# Estimation and Testing in A Structural Equation Model with Possibly Many Instruments

Yukitoshi Matsushita

2005

Statistical inference procedures in structural equation models are known to have poor finite sample properties in several contexts. It has been known that if instruments are only weakly correlated with the endogenous variables, classical normal and chi-square asymptotic approximations to the finite-sample distributions of GMM and IV statistics can be poor. In addition, in recent microeconometric applications some econometricians have used many instrumental variables in estimating an important structural equation. One empirical example of this kind often cited in econometric literatures is Angrist and Krueger (1991), where they used 178 instruments in one of their specifications. If the number of the instrumental variables is large efficiency can be improved, but it makes the usual inference procedures inaccurate.

Recently, the problems of weak and many instruments have received in-

creased attention by theoretical and applied researchers. Nelson and Startz (1990a,b), using a short sample and single instrument, showed that the distribution of the Two-Stage Least Squares (TSLS) estimetor can be strikingly nonnormal, biased seriously in the direction of the Ordinary Least Squares (OLS) estimator. On the other hand, Bound, Jaeger, and Baker (1995) showed that the properties of the TSLS estimator can be poor in the face of many weak instruments even when the sample size is huge, by analyzing the properties of TSLS in the context of Angrist and Krueger's (1991) regression of wages on education and exogenous variables.

In this thesis, several alternative asymptotic approximations are developed to improve the finite sample properties on the estimation and testing problems with possibly many instruments.

First, we focus on the finite sample properties of estimators. A method of asymptotic expansions of the distributions of a class of semi-parametric estimators is developed. (Chapter 3) To estimate the coefficients of a single equation in econometric models, GMM has been quite popular in the past two decades. In addition, the method of Empirical Likelihood (EL) is one of the alternatives which have been proposed recently, and has gotten some attention in the statistical and econometric literatures. It is shown that, under some conditions, the resulting formulas of the maximum empirical likelihood (MEL) estimator and the generalized method of moments (GMM) estimator are identical to those for the LIML estimator and the TSLS estimator respectively, except one term depending on the fourth order moments of the disturbances.

The distributions of the four different estimators (LIML, TSLS, MEL,

GMM) of one endogenous variable are evaluated numerically in Chapter 4. Since it is quite difficult to obtain the exact densities and cdf's of these estimators, the numerical information makes possible the comparison of properties of alternative estimation methods. Again, it is shown that the relation of the exact distributions of the MEL estimator and GMM estimator are quite similar to that of the distributions of the LIML and TSLS estimators. The distribution of the MEL estimator approaches normality faster than the distribution of the GMM estimator, and is centered at the parameter value in all cases. However, the tail of the distribution of the MEL estimator is relatively long, especially when the instruments are weak. The distribution of the GMM estimator is, on the other hand, less spread out, but can be seriously skewed when the number of the instruments is large.

In Chapter 5, we consider the testing problems when a large or moderate number of instruments are available. We develop an alternative approximation theory, "large K-asymptotics", allowing the number of instruments to grow with the number of observations. The properties of $t$-ratio statistics are explored under the large K-asymptotic theory. A modified $t$-ratio statistic from the asymptotic expansion is also proposed. It is shown that when the number of the instruments is large, the null distribution of the standard $t$-ratio is skewed and extremely deviated from the normal distribution. The null distribution of the large K $t$-ratio is closer to the normal distribution than the standard $t$-ratio, but still skewed and size distortion can be large for the one-sided test. The null distribution of the modified large K $t$-ratio is closest to the standard normal distribution. In addition, the power of the large K $t$-ratio test is shown to dominate the other test, and the difference is

substantial when the instruments are weak. When we know the sign of parameter from the economic theory, use of the modified large K $t$-atio statistic is recommended as a more accurate test procedure.

Finally, we make an asymptotic expansion of the distribution of the likelihood ratio (LR) statistic under the null hypothesis $H_0 : \beta = \beta_0$ under normality, and propose a modification of the LR test.(Chpater 6) The comparison of this modified LR test are made with Moreira's conditional likelihood ratio test and the large K $t$-test by Monte Carlo experiments. The Monte Carlo experiments show that, when the instruments are weak, the size properties of the LR test become quite poor, and the modified LR ($LR_m$) test improves upon the LR test when the number of the instruments is small and $\delta^2/K_2$ is more than one. However, both the modified LR test and the Moreira's CLR test can be size distorted when the number of the instruments is large. The large K $t$-test has the best size properties when the number of the instruments is large and $\delta^2/K_2$ is more than one.