j,l}^2)/2$. In the third step, we employ agglomerative hierarchical clustering (AHC). In the AHC procedure, each estimate starts in its own cluster, and at each step pairs of clusters are merged until a hierarchical tree is formed. As the last step, one can decide where to cut the hierarchical cluster tree to produce the clustering. We cut the tree by specifying the number of clusters G, and the algorithm automatically gives a unique clustering. By varying G from 1 to $G_{\max}$, we numerate all "reasonable" clusterings. A significant advantage of using AHC for clustering is its low computational cost. This algorithm leads to slightly different clustering results from C-Lasso, but the main results are qualitatively unchanged.

Weaker degree of heterogeneity

The degree of heterogeneity depends jointly on how far apart the coefficient values are and also how many groups there exist. We have examined how the performance of methods depends on the number of groups. Here we investigate the role of heterogeneity from a different perspective and consider the cases where the parameters are closer across groups while the number of groups remains the same. We fix the number of groups as in DGP 3 (strongly heterogeneity), but vary the size of coefficients, so that the discrepancy between individuals is smaller. In particular, we set the slope coefficients as

$$\beta_{i1}, \beta_{i2} = \begin{cases} b_1, & i = 1, \ldots, [N/4] \\ b_2, & i = [N/4]+1, \ldots, [N/2], \\ b_3, & i = [N/2]+1, \ldots, [3N/4], \\ b_4, & i = [3N/4]+1, \ldots, N, \end{cases} \quad \beta_{i3} = \begin{cases} b_1, & i = 1, \ldots, [N/5] \\ b_2, & i = [N/5]+1, \ldots, [2N/5], \\ b_3, & i = [2N/5]+1, \ldots, [3N/5], \\ b_4, & i = [3N/5]+1, \ldots, N, \end{cases}$$

The difference between the $b_1, \ldots, b_4$ determines the degree of heterogeneity. We consider two cases of $b = (b_1, b_2, b_3, b_4)$

$$b = (3,\ 3.25,\ 3.5,\ 3.75), \quad \text{and} \quad b = (3,\ 3.1,\ 3.2,\ 3.3).$$

Table S.5: Risk comparison: DGP 3 with smaller difference in slope coefficients

N	T	MPA	BPA	C-Lasso	SAIC	SBIC	AIC	BIC	Pool	SHK
					$b = (3,\ 3.25,\ 3.5,\ 3.75)$					
10	20	0.228	0.425	0.677	0.567	0.257	0.664	0.284	**0.213**	0.769
30	20	0.278	0.317	0.517	0.608	0.392	0.670	0.434	**0.139**	0.730
10	40	**0.309**	0.445	0.717	0.613	0.350	0.714	0.390	0.317	0.900
30	40	0.317	0.342	0.547	0.631	0.413	0.694	0.449	**0.248**	0.880
10	80	**0.435**	0.479	0.743	0.659	0.520	0.754	0.584	0.550	0.959
30	80	**0.370**	0.371	0.576	0.651	0.451	0.713	0.477	0.458	0.951
					$b = (3,\ 3.1,\ 3.2,\ 3.3)$					
10	20	0.151	0.388	0.651	0.512	0.166	0.615	0.182	**0.126**	0.748
30	20	0.238	0.265	0.489	0.581	0.329	0.647	0.380	**0.057**	0.708
10	40	0.175	0.393	0.662	0.518	0.176	0.621	0.190	**0.144**	0.883
30	40	0.248	0.276	0.503	0.594	0.299	0.660	0.350	**0.082**	0.860
10	80	0.215	0.399	0.678	0.534	0.209	0.634	0.218	**0.197**	0.945
30	80	0.281	0.294	0.525	0.601	0.322	0.672	0.366	**0.128**	0.935

The results are displayed in Table S.5. As expected, a smaller difference between the slope coefficients leads to a weaker degree of heterogeneity, and further favours the pooled estimator, even though the number of groups is unchanged.

S.4 Additional results of empirical applications

This section presents the estimated effect of determinants of sovereign CDS spreads using pooled and individual estimation. The first column of Table S.4 presents the pooled estimates, and the individual estimates are provided in the remaining columns.

Table S.6: Effects of CDS spreads determinants: Pooled and individual estimation

	Pooled estimation	Individual estimation					
		Brazil	Bulgaria	Chile	China	Hungary	Japan
lstock	−0.0651	−0.2389	−0.1244	0.0149	−0.0605	0.0190	−0.1453
	(0.0293)	(0.0416)	(0.0237)	(0.0333)	(0.0278)	(0.0385)	(0.0272)
fxrates	0.0172	−0.0548	0.0254	0.1541	−0.0324	0.0721	−0.0330
	(0.0266)	(0.0355)	(0.0292)	(0.0297)	(0.0232)	(0.0338)	(0.0220)
fxres	0.0174	−0.0990	0.0022	0.1393	0.0596	−0.0526	0.1327
	(0.0225)	(0.0215)	(0.0267)	(0.0207)	(0.0260)	(0.0234)	(0.0242)
gstock	0.1007	0.1360	−0.0094	0.0604	0.1350	0.1000	0.2377
	(0.0420)	(0.0441)	(0.0385)	(0.0428)	(0.0436)	(0.0399)	(0.0430)
trsy	−0.0836	−0.0504	−0.0773	−0.1699	−0.1467	−0.1188	0.0797
	(0.0307)	(0.0301)	(0.0300)	(0.0298)	(0.0309)	(0.0302)	(0.0340)
hy	−0.0769	−0.0792	−0.0562	−0.1875	−0.0644	−0.0187	0.0676
	(0.0292)	(0.0290)	(0.0282)	(0.0285)	(0.0296)	(0.0280)	(0.0289)
eqp	0.1662	0.0403	0.1713	0.0969	0.0926	0.3316	0.3166
	(0.0356)	(0.0354)	(0.0351)	(0.0354)	(0.0360)	(0.0350)	(0.0347)
volp	−0.1413	−0.0211	−0.2639	−0.1677	−0.0988	−0.0991	−0.2836
	(0.0221)	(0.0217)	(0.0213)	(0.0216)	(0.0226)	(0.0212)	(0.0214)
ef	−0.1131	−0.0373	−0.0564	−0.1157	−0.1215	−0.1381	−0.0974
	(0.0242)	(0.0241)	(0.0243)	(0.0234)	(0.0261)	(0.0234)	(0.0240)
bf	−0.0698	0.0107	−0.0746	−0.0750	−0.1325	−0.0471	0.0339
	(0.0225)	(0.0228)	(0.0218)	(0.0223)	(0.0228)	(0.0221)	(0.0222)

Note:

1. Standard deviations are given in the parentheses.
2. The individual estimates differ from those of Longstaff et al. (2011) because we use an updated sample with a longer time span, and include the lagged values of determinants as the explanatory variables.

Table S.4 (con't): Effects of CDS spreads determinants: Pooled and individual estimation

	Individual estimation							
	Korea	Malaysia	Philippines	Poland	Romania	Slovak	S.Afrika	Thailand
lstock	0.0326	−0.1231	0.1919	−0.0280	−0.0788	0.0395	−0.1763	−0.2832
	(0.0374)	(0.0339)	(0.0290)	(0.0396)	(0.0337)	(0.0245)	(0.0318)	(0.0330)
fxrates	0.0269	0.0046	0.1405	−0.0703	0.1099	0.0541	−0.0793	−0.1795
	(0.0339)	(0.0310)	(0.0277)	(0.0372)	(0.0323)	(0.0279)	(0.0272)	(0.0307)
fxres	0.1051	0.0434	0.1792	−0.0691	−0.0182	−0.1015	−0.0125	0.0488
	(0.0289)	(0.0237)	(0.0218)	(0.0253)	(0.0247)	(0.0218)	(0.0223)	(0.0245)
gstock	−0.0025	0.1967	−0.0129	0.0460	−0.0528	0.1218	0.1783	0.2598
	(0.0449)	(0.0428)	(0.0431)	(0.0426)	(0.0397)	(0.0406)	(0.0445)	(0.0428)
trsy	−0.1691	−0.0433	−0.0151	−0.0995	−0.0946	−0.2361	−0.0331	0.0158
	(0.0308)	(0.0302)	(0.0307)	(0.0303)	(0.0299)	(0.0314)	(0.0303)	(0.0300)
hy	−0.1977	−0.0835	−0.0718	−0.0321	−0.0188	−0.1178	−0.1132	−0.0902
	(0.0292)	(0.0291)	(0.0295)	(0.0289)	(0.0288)	(0.0292)	(0.0288)	(0.0283)
eqp	0.1837	0.2692	0.0530	0.2601	−0.0136	0.0348	0.1249	0.3055
	(0.0356)	(0.0360)	(0.0361)	(0.0360)	(0.0366)	(0.0357)	(0.0349)	(0.0349)
volp	−0.1482	−0.1266	−0.1142	−0.1746	−0.1581	−0.0610	−0.1138	−0.1652
	(0.0224)	(0.0219)	(0.0222)	(0.0219)	(0.0217)	(0.0224)	(0.0218)	(0.0216)
ef	−0.1215	−0.1415	−0.0834	−0.0791	−0.0743	−0.1188	−0.2237	−0.1437
	(0.0253)	(0.0249)	(0.0238)	(0.0243)	(0.0240)	(0.0240)	(0.0242)	(0.0232)
bf	−0.0569	−0.0789	−0.0418	−0.0723	−0.0716	−0.1666	−0.1455	−0.1229
	(0.0228)	(0.0231)	(0.0230)	(0.0231)	(0.0222)	(0.0234)	(0.0233)	(0.0224)

Note:

1. Standard deviations are given in the parentheses.
2. The individual estimates differ from those of Longstaff et al. (2011) because we use an updated sample with a longer time span, and include the lagged values of determinants as the explanatory variables.

Reference

Hansen BE. 2014. Model averaging, asymptotic risk, and regressor groups. *Quantitative Economics* 5: 495–530.

Maddala GS, Trost RP, Li H, Joutz F. 1997. Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimator. *Journal of Business & Economic Statistics* 15: 90–100.