ARTICLE IN PRESS

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

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

Minimum distance approach to inference with many instruments☆

Michal Kolesár *

Department of Economics and Woodrow Wilson School, Princeton University, United States

ARTICLE INFO

Article history:
Received 5 April 2014
Received in revised form 12 July 2017
Accepted 10 January 2018
Available online xxxx

JEL classification:
C13
C26
C36

Keywords:
Instrumental variables
Minimum distance
Incidental parameters
Random effects
Many instruments
Misspecification
Limited information maximum likelihood
Bias-corrected two-stage least squares

ABSTRACT

I analyze a linear instrumental variables model with a single endogenous regressor and many instruments. I use invariance arguments to construct a new minimum distance objective function. With respect to a particular weight matrix, the minimum distance estimator is equivalent to the random effects estimator of Chamberlain and Imbens (2004), and the estimator of the coefficient on the endogenous regressor coincides with the limited information maximum likelihood estimator. This weight matrix is inefficient unless the errors are normal, and I construct a new, more efficient estimator based on the optimal weight matrix. Finally, I show that when the minimum distance objective function does not impose a proportionality restriction on the reduced-form coefficients, the resulting estimator corresponds to a version of the bias-corrected two-stage least squares estimator. I use the objective function to construct confidence intervals that remain valid when the proportionality restriction is violated.

1. Introduction

This paper provides a principled and unified way of doing inference in a linear instrumental variables model with a single endogenous regressor and homoscedastic errors in which the number of instruments, k_n, is potentially large. To capture this feature in asymptotic approximations, I employ the many instrument asymptotics of Kunitomo (1980), Morimune (1983), and Bekker (1994) that allow k_n to increase in proportion with the sample size, n. I focus on the case in which collectively the instruments have substantial predictive power, so that the concentration parameter grows at the same rate as the sample size. I make no assumptions about the strength of individual instruments. I allow the rate of growth of k_n to be zero, in which case the asymptotics reduce to the standard few strong instrument asymptotics.

The presence of many instruments creates an incidental parameters problem (Neyman and Scott, 1948), as the number of first-stage coefficients, k_n, increases with the sample size. To directly address this problem, I use sufficiency and invariance arguments together with an assumption that the reduced-form errors are normally distributed to reduce the data to a pair of two-by-two matrices. In the absence of exogenous regressors, the first matrix can be written as $T = \begin{pmatrix} y & x \end{pmatrix}' P_Z \begin{pmatrix} y & x \end{pmatrix}/n$, where P_Z is the projection matrix of the instruments Z, and y and x are vectors corresponding to the outcome and the endogenous regressor. The second matrix, $S = \begin{pmatrix} y & x \end{pmatrix}'(I_n - P_Z)\begin{pmatrix} y & x \end{pmatrix}/(n - k_n)$, where I_n is the identity, corresponds to an estimator of the reduced-form covariance matrix. This solves the incidental parameters problem because the distribution of T and S depends on a fixed number of parameters even as $k_n \to \infty$: it depends on the first-stage coefficients only through the parameter λ_n, a measure of their collective strength.

I then drop the normality assumption and use a restriction on the first moment of T implied by the model to construct a minimum distance (MD) objective function. This restriction follows from the property of the instrumental variables model that the coefficients on the instruments in the first-stage regression are proportional to the coefficients in the reduced-form outcome regression. I use this MD objective function to derive three main results.

First, I show that minimizing the MD objective function with respect to the optimal weight matrix yields a new estimator of β, the coefficient on the endogenous regressor, that exhausts the information in T and S. In particular, this efficient MD estimator is

☆ Previously circulated under the title "Integrated Likelihood Approach to Inference with Many Instruments". I am deeply grateful to Guido Imbens and Gary Chamberlain for their guidance and encouragement. I also thank Joshua Angrist, Adam Guren, Whitney Newey, Jim Stock, Peter Phillips and participants at various seminars and conferences for their helpful comments and suggestions.

* Correspondence to: Department of Economics, Julis Romo Rabinowitz Building, Princeton University, Princeton, NJ 08544, United States.
E-mail address: mkolesar@princeton.edu.

https://doi.org/10.1016/j.jeconom.2018.01.004
0304-4076/

Please cite this article in press as: Kolesár M., Minimum distance approach to inference with many instruments. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.01.004.

asymptotically more efficient than the limited information maximum likelihood (LIML) estimator when the reduced-form errors are not normal. Standard errors can easily be constructed using the usual sandwich formula for asymptotic variance of minimum distance estimators.[1] The MD approach thus gives a simple practical solution to the many-instrument incidental parameters problem.

Second, I compare the MD approach to that based on the invariant likelihood—the likelihood, under normality, based on T and S. I show that, when combined with a particular prior on λ_n, the likelihood is equivalent to the random-effects (RE) quasi-maximum likelihood of Chamberlain and Imbens (2004), and that maximizing it yields LIML. Therefore, the random-effects estimator of β is in fact equivalent to LIML. Furthermore, I show that the RE estimator of the model parameters also minimizes the MD objective function with respect to a particular weight matrix. This weight matrix is efficient under normality, but not in general.

Third, I consider minimum distance estimation that leaves the first moment of T unrestricted. This situation arises, for instance, when the instrumental variables model is used to estimate potentially heterogeneous causal effects, as in Angrist and Imbens (1995). When the causal effect is heterogeneous, the reduced-form coefficients are no longer proportional, so that the first moment of T is unrestricted. In this case, the instrumental variables estimand β can be interpreted as a weighted average of the marginal effect of the endogenous variable on the outcome (Angrist et al., 2000). I show that the unrestricted minimum distance estimator coincides with a version of the bias-corrected two-stage least squares estimator (Nagar, 1959; Donald and Newey, 2001), and use the MD objective function to construct confidence intervals that remain valid when the proportionality restriction is violated.

The MD objective function is also helpful in deriving a specification test that is robust to many instruments. By testing the restriction on the first moment of T, I derive a new test that is similar to that of Cragg and Donald (1993), but with an adjusted critical value. The adjustment ensures that the test is valid under few strong as well as many instrument asymptotics that also allow for many regressors. In contrast, when the number of regressors is allowed to increase with the sample size, the size of the standard Sargan (1958) specification test converges to one, as does the size of the test proposed by Anatolyev and Gospodinov (2011).

The paper draws on two separate strands of literature. First, the literature on many instruments that builds on the work by Kunitomo (1980), Morimune (1983), Bekker (1994) and Chao and Swanson (2005). Like Anatolyev (2013), I relax the assumption that the dimension of regressors is fixed, and I allow them to grow with the sample size. Hahn (2002), Chamberlain (2007), Chioda and Jansson (2009), and Moreira (2009) focus on optimal inference with many instruments when the errors are normal and homoscedastic, and my optimality results build on theirs. Papers by Hansen et al. (2008), Anderson et al. (2010) and van Hasselt (2010) relax the normality assumption. Hausman et al. (2012), Chao et al. (2012), Chao et al. (2014) and Bekker and Crudu (2015) also allow for heteroscedasticity. An interesting new development is to employ shrinkage or regularization to solve the incidental parameters problem (see, for example, Belloni et al., 2012; Gautier and Tsybakov, 2014; or Carrasco, 2012). When combined with additional assumptions on the model, these shrinkage estimators can be more efficient than the efficient MD estimator proposed here.

Second, the literature on incidental parameters dating back to Neyman and Scott (1948). Lancaster (2000) and Arellano (2003) discuss the incidental parameters problem in a panel data context. Chamberlain and Moreira (2009) relate invariance and random effects approaches to the incidental parameters problem in a dynamic panel data model. My results on the relationship between these two approaches in an instrumental variables model build on theirs. Sims (2000) proposes a similar random-effects solution in a dynamic panel data model. Moreira (2009) proposes to use invariance arguments to solve the incidental parameters problem.

The remainder of this paper is organized as follows. Section 2 sets up the instrumental variables model, and reduces the data to the T and S statistics. Section 3 considers likelihood-based approaches to inference under normality. Section 4 relaxes the normality assumption and considers the MD approach to inference. Section 5 considers MD estimation without imposing proportionality of the reduced-form coefficients. Section 6 studies tests of overidentifying restrictions. Section 7 concludes. Proofs and derivations are collected in the Appendix. The supplemental appendix contains additional derivations.

2. Setup

In this section, I first introduce the model, notation, and the many instrument asymptotic sequence that allows both the number of instruments and the number of exogenous regressors to increase in proportion with the sample size. I then reduce the data to the low-dimensional statistics T and S, and define the minimum distance objective function.

2.1. Model and assumptions

There is a sample of individuals $i = 1, \dots, n$. For each individual, we observe a scalar outcome y_i, a scalar endogenous regressor x_i, ℓ_n-dimensional vector of exogenous regressors w_i, and k_n-dimensional vector of instruments z_i^*. The instruments and exogenous regressors are treated as non-random.

It will be convenient to define the model in terms of an orthogonalized version of the original instruments. To describe the orthogonalization, let W denote the $n \times \ell_n$ matrix of regressors with ith row equal to w_i', and let Z^* denote the $n \times k_n$ matrix of instruments with ith row equal to $z_i^{*\prime}$. Let $\tilde{Z} = Z^* - W(W'W)^{-1}W'Z^*$ denote the residuals from regressing Z^* onto W. Then the orthogonalized instruments $Z \in \mathbb{R}^{n\times k_n}$ are given by $Z = \tilde{Z}R^{-1}$, where the upper-triangular matrix $R \in \mathbb{R}^{k_n\times k_n}$ is the Cholesky factor of $\tilde{Z}'\tilde{Z}$. Now, by construction, the columns of Z are orthogonal to each other as well as to the columns of W.[2]

Denote the ith row of Z by z_i', and let $Y \in \mathbb{R}^{n\times 2}$ with rows (y_i, x_i) pool all endogenous variables in the model. The reduced form regression of Y onto Z and W can be written as

$$Y = Z\begin{pmatrix}\pi_{1,n} & \pi_{2,n}\end{pmatrix} + W\begin{pmatrix}\psi_{1,n} & \psi_{2,n}\end{pmatrix} + V, \tag{1}$$

where $V \in \mathbb{R}^{n\times 2}$ with rows $v_i' = (v_{1i}, v_{2i})$ pools the reduced-form errors, which are assumed to be mean zero and homoscedastic,

$$\mathbb{E}[v_i] = 0, \qquad \text{and} \quad \mathbb{E}[v_i v_i'] = \Omega. \tag{2}$$

The reduced-form coefficients on the instruments are assumed to satisfy a proportionality restriction, and the parameter of interest, β, corresponds to the constant of proportionality:

Assumption PR (*Proportionality Restriction*). $\pi_{1,n} = \pi_{2,n}\beta$.

The proportionality restriction implies that

$$y_i = x_i\beta + w_i'\beta_n^w + \epsilon_i, \tag{3}$$

where $\epsilon_i = v_{1i} - v_{2i}\beta$ is known as the structural error, and $\beta_n^w = \psi_{1,n} - \psi_{2,n}\beta$. This equation is known as the structural equation.

[1] Software implementing estimators and standard errors based on the MD objective function is available at https://github.com/kolesarm/ManyIV.

[2] This orthogonalization is sometimes called a standardizing transformation; see Phillips (1983) for discussion.

Please cite this article in press as: Kolesár M., Minimum distance approach to inference with many instruments. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.01.004.