Journal of Econometrics 186 (2015) 142–159

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Distribution theory of the least squares averaging estimator☆

Chu-An Liu*

Department of Economics, National University of Singapore, AS2 Level 6, 1 Arts Link, Singapore 117570, Singapore

ARTICLE INFO

Article history:
Received 13 August 2012
Received in revised form
2 June 2014
Accepted 6 July 2014
Available online 15 July 2014

JEL classification:
C51
C52

Keywords:
Local asymptotic theory
Model averaging
Model selection
Plug-in estimators

ABSTRACT

This paper derives the limiting distributions of least squares averaging estimators for linear regression models in a local asymptotic framework. We show that the averaging estimators with fixed weights are asymptotically normal and then develop a plug-in averaging estimator that minimizes the sample analog of the asymptotic mean squared error. We investigate the focused information criterion (Claeskens and Hjort, 2003), the plug-in averaging estimator, the Mallows model averaging estimator (Hansen, 2007), and the jackknife model averaging estimator (Hansen and Racine, 2012). We find that the asymptotic distributions of averaging estimators with data-dependent weights are nonstandard and cannot be approximated by simulation. To address this issue, we propose a simple procedure to construct valid confidence intervals with improved coverage probability. Monte Carlo simulations show that the plug-in averaging estimator generally has smaller expected squared error than other existing model averaging methods, and the coverage probability of proposed confidence intervals achieves the nominal level. As an empirical illustration, the proposed methodology is applied to cross-country growth regressions.

1. Introduction

In recent years, interest has increased in model averaging from the frequentist perspective. Unlike model selection, which picks a single model among the candidate models, model averaging incorporates all available information by averaging over all potential models. Model averaging is more robust than model selection since the averaging estimator considers the uncertainty across different models as well as the model bias from each candidate model. The central questions of concern are how to optimally assign the weights for candidate models and how to make inferences based on the averaging estimator. This paper investigates the averaging estimators in a local asymptotic framework to deal with these issues. The main contributions of the paper are the following: first, we characterize the optimal weights of the model averaging estimator and propose a plug-in estimator to estimate the infeasible optimal weights. Second, we investigate the focused information criterion (FIC; Claeskens and Hjort, 2003), the plug-in averaging estimator, the Mallows model averaging (MMA; Hansen, 2007), and the jackknife model averaging (JMA; Hansen and Racine, 2012). We show that the asymptotic distributions of averaging estimators with data-dependent weights are nonstandard and cannot be approximated by simulation. Third, we propose a simple procedure to construct valid confidence intervals to address the problem of inference post model selection and averaging.

In finite samples, adding more regressors reduces the model bias but causes a large variance. To yield a good approximation to the finite sample behavior, we follow Hjort and Claeskens (2003a) and Claeskens and Hjort (2008) and investigate the asymptotic distribution of averaging estimators in a local asymptotic framework where the regression coefficients are in a local $n^{-1/2}$ neighborhood of zero. This local asymptotic framework ensures the consistency of the averaging estimator while in general presents an asymptotic bias. Excluding some regressors with little information introduces the model bias but reduces the asymptotic variance. The trade-off between omitted variable bias and estimation variance remains in the asymptotic theory. Under drifting sequences of parameters, the asymptotic mean squared error (AMSE) remains finite and provides a good approximation to finite sample mean squared error. The $O(n^{-1/2})$ framework is canonical in the sense that both squared model biases and estimator variances have the same order $O(n^{-1})$. Therefore, the optimal model is the one that has the best trade-off between bias and variance in this context.

Under the local-to-zero assumption, we derive the asymptotic distributions of least squares averaging estimators with both fixed weights and data-dependent weights. We show that the submodel estimators are asymptotically normal and develop a model selection criterion, FIC, which is an unbiased estimator of the AMSE

☆ A previous version was circulated under the title "A Plug-In Averaging Estimator for Regressions with Heteroskedastic Errors".

* Tel.: +65 65166010.
E-mail address: ecslca@nus.edu.sg.

http://dx.doi.org/10.1016/j.jeconom.2014.07.002
0304-4076/

of the submodel estimator. The FIC chooses the model that achieves the minimum estimated AMSE. We extend the idea of FIC to the model averaging. We first derive the asymptotic distribution of the averaging estimator with fixed weights, which allows us to characterize the optimal weights under the quadratic loss function. The optimal weights are found by numerical minimization of the AMSE of the averaging estimator. We then propose a plug-in estimator of the infeasible optimal fixed weights and use these estimated weights to construct a plug-in averaging estimator of the parameter of interest. Since the estimated weights depend on the covariance matrix, it is quite easy to model the heteroskedasticity.

Estimated weights are asymptotically random, and this must be taken into account in the asymptotic distribution of the plug-in averaging estimator. This is because the optimal weights depend on the local parameters, which cannot be estimated consistently. To address this issue, we first show the joint convergence in distribution of all candidate models and the data-dependent weights. We then show that the asymptotic distribution of the plug-in estimator is a nonlinear function of the normal random vector. Under the same local asymptotic framework, we show that both MMA and JMA estimators have nonstandard asymptotic distributions.

The limiting distributions of averaging estimators can be used to address the important problem of inference after model selection and averaging. We first show that the asymptotic distribution of the model averaging t-statistic is nonstandard and not asymptotically pivotal. Thus, the traditional confidence intervals constructed by inverting the model averaging t-statistic lead to distorted inference. To address this issue, we propose a simple procedure for constructing valid confidence intervals. Simulations show that the coverage probability of traditional confidence intervals is generally too low, while the coverage probability of proposed confidence intervals achieves the nominal level.

In simulations, we compare the finite sample performance of the plug-in averaging estimator with other existing model averaging methods. Simulation studies show that the plug-in averaging estimator generally produces lower expected squared error than other data-driven averaging estimators. As an empirical illustration, we apply the least squares averaging estimators to cross-country growth regressions. Our estimator has the smaller variance of the log GDP per capita in 1960, though our regression coefficient of the log GDP per capita in 1960 is close to those of other estimators. Our results also find little evidence of the new fundamental growth theory.

The model setup in this paper is similar to that of Hansen (2007) and Hansen and Racine (2012). The main difference is that we consider a finite-order regression model instead of an infinite-order regression model. Hansen (2007) and Hansen and Racine (2012) propose the MMA and JMA estimators and demonstrate the asymptotic optimality in homoskedastic and heteroskedastic settings, respectively. However, it is difficult to make inferences based on their estimators since there is no asymptotic distribution available in either paper. By considering a finite-order regression model, we are able to derive the asymptotic distributions of the MMA and JMA estimators in a local asymptotic framework.

The idea of using the local asymptotic framework to investigate the limiting distributions of model averaging estimators is developed by Hjort and Claeskens (2003a) and Claeskens and Hjort (2008). Like them, we employ a drifting asymptotic framework and use the AMSE to approximate the finite sample MSE. We, however, consider a linear regression model instead of the likelihood-based model, and allow the errors to be both heteroskedastic and serially correlated. Furthermore, we characterize the optimal weights of the averaging estimator in a general setting and propose a plug-in estimator to estimate the infeasible optimal weights.

Other work on the asymptotic properties of averaging estimators includes Leung and Barron (2006), Pötscher (2006), and Hansen (2009, 2010, 2013). Leung and Barron (2006) study the risk bound of the averaging estimator under a normal error assumption. Pötscher (2006) analyzes the finite sample and asymptotic distributions of the averaging estimator for the two-model case. Hansen (2009) evaluates the AMSE of averaging estimators for the linear regression model with a possible structural break. Hansen (2010) examines the AMSE and forecast expected squared error of averaging estimators in an autoregressive model with a near unit root in a local-to-unity framework. Hansen (2013) studies the asymptotic risk of least squares averaging estimator in a nested model framework. Most of these studies, however, are limited to the two-model case and the homoskedastic framework.

There is a growing body of literature on frequentist model averaging. Buckland et al. (1997) suggest selecting the weights using the exponential AIC. Yang (2000, 2001), and Yuan and Yang (2005) propose an adaptive regression by mixing models. Hansen (2007) introduces the Mallows model averaging estimator for nested and homoskedastic models where the weights are selected by minimizing the Mallows criterion. Wan et al. (2010) extend the asymptotic optimality of the Mallows model averaging estimator for continuous weights and a non-nested setup. Liang et al. (2011) suggest selecting the weights by minimizing the trace of an unbiased estimator of mean squared error. Zhang and Liang (2011) propose an FIC and a smoothed FIC averaging estimator for generalized additive partial linear models. Hansen and Racine (2012) propose the jackknife model averaging estimator for non-nested and heteroskedastic models where the weights are chosen by minimizing a leave-one-out cross-validation criterion. Zhang et al. (2013) show the asymptotic optimality of the JMA estimator in the presence of lagged dependent variables. DiTraglia (2013) proposes a moment selection criterion and a moment averaging estimator for the GMM framework. In contrast to frequentist model averaging, there is a large body of literature on Bayesian model averaging; see Hoeting et al. (1999) and Moral-Benito (2013) for a literature review.

There is a large body of literature on inference after model selection, including Pötscher (1991), Kabaila (1995, 1998), and Leeb and Pötscher (2003, 2005, 2006, 2008, 2012). These papers point out that the coverage probability of the confidence interval based on the model selection estimator is lower than the nominal level. They also argue that no estimator for the conditional and unconditional distribution of post model selection can be uniformly consistent. In the model averaging literature, Hjort and Claeskens (2003a) and Claeskens and Hjort (2008) show that the traditional confidence interval based on normal approximations leads to distorted inference. Pötscher (2006) argues that the finite-sample distribution of the averaging estimator cannot be uniformly consistently estimated.

There are also alternatives to model selection and model averaging. Tibshirani (1996) introduces the LASSO estimator, a method for simultaneous estimation and variable selection. Zou (2006) proposes the adaptive LASSO approach and presents its oracle properties. Hansen et al. (2011) propose the model confidence set, which is constructed based on an equivalence test. White and Lu (2014) propose a new Hausman (1978) type test of robustness for the core regression coefficients. They also provide a feasible optimally combined GLS estimator.

The outline of the paper is as follows. Section 2 presents the regression model, the submodel, and the averaging estimator. Section 3 presents the asymptotic framework and assumptions. Section 4 introduces the FIC and the plug-in averaging estimator. Section 5 derives the distribution theory of FIC, plug-in, MMA, and JMA estimators, and proposes a procedure to construct valid confidence intervals for averaging estimators. Section 6 examines the finite sample properties of averaging estimators. Section 7 presents the empirical application, and Section 8 concludes the paper. Proofs are included in the Appendix.