Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Linear instrumental variables model averaging estimation

Luis F. Martins [a,b,*], Vasco J. Gabriel [b,c]

[a] Department of Quantitative Methods, ISCTE-IUL, Portugal
[b] Centre for International Macroeconomic Studies (CIMS), University of Surrey, UK
[c] NIPE-UM, Portugal

ARTICLE INFO

Article history:
Received 30 June 2012
Received in revised form 6 May 2013
Accepted 8 May 2013
Available online xxxx

Keywords:
Instrumental variables
Model selection
Model averaging
Model screening
Returns to education

ABSTRACT

Model averaging (MA) estimators in the linear instrumental variables regression framework are considered. The obtaining of weights for averaging across individual estimates by direct smoothing of selection criteria arising from the estimation stage is proposed. This is particularly relevant in applications in which there is a large number of candidate instruments and, therefore, a considerable number of instrument sets arising from different combinations of the available instruments. The asymptotic properties of the estimator are derived under homoskedastic and heteroskedastic errors. A simple Monte Carlo study contrasts the performance of MA procedures with existing instrument selection procedures, showing that MA estimators compare very favorably in many relevant setups. Finally, this method is illustrated with an empirical application to returns to education.

1. Introduction

In this paper, we consider model averaging (MA) estimation methods in the linear instrumental variables (IV) regression context. The model averaging estimator is a weighted average of individual estimates obtained using different lists of valid instruments. We propose obtaining empirical weights based on existing and well-established instrument selection criteria for IV models. This can be achieved by direct smoothing of information criteria arising from the estimation stage, as in Buckland et al. (1997), Burnham and Anderson (2002) and Hjort and Claeskens (2003). We show that the MA estimator is consistent and normally distributed with a specific closed-form expression for its asymptotic variance–covariance matrix. The proposed MA estimator and its first-order asymptotic properties are defined under the case of homoskedasticity and for general forms of heteroskedastic errors.

In many applications of IV estimation, there is often a large set of candidate variables that can be used as instruments. However, the properties of IV estimators are very sensitive to the choice (and the characteristics) of the instrument set. Indeed, instruments might be poorly correlated with the endogenous variables, which invalidates conventional inference procedures (Staiger and Stock, 1997; Stock and Wright, 2000). On the other hand, using many (potentially weak) instruments can improve efficiency and precision, but it can also lead to substantial deviations from the usual Gaussian asymptotic approximation (see Chao and Swanson, 2005, Han and Phillips, 2006, Hansen et al., 2008 and Newey and Windmeijer, 2009).

Thus, much of the literature has focused on procedures for the selection of the appropriate number and list of instruments. Donald and Newey (2001) propose a selection procedure such that an approximate mean-square error is minimized over all existing instruments deemed to be valid. It includes Two-Stage Least Squares (TSLS) limited information maximum likelihood and a bias adjusted version of the TSLS. On the other hand, Andrews (1999) developed GMM analogues of model selection criteria (MSC) based on the J-statistic in order to consistently select the largest set of valid moment conditions. Hall et al. (2007) suggest selecting instruments according to the relevant moment selection criterion (RMSC), based on the

* Correspondence to: Department of Quantitative Methods, ISCTE-IUL, Av. das Forças Armadas, 1649-026 Lisbon, Portugal. Tel.: +351 929659233; fax: +351 217903942.
E-mail addresses: luis.martins@iscte.pt (L.F. Martins), v.gabriel@surrey.ac.uk (V.J. Gabriel).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.05.008

entropy of the limiting distribution of the estimator, while Hall and Peixe (2003) propose canonical correlations information criteria (CCIC) for instrument selection (see also Lo and Ronchetti (2012) for an information and entropy-based approach to moment conditions estimation). As an alternative, Pesaran and Smith (1994) define a measure of the goodness of fit for IV regressions. They call it generalized R-squared, GR^2, and show that it ought to be computed based on the prediction errors.

Model selection entails choosing one of the estimated competing models under consideration. Testing competing, non-nested formulations, in which the outcome may not be the selection of one particular model, can be carried out using the tests of Smith (1992) and Smith and Ramalho (2002). Shrinkage methods, on the other hand, are an alternative to model selection. Caner (2009) proposes a LASSO-type GMM estimator, while Canay (2010) and Okui (2011) propose shrinkage-type estimators for linear models with many instruments. In fact, shrinkage estimators can be viewed as a special case of 'instrument averaging', in which some instruments receive weights approaching zero.

Here, we pursue the alternative approach of model averaging, in which parameter estimates are constructed based on a weighted average of estimates from a number of possible specifications. By making use of the information conveyed by otherwise discarded alternative specifications, model averaging as an estimation strategy may yield some gains in terms of bias and efficiency when compared to procedures that make use of a single set of instruments. Furthermore, our approach can cope with high-dimensional problems arising from large numbers of combinations of instruments, particularly when there is no clear indication as to which instruments should be discarded.

Our work is a natural extension of the literature, in which model averaging usually involves weights obtained from functions of model selection criteria, such as the BIC, AIC, etc. Indeed, there is a large literature on model averaging, both in the Bayesian tradition and in a frequentist context (see Claeskens and Hjort, 2008 for a review). In the latter framework, Hansen (2007) proposed a Mallows criterion for the selection of weights for averaging across least squares estimates obtained from a set of approximating models, in which regressors (or groups of regressors) are added sequentially. Liang et al. (2011), in turn, discuss optimal weight choice based on an unbiased estimator of the MA estimator's mean squared error (MSE), thus attaining good finite sample properties. On the other hand, Hansen and Racine (2012) consider a jackknife MA estimator, with weights based on a cross-validation criterion, which is asymptotically optimal under bounded heteroskedasticity of unknown form.

Model averaging in the linear IV context has seen some very recent developments. Kuersteiner and Okui (2010) suggest using Hansen's (2007) method as a first step to construct optimal instruments IV estimation with TSLS, LIML and Fuller estimators. The weights are chosen to minimize the approximate mean squared error (AMSE), as in Donald and Newey (2001). Koop et al. (2012), on the other hand, use a Bayesian model averaging approach to address different sources of uncertainty, such as the set of instruments, exogeneity restrictions, the validity of identifying restrictions and the set of exogenous regressors.

Nevertheless, our approach is distinct in that it averages estimates of the parameters of interest (rather than first-stage results as in Kuersteiner and Okui (2010)) and, in our case, the list of candidate models does not depend on ordered instruments from the full-instrument matrix. Indeed, with m instruments we can consider m models (each model including an extra instrument as in Hansen (2007)), but also any possible combination of these. This makes our approach more general and not restricted by how the instruments are ordered. Also, unlike TSLS kernel-based weighting as proposed by Canay (2010) and Okui (2011), our procedure does not depend on the choice of kernels or arbitrarily user-chosen smoothing parameters. Thus, we are able to combine the estimation of general moment condition models with one-step and information criteria-based model averaging estimation.

In fact, our paper is related to recent (and parallel) contributions. Lee and Zhou (2012) consider a different 'feasible' weighting scheme based on the strength of subsets of instruments (measured by the ratio of the first-stage R^2 and the Sargan statistic). Similarly, Chen et al. (2012) consider averaging moment condition estimators in a more general conditional estimation setup.

To study the properties of our estimators, we conduct a small-scale Monte Carlo experiment in which we contrast the performance of an averaging approach and that of an instrument selection strategy, showing that, in several setups, our model averaging estimation procedure outperforms the selection method of Donald and Newey (2001) in terms of median bias, absolute deviation and dispersion. Moreover, we illustrate empirically the use of MA procedures by examining returns to education, in which we show that even when both the sample and the number of instruments is quite large, our methods are flexible enough to cope with high-dimensional problems in an efficient way.

Next, Section 2 introduces the linear IV regression model and defines the instrument selection criteria. In Section 3, we introduce our model averaging approach, we discuss different procedures to obtain empirical weights and model screening as a strategy to narrow down the list of candidate specifications, thus reducing the computational burden. In Section 4, we derive the asymptotic properties of the model averaging estimator. A Monte Carlo simulation study providing evidence in support of our MA procedures is discussed in Section 5. An application to returns to schooling is considered in Section 6 and, finally, Section 7 concludes.

2. Definitions

Following the notation in Staiger and Stock (1997), the linear IV regression model is specified by a structural equation of interest

$$y = Y\beta + X\gamma + u, \tag{1}$$