

Article

A Combination Method for Averaging OLS and GLS Estimators [†]

Qingfeng Liu ^{1,*} and Andrey L. Vasnev ² 

¹ Department of Economics, Otaru University of Commerce, Otaru 047-8501, Japan

² Discipline of Business Analytics, The University of Sydney Business School, The University of Sydney, Sydney, NSW 2006, Australia; andrey.vasnev@sydney.edu.au

* Correspondence: qliu@res.otaru-uc.ac.jp

[†] The authors thank Okui Ryo, Mototsugu Shintani, Arihiro Yoshimura and the conference participants at the 2014 European Meeting of the Econometric Society and the 2013 Kansai Keiryō Keizaigaku Kenkyukai for their helpful comments. Liu acknowledges financial support from the JSPS Grant-in-Aid for Young Scientists (B) No. 25780148 and JSPS KAKENHI Grant (C) No. JP16K03590 and No. JP19K01582.

The authors are grateful to the Editor and two reviewers for their constructive comments.

Received: 6 June 2019; Accepted: 4 September 2019; Published: 9 September 2019



Abstract: To avoid the risk of misspecification between homoscedastic and heteroscedastic models, we propose a combination method based on ordinary least-squares (OLS) and generalized least-squares (GLS) model-averaging estimators. To select optimal weights for the combination, we suggest two information criteria and propose feasible versions that work even when the variance-covariance matrix is unknown. The optimality of the method is proven under some regularity conditions. The results of a Monte Carlo simulation demonstrate that the method is adaptive in the sense that it achieves almost the same estimation accuracy as if the homoscedasticity or heteroscedasticity of the error term were known.

Keywords: model averaging; OLS; GLS; combination method

1. Introduction

Model averaging has been developed as an alternative to model selection. In many situations, model-averaging methods perform better than alternative model-selection methods. The main reason for this is that model selection delivers a pretest estimator that has inferior properties, and its use can be harmful (see [Danilov and Magnus, 2004](#)). [Yuan and Yang \(2005\)](#) provided a detailed discussion on the choice between model averaging and model selection. As one of the pioneers of frequentist model averaging, [Hansen \(2007\)](#) proposed Mallows model averaging (MMA) based on the ordinary least-squares (OLS) estimator for linear regression models with homoscedastic errors. [Wan et al. \(2010\)](#) extended the results for non-nested models with homoscedastic errors. [Zhao et al. \(2018\)](#) is the most recent work in this area. For linear regression models with heteroscedastic errors, [Hansen and Racine \(2012\)](#), [Liu and Okui \(2013\)](#) and [Zhang et al. \(2013, 2015\)](#) proposed model averaging methods that are still based on the OLS estimator, while [Liu et al. \(2016\)](#) proposed a method based on the generalized least squares estimator (GLS). They demonstrated that their methods are optimal in the sense of [Li \(1987\)](#) for homoscedastic or heteroscedastic models. For model averaging in big datasets, [Xie \(2017\)](#) proposed the use of model screening (before averaging) in order to deal with the large number of candidate models/regressors.

However, all previous papers assumed that it was known whether the errors of the true data-generating process are homoscedastic or heteroscedastic. Due to this assumption, the previous averaging methods were based only on estimators constructed using the same estimation method,

either OLS or GLS estimators (with different regressor sets). This assumption can be unrealistic in empirical applications. Usually, researchers do not know the structure of the error term; therefore, this assumption leads to possible misspecification. A natural solution is to combine OLS and GLS estimators. Combinations of different methods are routinely used in the applied forecast combination literature. In a recent forecasting competition that included 100,000 series, Makridakis et al. (2018) found that, out of the 17 most accurate methods, 12 were combinations. All combinations used different models/methods varying from simple exponential smoothing models to sophisticated machine-learning algorithms.

We propose a combination method based on OLS and GLS estimators to reduce the risk of misspecification between homoscedastic and heteroscedastic linear models. More precisely, the proposed estimator is a weighted average of mixtures of OLS and GLS estimators. The OLS mixture is constructed using the MMA of Hansen (2007) or the heteroscedasticity-robust Cp(HRCp) model averaging of Liu and Okui (2013). The GLS mixture is constructed using the GLS model averaging (GLSMA) of Liu et al. (2016).

We propose the use of two criteria, MMA-GLSMA and HRCp-GLSMA, to choose the weight vector for combining estimators. The optimality of the chosen weight vector in the sense of Li (1987) is investigated. Our method works in situations with an unknown variance-covariance matrix of the error term if an estimate based on the nonparametric method k -nearest neighbours (k -NN) is used. The results of the simulation experiments show that our combination method is adaptive in the sense that it can achieve almost the same estimation accuracy as if the homoscedasticity or heteroscedasticity of the error term was known.

The rest of the paper is organized as follows: In Section 2, we describe the theoretical setup and introduce the new combination method with the criteria for choosing the weight vector. In Section 3, we investigate the optimality of the proposed criteria. Section 4 presents the results of the Monte Carlo simulations. Section 5 concludes the paper, and all proofs are provided in the Appendix A.

2. Method

Suppose that we have an independent random sample of (y_i, x_i) for $i = 1, \dots, n$, where $x_i = (x_{i1}, x_{i2}, \dots)'$ is a countably infinite real-valued vector and y_i is a real-valued scalar random variable generated from an infinite dimensional linear regression model:

$$y_i = \mu_i + e_i,$$

where:

$$\mu_i = \sum_{j=1}^{\infty} \theta_j x_{ij},$$

e_i is an unobserved error term that can be homoscedastic or heteroscedastic with $E(e_i|x_i) = 0$, $E(e_i^2) = \sigma^2$ and $E(e_i^2|x_i) = \sigma_i^2$ and θ_j for $j = 1, 2, \dots$ are unknown parameters. We also define $X \equiv (x_1, \dots, x_n)$, $Y \equiv (y_1, \dots, y_n)'$, $\mu \equiv (\mu_1, \dots, \mu_n)'$, $e \equiv (e_1, \dots, e_n)'$ and denote the variance-covariance matrix as $\Omega \equiv E(ee'|X) = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$. We state the theoretical results considering the distributions conditional on X and omit all notations for those conditional on X hereafter.

2.1. Infeasible Combination Estimator and Information Criteria

Suppose Ω is known; we have a candidate set of M_1 linear models with different numbers of independent variables for OLS estimation and a candidate set of M_2 linear models with different numbers of independent variables for GLS estimation. Then, we can obtain a set of OLS estimates $\hat{\beta}_m^{ols} \equiv P_m Y$ for $m = 1, \dots, M_1$ and a set of GLS estimates $\hat{\beta}_m^{gls} \equiv G_m Y$ for $m = 1, \dots, M_2$. Therein, $P_m = X_m(X_m'X_m)^{-1}X_m'$ is the projection matrix of the m^{th} regression model for OLS with $m = 1, \dots, M_1$,

and $G_m \equiv X_m(X_m'\Omega^{-1}X_m)^{-1}X_m'\Omega^{-1}$ with X_m being the independent variable matrix of the m^{th} regression model for GLS with $m = 1, \dots, M_2$. In this paper, we only consider the situation with nested models for both OLS and GLS estimators. This means that the m^{th} model is nested in the $(m + 1)^{\text{th}}$ model. Theoretical results may be extended to non-nested candidate models using the approach of [Wan et al. \(2010\)](#). Moreover, M_1 and M_2 can be fixed or go to infinity when the sample size n is increasing.

Based on those OLS and GLS estimates, we construct a combination estimator as follows:

$$\begin{aligned}\hat{\mu}(W) &= \sum_{m=1}^{M_1} w_m^{\text{ols}} \hat{\mu}_m^{\text{ols}} + \sum_{m=1}^{M_2} w_m^{\text{gls}} \hat{\mu}_m^{\text{gls}}, \\ &\equiv \hat{\mu}_{\text{ols}}(W_1) + \hat{\mu}_{\text{gls}}(W_2)\end{aligned}$$

where $W = (W_1', W_2') = (w_1^{\text{ols}}, \dots, w_{M_1}^{\text{ols}}, w_1^{\text{gls}}, \dots, w_{M_2}^{\text{gls}})'$ belongs to:

$$\mathcal{H} = \left\{ W \mid W \in [0, 1]^M, I'W = 1 \right\},$$

where $M = M_1 + M_2$ and I denote a $M \times 1$ vector having all elements equal to one.

In order to reduce the risk of the combination-estimation method proposed above, we need to select a suitable weight vector. To do that, in this subsection, we propose two versions, MMA-GLSMA and HRCp-GLSMA, with infeasible criteria. In the next subsection, we provide their feasible counterparts for the situation with unknown Ω values.

HRCp-GLSMA: The first information criterion for selecting a weight vector is defined as:

$$\bar{C}_n(W) = s_1^2 \left\{ \|Y - \hat{\mu}_{\text{ols}}(W_1^*)\|^2 + 2\text{tr}(P(W_1^*)\Omega) \right\} \quad (1)$$

$$+ s_2^2 \left\{ \|Y - \hat{\mu}_{\text{gls}}(W_2^*)\|^2 + 2\text{tr}(G(W_2^*)\Omega) \right\} \quad (2)$$

$$+ 2s_1s_2 \left\{ (Y - \hat{\mu}_{\text{ols}}(W_1^*))' (Y - \hat{\mu}_{\text{gls}}(W_2^*)) \right\} \quad (3)$$

$$+ \text{tr}(P(W_1^*)\Omega) + \text{tr}(G(W_2^*)\Omega) - e'e \quad (4)$$

where $s_1 \equiv \sum_{m=1}^{M_1} w_m^{\text{ols}}$, $s_2 \equiv \sum_{m=1}^{M_2} w_m^{\text{gls}}$, $W_1^* \equiv W_1/s_1$, $W_2^* \equiv W_2/s_2$, $P(W_1^*) \equiv \sum_{m=1}^{M_1} w_m^{\text{ols}} P_m$ and $G(W_2^*) \equiv \sum_{m=1}^{M_2} w_m^{\text{gls}} G_m$.

Note that:

$$\begin{aligned}\bar{C}_n(W) &= s_1^2 \text{HRC}_p(W_1^*) + s_2^2 C_{I_n}(W_2^*) \\ &+ 2s_1s_2 \left\{ (Y - \hat{\mu}_{\text{ols}}(W_1^*))' (Y - \hat{\mu}_{\text{gls}}(W_2^*)) \right\} \\ &+ \text{tr}(P(W_1^*)\Omega) + \text{tr}(G(W_2^*)\Omega) - e'e,\end{aligned}$$

where $\text{HRC}_p(W_1^*) = \|Y - \hat{\mu}_{\text{ols}}(W_1^*)\|^2 + 2\text{tr}(P(W_1^*)\Omega)$ is the HRCp model-averaging criterion proposed by [Liu and Okui \(2013\)](#) with the weight vector W_1^* and $C_{I_n}(W_2^*) = \|Y - \hat{\mu}_{\text{gls}}(W_2^*)\|^2 + 2\text{tr}(G(W_2^*)\Omega)$ is a GLSMA information criterion proposed by [Liu et al. \(2016\)](#) with the weight vector W_2^* . \bar{C} can be regarded as a combination of the HRCp and the GLSMA. Hence, we call \bar{C} the HRCp-GLSMA-type criterion.

MMA-GLSMA: Second, we propose an MMA-GLSMA-type criterion for weight selection. The infeasible MMA-GLSMA-type criterion is defined as follows:

$$\tilde{C}_n(W) = s_1^2 C_{\text{MMA}}(W_1^*) + s_2^2 C_{I_n}(W_2^*)$$

$$+2s_1s_2 \left\{ (Y - \hat{\mu}_{ols}(W_1^*))' (Y - \hat{\mu}_{gls}(W_2^*)) + \sigma^2 tr(P(W_1^*)) + tr(G(W_2^*)\Omega) - e'e \right\},$$

where $C_{MMA}(W_1^*) = \|Y - \hat{\mu}_{ols}(W_1^*)\|^2 + 2\sigma^2 tr(P(W_1^*))$ is the MMA criterion proposed by Hansen (2007) with the weight vector W_1^* .

Suppose the variance-covariance matrix Ω is known; we can then choose the weight by minimizing the criteria \bar{C}_n or \tilde{C}_n , as follows:

$$\bar{W} = \arg \min_{W \in H} \bar{C}_n$$

or

$$\tilde{W} = \arg \min_{W \in H} \tilde{C}_n.$$

However, since the variance-covariance matrix Ω is unknown, \bar{W} and \tilde{W} are infeasible.

2.2. Feasible Combination Estimator and Information Criteria

For a situation with unknown variance, a feasible combination estimator can be constructed using feasible GLS (FGLS) estimators. FGLS estimators and the feasible combination estimator are defined below:

$$\begin{aligned} \hat{\mu}^F(W) &= \sum_{m=1}^{M_1} w_m^{ols} \hat{\mu}_m^{ols} + \sum_{m=1}^{M_2} w_m^{gls} \hat{\mu}_m^{fgls}, \\ &\equiv \hat{\mu}_{ols}(W_1) + \hat{\mu}_{fgls}(W_2) \end{aligned}$$

where the FGLS estimator is $\hat{\mu}_m^{fgls} \equiv X_m(X_m'\hat{\Omega}^{-1}X_m)^{-1}X_m'\hat{\Omega}^{-1}Y$. Therein, the estimator $\hat{\Omega}$ is based on the k -NN estimator $\hat{\sigma}_i^2$ of Liu et al. (2016).

We propose two feasible counterparts of \bar{C}_n and \tilde{C}_n . The feasible HRCp-GLSMA-type criterion \bar{C}_n^F is defined as:

$$\begin{aligned} \bar{C}_n^F(W) &= s_1^2 \left\{ \|Y - \hat{\mu}_{ols}(W_1^*)\|^2 + 2\sum_{i=1}^n \hat{e}_i^2 p_{ii}(W_1^*) \right\} \\ &+ s_2^2 \left\{ \|Y - \hat{\mu}_{fgls}(W_2^*)\|^2 + 2tr(G^F(W_2^*)\hat{\Omega}) \right\} \\ &+ 2s_1s_2 \left\{ (Y - \hat{\mu}_{ols}(W_1^*))' (Y - \hat{\mu}_{fgls}(W_2^*)) \right. \\ &\left. + tr(P(W_1^*)\hat{\Omega}) + tr(G^F(W_2^*)\hat{\Omega}) - e'\hat{e} \right\}, \end{aligned} \tag{5}$$

and the feasible MMA-GLSMA-type criterion \tilde{C}_n^F is defined as:

$$\begin{aligned} \tilde{C}_n^F(W) &= s_1^2 \left\{ \|Y - \hat{\mu}_{ols}(W_1^*)\|^2 + 2\hat{\sigma}^2 tr(P(W_1^*)) \right\} \\ &+ s_2^2 \left\{ \|Y - \hat{\mu}_{fgls}(W_2^*)\|^2 + 2tr(G^F(W_2^*)\hat{\Omega}) \right\} \\ &+ 2s_1s_2 \left\{ (Y - \hat{\mu}_{ols}(W_1^*))' (Y - \hat{\mu}_{fgls}(W_2^*)) \right. \\ &\left. + \hat{\sigma}^2 tr(P(W_1^*)) + tr(G^F(W_2^*)\hat{\Omega}) - e'\hat{e} \right\}, \end{aligned} \tag{6}$$

where \hat{e}_i denotes the i^{th} element of $\hat{e} \equiv \sqrt{n/(n-k_L)}(I - P_L)Y$ with k_L denoting the number of independent variables in the largest OLS model:

$$G^F(W_2^*) = \sum_{m=1}^{M_2} w_m^{gls} X_m(X_m'\hat{\Omega}X_m)^{-1}X_m'\hat{\Omega}^{-1}.$$

P_L denotes the projection matrix of that model, and p_{ii} denotes the i^{th} diagonal element of $P(W_1^*)$. $\tilde{\varepsilon} \equiv (I - P_L)Y$, while $\hat{\sigma}^2 \equiv (n - k_L)^{-1} \tilde{\varepsilon}' \tilde{\varepsilon}$ is defined as one of the estimators suggested by Hansen (2007). $\hat{\Omega}$ is calculated by plugging in the k -NN estimator $\hat{\sigma}_{i^2}$ used in Liu et al. (2016).

3. Properties of the Criteria

The following lemma demonstrates the significant fact that the two infeasible criteria proposed above are unbiased estimates of the risk function $R(W) \equiv E(L(W))$ plus a constant; $L(W) \equiv \|\mu - \hat{\mu}(W)\|^2$ is the loss function.

Lemma 1. For any real-valued vector W , $E(\bar{C}_n(W)) = R(W) + c_1$ and $E(\tilde{C}_n(W)) = R(W) + c_2$, where c_1 and c_2 are constants.

Another useful property is that all criteria are asymptotically optimal in the sense of Li (1987). Proofs of the asymptotic optimality for all of the above-mentioned criteria can be performed by extending the proofs of Hansen (2007); Liu and Okui (2013); Liu et al. (2016). As an example, we demonstrate the optimality of the feasible MMA-GLSMA case method with the nonparametric estimator of Ω used in Liu et al. (2016).

To do that, we define the feasible loss function and the risk function as follows:

$$\begin{aligned} L^F(W) &\equiv \|\mu - \hat{\mu}^F(W)\|^2 \\ &= (\hat{\mu}^F(W) - \mu)'(\hat{\mu}^F(W) - \mu). \end{aligned}$$

We employ some notations and assumptions from Liu et al. (2016), reproduced in the Appendix, and add the following additional assumption:

Assumption 1'. As $n \rightarrow \infty$ and $k_l \rightarrow \infty$, $(k_l + k_l \sqrt{\sum_{j=1}^n b_{lj}^2}) / \zeta_n \rightarrow 0$, where $\zeta_n \equiv \inf_{W \in \mathcal{H}_n(N)} R(W)$, k_l is the number of regressors used in the regression model for the k -NN estimator and b_{li} denotes the approximation error of that model.

Assumption 1' guarantees that when the number of regressors used in the regression model (adjusted with the approximation errors of that model) increases with the sample size n , it increases at a rate slower than the lower bound of the risk across all possible weights. In practice, this assumption requires us to moderate the increase in the number of regressors k_l (relative to the sample size) to reduce the approximation errors.

Following Hansen (2007), we restrict the elements of the weighting vector to belong to set $\mathcal{H}_M(N) \equiv \{0, 1/N, 2/N, \dots, 1\}$ for some integer $N < \infty$.

Theorem 1. Under Assumption 1' and Assumptions 1–3, 6, and 10–14 of Liu et al. (2016), as $n \rightarrow \infty$, $k_l \rightarrow \infty$, $\kappa \rightarrow \infty$ and $k_L / \sqrt{n} \rightarrow 0$, we have:

$$\frac{L^F(\tilde{W})}{\inf_{W \in \mathcal{H}_M(N)} L(W)} \rightarrow_p 1,$$

where $\tilde{W} = \arg \min_{W \in \mathcal{H}_M(N)} \tilde{C}_n^F(W)$.

In other words, as the sample size increases, our method can achieve the infimum of the loss.

4. Simulation Study

To investigate the finite sample performance of the proposed MMA-GLSMA and HRCp-GLSMA versions and to compare them with other alternative methods, we performed a Monte Carlo simulation. Alternative methods include MMA, HRCp and GLSMA with C_{In}^F , where C_{In}^F is the feasible counterpart of C_{In} proposed in Liu et al. (2016).

The data-generating process (DGP) is:

$$y_i = \mu_i + e_i,$$

where:

$$\mu_i = \sum_{j=1}^{10,000} \theta_j x_{ij}$$

for $i = 1, \dots, n$, with $n = 50$. We used parameter $\theta_j = cj^{-1}$ with j truncated at 10,000 and a positive constant c . $x_{i1} = 1$ and $x_{ij} \sim N(0, 1)$ for $j \neq 1$ are independent with respect to i . We conducted three simulations. The first case was a simulation with a homoscedastic error term $e_i \sim N(0, 1)$. The second case was a mild heteroscedastic case with heteroscedastic error term $e_i \sim N(0, \sigma_i^2)$, with $\sigma_i = |x_{i2}|$. The third case was a strong heteroscedastic case with $\sigma_i = x_{i2}^2$. In all cases, e_i was independent with respect to i . The number of regressors in the largest approximation model or the number of nested models was $M = 10$. The m^{th} model contained the first m regressors, including the constant term. We varied c to change the R^2 of the DGP from 0.1–0.9 with an increment of 0.1.

We considered two cases of GLSMA with different estimation methods of σ_i , one based on the maximum likelihood estimation (MLE), and the other based on the nonparametric method k -NN. For details of these two cases, see Liu et al. (2016). Because the true specification of σ_i is usually unknown in practice, we misspecified σ_i for GLSMA. For the MLE-based method, we set $\sigma_i^2 = a + bx_{i3}^4$, where a and b are the unknown parameters to be estimated. For the nonparametric case, we only used x_{i3} and x_{i4} for k -NN estimation.

The number of replications for all simulations was 1000. We evaluated the performance of each method by the sample mean squared error (MSE) $= 1/1000 \sum_{k=1}^{1000} \|\hat{\mu}^{(k)} - \mu^{(k)}\|^2$, where $\hat{\mu}^{(k)}$ and $\mu^{(k)}$ are the realized vector of the estimated value $\hat{\mu}$ and true value μ in the k^{th} replication, respectively. The simulation results are shown in Tables 1–3. Figure 1 presents the same information relative to the MSE of the GLSMA method.

The results in Tables 1 and 2 show that our combination methods, MMA-GLSMA and HRCp-GLSMA, performed better than the alternatives (GLSMA, HRCp and MMA) when the error term was homoscedastic or had mild heteroscedasticity for $R^2 \leq 0.7$. When $R^2 \geq 0.8$, the performance of our methods was slightly worse than that of the alternative methods. For the homoscedastic case, the three alternative methods performed similarly. On the other hand, in the case of mild heteroscedasticity, GLSMA and HRCp performed better than MMA (which was expected, as MMA was designed for homoscedastic models).

Table 3 demonstrates that, when the heteroscedasticity of the error term was considerably strong, our combination method HRCp-GLSMA worked much better than the others when the MLE-based estimation of σ_i was used. However, MMA-GLSMA and HRCp-GLSMA became worse than GLSMA when σ_i was estimated using the nonparametric method.

Moreover, for most cases, GLSMA with the nonparametric estimator of σ_i outperformed GLSMA with the MLE-based estimator of σ_i . This can be explained by a characteristic of the k -NN method that we adopted. In the k -NN estimation, a large weight was placed on the i^{th} squared residual to estimate σ_i^2 ; therefore, even though σ_i^2 was misspecified, the estimate could catch the heteroscedasticity of the error term to some extent.

The aforementioned simulation results gave us the following indications for practical analysis. If we know that the heteroscedasticity of the data is considerably strong or the population R^2 is large, we should use the nonparametric GLSMA. Otherwise, it is preferable to choose an MMA-GLSMA or HRCp-GLSMA combination.

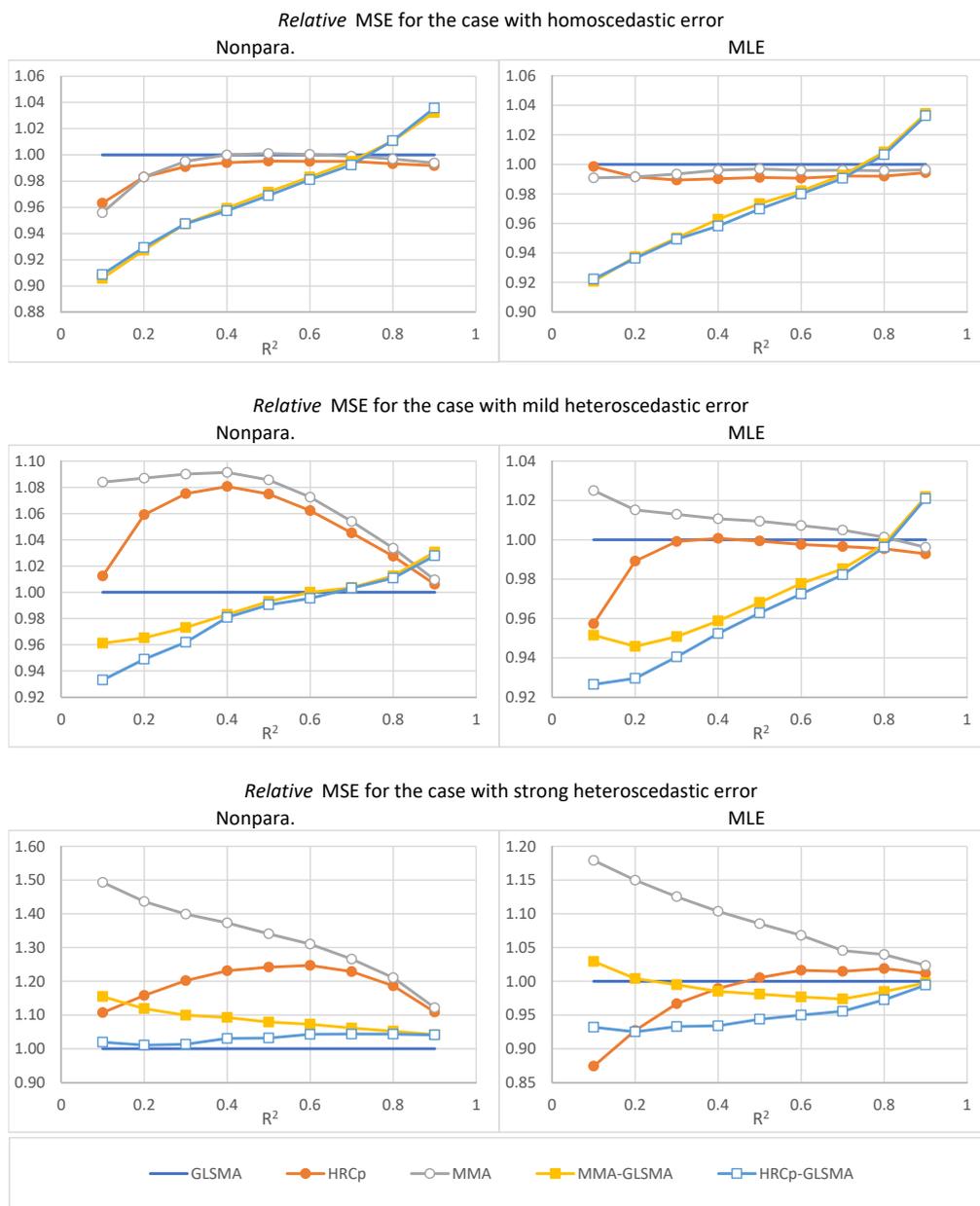


Figure 1. Performance of the methods relative to generalized least-squares model averaging (GLSMA): Heteroscedasticity-robust Cp (HRCp), Mallows model averaging (MMA), MMA-GLSMA and HRCp-GLSMA.

Table 1. Sample mean squared error (MSE) for the case with homoscedastic error.

	R ²	GLSMA	HRCp	MMA	MMA-GLSMA	HRCp-GLSMA
Nonpara.	0.1	6.80	6.55	6.50	6.16	6.18
	0.2	9.50	9.34	9.34	8.81	8.83
	0.3	12.20	12.09	12.14	11.56	11.56
	0.4	15.25	15.16	15.25	14.63	14.60
	0.5	19.06	18.97	19.08	18.52	18.47
	0.6	24.33	24.21	24.34	23.92	23.87
	0.7	32.64	32.48	32.61	32.48	32.39
	0.8	48.66	48.33	48.51	49.18	49.19
	0.9	95.66	94.88	95.07	98.76	99.08
MLE	0.1	6.56	6.55	6.50	6.04	6.05
	0.2	9.42	9.34	9.34	8.83	8.82
	0.3	12.22	12.09	12.14	11.61	11.60
	0.4	15.31	15.16	15.25	14.74	14.67
	0.5	19.14	18.97	19.08	18.63	18.56
	0.6	24.44	24.21	24.34	24.00	23.95
	0.7	32.74	32.48	32.61	32.50	32.43
	0.8	48.72	48.33	48.51	49.13	49.04
	0.9	95.42	94.88	95.07	98.70	98.56

Table 2. Sample MSE for the case with mild heteroscedastic error.

	R ²	GLSMA	HRCp	MMA	MMA-GLSMA	HRCp-GLSMA
Nonpara.	0.1	6.43	6.51	6.97	6.18	6.00
	0.2	8.62	9.13	9.37	8.32	8.18
	0.3	10.77	11.58	11.74	10.48	10.36
	0.4	13.02	14.07	14.21	12.80	12.77
	0.5	15.76	16.94	17.11	15.65	15.61
	0.6	19.45	20.66	20.86	19.45	19.36
	0.7	25.20	26.34	26.56	25.29	25.28
	0.8	36.05	37.04	37.26	36.50	36.44
	0.9	67.46	67.87	68.10	69.53	69.34
MLE	0.1	6.80	6.51	6.97	6.47	6.30
	0.2	9.23	9.13	9.37	8.73	8.58
	0.3	11.59	11.58	11.74	11.02	10.90
	0.4	14.06	14.07	14.21	13.48	13.39
	0.5	16.95	16.94	17.11	16.41	16.32
	0.6	20.71	20.66	20.86	20.25	20.14
	0.7	26.43	26.34	26.56	26.04	25.96
	0.8	37.21	37.04	37.26	37.13	37.07
	0.9	68.36	67.87	68.10	69.86	69.79

Table 4 and 5 give the averages of the estimated weights corresponding to the OLS and GLS parts for HRCp-GLSMA and MMA-GLSMA. Table 4 shows that for HRCp-GLSMA, when the heteroscedasticity became stronger, the average weights corresponding to the GLS part increased and the average weights corresponding to the OLS estimates decreased. Table 5 does not show a similar trend for MMA-GLSMA. This might be an explanation for why the performance of MMA-GLSMA was worse than that of HRCp-GLSMA.

5. Conclusions

In this paper, we proposed a combination method of OLS and GLS estimators. The proposed method reduced the risk of misspecification between homoscedastic and heteroscedastic models. The optimality of the criteria for choosing a weight vector was proven under some regularity conditions. We performed simulation experiments to investigate the finite sample property of our combination method. The results of the simulations demonstrate that our method was adaptive for homoscedasticity and heteroscedasticity. As mentioned previously, the proposed method was novel in that it combined estimators from two different estimation methods. Combining the different estimation methods can also combine the advantages of each. This idea could be useful and should be extended to the combination of other estimation methods.

Author Contributions: This is a collaborative project. The authors are listed in alphabetical order. The proofs are done by Q. Liu.

Funding: Q. Liu acknowledges financial support from the JSPS Grant-in-Aid for Young Scientists (B) No. 25780148 and JSPS KAKENHI Grant (C) No. JP16K03590 and No. JP19K01582.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

For the convenience of the readers of this journal, we list Assumptions 1–3, 6 and 10–14 and replicate Lemma 7 of Liu et al. (2016) here. Their notation of $R_{I_n}(W)$ coincides with our notation $R(W)$; x_{mi} denotes the i^{th} observation vector of the regressors of the m^{th} model, and $x_{m,j,i}$ is the j^{th} entry of x_{mi} .

Assumption 1. $E(|e_i|^{4(N+1)}) \leq c < \infty$ for some c .

Assumption 2. $\xi_n \equiv \inf_{W \in \mathcal{H}_n(N)} R_{I_n}(W) \rightarrow \infty$ as $n \rightarrow \infty$.

Assumption 3. $0 < \inf_i \sigma_i^2 \leq \sup_i \sigma_i^2 < \infty$.

Assumption 6. The maximum eigenvalue of $\sum_{i=1}^n x_{mi}x'_{mi}/n$ is bounded uniformly in n and m . There exists a $c > 0$ such that the minimum eigenvalue of $\sum_{i=1}^n x_{mi}x'_{mi}/n$ is greater than c for any n and m .

Assume that σ_i^2 is a function of a finite subset of x_i , denoted by z_i , so that $\sigma_i^2 = \sigma^2(z_i)$.

Assumption 10. $\sigma^2(\cdot)$ is differentiable. Let $\sigma'(\cdot)$ denote its first derivative. Then, $\sup_z \|\sigma'(z)\| < \infty$. Moreover, the support of z_i is bounded, and the density of z_i is bounded from below.

Let κ be the tuning parameter for the k -NN estimator. Let c_i , $1 \leq i \leq n$ be such that $c_i > 0$ for $1 \leq i \leq \kappa$, $c_i = 0$ for $i > \kappa$ and $\sum_{i=1}^{\kappa} c_i = 1$.

Assumption 11. $\lim_{n \rightarrow \infty} \max_{1 \leq i \leq \kappa} \kappa c_i < \infty$.

Assumption 12. $\lim_{n \rightarrow \infty} \sum_{i=1}^n x_{m,j,i}^4/n$ is bounded uniformly in m and j .

Assumption 13. There exists a $C < \infty$ such that $\lim_{n \rightarrow \infty} \sum_{i=1}^n \mu_i^4/n < C$.

Assumption 14. Define $v = 4(N + 1)$ and:

$$A_n = \frac{n^{2/v}}{\sqrt{\kappa}} + \frac{k_l}{\sqrt{\kappa}} + \frac{k_l \sqrt{\sum_{j=1}^n b_{lj}^2}}{\sqrt{\kappa}} + \frac{\kappa^{1/q}}{n^{1/q}}.$$

As $n \rightarrow \infty$, $k_l \rightarrow \infty$ and $\kappa \rightarrow \infty$, it is the case that $k_M A_n \rightarrow 0$, $k_M^2 A_n / \xi_n \rightarrow 0$, $nk_M A_n^2 / \xi_n \rightarrow 0$ and $\sqrt{n} k_M A_n / \xi_n \rightarrow 0$.

Lemma 7 of Liu et al. (2016). Suppose that Assumptions 1, 3, 5, 6 and 11 of Liu, Okui and Yoshimura (2016) hold. Then, as $n \rightarrow \infty$, $k_l \rightarrow \infty$ and $\kappa \rightarrow \infty$, with $k_l^2 / \kappa \rightarrow 0$ and $k_l^2 \sum_{j=1}^n b_{lj}^2 / \kappa \rightarrow 0$, we have:

$$\max_i |\hat{\sigma}_i^2 - \tilde{\sigma}_i^2| = O_p \left(\frac{k_l}{\sqrt{\kappa}} \left(1 + \sqrt{\sum_{j=1}^n b_{lj}^2} \right) \right).$$

Proof of Lemma 1. We have:

$$\begin{aligned}
 R(W) &= E \|\mu - \hat{\mu}(W)\|^2 \\
 &= E \left\| s_1\mu - s_1\hat{\mu}_{ols}(W_1^*) + s_2\mu - s_2\hat{\mu}_{gls}(W_2^*) \right\|^2 \\
 &= E \left[s_1^2 \|\mu - \hat{\mu}_{ols}(W_1^*)\|^2 \right] \tag{A1}
 \end{aligned}$$

$$+ E \left[2s_1s_2 (\mu - \hat{\mu}_{ols}(W_1^*))' (\mu - \hat{\mu}_{gls}(W_2^*)) \right] \tag{A2}$$

$$+ E \left[s_2^2 \|\mu - \hat{\mu}_{gls}(W_2^*)\|^2 \right]. \tag{A3}$$

According to Liu and Okui (2013); Liu et al. (2016), it is obvious that the squared terms (1) and (2) are unbiased estimates of the terms (A1) and (A3) plus a constant.

In order to estimate the cross-product term (A2), we can observe that:

$$\begin{aligned}
 (Y - \hat{\mu}_{ols}(W_1^*))' (Y - \hat{\mu}_{gls}(W_2^*)) &= (\mu - \hat{\mu}_{ols}(W_1^*))' (\mu - \hat{\mu}_{gls}(W_2^*)) + e'e \\
 &\quad + \langle e, (I - P(W_1^*))\mu \rangle + \langle e, (I - G(W_2^*))\mu \rangle \\
 &\quad - e'P(W_1^*)e - e'G(W_2^*)e
 \end{aligned}$$

and $E \langle e, (I - P(W_1^*))\mu \rangle = E \langle e, (I - G(W_2^*))\mu \rangle = 0$. Therefore, the sum of the terms (3) and (4) unbiasedly estimates the cross-product term (A2). □

Proof of Theorem 1. We define $L_{ols}(W) \equiv (\mu - \hat{\mu}_{ols}(W))' (\mu - \hat{\mu}_{ols}(W))$ and $L_{fgls}(W) \equiv (\mu - \hat{\mu}_{fgls}(W))' (\mu - \hat{\mu}_{fgls}(W))$. Note that:

$$\begin{aligned}
 \tilde{C}_n^F(W) - L^F(W) &= s_1^2 \left[C_{MMA}^F(W_1^*) - L_{ols}(W_1^*) \right] + s_2^2 \left[C_{In}^F(W_2^*) - L_{fgls}(W_2^*) \right] \\
 &\quad + 2s_1s_2 \left\{ \langle e, (I - P(W_1^*))\mu \rangle + \langle e, (I - G^F(W_2^*))\mu \rangle \right. \\
 &\quad + \hat{\sigma}^2 tr(P(W_1^*)) + tr(G^F(W_2^*)\hat{\Omega}) - e'P(W_1^*)e - e'G^F(W_2^*)e \\
 &\quad \left. + e'e - \hat{e}'\hat{e} \right\},
 \end{aligned}$$

where $C_{MMA}^F = C_{MMA}(W_1^*) = \|\mu - \hat{\mu}_{ols}(W_1^*)\|^2 + 2\hat{\sigma}^2 tr(P(W_1^*))$ and $C_{In}^F(W_2^*) = \|\mu - \hat{\mu}_{gls}(W_2^*)\|^2 + 2tr(G^F(W_2^*)\hat{\Omega})$.

Using the results of Hansen (2007); Liu et al. (2016), it can be shown that the supremum with respect to $W \in \mathcal{H}_M(N)$ of all the absolute values of the terms except $e'e - \hat{e}'\hat{e}$ divided by risk function $R(W)$ go to zero in probability. We only need to show:

$$\sup_W \left| \frac{e'e - \hat{e}'\hat{e}}{R(W)} \right| \rightarrow_p 0.$$

This can easily be done by modifying Lemma 7 of Liu et al. (2016). By replacing s_{ij} , the weight defined for k -NN in Liu et al. (2016) with one, we have:

$$\begin{aligned}
 \sup_W \left| \frac{e'e - \hat{e}'\hat{e}}{R(W)} \right| &\leq |e'e - \hat{e}'\hat{e}| / \xi_n + (k_L / (n - k_L)) \hat{e}'\hat{e} / \xi_n \\
 &= O_p \left(k_l + k_l \sqrt{\sum_{j=1}^n b_{lj}^2} \right) / \xi_n + o_p(1)
 \end{aligned}$$

→ 0.

The proof is complete. □

References

- Danilov, Dmitry, and Jan R. Magnus. 2004. On The Harm That Ignoring Pretesting Can Cause. *Journal of Econometrics* 122: 27–46. [CrossRef]
- Hansen, Bruce E. 2007. Least Squares Model Averaging. *Econometrica* 75: 1175–89. [CrossRef]
- Hansen, Bruce E., and Jeffrey S. Racine. 2012. Jackknife Model Averaging. *Journal of Econometrics* 167: 38–46. [CrossRef]
- Li, Ker-Chau. 1987. Asymptotic Optimality for C_p , C_L , Cross-Validation and Generalized Cross-Validation: Discrete Index Set. *Annals of Statistics* 15: 958–75. [CrossRef]
- Liu, Qingfeng, and Ryo Okui. 2013. Heteroscedasticity-Robust C_p Model Averaging. *Econometrics Journal* 16: 463–72. [CrossRef]
- Liu, Qingfeng, Ryo Okui, and Arihiro Yoshimura. 2016. Generalized Least Squares Model Averaging. *Econometric Reviews* 35: 1692–752. [CrossRef]
- Makridakis, Spyros, Evangelos Spiliotis, and Vassilios Assimakopoulos. 2018. The M4 Competition: Results, Findings, Conclusion and Way Forward. *International Journal of Forecasting* 34: 802–8. [CrossRef]
- Wan, Alan TK, Xinyu Zhang, and Guohua Zou. 2010. Least Squares Model Averaging by Mallows Criterion. *Journal of Econometrics* 156: 277–83. [CrossRef]
- Xie, Tian. 2017. Heteroscedasticity-Robust Model Screening: A Useful Toolkit for Model Averaging in Big Data Analytics. *Economics Letters* 151: 119–22. [CrossRef]
- Yuan, Zheng, and Yuhong Yang. 2005. Combining Linear Regression Models: When and How? *Journal of the American Statistical Association* 100: 1202–14. [CrossRef]
- Zhang, Xinyu, Alan TK Wan, and Guohua Zou. 2013. Model Averaging by Jackknife Criterion in Models with Dependent Data. *Journal of Econometrics* 174: 82–94. [CrossRef]
- Zhang, Xinyu, Guohua Zou, and Raymond J. Carroll. 2015. Model Averaging Based on Kullback-Leibler Distance. *Statistica Sinica* 25: 1583–98. [CrossRef] [PubMed]
- Zhao, Shangwei, Aman Ullah, and Xinyu Zhang. 2018. A Class of Model Averaging Estimators. *Economics Letters* 162: 101–6. [CrossRef]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).