Journal of Econometrics 188 (2015) 40–58

Contents lists available at ScienceDirect

Journal of Econometrics

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

Jackknife model averaging for quantile regressions[☆]

Xun Lu[a], Liangjun Su[b,*]

[a] Department of Economics, Hong Kong University of Science & Technology, Hong Kong

[b] School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore

ARTICLE INFO

Article history:
Received 4 September 2012
Received in revised form
14 June 2014
Accepted 24 November 2014
Available online 11 April 2015

JEL classification:
C51
C52

Keywords:
Final prediction error
High dimensionality
Model averaging
Model selection
Misspecification
Quantile regression

ABSTRACT

In this paper we consider model averaging for quantile regressions (QR) when all models under investigation are potentially misspecified and the number of parameters is diverging with the sample size. To allow for the dependence between the error terms and regressors in the QR models, we propose a jackknife model averaging (JMA) estimator which selects the weights by minimizing a leave-one-out cross-validation criterion function and demonstrate its asymptotic optimality in terms of minimizing the out-of-sample final prediction error. We conduct simulations to demonstrate the finite-sample performance of our estimator and compare it with other model selection and averaging methods. We apply our JMA method to forecast quantiles of excess stock returns and wages.

1. Introduction

In practice researchers are often confronted with a large number of candidate models and are not sure which model to use. Model selection helps to choose a single optimal model, ignores the information in other models, and often produces a rather unstable estimator in applications despite the fact that it has a long history and nice theoretical properties in both statistics and econometrics literature.[1] As an alternative to model selection, model averaging, on the other hand, seeks to obtain a combined estimator by taking the weighted average of the estimators obtained from all candidate models under investigation. It allows researchers to diversify, account for model uncertainty, and improve out-of-sample performance.

Model averaging can be classified as Bayesian model averaging (BMA) and frequentist model averaging (FMA). See Hoeting et al. (1999) for an overview on BMA and Moral-Benito (2015) for a recent overview on both BMA and FMA. FMA has a relatively shorter history than BMA. Buckland et al. (1997) and Burnham and Anderson (2002, ch. 6) construct model averaging weights based on the values of AIC or BIC scores. Yang (2001) and Yuan and Yang (2005) propose a model averaging method known as adaptive regression by mixing (ARM). In a local asymptotic framework Hjort and Claeskens (2003) and Claeskens and Hjort (2008, ch. 7) study the asymptotic properties of the FMA maximum likelihood estimator by studying perturbations around a given narrow model in certain directions. Other works on the asymptotic property of averaging estimators include Leung and Barron (2006), Pötscher (2006), Hansen (2009, 2010), and Liu (2015). In particular, Liu (2015) proposes a plug-in estimator of the optimal weights by minimizing the sample analog of the asymptotic mean squared error (MSE) for linear regression models. In a similar spirit, Liang et al. (2011) derive an exact unbiased estimator of the MSE of the model average estimator and propose selecting the weights that minimize the trace of the MSE estimate of focus parameters.

In a seminal article, Hansen (2007) proposes selecting the model weights in least squares model averaging by minimizing Mallows' criterion over a set of discrete weights. The justification of this method lies in the fact that the Mallows' criterion is

☆ The authors gratefully thank the co-editor, the associate editor, and two anonymous referees for their many helpful comments. They are also indebted to Peter C. B. Phillips for his constructive comments on the paper and valuable discussions on the subject matter. Su gratefully acknowledges the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021.

* Corresponding author. Tel.: +65 6828 0386.
E-mail addresses: xunlu@ust.hk (X. Lu), ljsu@smu.edu.sg (L. Su).

[1] It is well known that a small perturbation of the data can result in selecting a very different model. As a consequence, estimators of the regression function based on model selection often have larger variance than usual. See Yang (2001).

http://dx.doi.org/10.1016/j.jeconom.2014.11.005
0304-4076/

asymptotically equivalent to the squared error so that the Mallows model averaging (MMA) estimator is asymptotically optimal in terms of minimizing the MSE. Thus his approach marks a significant step toward the development of optimal weight choice in the FMA estimator. Hansen (2008, 2009, 2010) extend his MMA method to the forecast combination literature, to models with structural break, and to models with a near unit root, respectively. Note that Hansen (2007) only considers nested models and his MMA estimator does not allow for (conditional) heteroskedasticity. Wan et al. (2010) extend Hansen's MMA estimator to allow for non-nested model and selection of continuous weights in a unit simplex. Liu and Okui (2013) extends Hansen's MMA estimator to allow for heteroskedasticity and non-discrete weights. To allow for both non-nested models and heteroskedasticity, Hansen and Racine (2012) propose jackknife model averaging (JMA) for least squares regression when the weights are selected by minimizing a leave-one-out cross-validation criterion function. Zhang et al. (2013) extend JMA to models with dependent data. In the case of instrument uncertainty, Kuersteiner and Okui (2010) apply the MMA approach to the first stage of the 2SLS, LIML and FIML estimators. In contrast, Lee and Zhou (2011) take an average over the second stage estimators. Sueishi (2010) proposes a new simultaneous model and instrument selection method for IV models based on 2SLS estimation when the true model is of infinite dimension.

Almost all of the above papers on FMA focus on the least squares regression and MSE criterion. The only exceptions are Hjort and Claeskens (2003) and Claeskens and Hjort (2008) who concentrate on the likelihood framework but with the MSE criterion too. The MSE criterion seems natural in the least squares regression framework because it balances the asymptotic bias and variance in a nice way. Nevertheless, it is also interesting to apply the idea of FMA to different contexts where MSE may not be the best criterion choice.

In this paper we extend the JMA of Hansen and Racine (2012) to the quantile regression (QR) framework. QR provides much more information about the conditional distribution of a response variable than the traditional conditional mean regression. Since the seminal paper of Koenker and Bassett (1978), QR has attracted huge attention in the literature. Just as in the least squares regression, model selection and model averaging can play an important role in the QR model building process. There is a growing literature on model selection for QR models or more generally, M-estimation. For example, Hurvich and Tsai (1990) develop a small sample criterion for the selection of LAD regression models; Machado (1993) and Burman and Nolan (1995) consider variants of the Schwarz information criterion (SIC) and Akaike information criterion (AIC), respectively, for M-estimation which includes the QR as a special case; Koenker et al. (1994) consider using SIC in QR models. More recently, Wu and Liu (2009) study variable selection in penalized QR (see Su and Zhang (2014) for an overview on this); Belloni and Chernozhukov (2011) consider l_1-penalized QR in high-dimensional sparse models. Nevertheless, to the best of our knowledge, there still is a lack of an FMA method in the QR framework. This work seeks to fill this gap. It is well known that quantile estimates tend to be unstable when the quantile index is very high or very low (say, close to 0.95 or 0.05). This implies that model averaging can certainly play an important role in this case.

To proceed, it is worth mentioning that the major motivation for model averaging is to address the problem of *model uncertainty* for forecasting. Kapetanios et al. (2008) provide compelling reasons for using model averaging for the purpose of forecasting. They consider two broad cases: one is when the model that generates the data belongs to the class of candidate models, and the other, which is perhaps more relevant in empirical applications, is when the true model does not belong to the class of models under consideration. In the first case, model averaging addresses the issue that the chosen model is not necessarily the true model, and by assigning probabilities to various models can yield an out-of-sample forecast that is robust to model uncertainty. In the second case, it is impossible that the chosen model could capture all the features of the true model, which makes the motivation for model averaging even stronger because it has been well documented in the forecasting literature that forecasts from different models can inform the overall forecast in different ways and tend to outperform individual forecasts significantly. Admittedly, forecasting a variable of interest and discovering the true model (or true structural/causal relation) can be quite different objectives in econometrics. As Ng (2013) puts it in her abstract, "*(i)rrespective of the model size, there is an unavoidable tension between prediction accuracy and consistent model determination.*" Consistent model selection of the true model, if existing, does not necessarily lead to a model that yields minimum forecast error. The main purpose of this paper is to provide a FMA method for the purpose of forecasting a variable of interest under the check loss function but not to discover the underlying true model because it is possible in practice that none of the models considered is the true model or even close to the truth.

Since we use the check loss function as a base for model averaging, we do not have the usual bias–variance decomposition for the MSE-based evaluation criterion, and it is difficult to define a Mallows-type criterion for the QR model averaging as in Hansen (2007).[2] For this reason, we focus on the extension of Hansen and Racine's (2012) JMA to the QR framework. Such an extension is not trivial for several reasons. First, there is no closed form solution for QR, and the asymptotic properties of jackknife QR estimators are not well studied in the literature. Second, since we do not adopt the local asymptotic framework, it is possible that all the models under investigation are incorrectly specified even asymptotically. The literature on QR under misspecification is quite limited. Third, we allow the number of parameters in the QR models to diverge with the sample size n, which also complicates the analysis of QR estimators under model misspecification. We shall study the consistency and asymptotic normality of QR estimators for a single potentially misspecified QR model with a diverging number of parameters, and then study the uniform consistency of the leave-one-out QR estimators. These results are needed in order to establish the asymptotic optimality of our JMA estimator. Fourth, we also allow the number of the candidate models to increase with the sample size at a suitable polynomial rate.

We conduct Monte Carlo simulations to compare the finite sample performance of our JMA QR estimators with other model averaging and model selection methods, such as those based on AIC and BIC. We find that our JMA QR estimators clearly dominate other methods for the 0.05th conditional quantile regression. For the conditional median regression, there is no clearly dominating method, but JMA QR estimators perform well in most of the cases. We apply our new method to predict the conditional quantiles of excess stock returns and wages.

The rest of the paper is structured as follows. Section 2 proposes the quantile regression model averaging estimator. We study the asymptotic properties of the quantile regression estimators and the asymptotic optimality of our jackknife selected weight vector in Section 3. Section 4 reports the Monte Carlo simulation results. In Section 5 we apply the proposed method to predict conditional quantiles of excess stock returns and wages. Section 6 concludes. The proofs of the main results in Section 3 are relegated

[2] Alternatively, one can continue to adopt the MSE as an evaluation criterion for QR estimators. It remains unknown whether Hansen's MMA has a straightforward extension to QR.