Generalized C_p Model Averaging for Heteroskedastic Models

Qingfeng Liu*
Department of Economics, Otaru University of Commerce
5-21, Midori 3-chome, Otaru, Hokkaido 047-8501, Japan

This paper extends Hansen (2007), which proposed a Mallows model averaging (MMA) estimator for models with homoskedastic errors. The weights of the models for the MMA estimator are determined by minimizing a criterion similar to Mallows' C_p (MC). Our extension is a generalization of the MMA method. The GC method works for both homoskedastic and heteroskedastic errors not only as a model averaging method but also as a model selection method. For heteroskedastic situations, Andrews (1991) showed asymptotic optimality for a model selection criterion based on MC. However, Andrews (1991) did not provide a feasible form of this criterion, because of the difficulty associated with the consistent estimation of the covariance matrix. We provide a way to avoid the estimate of the covariance matrix, and are thus able to propose a feasible form of the GC method. Under some regularity conditions, we show that the GC method has asymptotic optimality not only as a model averaging method but also as a model selection method for models with heteroskedastic errors.

Hansen (2007) proposed an MMA estimator. In his setup, the regressors are assumed to be ordered, and the candidate regression models are assumed to be nested. Wan, Zhang, and Zou (2010) extended the results of Hansen (2007), by removing these assumptions. Our setup is similar to Wan, Zhang, and Zou (2010). The following is our model:

$$y_i = \mu_i + e_i,\ \mu_i = \sum_{j=1}^{\infty} \theta_j x_{ij},\ E\left(e_i | x_i\right) = 0, \tag{1}$$

for $i = 1, \cdots, n$, where y_i is a real-valued scalar, $x_i = (x_{i1}, x_{i2}, \cdots)$ is a countably infinite real-valued vector, μ_i is assumed to be converging in mean square, and $E\mu_i^2 < \infty$. Our results almost all are condational on x_i, for simplicity, we omit the conditinal expression in some cases hereafter. The most important difference between our setup and that of Hansen (2007) and Wan, Zhang, and Zou (2010) is that in their setup, the error term e_i is assumed to be homoskedastic and not heteroskedastic as in our setup. We assume that e_i is independent over i and $E\left(e_i^2 | x_i\right) = \sigma_i^2$.

*Associate Professor, Department of Economics, Otaru University of Commerce (Tel & Fax: +81 134 27 5312, E-mail: qliu@res.otaru-uc.ac.jp).

The model averaging criterion is defined as follows:

$$GC_n = \|Y - P(W)Y\|^2 + 2tr[\Omega P(W)], \tag{2}$$

where Ω is an $n \times n$ diagonal matrix whose ii element is σ_i^2. Then, the estimator of the optimal weight vector is derived as $\hat{W}_{GC} = \arg\min_{W \in \mathcal{H}_n} GC_n$.

The feasible form of GC is

$$\widehat{GC}_n \equiv \|Y - P(W)Y\|^2 + 2\sum_{i=1}^{n} \hat{e}_i^2 p_{ii}(W), \tag{3}$$

with $\hat{W}_{\widehat{GC}_n} \equiv \arg\min_{W \in \mathcal{H}_n} \widehat{GC}_n$.

Our aim is to show the optimality of $\hat{W}_{GC}$ under some regularity conditions. We define the loss function and the risk function as $L_n(W) = \|\hat{\mu}(W) - \mu\|^2$and $R_n(W) = E(L_n(W)|X)$ respectively. Then, optimality implies

$$\frac{L_n\left(\hat{W}_{GC_n}\right)}{\inf_{W \in \mathcal{H}_n} L_n(W)} \rightarrow_p 1. \tag{4}$$

We get the following theorems on the optimalities of $\hat{W}_{GC}$ and $\hat{W}_{\widehat{GC}_n}$.

Theorem 1 *For $\xi_n \equiv \inf_{W \in \mathcal{H}_n} R_n(W)$ and some integer $1 \leq G < \infty$, if*

$$E\left(e_i^{4G}|x_i\right) \leq \kappa < \infty, \tag{5}$$

$$M\xi_n^{-2G} \sum_{m=1}^{M} \left(R_n\left(W_m^0\right)\right)^G \rightarrow 0, \tag{6}$$

and $0 < \inf_i \sigma_i^2 \leq \sup_i \sigma_i^2 < \infty$, then $\frac{L_n\left(\hat{W}_{GC_n}\right)}{\inf_{W \in \mathcal{H}_n} L_n(W)} \xrightarrow{p} 1$, where W_m^0 is a vector whose mth element is one and all other elements are zeros.

Theorem 2 *When $\sum_{i=1}^{n} \hat{e}_i^2 p_{ii}(W)$ is used instead of $tr[\Omega P(W)]$, Theorem 2 is valid if*

$$0 < \lim n^{-1} \sum_{i=1}^{n} \sigma_i^2 = \overline{\sigma^2} < \infty, \tag{7}$$

$$\mu'\mu/n = O(1), \tag{8}$$

$$\max_{1 \leq m \leq M} \max_{1 \leq i \leq n} p_{m,ii} = O\left(n^{-1/2}\right), \tag{9}$$

$$\frac{\tilde{p}e'e}{\xi_n} \xrightarrow{p} 0, \tag{10}$$

$$\lim \lambda_{\max}(n) = \infty, \ \ \log(\lambda_{\max}(n)) = O\left(n^{1/2}\right), \tag{11}$$

where $\tilde{p} \equiv \sup_{W \in \mathcal{H}_n} \max_{1 \leq i \leq n}(p_{ii}(W))$, $\lambda_{\max}(n)$ is the maximum eigenvalue of $\tilde{X}'\tilde{X}$ with $\tilde{X}$ denoting the matrix of the regressors of the largest model, and $p_{m,ii}$ is the ith diagonal element of $P_{(m)}$.