Journal of Econometrics 186 (2015) 427–442

Contents lists available at ScienceDirect

Journal of Econometrics

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

Regularized LIML for many instruments[☆]

Marine Carrasco[a], Guy Tchuente[b,*]

[a] University of Montreal, CIREQ, CIRANO, Canada

[b] University of Kent, School of Economics, Keynes College, Canterbury, Kent CT2 7NP, United Kingdom

ARTICLE INFO

Article history:
Available online 5 March 2015

JEL classification:
C13

Keywords:
Heteroskedasticity
High-dimensional models
LIML
Many instruments
MSE
Regularization methods

ABSTRACT

The use of many moment conditions improves the asymptotic efficiency of the instrumental variables estimators. However, in finite samples, the inclusion of an excessive number of moments increases the bias. To solve this problem, we propose regularized versions of the limited information maximum likelihood (LIML) based on three different regularizations: Tikhonov, Landweber–Fridman, and principal components. Our estimators are consistent and asymptotically normal under heteroskedastic error. Moreover, they reach the semiparametric efficiency bound assuming homoskedastic error. We show that the regularized LIML estimators possess finite moments when the sample size is large enough. The higher order expansion of the mean square error (MSE) shows the dominance of regularized LIML over regularized two-staged least squares estimators. We devise a data driven selection of the regularization parameter based on the approximate MSE. A Monte Carlo study and two empirical applications illustrate the relevance of our estimators.

1. Introduction

The problem of many instruments is a growing part of the econometric literature. This paper considers the efficient estimation of a finite dimensional parameter in a linear model where the number of potential instruments is very large or infinite. Many moment conditions can be obtained from nonlinear transformations of an exogenous variable or from using interactions between various exogenous variables. One empirical example of this kind often cited in econometrics is Angrist and Krueger (1991) who estimate returns to schooling using many instruments, Dagenais and Dagenais (1997) also estimate a model with errors in variables using instruments obtained from higher-order moments of available variables. The use of many moment conditions improve the asymptotic efficiency of the instrumental variables (IV) estimators. For example, Hansen et al. (2008) have recently found that in an application from Angrist and Krueger (1991), using 180 instruments, rather than 3 shrinks correct confidence intervals substantially toward those of Kleibergen (2002). It has been observed that in finite samples, the inclusion of an excessive number of moments may result in a large bias (Andersen and Sorensen, 1996).

To solve the problem of many instruments efficiently, Carrasco (2012) proposed an original approach based on regularized two-stage least-squares (2SLS). However, such a regularized version is not available for the limited information maximum likelihood (LIML). Providing such an estimator is desirable, given LIML has better properties than 2SLS (see e.g. Hahn and Inoue (2002), Hahn and Hausman (2003), and Hansen et al., 2008). In this paper, we propose a regularized version of LIML based on three regularization techniques borrowed from the statistic literature on linear inverse problems (see Kress (1999) and Carrasco et al. (2007)). The three regularization techniques were also used in Carrasco (2012) for 2SLS. The first estimator is based on Tikhonov (ridge) regularization. The second estimator is based on an iterative method called Landweber–Fridman. The third regularization technique, called spectral cut-off or principal components, is based on the principal components associated with the largest eigenvalues. In our paper, the number of instruments is not restricted and may be smaller or larger than the sample size or even infinite. We also allow for a continuum of moment restrictions. We restrict our attention to the case where the parameters are strongly identified and the estimators converge at the usual $\sqrt{n}$ rate. However, a subset of instruments may be irrelevant.

We show that the regularized LIML estimators are consistent and asymptotically normal under heteroskedastic error. Moreover, they reach the semiparametric efficiency bound in presence of homoskedastic error. We show that the regularized LIML has finite

[☆] The authors thank a co-editor and two referees for insightful comments. They thank the participants of CIREQ conference on High Dimensional Problems in Econometrics (Montreal, May 2012), of the conference in honor of Jean-Pierre Florens (Toulouse, September 2012), of the seminars at the University of Rochester, the University of Pennsylvania, and Queen's University for helpful comments.

* Corresponding author. Tel.: +44 1227827249.
E-mail addresses: marine.carrasco@umontreal.ca (M. Carrasco), g.tchuente@kent.ac.uk (G. Tchuente).

http://dx.doi.org/10.1016/j.jeconom.2015.02.018

first moments provided the sample size is large enough. This result is in contrast with the fact that standard LIML does not possess any moments in finite sample.

Following Nagar (1959), we derive the higher-order expansion of the mean-square error (MSE) of our estimators and show that the regularized LIML estimators dominate the regularized 2SLS in terms of the rate of convergence of the MSE. Our three estimators involve a regularization or tuning parameter, which needs to be selected in practice. The expansion of the MSE provides a tool for selecting the regularization parameter. Following the same approach as in Donald and Newey (2001), Okui (2011), and Carrasco (2012), we propose a data-driven method for selecting the regularization parameter, α, based on a cross-validation approximation of the MSE. We show that this selection method is optimal in the sense of Li (1986, 1987), meaning that the choice of α using the estimated MSE is asymptotically as good as if minimizing the true unknown MSE.

The simulations show that the regularized LIML is better than the regularized 2SLS in almost every case. Simulations show that the LIML estimator based on Tikhonov and Landweber–Fridman regularizations often have smaller median bias and smaller MSE than the LIML estimator based on principal components and than the LIML estimator proposed by Donald and Newey (2001).

There is a growing amount of articles on many instruments and LIML. The first papers focused on the case where the number of instruments, L, grows with the sample size, n, but remains smaller than n. In this case, the 2SLS estimator is inconsistent while LIML is consistent (see Bekker (1994), Chao and Swanson (2005), Hansen et al. (2008), among others). Hausman et al. (2012) and Chao et al. (2012) give modified LIML estimators which are robust to heteroskedasticity in the presence of many weak instruments. Modifications of GMM have been considered by Canay (2010) and Kuersteiner (2012) who consider kernel weighted GMM estimators and Okui (2011) who uses shrinkage. Recently, some work has been done in the case where the number of instruments exceeds the sample size. Bai and Ng (2010) and Kapetanios and Marcellino (2010) assume that the endogenous regressors depend on a small number of factors which are exogenous, they use estimated factors as instruments. Belloni et al. (2012a) assume the approximate sparsity of the first stage equation and apply an instrument selection based on Lasso. Recently, Hansen and Kozbur (2014) propose a ridge regularized jackknife instrumental variable estimator in the presence of heteroskedasticity which does not require sparsity and provide tests with good sizes. The paper which is the most closely related to ours is that by Donald and Newey (2001) (DN henceforth) which selects the number of instruments by minimizing an approximate MSE. Our method assumes neither a strong factor structure, nor an exactly sparse first stage equation. However, it assumes that the instruments are sufficiently correlated among themselves so that the trace of the instruments covariance matrix is finite and hence the eigenvalues of the covariance matrix decrease to zero sufficiently fast.

The paper is organized as follows. Section 2 presents the three regularized LIML estimators and their asymptotic properties. Section 3 derives the higher order expansion of the MSE of the three estimators. In Section 4, we give a data-driven selection of the regularization parameter. Section 5 presents a Monte Carlo experiment. Empirical applications are examined in Section 6. Section 7 concludes. The proofs are collected in Appendix.

2. Regularized version of LIML

This section presents the regularized LIML estimators and their properties. We show that the regularized LIML estimators are consistent and asymptotically normal in presence of heteroskedastic error and they reach the semiparametric efficiency bound assuming homoskedasticity. Moreover, we establish that, under some conditions, they have finite moments.

2.1. Presentation of the estimators

The model is

$$\begin{cases} y_i = W_i'\delta_0 + \varepsilon_i \\ W_i = f(x_i) + u_i \end{cases} \quad (1)$$

$i = 1, 2, \ldots, n$. The main focus is the estimation of the $p \times 1$ vector δ_0. y_i is a scalar and x_i is a vector of exogenous variables. W_i is correlated with ε_i so that the ordinary least-squares estimator is not consistent. Some rows of W_i may be exogenous, with the corresponding rows of u_i being zero. A set of instruments, Z_i, is available so that $E(Z_i\varepsilon_i) = 0$. The estimation of δ is based on the orthogonality condition:

$$E[(y_i - W_i'\delta)Z_i] = 0.$$

Let $f(x_i) = E(W_i|x_i) \equiv f_i$ denote the $p \times 1$ reduced form vector. The notation $f(x_i)$ covers various cases. $f(x_i)$ may be a linear combination of a large dimensional (possibly infinite dimensional) vector x_i. Let $Z_i = x_i$, then $f(x_i) = \beta' Z_i$ for some $L \times p\beta$. Some of the coefficients β_j may be equal to zero, in which case the corresponding instruments Z_j are irrelevant. In that sense, $f(x_i)$ may be sparse as in Belloni et al. (2012b). The instruments have to be strong as a whole but some of them may be irrelevant. We do not consider the case where the instruments are weak (case where the correlation between W_i and Z_i converges to zero at the $\sqrt{n}$ rate) and the parameter δ is not identified as in Staiger and Stock (1997). We do not allow for many weak instruments (case where the correlation between W_i and Z_i declines to zero at a faster rate than $\sqrt{n}$ and the number of instruments Z_i grows with the sample size) considered by Newey and Windmeijer (2009) among others.

The model allows for x_i to be a few variables and Z_i to approximate the reduced form $f(x_i)$. For example, Z_i could be a power series or splines (see Donald and Newey, 2001).

As in Carrasco (2012), we use a general notation which allows us to deal with a finite, countable infinite number of moments, or a continuum of moments. The estimation is based on a set of instruments $Z_i = \{Z(\tau; x_i) : \tau \in S\}$ where S is an index set. Examples of Z_i are the following.

- Assume $Z_i = x_i$ where x_i is a L-vector with a fixed L. Then $Z(\tau; x_i)$ denotes the τth element of x_i and $S = \{1, 2, \ldots, L\}$.
- $Z(\tau; x_i) = (x_i)^{\tau-1}$ with $\tau \in S = \mathbb{N}$, thus we have infinite countable instruments.
- $Z(\tau; x_i) = exp(i\tau' x_i)$ where $\tau \in S = \mathbb{R}^{dim(x_i)}$, thus we have a continuum of moments.

It is important to note that throughout the paper, the number of instruments, L, of Z_i is either fixed or infinite and L is always independent of T. We view L as the number of instruments available to the econometrician and the econometrician uses all these instruments to estimate the parameters. We need to define a space of reference in which elements such that $E(W_i Z(\tau; x_i))$ are supposed to lie. We denote $L^2(\pi)$ the Hilbert space of square integrable functions with respect to π where π is a positive measure on S. $\pi(\tau)$ attaches a weight to each moments indexed by τ. π permits to dampen the effect of some instruments. For instance, if $Z(\tau; x_i) = exp(i\tau' x_i)$, it makes sense to put more weight on low frequencies (τ close to 0) and less weight on high frequencies (τ large). In that case, a π equal to the standard normal density works well as shown in Carrasco et al. (2007).

We define the covariance operator K of the instruments as

$$K : L^2(\pi) \to L^2(\pi)$$

$$(Kg)(\tau_1) = \int E(Z(\tau_1; x_i)\overline{Z(\tau_2; x_i)})g(\tau_2)\pi(\tau_2)d\tau_2$$

where $\overline{Z(\tau_2; x_i)}$ denotes the complex conjugate of $Z(\tau_2; x_i)$. K is assumed to be a nuclear (also called trace-class) operator which