Computational Statistics and Data Analysis 55 (2011) 1331–1341

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Weighted average least squares estimation with nonspherical disturbances and an application to the Hong Kong housing market

Jan R. Magnus[a], Alan T.K. Wan[b,*], Xinyu Zhang[c]

[a] Department of Econometrics & Operations Research, Tilburg University, Netherlands
[b] Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong
[c] Academy of Mathematics & Systems Science, Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Article history:
Received 17 January 2010
Received in revised form 22 September 2010
Accepted 22 September 2010
Available online 29 September 2010

Keywords:
Model averaging
Bayesian analysis
Monte Carlo
Housing demand

ABSTRACT

The recently proposed 'weighted average least squares' (WALS) estimator is a Bayesian combination of frequentist estimators. It has been shown that the WALS estimator possesses major advantages over standard Bayesian model averaging (BMA) estimators: the WALS estimator has bounded risk, allows a coherent treatment of ignorance and its computational effort is negligible. However, the sampling properties of the WALS estimator as compared to BMA estimators are heretofore unexamined. The WALS theory is further extended to allow for nonspherical disturbances, and the estimator is illustrated with data from the Hong Kong real estate market. Monte Carlo evidence shows that the WALS estimator performs significantly better than standard BMA and pretest alternatives.

1. Introduction

In standard econometric practice, regression diagnostics such as the t-ratio are used to select a model, after which the unknown parameters are estimated in the selected model. These estimates are then presented as if they were unbiased and the standard errors as if they represented the true standard deviations. This is clearly wrong, even in the unlikely event that the selected model happens to be the data-generating process.

The traditional role of the t-ratio in model selection is itself suspect. The t-ratio was developed in the context of hypothesis testing, not in the context of model selection. While it may make good sense to reject a null hypothesis only when 95% of the evidence is in favor of the alternative, it does not make nearly as good sense to only prefer one model over another when 95% of the evidence points that way.

Even when we accept the t-ratio as a model selection tool, there is the problem of pretesting. We are ignoring the fact that model selection and estimation are a combined effort and that the model selection part influences the properties of our estimators. These properties depend not only on the stochastic nature of our framework, but also on the way the model has been selected. If we ignore the model selection part, then our reported properties are conditional rather than unconditional. Focusing on the general-to-specific and specific-to-general model selection procedures, Danilov and Magnus (2004a) derived the unconditional moments of the least squares pretest estimators, and Danilov and Magnus (2004b) obtained the corresponding forecast moments. In a series of papers Leeb and Pötscher (2003, 2005, 2006, 2008) also studied conditional distributions of estimators subsequent to a range of model selection strategies. All these studies conclude that the problems associated with ignoring model selection can be very serious. For example, Danilov and Magnus (2004a)

* Corresponding author. Tel.: +852 27887146; fax: +852 27888560.
E-mail addresses: magnus@uvt.nl (J.R. Magnus), msawan@cityu.edu.hk (A.T.K. Wan), xinyu@amss.ac.cn (X. Zhang).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2010.09.023

showed that the mean squared error associated with the specific-to-general pretest procedure is unbounded, and that the errors in not reporting the correct moments can be large.

But even if we could fully understand the properties of the pretest estimator and take them into account, the problem remains that the pretest estimator itself is a poor estimator. This is primarily because the pretest estimator is 'kinked', which implies that it is inadmissible (Magnus, 2002, Theorem A.6). In addition to this theoretical problem, the pretest estimator is also uncomfortable intuitively. We may be used to taking one estimator when $t = 1.95$ and another when $t = 1.97$, but it remains difficult to justify. Would it not be more comfortable to abandon this 'kinked' estimator and use a 'continuous' estimator instead? A 'continuous' estimate of a specific parameter would consider the estimates from all models in a certain model space and then weigh them according to their statistical strengths, possibly in addition to prior information. This is the idea underlying model averaging.

The aim of model averaging is not to find the best possible model but rather to find the best estimates of the parameters of interest. Instead of selecting a single 'winning' model, we use information from all models, but of course some models are more reliable than others, based on data and priors. Model uncertainty is then incorporated in the properties of our estimators in a natural way. Bayesian model averaging (BMA) offers a natural framework (Leamer, 1978) and has been used widely, especially in growth econometrics. Raftery et al. (1997) and Hoeting et al. (1999) provide useful literature summaries of BMA. Recent applications of BMA can be found in Peña and Redondas (2006) and Ouysse and Kohn (2010). There has also been a rising interest in model averaging from a frequentist perspective and several frequentist model average estimators have been proposed. For example, Buckland et al. (1997) suggest mixing models based on the AIC or BIC scores of the competing models; Yang (2001, 2003) propose a frequentist-based adaptive regression mixing method; Hjort and Claeskens (2003) provide a likelihood-based local misspecification framework for analyzing the asymptotic properties of model average estimators; Hansen (2007, 2008) and Wan et al. (2010) suggest a Mallows criterion for selecting weights in a model average estimator; and Schomaker et al. (2010) develop frequentist model averaging schemes in the face of missing observations. The recent monograph by Claeskens and Hjort (2008) covers much of the progress that has been made in this direction. In addition, subset selection has been proposed in the frequentist model averaging literature, in order to narrow down the list of candidate models for averaging and thus save computing time; see, for example, Yuan and Yang (2005), Claeskens et al. (2006) and Buchholz et al. (2008).

Recently, Magnus et al. (2010), hereafter MPP, proposed a weighted average least squares (WALS) estimator based on the equivalence theorem developed in Magnus and Durbin (1999) and Danilov and Magnus (2004a). The WALS estimator is a Bayesian combination of frequentist estimators, and it possesses some important advantages over standard BMA techniques. First, in contrast to standard BMA estimators that adopt normal priors leading to unbounded risk, WALS uses priors from the Laplace distribution and thus generates bounded risk. Second, the use of Laplace priors implies a coherent treatment of ignorance. Third, WALS requires a trivial computational effort because computing time is linear in the number of regressors rather than exponential as in BMA or standard frequentist model averaging.

MPP discuss WALS estimation in the context of growth models, where there is a large number of potentially relevant explanatory variables, while Wan and Zhang (2009) apply WALS in a tourism study. In both cases the disturbances are assumed to be identically and independently distributed. This assumption is not realistic in most applications, and one purpose of the current paper is to develop WALS in the more general framework of nonspherical disturbances. The main purpose of the paper is to provide Monte Carlo evidence on the performance of WALS in a realistic set-up. In MPP it is shown that WALS works well, has intuitive appeal, is easy to compute, and produces realistic results, but we do not know yet whether the estimates are close to the true parameter values, neither do we know whether the standard errors capture the combined model selection and estimation errors adequately. This can only be found out through simulations and the current study attempts to provide a representative subset of such simulations.

As our empirical framework we choose a hedonic housing price model with data from Hong Kong, one of the world's most buoyant real estate markets. There typically exists a wide range of model specifications in hedonic housing price modeling, and this makes model uncertainty an important and challenging issue. We explore the root mean squared error (RMSE) of the WALS estimator and contrast this with the RMSE of the BMA estimator and a *stepwisefit* pretest estimator. We conclude that the WALS estimator has smaller RMSE than BMA and other model selection alternatives over a large portion of the parameter space, and is therefore a potentially useful estimator, not only when the number of regressors is large, but in any linear regression problem.

Our theoretical framework is the linear regression model

$$y = X_1\beta_1 + X_2\beta_2 + u = X\beta + u, \tag{1}$$

where y $(n \times 1)$ is the vector of observations on the dependent variable, X_1 $(n \times k_1)$ and X_2 $(n \times k_2)$ are matrices of nonrandom regressors, u is a random vector of unobservable disturbances, and β_1 and β_2 are unknown parameter vectors. The columns of X_1 are called focus regressors (those we want in the model on theoretical or other grounds), while the columns of X_2 are called auxiliary regressors (those we are less certain of). We assume that $k_1 \geq 0,\ k_2 \geq 0,\ 1 \leq k := k_1 + k_2 \leq n - 1$, and that $X := (X_1 : X_2)$ has full column-rank. Model selection takes place over the auxiliary regressors only. Since each of the k_2 auxiliary regressors can either be included or not, we have 2^{k_2} models to consider. In contrast to common practice in the model averaging literature we shall not assume that the disturbances are identically and independently distributed. Instead we assume that

$$u \sim \mathrm{N}(0, \Omega(\theta)), \tag{2}$$