Consequences and detection of invalid exogeneity conditions

Niemczyk, J.

Chapter 1

General Introduction and Overview

1.1 Endogenous regressors

In applied econometrics linear models are still widely utilized. The Ordinary Least Squares (OLS) estimation technique is the easiest one to obtain and to start the analysis with. However, when some of the regressors are endogenous (contemporaneously correlated with the regression error term) then, according to established wisdom, OLS is inappropriate to apply due to its inconsistency and then different methods are called for. Endogeneity of regressors could arise for several reasons, for example, measurement error, omitted explanatory variables, simultaneous equations, functional form misspecification.

Under appropriate distributional assumptions, OLS can be viewed as an application of Maximum Likelihood (ML) which is known for its optimal properties: it is asymptotically normally distributed; it is consistent and has minimal asymptotic variance. The ML estimator is derived from a full specification of the likelihood function which contains all the relevant information about the joint distribution of the observations on the dependent variable conditional on the exogenous or predetermined explanatory variables. Maximizing the likelihood function yields the desired estimator. ML heavily relies on the distributional assumptions of the sample at hand and in general could be computationally troublesome. Only under the normality assumption 'Quasi' ML corresponds in a linear model to OLS estimation.

One of the methods that does not require full knowledge of the conditional distribution of the dependent variable whereas it is designed to cope with possible endogeneity of the regressors in the linear model is Instrumental Variable (IV) estimation, also known as Two Stage Least Squares (2SLS). According to Stock and Trebbi (2003) it was invented by Phillip G. Wright (1928). It requires a researcher to posses a sufficient number of extra variables (instruments) possibly including some of the explanatory variables that are contemporaneously uncorrelated with the error term. This 'uncorrelatedness' establishes so called orthogonality conditions which allow to estimate the parameters of the model consistently.

Both OLS and IV estimators fall within a broader class of estimators introduced to econometrics by Hansen (1982) known as Generalized Method of Moments (GMM), see also Hall (2005). It is a more flexible estimation method than ML: it does not require to have full knowledge of the true underlying conditional distribution. Instead it relies on a set of moment conditions that are assumed to hold in the population of interest. GMM finds the parameter estimates by matching the sample moments with the population moments. The number of moment conditions can exceed the number of parameters to be estimated, but should not be smaller. If the population moment conditions are true, GMM can be shown to be consistent and asymptotically normal under fairly weak regularity conditions. GMM will make optimal use of all the available moment conditions in the sense that it is asymptotically efficient within its class when it uses an asymptotically optimal weighting matrix. However, it has also been shown that in finite samples it may be beneficial to abstain from exploiting weakly identified moment conditions, and that using very many instrumental variables, although beneficial from a standard asymptotic point of view, may lead to serious bias in small samples, see Donald and Newey (2001).

Instrumental variables (which form the moment conditions for the estimation), apart from being uncorrelated with the error term (being valid), should also explain the endogenous explanatory variables well, that is jointly they should be strongly correlated with the endogenous regressors (they should not be weak). Nelson and Startz (1990a,b) have shown that IV can have poor finite sample properties when instruments are 'weak'. Since then, a substantial body of literature has been written on the issue of weak instruments. In particular, Bound, Jaeger, and Baker (1995) illustrate that the empirical study of Angrist and Krueger (1991) is affected by the weak instruments problem. They also show that IV when using weak instruments will be biased in finite samples in the direction of the inconsistency of OLS. Therefore, in the presence of weak instruments Hausman (1978) type tests will have poor power in detecting endogeneity of regressors, because they are based on comparing OLS, which is inappropriate under endogeneity, with consistent IV. If the two estimators differ substantially we can suspect that some of the regressors are endogenous. Hall, Rudebusch, and Wilcox (1996) propose the use of canonical correlation criteria to measure instruments relevance. Staiger and Stock (1997) derive for the simultaneous equations model (SEM) the asymptotic distribution theory for various IV statistics when the correlation of instruments and endogenous regressor is 'local to zero' and propose a practical guideline for applied research when the instruments are weak. Anderson and Rubin (1949), Kleibergen (2002) and Moreira (2003) propose parameter tests that have a size which is invariant to the weak instruments problem. The studies on weak instruments are mainly dealing with the detection of the potential problems of weak identification and devising inference methods that are robust under weak instruments. Stock, Wright, and Yogo (2002) and more recently Andrews and Stock (2007) give comprehensive surveys on the weak instruments literature.

The finite sample properties of the optimally-weighted two step GMM estimator can be poor, see for example Hansen, Heaton, and Yaron (1996). Attractive alternative estimators (that similar to GMM exploit underlying moment conditions) have been proposed: Empirical Likelihood (EL) by Qin and Lawless (1994), see also Imbens (1997); Exponential Tilting (ET) by Kitamura and Stutzer (1997); and the Continuous Updated Estimator (CUE) by Hansen, Heaton, and Yaron (1996). Smith (1997) and Newey and Smith (2004) show that those estimators fall into the broader family of Generalized Empirical Likelihood (GEL) estimators. The study of Newey and Smith (2004) suggests that GEL will have better finite sample properties relative to GMM. In practice, assessing finite sample properties of GEL type estimators and test statistics in a Monte Carlo setup can be burdensome due to computational difficulties in obtaining both fast and accurately the solution to the saddle point problem that GEL poses.

Of course GMM (and also its generalizations offered by GEL and its specializations provided by IV and OLS) lose their attractive asymptotic properties when invalid moment conditions are being exploited. In order to prevent that from happening instrument validity tests have been developed. The most common tests for testing the validity of the extra moment conditions, so-called overidentifying restrictions tests, are the Sargan test for IV, see Sargan (1958), the J test for GMM, see Hansen (1982) and the Likelihood Ratio type tests for GEL, see Smith (1997). The consistency of the overidentifying restrictions tests for GEL was proven by Smith (1997). Optimality of EL for testing moment conditions has been shown by Kitamura (2001). However, it is impossible to test the validity of all the exploited moment conditions. Only under the untestable assumption of having a number of valid instruments equal to the number of unknown coefficients in the model the validity of any additional instruments can be tested by overidentification tests. Hence, it seems most likely that in practice it will often happen that GMM or IV are being exploited with invalid instruments leading to inconsistent estimators. The consequences of using invalid instruments has been analyzed by Hendry (1979), Maasoumi and Phillips (1982), Hall and Inoue (2003). Andrews (1997, 1999) proposes and proves the consistency of several strategies to find from a (sufficiently large) set of instruments the largest subset of valid moment conditions. Some of those selection procedures find the optimal set of moment conditions by minimizing an appropriate criterion function (over possible subsets of moment conditions). Other strategies are upward or downward searching procedures.

1.2 Some practical research questions

Still many research questions are waiting for answers that would help practitioners, despite the great many contributions in the literature regarding the consequences for various inference methods both in large samples and in small samples: regarding the presence of endogenous explanatory variables in econometric models; the possible use of invalid instruments; and the qualities of tests to detect endogenous regressors and possibly invalid instruments. Below we mention those research questions that will be addressed in this thesis.

1. What are the actual consequences both asymptotically and in finite samples of using

OLS when some of the regressors are actually endogenous? What is the magnitude of bias (inconsistency), how is the distribution of OLS and its variance affected?

2. What are the consequences both asymptotically and in finite samples of using IV when some of the instruments are in fact invalid, and in what way does weakness of instruments affect these consequences?

3. Can classic detection methods for the (in)validity of instruments, like the Sargan test or Hansen's J-test, be improved by bootstrapping them or by replacing them with similar statistics based on GEL type estimators? Are these tests vulnerable regarding instrument weakness?

4. Can sequential selection methods for the detection of valid instruments be improved by using incremental versions of Sargan or J tests? The incremental version of the overidentifying restriction test tests the validity of a subset of instruments by taking the difference between test statistics of the validity of a set of instruments and of a subset of those instruments.

The above issues will be examined in this thesis mostly for single equation models, where the model specification is mostly linear and all the variables are stationary and either endogenous or exogenous. Often our analysis is valid too for dynamic relationships, which may also contain predetermined regressors. However, we exclude non-stationarity of variables. Hence, most of our results pertain to analyzing cross-sectional data.

1.3 Overview of this thesis

In practice structural equations are often estimated by least-squares, thus neglecting any simultaneity. It is examined in Chapter 2[1] why this may often be justifiable and when. Assuming data stationarity and existence of the first four moments of the disturbances we

[1] This is an updated and corrected version of Kiviet and Niemczyk (2007), see Kiviet and Niemczyk (2009a). The corrections concern the following: (a) the formulation and proof of the main result has been adapted and clarifies now that it produces the conditional asymptotic distribution of inconsistent OLS in linear models; (b) also the unconditional asymptotic distribution is derived; (c) the illustrations now compare both conditional and unconditional distributions, both asymptotically and in finite samples.

study the limiting distribution of the OLS estimator in a linear simultaneous equations model. In simple static models we compare the asymptotic efficiency of this inconsistent estimator with that of consistent simple IV estimators and depict cases where – due to relative weakness of the instruments or mildness of the simultaneity – the inconsistent estimator is more precise. In addition, we examine by simulation to what extent these first-order asymptotic findings are reflected in finite samples, taking into account non-existence of moments of the IV estimator. In all comparisons we distinguish between conditional and unconditional (asymptotic) distributions.

In Chapter 3[2], we examine IV estimation when instruments may be invalid. This is relevant because validity of the initial just-identification restrictions is untestable. Moreover, tests for the validity of additional instruments, so-called over-identification restriction tests, have limited power when samples are small, especially when instruments are weak. Conditioning on genuinely predetermined possibly latent variables, we find the limiting normal distribution of inconsistent IV, expressed in parameters and data moments. This provides a first-order asymptotic approximation to the density in finite samples. For a specific simple class of models we compare this approximation and its unconditional counterpart with the simulated empirical distribution over almost the full parameter space, which is expressed in measures for: model fit, simultaneity, instrument invalidity and instrument weakness. Our major findings are that for the accuracy of large sample asymptotic approximations instrument weakness is much more detrimental than instrument invalidity. And, IV estimators obtained from strong but possibly invalid instruments are usually much closer to the true parameter values than those obtained from valid but weak instruments.

In the remaining part of the thesis we focus on instrumental variables selection methods. In Chapter 4, we investigate in a Monte Carlo setup several versions of classic GMM tests and Likelihood Ratio like tests based on GEL estimators for overidentifying restrictions. We also investigate incremental versions of those tests. We examine several bootstrap methods to obtain better finite sample critical values other than the standard ones based on the asymptotic null distribution of the tests. Those bootstrap methods

[2] This is based on Kiviet and Niemczyk (2009b)

include Hall and Horowitz (1996) who replicate the test statistics in a way that the bootstrap moments are exactly satisfied in the bootstrap samples. We also examine methods proposed by Brown and Newey (2002) who bootstrap the data using the implied probabilities obtained from GEL estimation, which then leads to the sample moments being satisfied in the bootstrap population. In addition we investigate a simple residual type bootstrap of the data. We find that for linear models the adapted Hall (2000) version of the Sargan test bootstrapped according to Hall and Horowitz (1996) performs very well (even in models with heteroscedastic errors) in comparison to the other tests we analyze. This is peculiar because the Sargan test is in principle designed for homoscedastic errors only. The standard χ^2 asymptotic critical values are derived under homoscedasticity.

In Chapter 5 we propose several ways to apply those overidentifying restrictions tests to detect the invalid instruments from a possibly very large set that contains a number of valid instruments equal to or larger than the number of unknown coefficients plus one. The selection procedures are sequential and differ in the way how the overidentifying restrictions tests are used. We compare the performance of our selection procedures and two of Andrews's (1999) procedures in a Monte Carlo setup. The first two of our procedures are computationally feasible, that is for a large set of moment restrictions to search from they will provide an answer within a reasonable time limit. We find that all the selection procedures are vulnerable to weak instruments and that the sample size and instrument invalidity should be substantial in order for the procedures to show good power in detecting the invalid instruments. Nonetheless, for moderate invalidity of instruments the resulting distribution of the estimates (over the instruments found by any of the selection procedures) is almost centered around the true value. Finally, we analyze the empirical data of Angrist and Krueger (1991), where we have as many as 180 potential instruments to search from. For several model specifications our selection procedure, which utilizes the incremental version of the test, finds that some of the instruments are invalid. That could be explained by the possibility that the instruments based on the quarter of birth of an individual are indeed correlated with some excluded characteristics that do affect the earnings of individuals.

In the last section of Chapter 5, the practical usefulness - despite their infeasibility

- of the theoretical results of Chapters 2 and 3 will be illustrated on one of the Angrist and Krueger specifications. We will produce an alternative inference based on making varying assumptions on the degree of simultaneity, and when external instruments have been used, their possible degree of invalidity. We will demonstrate that this inference, which is based on an assumption regarding hard to establish nuisance parameters, allows a useful sensitivity analysis. We demonstrate that for these data, such inference based on an unfeasible bias corrected OLS estimator is more attractive than for the unfeasible bias corrected IV estimator, due to the weakness of the instruments.

At the end there is a summary of the thesis in English and in Dutch.