Estimation and Inference with Weak Instruments and Near Exogeneity

by

Ying Fang

B.A., Fudan University, 1996

M.A., Fudan University, 1999

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH

Department of Economics

This dissertation was presented

by

Ying Fang

It was defended on

April 19, 2006

and approved by

Mehmet Caner, Associate Professor, Department of Economics, University of Pittsburgh

Daniel Berkowitz, Associate Professor, Department of Economics, University of Pittsburgh

Jean-Francois Richard, Professor, Department of Economics, University of Pittsburgh

Karen Clay, Assistant Professor, Heinz School of Public Policy and Management

Nese Yildiz, Assistant Professor, Department of Economics, University of Pittsburgh

Dissertation Advisor: Mehmet Caner, Associate Professor, Department of Economics,

University of Pittsburgh

Estimation and inference with Weak Instruments and Near Exogeneity

Ying Fang, PhD

University of Pittsburgh, 2006

Empirical economic studies are often confronted by the joint problem of weak instruments and near exogeneity, such as labor economics and empirical economic growth theory. This dissertation presents new evidence and solutions on estimation and inference with weak instruments and near exogeneity. Chapter 1 reexamines the effect of institutions on economic performance in Acemoglu, Johnson and Robinson (2001) where the measurement of current institutions is instrumented by European settler mortality rates. Since many economists argue that the settler mortality rates can possibly affect economic performance through other channels, I reexamine the effect of institutions by considering near exogeneity. I provide some evidence to show that the effect of institutions is not significant in many regression specifications when the settler mortality rates are used as the main instrument. Chapter 2 studies estimation and inference with weak instruments and near exogeneity in a linear simultaneous equations model. I show that near exogeneity can exaggerate asymptotic bias of the TSLS and the LIML estimators. When using critical values from chi-square distributions, Anderson-Rubin and Kleibergen tests under exogeneity have a large size distortion. I propose the delete-d jackknife based Anderson-Rubin and Kleibergen tests to automatically reduce the size distortion in finite samples without a need for any pretest of exogeneity. Chapter 3 extends estimation and inference with weak identification and near exogeneity into a GMM framework with instrumental variables. A GMM framework allows nonlinear and nondifferentiable moment conditions. I examine asymptotic results of one-step GMM estimator, two-step efficient GMM estimator and continuously updating estimator with weak identification and near exogeneity. Near exogeneity can produce relatively large bias for all these estimators. The Anderson-Rubin type and the Kleibergen type tests under near exogeneity converge in distribution to nonstandard distributions, which creates large size distortion when using critical values from chi-square distributions. The delete-d jackknife based approach can reduce the size distortion.

TABLE OF CONTENTS

LIST OF TABLES

PREFACE

This dissertation is the result of two and a half years of work whereby I have been accompanied and supported by many people.

First of all, I would like to extend my wholehearted thanks to my advisor Prof. Mehmet Caner. I will always remember the happy time we worked and enjoyed together. I would like to express my gratitude to Professors Daniel Berkowitz, Oliver Board, Karen Clay, David DeJong, Taisuke Otsu, Jean-Francois Richard, and Nese Yildiz for their helpful conversations and suggestions. I also wish to thank Xin He and Yong Sui for their helpful discussions.

Finally, all my gratitude goes to my parents and my fiancée Ms. Zheng Jing. Your affectionate encouragement has been constantly guiding me ahead to the final pilgrimage. This dissertation is dedicated to all of you. Without you, I will be nothing.

1.0 INTRODUCTION

Estimation and inference with instrumental variables (IV) have wide applications in empirical studies. In order to justify the IV method, it should satisfy two important criteria. One is called "instrument exogeneity", which means that instruments excluded from the structural equation should be uncorrelated with the structural errors. The other is called "instrument relevance", which requires that instruments should be strongly correlated with the endogenous explanatory variables. Finding valid instruments to satisfy the two criteria is not an easy job.

In an influential empirical study of labor economics, Angrist and Krueger (1991) use quarter of birth as an instrument for education to estimate the impact of compulsory schooling laws on earnings. They argue that children's quarter of birth is random, so it is uncorrelated with ability and should be exogenous. Because of compulsory laws, average education is generally longer for children born near the end of the year than for children born early in the year, which means that quarter of birth is correlated with educational attainment. Based on large samples (329,000 observations or more) from the U.S. census, they estimate the return to education by the TSLS procedure, using as instruments for education a set of three quarter-of-birth dummies interacted with fifty state-of-birth dummies and nine year-of-birth dummies respectively. But Bound, Jaeger and Baker (1995) point out that the instruments used in Angrist and Krueger's paper are weak and nearly exogenous in which case the resulting estimation and inference are misleading. Many authors work on improving inference under weak instruments; see, for example, Staiger and Stock(1997), Dufour(1997), Kleibergen(2002), Moreira(2003), among others.

Instrument exogeneity is another important criterion for valid instruments. In empirical studies, the validity of instrument exogeneity is mainly based on economic reasoning. But unfortunately, it is almost impossible to control for all possible variables that might be correlated with instruments and dependent variables. As a result, the instruments might catch the effect on dependent variables through other channels. It is hard to argue that instruments are exogenous in

empirical studies. For example, Acemoglu, Johnson, and Robinson (2001) estimate the effect of institutions on economic performance by using as instrument the logarithm of the European settler mortality rates. They argue that the settler mortality rate more than 100 years ago is strongly correlated with current institutions in the countries colonized by Europeans in the history. The mortality rates expected by the first European settlers determined the settlement decision and then influenced the colonization strategy: introducing "extractive states" (bad institution) or "Neo-Europes" (good institution). In a study of whether a reversal in relative incomes among the former European colonies reflects changes in the institutions resulting from European colonialism, Acemoglu, Johnson, and Robinson (2002) use data on urbanization and population density in 1500 to proxy for economic prosperity. In order to test whether population density or urbanization in 1500 affects income today only through institutions, the settler mortality rate is used as instruments again. But Glaeser, La Porta, Lopez-De-Silanes, and Shleifer (2004) argue that the settler mortality rate is not an exogenous instrument because the mortality rate might affect today's income through other channels, for example, the human capital. This is a problem of near exogeneity where the instruments are weakly correlated with the structural errors. Due to the nature that it is almost impossible to control for all possible variables that might be correlated with instruments and dependent variables of interest, the problem of near exogeneity is prevalent in empirical studies. Angrist (1990) estimates the effect of veteran status on civilian earnings by using as instruments the draft lottery numbers. But Wooldridge (2002) argues that the draft lottery numbers might be correlated with the structural errors if education is not controlled in the earnings equation. Bound, Jaeger, and Baker (1995) argue that the instruments used by Angrist and Krueger (1991) are not only weak but also suffer from near exogeneity.

This dissertation presents new evidence and solutions on estimation and inference with weak instruments and near exogeneity. Chapter 1 reexamines the effect of institutions on economic performance in Acemoglu, Johnson and Robinson (2001) by considering near exogeneity. I provide some evidence to show that the effect of institutions is not significant in many regression specifications when the settler mortality rates are used as the main instrument. Chapter 2 studies estimation and inference with weak instruments and near exogeneity in a linear simultaneous equations model. I show that near exogeneity can exaggerate asymptotic bias of the TSLS and the LIML estimators. When using critical values from chi-square distributions, Anderson-Rubin

and Kleibergen tests under exogeneity have a large size distortion. I propose the delete-d jackknife based Anderson-Rubin and Kleibergen tests to automatically reduce the size distortion in finite samples without a need for any pretest of exogeneity. Chapter 3 extends estimation and inference with weak identification and near exogeneity into a GMM framework with instrumental variables. A GMM framework allows nonlinear and nondifferentiable moment conditions. I examine asymptotic results of one-step GMM estimator, two-step efficient GMM estimator and continuously updating estimator with weak identification and near exogeneity. Near exogeneity can produce relatively large bias for all these estimators. The Anderson-Rubin type and the Kleibergen type tests under near exogeneity converge in distribution to nonstandard distributions, which creates large size distortion when using critical values from chi-square distributions. The delete-d jackknife based approach can reduce the size distortion

2.0 REEXAMINING THE EFFECT OF INSTITUTIONS BY CONSIDERING NEAR EXOGENEITY

In empirical studies, instrumental variables have wide applications by using the exogenous variance of instruments to estimate the effect of endogenous variables. In order for the valid use of the instrumental variables method, it requires a strict orthogonality condition between instrumental variables and the error terms in the structural equations. The validity of instrument exogeneity is mainly based on economists' knowledge about a specific economic issue at hand.

But unfortunately, it is almost impossible to control for all possible variables that might be correlated with instruments and dependent variables. As a result, the instruments might catch the effect on dependent variables through other channels. It is hard to argue that instruments are exogenous in empirical studies. For example, Acemoglu, Johnson, and Robinson (2001) estimate the effect of institutions on economic performance by using as instrument the logarithm of the European settler mortality rates. They argue that the settler mortality rate more than 100 years ago is strongly correlated with current institutions in the countries colonized by Europeans in the history. The mortality rates expected by the first European settlers determined the settlement decision and then influenced the colonization strategy: introducing "extractive states" (bad institution) or "Neo-Europes" (good institution). In a study of whether a reversal in relative incomes among the former European colonies reflects changes in the institutions resulting from European colonialism, Acemoglu, Johnson, and Robinson (2002) use data on urbanization and population density in 1500 to proxy for economic prosperity. In order to test whether population density or urbanization in 1500 affects income today only through institutions, the settler mortality rate is used as instruments again. But Glaeser, La Porta, Lopez-De-Silanes, and Shleifer (2004) argue that the settler mortality rate is not an exogenous instrument because the mortality rate might affect today's income through other channels, for example, the human capital. This is a problem of near exogeneity where the instruments are weakly correlated with

the structural errors. Due to the nature that it is almost impossible to control for all possible variables that might be correlated with instruments and dependent variables of interest, the problem of near exogeneity is prevalent in empirical studies. For example, Angrist (1990) estimates the effect of veteran status on civilian earnings by using as instruments the draft lottery numbers. But Wooldridge (2002) argues that the draft lottery numbers might be correlated with the structural errors if education is not controlled in the earnings equation. Bound, Jaeger, and Baker (1995) argue that the instruments used by Angrist and Krueger (1991) are not only weak instruments but also suffer from near exogeneity.

The usual overidentification tests, like the Sargen test and the J test, cannot solve the problem of near exogeneity satisfactorily. First, these overidentification tests usually have power problem in finite samples. Even the instruments pass through these overidentification tests, we cannot blindly assume a zero correlation between instruments and structural errors. Second, finding instruments is a creative but very tough job. It's statistically impossible to test the instrument exogeneity in the case of just-identification. Last, these overidentification tests cannot apply when there is a joint problem of near exogeneity and weak instruments.

One of the most widely test statistics used in empirical studies is the t -statistic. For instance, it's a routine to use the t -statistic to test whether an estimator of interest is significant away from zero. We show that under the t -statistic has a large size distortion even when there is a slight violation of the orthogonality condition. The subsampling based or the delete- d jackknife based t -statistic cannot help to solve the size problem. We propose the subsampling based or the delete- d jackknife based Anderson-Rubin test in empirical studies under near exogeneity. We reexamine the estimates in Acemoglu, Johnson and Robinson (2001) by considering the effect of near exogeneity.

This paper is organized as follows. Section 2.1 examines the large sample property of the t -statistic and the Anderson-Rubin test and their corresponding resampling based versions when the knife-edge exogeneity assumption is slightly violated. Section 2.2 provides an reexamination of the estimates in Acemoglu, Johnson and Robinson (2001) by considering the effect of near exogeneity. Section 2.3 conducts simulations to compare their finite sample performance, and Section 2.4 concludes. Appendix is included in last section.

2.1 THE EFFECT OF VIOLATION OF EXOGENEITY ASSUMPTION

In this section, we consider a linear simultaneous equations model (Hausman, 1983; Phillips, 1983) which is popular in empirical studies when instrumental variables are used,

$$y = Y\beta + u$$

$$Y = Z\Pi + V$$

where y and Y are respectively an $N \times 1$ vector and an $N \times m$ matrix of endogenous variables, Z is an $N \times K$ matrix of instruments, u is an $N \times 1$ vector of structural errors, V is an $N \times m$ matrix of reduced form errors, and errors have zero means and finite variance. The β and Π are respectively an $m \times 1$ unknown parameter vector and a $K \times m$ unknown matrix of parameters. Note that we require $K \geq m$. Other covariates can be added into Equations (100) and (110). We can always use the Frisch-Waugh-Lovell Theorem (see Davidson and MacKinnon, 1993, p19) to project out these covariates so the above equations give a simple linear model without loss of generality. The first equation is a structural equation and the second equation is a reduced form equation.

To estimate β properly we need valid instruments. Valid instruments depend on two criteria. First, the instruments should be well correlated with endogenous variables, i.e., instrument relevance. The second and more difficult criterion to satisfy is the assumption of exogeneity of instruments. This means that the covariance between instruments and structural errors is zero ($covZ_i'u_i = 0$). Even for the most carefully chosen instruments in empirical studies, it is almost impossible to argue a strict exogeneity condition.

We are interested in estimation and inference about β when there exits a small correlation between instruments and structural errors. We model this small correlation as near exogeneity which is a local to zero setup such that

$$E[Z_i'u_i] = C/\sqrt{N}$$

where C is a fixed $K \times 1$ vector. This is used as our main assumption in the paper. The correlation between instruments and structural errors shrinks toward zero as the sample size N

grows large. Fang (2005) considers near exogeneity in a linear simultaneous equations model. Caner (2005) considers near exogeneity in nonlinear moment restrictions in generalized empirical likelihood estimators. Near exogeneity is a more realistic assumption in applied works than the knife-edged assumption that requires a zero correlation between instruments and the error terms in the structural equation. We show that even a slight violation of this orthogonality condition can lead to a large size distortion of the *t*-statistic and the limiting distribution is different.

2.1.1 The t-statistic under Near Exogeneity

Consider the TSLS estimator $\hat{\beta}_{TSLS}$ in a linear model,

$$\hat{\beta}_{TSLS} = (Y'P_Z Y)^{-1}(Y'P_Z y).$$

where $P_Z = Z(Z'Z)^{-1}Z'$. Under near exogeneity assumption, it is easy to show that the TSLS estimator $\hat{\beta}_{TSLS}$ is consistent and converges to a normal distribution with a nonzero mean. The main reason why we can obtain consistent estimator under near exogeneity is that the correlation between instruments and structural errors shrinks toward zero at the rate of the square root of N when the sample size N grows infinity. The nonzero mean is due to the fact of near exogeneity. For these details, see Lemma 1 in Appendix 1.

Now, consider the t -statistic in the two-stage least squares method which is heavily used in the empirical literature. We want to test

$$H_0 : \beta_{i,TSLS} = \beta_{i,0}$$

against

$$H_1 : \beta_{i,TSLS} \neq \beta_{i,0}$$

The t -statistic is given by

$$t = \frac{\hat{\beta}_{i,TSLS} - \beta_{i,0}}{\sqrt{avar(\hat{\beta}_{i,TSLS})}}$$

where $i = 1, 2, \ldots, m$, $avar(\hat{\beta}_{i,TSLS}) = \hat{\sigma}_u^2 [(Y'Z)(Z'Z)^{-1}(Z'Y)]_{ii}^{-1}$ and $\hat{\sigma}_u^2 = \frac{1}{N-K-m}(y - Y\hat{\beta}_{TSLS})'(y - Y\hat{\beta}_{TSLS})$.

Under near exogeneity, the t -statistic converges in distribution to a normal distribution with nonzero mean which is showed in Theorem 1 in Appendix 1. When $C = 0$, we can obtain a standard normal distribution for the t -statistic under the exogeneity assumption. Near exogeneity shifts the asymptotic distribution to the right when $C > 0$. Using critical values from the standard normal distribution can lead to a large overrejection in finite samples. Table 1 shows the size distortion of t -statistic under near exogeneity from the simulation. When the correlation between instruments and structural errors is 0.15 , the actual size can be 39.1% while the nominal size is just 10% . This means that the t -statistic can overreject a true null hypothesis in empirical studies when there is near exogeneity problem.

The employment of the t -statistic heavily relies on the exogeneity condition. A slight violation of the exogeneity condition like the near exogeneity assumption can exaggerate the size distortion immensely. We also consider whether the resampling versions of the t -statistic can correct the size problem. The delete- d jackknife based t -statistic can be constructed by following the steps described in the appendix and Theorem 2 in Appendix 1 summarizes the limiting results of the delete- d jackknife based t -statistic. The delete- d jackknife based t -statistic cannot replicate the near exogeneity effect in the limiting distribution and the simulation in Table 2 shows that it works very bad in finite samples.

Note that bootstrap cannot be a solution to near exogeneity problem since it cannot replicate the correlation between instruments and structural errors in bootstrap samples. Subsampling cannot replicate such a correlation either. This can be seen in papers by Caner (2006) and Fang (2005). We should be very careful in empirical studies when using instruments and making inference based on the t -statistic. But, we will show that the delete- d jackknife method works very well in the case of the Anderson-Rubin test in the next section.

2.1.2 The Anderson-Rubin Test

In the last section, we show that the t -statistic has a large size distortion under near exogeneity. Another weakness of the t -statistic is the nonstandard limiting distribution when the nuisance parameter Π is close to zero, which is called weak instruments in the literature (Staiger and Stock, 1997). The nonstandard distribution is due to the fact that the t -statistic depends on the TSLS estimator and the TSLS estimator with weak instruments is inconsistent. Instead of the t -statistic, we propose the delete- d jackknife based Anderson-Rubin test. We show that the delete- d jackknife based Anderson-Rubin test is not only robust to weak instruments but also has only a slightly liberal limit compared to the regular asymptotics when there is near exogeneity. The size performance of the delete- d jackknife based Anderson-Rubin test in finite samples is summarized in Table 4.

We first examine the Anderson-Rubin test (Anderson and Rubin, 1949) under near exogeneity. The test is given by

$$AR(\beta_0) = (y - Y\beta_0)'P_Z(y - Y\beta_0) / [(y - Y\beta_0)'M_Z(y - Y\beta_0)/(N - K - m)].$$

where $M_Z = I_N - P_Z$ and I_N is an identity matrix with dimension N . We test $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$.

The Anderson-Rubin test is robust to weak instruments since the test itself does not use any information about the estimator of the first-stage parameter $\hat{\Pi}$. We know that under weak instruments the first-stage parameter Π cannot be consistently estimated. Under the null hypothesis of $\beta = \beta_0$, the test converges in distribution to a chi-square distribution with degrees of freedom K , the number of instruments. Moreira (2003) shows that the Anderson-Rubin test is uniformly most powerful among the class of unbiased tests when $K = m$.

The Anderson-Rubin test is also affected asymptotically by near exogeneity problem. Theorem 3 in Appendix 1 shows that the test converges in distribution to a noncentral chi-square distribution. The limit of the test depends on the nuisance parameter C which comes from near exogeneity.

We obtain the result of the test as a chi-square distribution when $C = 0$. Near exogeneity leads to a distortion in size when we use critical values from the chi-square distribution with degrees of freedom K. This can be showed by simulation results summarized in Table 3. When the correlation between instruments and structural errors is in the range of 0.10 and 0.15, the actual sizes are between 20% and 35% while the nominal size is just 10%.

Our strategy is to use the delete- d jackknife procedure to mimic the noncentral chi-square distribution defined in the appendix.

To introduce the delete- d jackknife based Anderson-Rubin test, we need to explain how we resample the sample data. First, d observations are randomly chosen without replacement from all of the sample observation, and then we form a subsample by deleting these d observations from the whole sample. Given d, the block size b of the subsample is $N - d$. Let y_b, Y_b and Z_b are respectively subvectors or submatrixes of y, Y and Z. So y_b is a $b \times 1$ vector, Y_b is a $b \times m$ matrix, and Z_b is a $b \times K$ matrix. These variables are denoted with subscript b because y_b, Y_b and Z_b represent randomly resampled data with block size $b = N - d$ from all sample observations without replacement. Let $d = \gamma N$ and then $b = (1 - \gamma)N$. Various γ 's will be tried in simulations, $0 < \gamma < 1$. From simulation results summarized in Table 4, we see that $\gamma = \frac{3}{4}$ works very well in finite samples. This is also suggested by Wu (1990). Given γ and N, the number of such blocks (denoted by N_b) we can generate is $N_b = \binom{N}{\gamma N}$. When $N = 64$ and $\gamma = \frac{3}{4}$, then $b = 16$ and $N_b = \binom{64}{48}$ which is a very large number. In the simulations, we use 1000 such random blocks. Next, we need to compute the Anderson-Rubin test in each block.

Denote by AR_S the delete- d jackknife based Anderson-Rubin test

$$AR_S(\beta_0) = (y_b - Y_b\beta_0)' P_{Z_b}(y_b - Y_b\beta_0) \big/ [(y_b - Y_b\beta_0)' M_{Z_b}(y_b - Y_b\beta_0)/(b - K - m)]$$

The delete- d jackknife based Anderson-Rubin test can be implemented by following steps:

Step 1: Randomly choose d observations from the sample without replacement, where $d < N$;

Step 2: Given d, the block size is $b = N - d$. Our subsample data are y_b, Y_b, and Z_b. Compute the delete- d jackknife based Anderson-Rubin test defined as above by using sample observations not deleted by Step 1 and the null hypothesis that $\beta = \beta_0$;

Step 3: Replicate Step 1 & 2 by at least 1000 times and sort these computed delete- d jackknife based Anderson-Rubin test;

Step 4: Use the 90% quantile as the data-dependent critical value. The delete- d jackknife based Anderson-Rubin test rejects the null hypothesis when the value of the Anderson-Rubin test for all sample observations, $AR(\beta_0)$ defined in (160), is larger than the data-dependent critical value.

Theorem 4 in Appendix 1 gives the limiting result of this delete- d jackknife based Anderson-Rubin test. It also converges in distribution to a noncentrality chi-square distribution. The noncentrality parameter is a fraction of the noncentrality parameter found in Theorem 3, which means the delete- d jackknife based Anderson-Rubin test is slightly liberal in large samples. By increasing the block size we can expect to reduce the size distortion due to near exogeneity. We can also observe this fact from simulations summarized in Table 4. Wu (1990) suggests $\frac{1}{4} \leq \lambda \leq \frac{3}{4}$ for delete- d jackknife. We propose $\gamma = \frac{3}{4}$ in finite samples. When $\gamma = \frac{3}{4}$, the block size $b = 16$. The actual sizes are 7.4 and 11.8 respectively when the correlation between instruments and structural errors is between 0.10 and 0.15. Compared to actual sizes in Table 3, we can see that the delete- d jackknife based Anderson-Rubin test can reduce the oversize problem under near exogeneity.

2.2 AN APPLICATION TO AJR (2001)

In this section, we reexamine Acemoglu, Johnson, and Robinson's (2001) estimates of the effects of institutions on economic performance by the foregoing results. One of their main contributions is to exploit the effect on economic performance by using European settler

mortality rates as an instrument for current institution. They argue that the European mortality rates determined the settlement decisions and then the early institutions in the countries colonized by Europeans. Since institutions persisted even after independence, the authors utilize the source of variation in European settler mortality rates as an instrument for current institutions, and then use the TSLS methods to estimate the effect of institutions on economic performance. The linear regression model used in their paper can be summarized as follows:

$$\log y_i = \mu + \alpha R_i + X_i'\gamma + \varepsilon_i$$
$$R_i = \varsigma + \beta \log M_i + X_i'\delta + \upsilon_i$$

where y_i is income per capita in country i, R_i is the measurement of institutions, an index of protection against expropriation, X_i is a vector of other covariates, M_i is the European settler mortality rate in $1,000$ mean strength, ε_i and υ_i are random errors. The logarithm of European settler mortality rates is the only instrument and other covariates which appear in the first-stage regression also appear in the second-stage regression.

Although most economists agree that the effect of institutions on economic performance is important and significant, it's still far from clear among economists that the instrument, the European settler mortality rates, is exogenous in this model with only sixty-four observations.

In empirical studies, one of the most important inference procedures is to use the t -statistic to test whether the estimator is significant away from zero. The null hypothesis is $H_0 : \alpha = 0$ against $H_1 : \alpha \neq 0$. The estimator is regarded as significant when the t -statistic rejects the null hypothesis. However, with the problem of near exogeneity, both Theorem 1 and Table 1 show the strong evidence of overrejection. Acemoglu, Johnson and Robinson (2001) obtain strongly significant estimators in all of their specifications. We reexamine their results by the delete- d jackknife based Anderson-Rubin test which consider the effect of near exogeneity.

The delete- d jackknife based Anderson-Rubin statistic tests the null hypothesis $H_0 : \alpha = 0$ against $H_1 : \alpha \neq 0$. Given a block size b , we randomly draw 1000 blocks with the block size b from 64 data observations. We calculate the value of the Anderson-Rubin statistic in each block under the null hypothesis, and then we can obtain the empirical distribution of the Anderson-Rubin test with 1000 various values. The data-dependent critical value is the top

90% quantile from the smallest value to the largest value. Next we compute the Anderson-Rubin test $AR(\beta_0)$ of the whole sample when $\beta_0 = 0$. Then the delete-d jackknife based Anderson-Rubin test rejects the null hypothesis with size 10% when $AR(\beta_0)$ is larger than the data-dependent critical values. The p-values of the delete-d jackknife based Anderson-Rubin test is calculated as the probability that the value of the Anderson-Rubin test $AR_S(\beta_0)$ computed in each block is larger than the value of the Anderson-Rubin test $AR(\beta_0)$ computed in the whole sample. From Table 4, we observe that the number of rejections of the test increases as the block size shrinks and $b = 16$, 18, 20 and 22 is the good range of the block sizes among various choices when the sample size is 64. When the block size is larger than 22, the test is very conservative. When the block size is smaller than 14, the test is overrejected under near exogeneity in simulations done by Fang (2005).

In Tables 5-8, we calculate p-values of regular t-statistic and the delete-d jackknife based Anderson-Rubin test when the block size is 16. For other block sizes, the results are reported in Tables 9-12. We observe that the results under $b = 16$ are not changing with other block sizes.

Table 5a repeats the baseline regressions in Acemoglu, Johnson and Robinson (2001). In column (1), the mortality rate is the only instrument and in column (2), latitude is added as a control variable. Columns (1) and (2) correspond columns (1) and (2) of Table 4 in Acemoglu, Johnson and Robinson (2001). In column (1) of Table 5b, the *p*-value of the resampling based AR test is 0.078, which shows that institutions are significant at 10% level. But, when the latitude is controlled in the regression, the p-value of the delete-d jackknife based AR test increases to 0.146, which is not significant at 10% level. In column (3), we add Asia dummy, Africa dummy and other continent dummy as the controlled variables and in column (4), the latitude is added. Columns (3) and (4) correspond columns (7) and (8) of Table 4 in Acemoglu, Johnson and Robinson (2001). Both columns show that the delete-d jackknife based AR test has large p-values and cannot reject the null hypothesis.

Tables 6-8 examine the robust tests by adding additional controls. In Table 6, the British\French colonial dummies or the French legal origin dummy is added into the regressions, which

corresponds columns (1),(2), (5) and (6) respectively of Table 5 in Acemoglu, Johnson and Robinson (2001). Our results show that the delete- d jackknife based Anderson-Rubin test has a large p -value and cannot reject the null hypothesis. In Table 7, religion variables or ethnolinguistic fragmentation is added into the regressions as the additional covariates, which corresponds columns (7) and (8) of Table 5 and columns (7) and (8) of Table 6 respectively in Acemoglu, Johnson and Robinson (2001). The delete- d jackknife based Anderson-Rubin test cannot reject the null hypothesis again. In Table 8, some geographically-related health variables, such as malaria, life expectancy and infant mortality, are added as additional controls, which corresponds columns (1)-(6) of Table 7 in Acemoglu, Johnson and Robinson (2001). We observe that the delete- d jackknife based Anderson-Rubin test cannot reject the null hypothesis.

By considering the effect of near exogeneity, we only observe two significant cases of the institution estimator. One is the most simple case where nothing is controlled except the mortality rate used as an instrument. The delete- d jackknife based Anderson-Rubin test has a p -value of 0.078 (see column (1) in Table 5). The other is a comprehensive specification where the latitude, the British\French colonial dummy, the French legal origin dummy, and religion variables are added into the regression simultaneously. The delete- d jackknife based Anderson-Rubin test has a p -value of 0.050 (see column (3) in Table 7).

We also compute p -values of the delete- d jackknife based Anderson-Rubin test under various block sizes, which are summarized in Tables 9-12. We use block sizes $b = \{12, 14, 20, 24, 28, 30, 32\}$. From simulations summarized in Table 4, we know that there is size distortion when $b = 12$ if the correlation between instruments and structural errors is in the range of 0.10 and 0.15 . Except column (1) in Table 5 and column (3) in Table 7, we observe that the delete- d jackknife based Anderson-Rubin test cannot reject the null hypothesis even when $b = 12$, which is a very strong evidence to show that the TSLS estimator is not significant when the mortality rate is used as an instrument for the institution. The results in Table 5-8 are robust to change in block size.

2.3 MONTE CARLO SIMULATION

We consider the linear simultaneous equations model defined in (100) and (110). Since there is only one endogenous variable Y, we set $l = 1$ (the just-identified case). We also examine the overidentified case when $l = 2$. Since the results from overidentification are very similar to those from just-identification, we only report the results from the just-identification in the paper. The β is the only structural parameter and we set the true value $\beta_0 = 0$. The N is the sample size and we set $N = 64$ in order to conduct comparisons of tests' performance in finite samples. The data (Z_i, u_i, V_i) is iid which are generated from a joint normal distribution $N(0, \Lambda)$.

When $l = 1$,

$$\Lambda = \begin{pmatrix} 1 & covZ_iu_i & 0 \\ covZ_iu_i & 1 & covV_iu_i \\ 0 & covV_iu_i & 1 \end{pmatrix}.$$

where $covV_iu_i$ measures the endogeneity of Y, which takes values of 0.25. When $l = 1$, $covZ_iu_i$ measures the degree of near exogeneity which takes values of 0, 0.10 or 0.15. The data generated from above also differ over the value of Π. The vector Π controls the quality of instruments. We set $\Pi = 0$, 0.1, or 1 in all cells of the vector to respectively represent nonidentification, weak instruments and strong instruments. In each simulation, the nominal size is 10%.

Table 1 shows the size distortion of the regular t-statistic under various degrees of near exogeneity when instruments are strong. When the correlation between instruments and structural errors is zero, the actual size of the t-statistic is close to the nominal size 10%. When the correlation is not zero, we can observe a size distortion and the size distortion increases immensely as the correlation increases.

Table 2 lists the actual sizes of the subsampling based or delete-d jackknife based t-statistic under near exogeneity when instruments are strong. We choose the block size

$b = \{8, 12, 16, 24, 32\}$. The simulation shows that under various choices of the block size the t_S -statistic always has a larger size distortion under near exogeneity than the regular asymptotics listed in Table 1. Theorem 2 can interpret the difference in finite sample performances. The limiting distribution of the t_S -statistic is a normal distribution with zero mean and whose variance is less than 1 . Compared to the standard normal distribution, the data-dependent critical values obtained from Theorem 2 are asymptotically less than the one from standard normal distribution, so we can observe large rejections under near exogeneity for the t_S -statistic.

Table 3 shows the size property of the Anderson-Rubin test under near exogeneity. When the correlation between instruments and structural errors is zero, the Anderson-Rubin test works very well. As Theorem 3 predicts, the Anderson-Rubin test has a large size distortion under near exogeneity. When the correlation between instruments and structural errors is 0.10 , the actual size can be 22.6% . When the correlation is 0.15 , the actual size can be 34.4% . Table 3 shows that we cannot use the Anderson-Rubin test based on chi-square critical values under near exogeneity.

Table 4 compares the size property of the delete- d jackknife based Anderson-Rubin test under near exogeneity for various choices of the block size b . We choose the block size $b = \{12, 14, 16, 18, 20, 22, 24, 28, 30, 32\}$. We also do simulations for $b = \{6, 8, 10\}$ which shows size distortion in finite samples; see Caner (2005) and Fang (2005). Since in practice the delete- d jackknife uses moderately sized blocks, we report results with $b = \{12, 14, 16, 18, 20, 22, 24, 28, 30, 32\}$. There are two parts in Table 4 which show the results under strong instruments and weak instruments respectively. We can observe that the results in two parts are very similar because the quality of instruments cannot affect the behavior of the test. When the block size is large, for example, $b = 32$, the AR_S -statistic is very conservative. For example, the actual sizes are 0.3 , 1.3 and 2.3 respectively when the correlation between instruments and structural errors are 0 , 0.1 and 0.15 . When the block size shrinks, we can observe more rejections. When $b = 16$, the actual size is 7.4% when

$covZ_i u_i = 0.10$ and the actual size is 11.8% when $covZ_i u_i = 0.15$. Note that when $b = 16$, $d = 48$ and $\lambda = \frac{3}{4}$. When b is smaller than 16, we can observe overrejection. For example, when $b = 12$, the actual size is 0.24 when the correlation between instruments and structural errors is 0.15. Table 4 also shows that the AR_S-statistic is very undersized when $covZ_i u_i = 0$ but by the choice of the right block size it works much better than the regular Anderson-Rubin test when the degree of near exogeneity is between 0.10 and 0.15. We suggest $b = 16$ in practice based on the simulation results summarized in Table 4. When $b = 16$, $\gamma = \frac{3}{4}$ which is also suggested by Wu (1989). When we begin with $b = 32$, from Table 4 we know that it is very conservative. When we increase the block size, we observe more rejections. The block sizes 16, 18, 20 and 22 provide good size performance. When the block size is larger than 16, we observe less size distortion.

2.4 CONCLUSIONS

This paper examines the size property of the t-statistic when there exists a slight violation of the exogeneity assumption in a linear simultaneous equations model. We show a large size distortion of the t-statistic in finite samples under near exogeneity. The subsampling based or the delete-d jackknife based t-statistic works even worse in finite samples than the regular asymptotics because the resampling procedures cannot catch the drift term from near exogeneity but produce smaller variance than the standard normal distribution. We propose the subsampling based or the delete-d jackknife based Anderson-Rubin test under near exogeneity. We find that the sizes of the test are liberal by the choice of the block size. We propose $\gamma = \frac{3}{4}$ to choose the block size in practice. Since the actual size increases as the block size shrinks, we are more confident to reject the null hypothesis in a large block size than that in a small block size, and we are also more confident unable to reject the null hypothesis in a small block size than that in a large block size. We use our method to reexamine the estimates in Acemoglu, Johnson and

Robinson (2001). We find that in most cases there exists strong evidence that the TSLS estimator by using the mortality rate as the instrument is not significant away from zero.

2.5 APPENDIX

2.5.1 Appendix 1

In the beginning of this appendix, we first list near exogeneity assumption and some moment conditions that are required to obtain the theorems in the paper. Assumptions 1 and 2 are sufficient for Lemma 1, Theorem 1 and Theorem 3. Assumptions 1 and 3 are sufficient for Theorem 2 and Theorem 4.

Assumption 1 Near Exogeneity $E[Z_i' u_i] = C/\sqrt{N}$, where C is a fixed $K \times 1$ vector.

Assumption 2: The following limits hold jointly when the sample size N converges to infinity:

(a) $(u'u/N, V'u/N, V'V/N) \xrightarrow{p} (\sigma_u^2, \Sigma_{Vu}, \Sigma_{VV})$, where σ_u^2 , Σ_{Vu} and Σ_{VV} are respectively a 1×1 scalar, an $m \times 1$ vector and an $m \times m$ matrix.

(b) $Z'Z/N \xrightarrow{p} Q_{ZZ}$ where Q_{ZZ} is a positive definite $K \times K$ matrix.

(c) $(Z'u/\sqrt{N}, Z'V/\sqrt{N}) \xrightarrow{d} (\widetilde{\Psi}_{Zu}, \Psi_{ZV})$, and

$$\begin{pmatrix} \widetilde{\Psi}_{Zu} \\ vec\Psi_{ZV} \end{pmatrix} \sim N\left(\begin{pmatrix} C \\ 0 \end{pmatrix}, \Sigma \otimes Q \right) \text{, where } \Sigma = \begin{pmatrix} \sigma_u^2 & \Sigma_{Vu}' \\ \Sigma_{Vu} & \Sigma_{VV} \end{pmatrix}.$$

These convergences in Assumption 2 are not primitive assumptions but hold under weak primitive conditions. Parts (a) and (b) follow from the weak law of large numbers, and Part (c) follows from triangular arrays central limit theorem. Instead of a mean zero normal distribution in Staiger and Stock (1997), the $\widetilde{\Psi}_{Zu}$ in (c) is a normal distribution with nonzero mean, which is a drift term C coming from the near exogeneity assumption. For any independent sequence $Z_i' u_i$, if $E[Z_i' u_i]^{2+\delta} < \Delta < \infty$ for some $\delta > 0$ for all $i = 1, 2, 3, \ldots, N$, then Liapunov's theorem leads to the limiting results in (c); see Davidson (1994).

Assumption 3: Define

$$\sigma_b = E(u_b' u_b / b)$$

and

$$Q_b = E(Z_b' Z_b / b)$$

Assume the following conditions hold jointly for $\delta > 0$,

(a) $E|z_{b,i} u_b|^{2+\delta} < \Delta_1 < \infty$ for all $b < N$ and all $1 \leq i \leq K$

(b) $E|z_{b,i} z_{b,j}|^{1+\delta} < \Delta_2 < \infty$ for all $b < N$ and all $1 \leq i, j \leq K$

(c) $E\left|u_b^2\right|^{1+\delta} < \Delta_3 < \infty$ for all $b < N$

(d) $\sigma_b \to \sigma_u^2 > 0$ uniformly as $b \to \infty$

(e) $Q_b \to Q_{ZZ}$ uniformly and uniformly positive definite as $b \to \infty$

Lemma 1 Suppose that Assumption 1 and 2 hold for a linear simultaneous equations model, then the TSLS estimator $\hat{\beta}_{TSLS}$ *is consistent and*

$$\sqrt{N}(\hat{\beta}_{TSLS} - \beta_0) \xrightarrow{d} N((\Pi' Q_{ZZ} \Pi)^{-1} \Pi' C, \sigma_u^2 (\Pi' Q_{ZZ} \Pi)^{-1})$$

where $u'u/N \to E(u_i^2) = \sigma_u^2$, $Z'Z/N \to E(Z_i' Z_i) = Q_{ZZ}$.

The proof is given in the Appendix 2.

Lemma 1 summarizes the limiting results of the TSLS estimator under near exogeneity. The reason why we can obtain a consistent estimator under near exogeneity is because the correlation between instruments and structural errors shrinks toward zero asymptotically. When $C = 0$, we can obtain the regular results of the TSLS estimator under the orthogonality condition. Instead of a normal distribution with a zero mean, near exogeneity can shift the distribution away from the zero mean. The nonzero mean depends on an unknown local to zero parameter C which is impossible to be estimated consistently (Andrews, 2000).

Theorem 1 Suppose that Assumption 1 and 2 hold for a linear simultaneous equations model, then

$$t \xrightarrow{d} N[\sigma_u^{-1}(\Pi' Q_{ZZ}\Pi)^{-1/2}\Pi' C, 1]$$

where σ_u *is the square root of* σ_u^2.

The proof is given in the Appendix 2.

Next, consider whether the resampling versions of the t-statistic can correct the size problem under near exogeneity in large samples. Denote by t_S the delete-d jackknife based t-statistic,

$$t_S = \frac{\hat{\beta}_{S,TSLS} - \hat{\beta}_{TSLS}}{\sqrt{avar(\hat{\beta}_{S,TSLS})}}$$

where $\hat{\beta}_{S,TSLS}$ is the delete-d jackknife based estimator and $avar(\hat{\beta}_{S,TSLS})$ is the estimated variance of the corresponding TSLS estimators. The delete-d jackknife based t-statistic using 10% as the nominal size can be implemented by the following steps:

Step 1: Randomly choose d observations from the sample without replacement, where $d < N$;

Step 2: Given d, the block size $b = N - d$. Compute the TSLS estimator and the corresponding estimated variance by using sample observations not deleted by Step 1, and then compute the delete-d jackknife based t-statistic;

Step 3: Replicate Step 1 & 2 by at least 1000 times and sort these computed delete- d jackknife based t -statistics; Use the 90% quantile as the data-dependent critical value;

Step 4: The delete- d jackknife based t -statistic rejects the null hypothesis when the sample value of the t -statistic is larger than the data-dependent critical value which is found in Step 3.

In order to construct asymptotic results, the delete- d jackknife requires that $d = \gamma N$, where $0 < \gamma < 1$ and N grows to infinity (Shao and Wu, 1989). The following theorem provides the limiting results of the t_S -statistic under near exogeneity.

Theorem 2 Suppose that Assumption 1 and 3 hold for a linear simultaneous equations model, then

$$t_S \xrightarrow{d} N[0, 2 - \gamma - 2\sqrt{1-\gamma}]$$

where $\gamma = d/N$ *and* $0 < \gamma < 1$.

The proof is given in the Appendix 2. Theorem 2 summarizes limiting results of the delete- d jackknife based t -statistic under near exogeneity. We obtain the delete- d jackknife based t -statistic when $0 < \gamma < 1$. The limiting distribution defined above is obviously not the limiting distribution of the t-statistic under near exogeneity, and there is no drift correction. Since the variance of the distribution above is less than 1 , we expect a larger size distortion under near exogeneity than the regular standard normal asymptotics.

Theorem 3 Suppose that Assumption 1 and 2 hold for a linear simultaneous equations model, then under the null hypothesis of $\beta = \beta_0$,

$$AR(\beta_0) \xrightarrow{d} \chi^2_K(\varsigma)$$

where $\chi^2_K(\varsigma)$ *is a noncentral chi-square distribution with degrees of freedom* K *and the*

noncentral parameter $\varsigma = C'\Omega^{-1}C$ *, and* $\Omega = \sigma_u^2 \otimes Q_{ZZ}$.

The proof of the theorem is given by Fang (2005) in a linear simultaneous equations model. Caner (2005) provides the proof of an Anderson-Rubin type test in a generalized empirical likelihood model.

Theorem 4 Suppose that Assumption 1 and 3 hold for a linear simultaneous equations model, then under the null hypothesis of $\beta = \beta_0$ *,*

$$AR_S(\beta_0) \xrightarrow{d} \chi_K^2(\tilde{\varsigma})$$

where $\chi_K^2(\tilde{\varsigma})$ *is a noncentral chi-square distribution with degrees of freedom* K *and the noncentral parameter* $\tilde{\varsigma} = (1-\gamma)C'\Omega^{-1}C$ *, and* $\gamma = d/N$.

The proof of the theorem is given by Fang (2005) in a linear simultaneous equations model. Caner (2005) provides the proof of an Anderson-Rubin type test in a generalized empirical likelihood model.

2.5.2 Appendix 2

Proof of Lemma 1 The TSLS estimator is s defined as,

$$\hat{\beta}_{TSLS} = (Y'P_ZY)^{-1}(Y'P_Zy).$$

So we have

$$\sqrt{N}(\hat{\beta}_{TSLS} - \beta_0)$$
$$= [(\frac{Y'Z}{N})(\frac{Z'Z}{N})^{-1}(\frac{Z'Y}{N})]^{-1}[(\frac{Y'Z}{N})(\frac{Z'Z}{N})^{-1}(\frac{Z'u}{\sqrt{N}})]$$

By Assumption 2, we can obtain that

$$[(\frac{Y'Z}{N})(\frac{Z'Z}{N})^{-1}(\frac{Z'Y}{N})]^{-1}$$
$$\xrightarrow{p} (\Pi' Q_{ZZ}\Pi)^{-1}$$

Now, we consider

$$\frac{Z'u}{\sqrt{N}} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[Z_i'u_i - E(Z_i'u_i)] + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}E(Z_i'u_i)$$

By the triangular array central limit theorem, we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[Z_i'u_i - E(Z_i'u_i)] \xrightarrow{d} N[0, \sigma_u^2 Q_{ZZ}].$$

By the triangular array weak law of large number and Assumption 1, we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}E(Z_i'u_i) \xrightarrow{p} C.$$

Combining above results, we obtain

$$\frac{Z'u}{\sqrt{N}} \xrightarrow{d} N[C, \sigma_u^2 Q_{ZZ}]$$

Then the result in the lemma follows. *Q.E.D.*

Proof of Theorem 1 The result in the theorem directly follows from Lemma 1. *Q.E.D.*

Proof of Theorem 2 A resampling based t-statistic is defined as,

$$t_S = \frac{\hat{\beta}_{S,TSLS} - \hat{\beta}_{TSLS}}{\sqrt{avar(\hat{\beta}_{S,TSLS})}}$$

where

$$avar(\hat{\beta}_{S,TSLS}) = \hat{\sigma}^2_{u,b}[(Y_b'Z_b)(Z_b'Z_b)^{-1}(Z_b'Y_b)]^{-1},$$

and

$$\hat{\sigma}^2_{u,b} = (y_b - Y_b \hat{\beta}_{S,TSLS})'(y_b - Y_b \hat{\beta}_{S,TSLS})/(b - K - m).$$

By Assumption 3 and weak law of large number (Fang, 2005), we have

$$\hat{\sigma}^2_{u,b} \xrightarrow{p} \sigma^2_u$$

and

$$[(\frac{Y_b'Z_b}{b})(\frac{Z_b'Z_b}{b})^{-1}(\frac{Z_b'Y_b}{b})]^{-1}$$
$$\xrightarrow{p} (\Pi' Q_{ZZ}\Pi)^{-1}.$$

The t_S -statistic can be rewritten as

$$t_S = \frac{(\hat{\beta}_{S,TSLS} - \beta_0) - (\hat{\beta}_{TSLS} - \beta_0)}{\sqrt{avar(\hat{\beta}_{S,TSLS})}}$$

Consider the first term in the above equation,

$$\sqrt{b}(\hat{\beta}_{S,TSLS} - \beta_0)$$
$$= [(\frac{Y_b'Z_b}{b})(\frac{Z_b'Z_b}{b})^{-1}(\frac{Z_b'Y_b}{b})]^{-1}[(\frac{Y_b'Z_b}{b})(\frac{Z_b'Z_b}{b})^{-1}(\frac{Z_b'u_b}{\sqrt{b}})]$$

We know that by Assumption 3 and the triangular array central limit theorem,

$$\frac{Z_b'u_b}{\sqrt{b}} = \frac{1}{\sqrt{b}}\sum_{i=1}^{b}[Z_{b,i}u_{b,i} - E(Z_{b,i}u_{b,i})] + \frac{1}{\sqrt{b}}\sum_{i=1}^{b}E(Z_{b,i}u_{b,i})$$
$$\xrightarrow{d} N[0, \sigma_u^2 Q_{ZZ}] + (\sqrt{1-\lambda})C$$
$$= N[(\sqrt{1-\lambda})C, \sigma_u^2 Q_{ZZ}].$$

So we have

$$\frac{\sqrt{b}\,(\hat{\beta}_{S,TSLS} - \beta_0)}{\sqrt{\sigma_u^2(\Pi' Q_{ZZ}\Pi)^{-1}}} \xrightarrow{d} N[\delta_C, 1]$$

where

$$\delta_C = \sigma_u(\Pi' Q_{ZZ}\Pi)^{-1/2}\Pi'(\sqrt{1-\lambda})C$$

By the similar method, noting that $\sqrt{b} = \sqrt{1-\gamma} \times \sqrt{N}$ we can obtain that

$$\frac{\sqrt{b}\,(\hat{\beta}_{TSLS} - \beta_0)}{\sqrt{\sigma_u^2(\Pi' Q_{ZZ}\Pi)^{-1}}} \xrightarrow{d} N[\delta_C, (1-\gamma)]$$

Then the result in the theorem follows. *Q.E.D.*

2.5.3 Appendix 3

Table 2-1: Sizes of the t -statistic ($\Pi = 1$)

	$CovZ_i'u_i =$	$CovZ_i'u_i =$	$CovZ_i'u_i =$
Actual Size	9.4	26.4	39.1

Note: The data generating process of the simulation is based on Λ *with strong instruments (* $\Pi = 1$ *). The sample size is* $N = 64$ *and the nominal size is* 10% *.*

Table 2-2: Sizes of the t_S -statistic ($\Pi = 1$)

	$CovZ_i'u_i =$	$CovZ_i'u_i =$	$CovZ_i'u_i =$
$b = 8$	20.1	32.8	42.5
$b = 12$	21.5	30.4	45.8
$b = 16$	22.1	29.3	54.6
$b = 24$	23.3	35.1	48.4
$b = 32$	27.1	41.6	54.7

Note: The data generating process of the simulation is based on Λ *with strong instruments (* $\Pi = 1$ *). The sample size is* $N = 64$ *and the nominal size is* 10% *. The* t_S *-statistic is defined in (140) and* b *represents the block size. We compute the actual sizes of the* t_S *-statistic when* $b = \{8, 12, 16, 24, 32\}$ *.*

Table 2-3: Sizes of the Anderson-Rubin test

	$CovZ_i'u_i =$	$CovZ_i'u_i =$	$CovZ_i'u_i =$
$\Pi = 1$ (strong instruments)			
Actual size	9.7	21.8	33.5
$\Pi = 0.1$ (weak instruments)			
Actual size	10.1	22.6	34.4
$\Pi = 0$ (nonidentification)			
Actual size	9.3	22.2	33.6

Note: The data generating process of the simulation is based on Λ. Π *represents the quality of instruments. The sample size is* $N = 64$ *and the nominal size is* 10% *. The Anderson-Rubin test is computed as defined in (160).*

Table 2-4: Sizes of the delete- d jackknife based Anderson-Rubin test

	$Cov Z_i' u_i =$	$Cov Z_i' u_i =$	$Cov Z_i' u_i =$
Part A: $\Pi = 1$ (strong instruments)			
$b = 12$	3.8	10.1	20.2
$b = 14$	3.1	8.5	16.1
$b = 16$	2.0	7.4	11.8
$b = 18$	2.1	6.0	13.5
$b = 20$	1.4	5.3	10.4
$b = 22$	0.7	4.0	9.5
$b = 24$	1.1	3.3	8.4
$b = 26$	0.8	2.5	6.0
$b = 28$	0.2	1.9	3.7
$b = 30$	0.5	1.5	3.2
$b = 32$	0.3	1.3	2.3
Part B: $\Pi = 0.1$ (weak instruments)			
$b = 12$	3.2	9.4	19.4
$b = 14$	3.1	9.7	17.3
$b = 16$	2.0	7.2	13.0
$b = 18$	2.3	6.7	9.4
$b = 20$	1.4	3.8	10.7
$b = 22$	1.6	3.1	7.0
$b = 24$	0.8	3.4	7.0
$b = 26$	1.0	3.2	5.2
$b = 28$	0.7	1.7	3.6
$b = 30$	0.6	0.8	3.3
$b = 32$	0.1	1.2	2.6

Note: The data generating process of the simulation is based on Λ. Π represents the quality of instruments. The sample size is $N = 64$ and the nominal size is 10%. The b represents the block size and $b = N - d$. We compute the delete- d jackknife based Anderson-Rubin test defined in Section 3 with various blocks.

Table 2-5: Baseline regressions

	(1)	(2)	(3)	(4)
Table 5a: Two-Stage Least Squares				
Average protection against expropriation risk 1985-1995	0.94	1.00	0.98	1.10
	(0.16)	(0.22)	(0.30)	(0.46)
Latitude		–0.65		–1.20
		(1.34)		(1.8)
Asia dummy			–0.92	–1.10
			(0.40)	(0.52)
Africa dummy			–0.46	–0.44
			(0.36)	(0.42)
"Other" continent dummy			–0.94	–0.99
			(0.85)	(1.0)
Table 5b: t -statistic and AR_S -statistic				
t -statistic and p -values				
Average protection against expropriation risk 1985-1995	5.875	4.545	3.266	2.391
	[< 0.000]	[< 0.000]	[0.001]	[0.017]
delete- d jackknife based AR_S and p -values				
Average protection against expropriation risk 1985-1995	12.812	6.847	0.446	0.635
	[0.078]	[0.146]	[0.272]	[0.280]

Note: The dependent variable in columns (1)-(4) is $\log$ *GDP per capita in 1995. Table 5a reports two-stage least squares estimates of institutions, instrumenting for protection against expropriation risk using* $\log$ *settler mortality. The results in Table 5a are replicated from Acemoglu, Johnson, and Robinson (2001, p1386). The numbers in parentheses are the standard errors of coefficient estimators. Table 5b reports values of* t *-statistic and delete-* d *jackknife based Anderson-Rubin test respectively. The numbers in brackets are their associated* p *-values. We use* $b = 16$ *to compute the delete-* d *jackknife based Anderson-Rubin test.*

Table 2-6: Robustness-1

	(1)	(2)	(3)	(4)
Table 6a: Two-Stage Least Squares				
Average protection against	1.10	1.16	1.10	1.20
expropriation risk 1985-1995	(0.22)	(0.34)	(0.19)	(0.29)
Latitude		-0.75		-1.10
		(1.70)		(1.56)
British colonial dummy	-0.78	-0.80		
	(0.35)	(0.39)		
French colonial dummy	-0.12	-0.06		
	(0.35)	(0.42)		
French legal origin dummy			0.89	-0.96
			(0.32)	(0.39)
Table 6b: t -statistic and AR_S -statistic				
t -statistic and p -values				
Average protection against	5.00	3.441	5.789	4.137
expropriation risk 1985-1995	[<0.000]	[<0.000]	[<0.000]	[<0.000]
delete- d jackknife based AR_S and p -values				
Average protection against	0.796	2.116	3.234	3.096
expropriation risk 1985-1995	[0.357]	[0.221]	[0.174]	[0.198]

Note: The dependent variable in columns (1)-(4) is $\log$ *GDP per capita in 1995. Table 6a reports two-stage least squares estimates of institutions, which are replicated from Acemoglu, Johnson, and Robinson (2001, p1389). The numbers in parentheses are the standard errors of coefficient estimators. Table 6b reports values of* t *-statistic and delete-* d *jackknife based Anderson-Rubin test respectively. The numbers in brackets are their associated* p *-values. We use* $b = 16$ *to compute the delete-* d *jackknife based Anderson-Rubin test.*

Table 2-7: Robustness-2

	(1)	(2)	(3)	(4)	(5)
Table 7a: Two-Stage Least Squares					
Average protection against expropriation risk 1985-1995	0.92	1.00	1.10	0.74	0.79
	(0.15)	(0.25)	(0.29)	(0.13)	(0.17)
Latitude		–0.94	–1.70		–0.89
		(1.50)	(1.6)		(1.00)
British colonial dummy					
French colonial dummy			0.02		
			(0.69)		
French legal origin dummy			0.51		
			(0.69)		
p -values for religion variables	[0.001]	[0.004]	[0.42]		
Ethnolinguistic fragmentation				–1.00	–1.10
				[0.32]	[0.34]
Table 7b: t -statistic and AR_S -statistic					
t -statistic and p -values					
Average protection against expropriation risk 1985-1995	6.133	4.00	3.793	5.692	4.647
	[< 0.000]	[< 0.000]	[< 0.000]	[< 0.000]	[< 0.000]
delete- d jackknife based AR_S and p -values					
Average protection against expropriation risk 1985-1995	1.054	0.233	1.278	0.187	0.142
	[0.123]	[0.373]	[0.050]	[0.450]	[0.576]

Note: The results in Table 7a are replicated from Acemoglu, Johnson, and Robinson (2001, p1389 and p1390). The religion variables are percentage of population that are Cathoics, Muslims, and "other" religions. Protestant is the base case. Table 7b reports values of t -statistic and delete- d jackknife based Anderson-Rubin test respectively. The numbers in brackets are their associated p -values. We use $b = 16$ to compute the delete- d jackknife based AR_S test.

Table 2-8: Robustness-3

	(1)	(2)	(3)	(4)	(5)	(6)
Table 8a: Two-Stage Least Squares						
Average protection against expropriation risk 1985-1995	0.69	0.72	0.63	0.68	0.55	0.56
	(0.25)	(0.30)	(0.28)	(0.34)	(0.24)	(0.31)
Latitude		–0.57		–0.53		–0.1
		(1.04)		(0.97)		(0.95)
Malaria in 1994	–0.57	–0.60				
	(0.47)	(0.47)				
Life expectancy			0.03	0.03		
			(0.02)	(0.02)		
Infant mortality					–0.01	–0.01
					(0.005)	(0.006)
Table 8b: t -statistic and AR_S -statistic						
t -statistic and p -values						
Average protection against expropriation risk 1985-1995	2.76	2.40	2.25	2.00	2.291	1.806
	[0.006]	[0.016]	[0.024]	[0.046]	[0.022]	[0.071]
delete- d jackknife based AR_S and p -values						
Average protection against expropriation risk 1985-1995	0.404	0.031	4.090	4.013	0.891	0.432
	[0.291]	[0.648]	[0.279]	[0.235]	[0.171]	[0.255]

Note: The dependent variable in columns (1)-(6) is $\log$ *GDP per capita in 1995. Table 8a reports two-stage least squares estimates of institutions, instrumenting for protection against expropriation risk using* $\log$ *settler mortality. The results in Table 8a are replicated from Acemoglu, Johnson, and Robinson (2001, p1392). The numbers in parentheses are the standard errors of coefficient estimators. Table 8b reports values of* t *-statistic and delete-* d *jackknife based Anderson-Rubin test respectively. The numbers in brackets are their associated* p *-values. We use* $b = 16$ *to compute the delete-* d *jackknife based Anderson-Rubin test.*

Table 2-9: Baseline regressions under various block sizes

	(1)	(2)	(3)	(4)
Table 9a: Two-Stage Least Squares				
Average protection against	0.94	1.00	0.98	1.10
expropriation risk 1985-1995	(0.16)	(0.22)	(0.30)	(0.46)
Latitude		–0.65		–1.20
		(1.34)		(1.8)
Asia dummy			–0.92	–1.10
			(0.40)	(0.52)
Africa dummy			–0.46	–0.44
			(0.36)	(0.42)
"Other" continent dummy			–0.94	–0.99
			(0.85)	(1.0)
Table 9b: delete- d jackknife based AR s test under various b 's				
$AR(\beta_0)$	12.812	6.847	0.446	0.635
$b = 32$	[0.156]	[0.211]	[0.290]	[0.407]
$b = 30$	[0.157]	[0.206]	[0.276]	[0.397]
$b = 28$	[0.156]	[0.199]	[0.309]	[0.387]
$b = 24$	[0.202]	[0.198]	[0.284]	[0.351]
$b = 20$	[0.133]	[0.176]	[0.277]	[0.336]
$b = 14$	[0.097]	[0.117]	[0.247]	[0.265]
$b = 12$	[0.074]	[0.073]	[0.238]	[0.217]

Note: The dependent variable in columns (1)-(4) is $\log$ *GDP per capita in 1995. Table 9a reports two-stage least squares estimates of institutions, instrumenting for protection against expropriation risk using* $\log$ *settler mortality. The results in Table 9a are replicated from Acemoglu, Johnson, and Robinson (2001, p1386). The numbers in parentheses are the standard errors of coefficient estimators. Table 9b reports* p *-values (in brackets) of the delete-* d *jackknife based Anderson-Rubin test under various block sizes.* $AR(\beta_0)$ *represents the sample value of the Anderson-Rubin test when* $\beta_0 = 0$ *.*

Table 2-10: Robustness-1 under various block sizes

	(1)	(2)	(3)	(4)
Table 10a: Two-Stage Least Squares				
Average protection against	1.10	1.16	1.10	1.20
expropriation risk 1985-1995	(0.22)	(0.34)	(0.19)	(0.29)
Latitude		–0.75		–1.10
		(1.70)		(1.56)
British colonial dummy	–0.78	–0.80		
	(0.35)	(0.39)		
French colonial dummy	–0.12	–0.06		
	(0.35)	(0.42)		
French legal origin dummy			0.89	–0.96
			(0.32)	(0.39)
Table 10b: delete- d jackknife based AR s test under various b 's				
$AR(\beta_0)$	0.796	2.116	3.234	3.096
$b = 32$	[0.279]	[0.233]	[0.340]	[0.392]
$b = 30$	[0.298]	[0.239]	[0.313]	[0.390]
$b = 28$	[0.297]	[0.250]	[0.300]	[0.394]
$b = 24$	[0.293]	[0.225]	[0.260]	[0.333]
$b = 20$	[0.308]	[0.249]	[0.238]	[0.250]
$b = 14$	[0.325]	[0.235]	[0.136]	[0.149]
$b = 12$	[0.322]	[0.197]	[0.121]	[0.114]

Note: The dependent variable in columns (1)-(4) is $\log$ *GDP per capita in 1995. Table 10a reports two-stage least squares estimates of institutions, instrumenting for protection against expropriation risk using* $\log$ *settler mortality. The results in Table 10a are replicated from Acemoglu, Johnson, and Robinson (2001, p1389). The numbers in parentheses are the standard errors of coefficient estimators. Table 10b reports* p *-values (in brackets) of the delete-* d *jackknife based Anderson-Rubin test under various block sizes.* $AR(\beta_0)$ *represents the sample value of the Anderson-Rubin test when* $\beta_0 = 0$ *.*

Table 2-11: Robustness-2 under various block sizes

	(1)	(2)	(3)	(4)	(5)
Table 11a: Two-Stage Least Squares					
Average protection against expropriation risk 1985-1995	0.92	1.00	1.10	0.74	0.79
	(0.15)	(0.25)	(0.29)	(0.13)	(0.17)
Latitude		–0.94	–1.70		–0.89
		(1.50)	(1.6)		(1.00)
British colonial dummy					
French colonial dummy			0.02		
			(0.69)		
French legal origin dummy			0.51		
			(0.69)		
p -values for religion variables	[0.001]	[0.004]	[0.42]		
Ethnolinguistic fragmentation				–1.00	–1.10
				[0.32]	[0.34]
Table 11b: delete- d jackknife based AR s test under various b 's					
$AR(\beta_0)$	1.054	0.233	1.278	0.187	0.142
$b = 32$	[0.151]	[0.369]	[0.164]	[0.423]	[0.562]
$b = 30$	[0.155]	[0.371]	[0.143]	[0.459]	[0.594]
$b = 28$	[0.148]	[0.421]	[0.124]	[0.455]	[0.566]
$b = 24$	[0.144]	[0.369]	[0.082]	[0.468]	[0.585]
$b = 20$	[0.118]	[0.391]	[0.066]	[0.454]	[0.575]
$b = 14$	[0.104]	[0.373]	[0.059]	[0.473]	[0.568]
$b = 12$	[0.120]	[0.311]	[0.025]	[0.463]	[0.582]

Note: The results in Table 11a are replicated from Acemoglu, Johnson, and Robinson (2001, p1389 and p1390). Table 11b reports p -values (in brackets) of the delete- d jackknife based Anderson-Rubin test under various block sizes. $AR(\beta_0)$ represents the sample value of the Anderson-Rubin test when $\beta_0 = 0$.

Table 2-12: Robustness-3 under various block sizes

	(1)	(2)	(3)	(4)	(5)	(6)
Table 12a: Two-Stage Least Squares						
Average protection against	0.69	0.72	0.63	0.68	0.55	0.56
expropriation risk 1985-1995	(0.25)	(0.30)	(0.28)	(0.34)	(0.24)	(0.31)
Latitude		–0.57		–0.53		–0.1
		(1.04)		(0.97)		(0.95)
Malaria in 1994	–0.57	–0.60				
	(0.47)	(0.47)				
Life expectancy			0.03	0.03		
			(0.02)	(0.02)		
Infant mortality					–0.01	–0.01
					(0.005)	(0.006)
Table 12b: delete-d jackknife based ARs test under various b's						
$AR(\beta_0)$	0.404	0.031	4.090	4.013	0.891	0.432
$b = 32$	[0.243]	[0.645]	[0.393]	[0.381]	[0.219]	[0.383]
$b = 30$	[0.220]	[0.655]	[0.366]	[0.360]	[0.209]	[0.340]
$b = 28$	[0.217]	[0.668]	[0.354]	[0.336]	[0.232]	[0.340]
$b = 24$	[0.259]	[0.642]	[0.327]	[0.329]	[0.197]	[0.319]
$b = 20$	[0.257]	[0.632]	[0.331]	[0.275]	[0.193]	[0.300]
$b = 14$	[0.281]	[0.656]	[0.228]	[0.181]	[0.152]	[0.246]
$b = 12$	[0.219]	[0.636]	[0.180]	[0.119]	[0.119]	[0.216]

Note: The dependent variable in columns (1)-(6) is $\log$ *GDP per capita in 1995. Table 12a reports two-stage least squares estimates of institutions, instrumenting for protection against expropriation risk using* $\log$ *settler mortality. The results in Table 12a are replicated from Acemoglu, Johnson, and Robinson (2001, p1392). The numbers in parentheses are the standard errors of coefficient estimators. Table 12b reports* p *-values (in brackets) of the delete-*d *jackknife based Anderson-Rubin test under various block sizes.* $AR(\beta_0)$ *represents the sample value of the Anderson-Rubin test when* $\beta_0 = 0$ *.*

3.0 INSTRUMENTAL VARIABLES REGRESSION WITH WEAK INSTRUMENTS AND NEAR EXOGENEITY

The linear instrumental variables (IV) regression has wide applications in empirical studies. In the linear simultaneous equations model, to justify the IV method, it should satisfy two important criteria. One is called "instrument exogeneity", which means that instruments excluded from the structural equation should be uncorrelated with the structural errors. The other is called "instrument relevance", which requires that instruments should be strongly correlated with the endogenous explanatory variables. Finding valid instruments to satisfy the two criteria is not an easy job. For example, the problem of weak instruments, which means that instruments are weakly correlated with endogenous explanatory variables, has recently received a lot of attention by both theoretical and empirical researchers (Stock, Wright and Yogo, 2002). If instruments are weak, then the limits of the sample distributions of the two-stage least square (TSLS) estimator and the limited information maximum likelihood (LIML) estimator are in general nonstandard, and the resulting conventional hypothesis tests and confidence intervals are not reliable.

In an influential empirical study of labor economics, Angrist and Krueger (1991) use quarter of birth as an instrument for education to estimate the impact of compulsory schooling laws on earnings. They argue that children's quarter of birth is random, so it is uncorrelated with ability and should be exogenous. Because of compulsory laws, average education is generally longer for children born near the end of the year than for children born early in the year, which means that quarter of birth is correlated with educational attainment. Based on large samples (329,000 observations or more) from the U.S. census, they estimate the return to education by the TSLS procedure, using as instruments for education a set of three quarter-of-birth dummies interacted with fifty state-of-birth dummies and nine year-of-birth dummies respectively. But Bound, Jaeger and Baker (1995) point out that the instruments used in Angrist and Krueger's paper are weak and nearly exogenous in which case the resulting estimation and inference are misleading.

Many authors work on improving inference under weak instruments; see, for example, Staiger and Stock(1997), Dufour(1997), Kleibergen(2002), Moreira(2003), among others.

Instrument exogeneity is another important criterion for valid instruments. In empirical studies, the validity of instrument exogeneity is mainly based on economic reasoning. But unfortunately, it is almost impossible to control for all possible variables that might be correlated with instruments and dependent variables. As a result, the instruments might catch the effect on dependent variables through other channels. It is hard to argue that instruments are exogenous in empirical studies. For example, Acemoglu, Johnson, and Robinson (2001) estimate the effect of institutions on economic performance by using as instrument the logarithm of the European settler mortality rates. They argue that the settler mortality rate more than 100 years ago is strongly correlated with current institutions in the countries colonized by Europeans in the history. The mortality rates expected by the first European settlers determined the settlement decision and then influenced the colonization strategy: introducing "extractive states" (bad institution) or "Neo-Europes" (good institution). In a study of whether a reversal in relative incomes among the former European colonies reflects changes in the institutions resulting from European colonialism, Acemoglu, Johnson, and Robinson (2002) use data on urbanization and population density in 1500 to proxy for economic prosperity. In order to test whether population density or urbanization in 1500 affects income today only through institutions, the settler mortality rate is used as instruments again. But Glaeser, La Porta, Lopez-De-Silanes, and Shleifer (2004) argue that the settler mortality rate is not an exogenous instrument because the mortality rate might affect today's income through other channels, for example, the human capital. This is a problem of near exogeneity where the instruments are weakly correlated with the structural errors. Due to the nature that it is almost impossible to control for all possible variables that might be correlated with instruments and dependent variables of interest, the problem of near exogeneity is prevalent in empirical studies. Angrist (1990) estimates the effect of veteran status on civilian earnings by using as instruments the draft lottery numbers. But Wooldridge (2002) argues that the draft lottery numbers might be correlated with the structural errors if education is not controlled in the earnings equation. Bound, Jaeger, and Baker (1995) argue that the instruments used by Angrist and Krueger (1991) are not only weak but also suffer from near exogeneity.

This paper examines asymptotic properties of estimation and inference with the joint problem of

weak instruments and near exogeneity in a linear simultaneous equations model. Near exogeneity is modeled as a local to zero correlation between instruments and structural errors. This research is partly motivated by the argument that even a weak correlation between instruments and structural errors under weak instruments can lead to a large inconsistency in IV estimates (Bound, Jaeger, and Baker, 1995). Estimation and inference with weak instruments have received more and more attention since the paper by Angrist and Krueger (1991), and some test statistics have been developed that are robust against weak instruments. This paper is the first one to study the estimation and inference in a linear IV framework that allows weak instruments and near exogeneity to occur at the same time. Caner (2005) studies the generalized empirical likelihood estimators with near exogeneity and weak instruments.

This paper obtains the limits of the TSLS estimator and the LIML estimator with weak instruments and near exogeneity. We show that the asymptotic bias may be larger than that in Staiger and Stock (1997) where only weak instruments occur. We show that the Anderson-Rubin test (Anderson and Rubin, 1949) and the Kleibergen test (Kleibergen, 2002) which are robust against weak instruments are no longer asymptotically pivotal with near exogeneity. Using critical values from the chi-square distribution leads to a serious size distortion. Moreira (2003) develops a conditional likelihood ratio test which has a correct size with weak instruments. We show that the conditional likelihood ratio test does not work under weak instruments and near exogeneity since the conditional distribution depends upon an unknown parameter. The conditional test using critical values obtained from simulating the conditional distribution ignoring the unknown parameter cannot be *similar* in general.

To correct asymptotically the sizes of tests under weak instruments and near exogeneity, we employ the resampling based Anderson-Rubin and Kleibergen tests. We use data-dependent critical values obtained from resampling instead of those obtained from the regular chi-square distributions. We propose the delete- d jackknife based Anderson-Rubin and Kleibergen tests to correct size distortion in finite samples under weak instruments and near exogeneity.

This paper is organized as follows. Section 3.1 introduces the model and some assumptions. Section 3.2 provides limits of the TSLS estimator and the LIML estimator with weak instruments and near exogeneity. The problem of testing and inference with weak instruments and near exogeneity is analyzed in Section 3.3. Section 3.4 derives the resampling based tests. Section 3.5 discusses the size properties in finite samples by using Monte Carlo simulation, and Section 3.6

concludes. Appendix is included in Section 3.7.

3.1 THE MODEL AND ASSUMPTIONS

In this article, we consider a linear simultaneous equations model (Hausman, 1983; Phillips, 1983),

$$y = Y\beta + u$$
$$Y = Z\Pi + V$$

Instrumental Variables Regression with Weak Instruments and Near ExogeneityInstrumental Variables Regression with Weak Instruments and Near ExogeneityInstrumental Variables Regression with Weak Instruments and Near ExogeneityInstrumental Variables Regression with Weak Instruments and Near ExogeneityInstrumental Variables Regression with Weak Instruments and Near ExogeneityInstrumental Variables Regression with Weak Instruments and Near ExogeneityInstrumental Variables Regression with Weak Instruments and Near Exogeneitywhere y and Y are respectively an $N \times 1$ vector and an $N \times m$ matrix of endogenous variables, Z is an $N \times K$ matrix of instruments, u is an $N \times 1$ vector of structural errors, V is an $N \times m$ matrix of reduced form errors, and β and Π are respectively an $m \times 1$ unknown parameter vector and a $K \times m$ unknown matrix of parameters. Note that we require $K \geq m$. We are interested in estimation and inference about β with weak instruments and near exogeneity. Assumption 1 and 2 give the models with weak instruments and near exogeneity considered in this paper.

Assumption 1: $\Pi = \Pi_N = C_1/\sqrt{N}$, where C_1 is a fixed $K \times m$ matrix.

Assumption 2: $E[Z_i' u_i] = C_2/\sqrt{N}$, where C_2 is a fixed $K \times 1$ vector.

Assumption 1 benefits from Staiger and Stock (1997), which models weak instruments as local

to zero on the reduced form coefficients. This means that the instruments Z are weakly correlated with the endogenous variables Y when the sample size N tends to infinity. Assumption 1 is widely used in weak instruments literature; see Stock, Wright and Yogo (2002). Assumption 1 also includes the case of nonidentification when it allows C_1 to be a matrix of zeros. Assumption 2 models near exogeneity, which means that the instruments Z are not weakly exogenous (Engle, Hendry and Richard, 1983) and the correlation is local to zero as the sample size N grows large. Caner (2005) considers near exogeneity in nonlinear moment restrictions in generalized empirical likelihood estimators. We observe a trade-off between weak instruments and near exogeneity. As the sample size N grows large, the reduced form coefficient Π_N tends to being unidentified but instruments tend to exogeneity. Assumption 1 and 2 model the idea of Bound, Jaeger, and Baker (1995) that "if the instruments are only weakly correlated with the endogenous explanatory variable then even a weak correlation between the instruments and the error in the original equation can lead to a large inconsistency in IV estimates". However, in this paper, we emphasize not only the problem of estimation but also the problem of inference with weak instruments and near exogeneity.

In order to construct asymptotic results, we need following assumptions with respect to error terms and instruments. These assumptions are standard in the literature and can be obtained under standard moment conditions.

Assumption 3: The following limits hold jointly when the sample size N converges to infinity:

(a)

$$(u'u/N, V'u/N, V'V/N) \xrightarrow{p} (\sigma_u^2, \Sigma_{Vu}, \Sigma_{VV})$$

where σ_u , Σ_{Vu} and Σ_{VV} are respectively a 1×1 scalar, an $m \times 1$ vector and an $m \times m$ matrix.

(b)

$$Z'Z/N \xrightarrow{p} Q$$

where Q is a positive definite $K \times K$ matrix.

(c)

$(Z'u/\sqrt{N}, Z'V/\sqrt{N}) \xrightarrow{d} (\widetilde{\Psi}_{Zu}, \Psi_{ZV})$

and

$$\begin{pmatrix} \widetilde{\Psi}_{Zu} \\ vec\Psi_{ZV} \end{pmatrix} \sim N\left(\begin{pmatrix} C_2 \\ 0 \end{pmatrix}, \Sigma \otimes Q \right)$$

where

$$\Sigma = \begin{pmatrix} \sigma_u^2 & \Sigma_{Vu}' \\ \Sigma_{Vu} & \Sigma_{VV} \end{pmatrix}.$$

These convergences in Assumption 1 are not primitive assumptions but hold under weak primitive conditions. Parts (a) and (b) are taken from Staiger and Stock (1997), which follow from the weak law of large numbers. Part (c) follows from triangular array central limit theorem. The $\widetilde{\Psi}_{Zu}$ in (c) is, rather than a mean zero normal distribution in Staiger and Stock (1997), a normal distribution with nonzero mean, which is a drift term C_2 coming from the near exogeneity assumption. For any independent sequence $Z_i'u_i$, if $E[Z_i'u_i]^{2+\delta} < \Delta < \infty$ for some $\delta > 0$ for all $i = 1, 2, 3, \ldots, N$, then Liapunov's theorem leads to the limiting results in (c); see Davidson (1994).

We use the following definitions and notation in the paper. Let $\bar{Y} = (y \quad Y)$ and let I denote the identity matrix. Let $P_W = W(W'W)^{-1}W'$ a projection on a full rank matrix W and $M_W = I - P_W$ a projection on the space orthogonal to W, where W is a general $a \times b$ matrix with $a \geq b$. Let $P_W^{1/2} = (W'W)^{-1/2}W'$.

For comparability, we follow the additional definitions and notation provided by Staiger and Stock (1997). Define $\rho = \Sigma_{VV}^{-1/2\prime}\Sigma_{Vu}\sigma_u^{-1}$; $\lambda = Q^{1/2}C_1\Sigma_{VV}^{-1/2}$; $z_V = Q^{-1/2}\Psi_{ZV} \quad \Sigma_{VV}^{-1/2}$; and $z_u = Q^{-1/2}\Psi_{Zu}\sigma_u^{-1}$ where Ψ_{Zu} is a normal distribution with zero mean and variance $\Omega = \sigma_u^2 \otimes Q$. Ψ_{Zu} is the centered version of $\widetilde{\Psi}_{Zu}$, which implies $\Psi_{Zu} = \widetilde{\Psi}_{Zu} - C_2$. Note that the Ψ_{Zu} defined here is the same as the Ψ_{Zu} defined in Staiger and Stock (1997) for the model defined by (1) and (2) in their paper. Furthermore, let $v_1 = (\lambda + z_V)'(\lambda + z_V)$ and

$v_2 = (\lambda + z_V)' z_u.$

3.2 ESTIMATION: LIMITING RESULTS AND ASYMPTOTIC BIAS

In this section we derive the limits of the TSLS estimator and the LIML estimator under weak instruments and near exogeneity.

First, under Assumptions 1, 2 and 3, we show that limiting results of sample moments are different from those under weak instruments. The following lemma provides useful limiting results we need in this section.

Lemma 1 Suppose that Assumptions 1, 2 and 3 hold for a linear simultaneous equations model, then the following limits hold jointly as $N \to \infty$,

(a)

$$\{Y'u/N, Y'Y/N\} \xrightarrow{p} \{\Sigma_{Vu}, \Sigma_{VV}\};$$

(b)

$$\{P_Z^{1/2} u, P_Z^{1/2} V\} \xrightarrow{d} \{\widetilde{z_u}\sigma_u, z_V \Sigma_{VV}^{1/2}\}$$

where

$$\widetilde{z_u} = z_u + Q^{-1/2} C_2 \sigma_u^{-1}$$

and

$$(\widetilde{z_u}' \ vec(z_V)')' \xrightarrow{d} N\left[\begin{pmatrix} Q^{-1/2} C_2 \sigma_u^{-1} \\ 0 \end{pmatrix}, \bar{\Sigma} \otimes I_K\right];$$

where $\bar{\Sigma}$ *is the* $(m+1) \times (m+1)$ *matrix with* $\bar{\Sigma}_{11} = 1, \bar{\Sigma}_{22} = I_m, \bar{\Sigma}_{12} = \rho'$ *and* $\bar{\Sigma}_{21} = \rho$.

(c)

$$P_Z^{1/2} Y \xrightarrow{d} (\lambda + z_V)\Sigma_{VV}^{1/2};$$

(d)

$$\{Y'P_Z u, Y'P_Z Y, u'P_Z u\} \xrightarrow{d} \{\Sigma_{VV}^{1/2\prime} \widetilde{v_2} \sigma_u, \Sigma_{VV}^{1/2\prime} v_1 \Sigma_{VV}^{1/2}, \sigma_u^2 \widetilde{z_u}' \widetilde{z_u}$$

where $\widetilde{v_2} = (\lambda + z_V)' \widetilde{z_u}$.

All the proofs are given in the appendix.

Comparing Lemma 1 with the analogous lemmas in Staiger and Stock (1997, Lemma A1) and Wang and Zivot (1998, Lemma 1), we observe that the difference comes from the fact that $\widetilde{z_u}$ in our lemma replaces z_u in the previous lemmas in the weak instruments literature. Obviously, the drift term $Q^{-1/2}C_2\sigma_u^{-1}$ in $\widetilde{z_u}$ is the asymptotic bias from near exogeneity. As the sample size N grows large, it seems that the correlation between instruments and structural errors tends to zero and instruments achieve weak exogeneity, but the convergence rate is at the square root of the sample size N, which is slower than the case of weak exogeneity. As a result, we observe that the asymptotic bias in Lemma 1 depends on the nuisance parameter C_2 . .

Consider a linear simultaneous equations model defined by (101) and (111), the TSLS estimator of β_{TSLS} is

$$\hat{\beta}_{TSLS} = (Y'P_ZY)^{-1}(Y'P_Zy)$$

and the LIML estimator of β_{LIML} is

$$\hat{\beta}_{LIML} = [Y'(I - kM_Z)Y]^{-1}[Y'(I - kM_Z)y]$$

where k is the smallest root of the determinantal equation

$$\left|\bar{Y}'\bar{Y} - k\bar{Y}'M_Z\bar{Y}\right| = 0.$$

Let $\hat{\sigma}_u^2 = \hat{u}'\hat{u}/(N - K - m)$ where $\hat{u} = y - Y\hat{\beta}$ is the estimated error. The following theorem extends the limiting results of the general $k - class$ estimators under weak instruments in Staiger and Stock (1997) to a general case combining weak instruments with near exogeneity. Note that the most popular $k - class$ estimators are the TSLS estimator when $k = 1$ and the LIML estimator when k is defined above.

Theorem 1 Suppose that Assumptions 1, 2 and 3 hold for a linear simultaneous equations model, then the following limits hold jointly as the sample size $N \to \infty$,

(a)

$$\hat{\beta}_{TSLS} - \beta_0 \xrightarrow{d} \sigma_u \Sigma_{VV}^{-1/2} v_1^{-1} \widetilde{v_2}.$$

(b)

$$\hat{\sigma}_{TSLS} \xrightarrow{d} \sigma_u^2 S_1(b_{TSLS})$$

where

$$S_1(b_{TSLS}) = 1 - 2\rho' b_{TSLS} + b'_{TSLS} b_{TSLS}$$

and $b_{TSLS} = v_1^{-1} \widetilde{v_2}$.

(c)

$$\hat{\beta}_{LIML} - \beta_0 \xrightarrow{d} \sigma_u \Sigma_{VV}^{-1/2} (v_1 - \kappa I_m)^{-1} (\widetilde{v_2} - \kappa\rho)$$

where $N(k-1) \Rightarrow \kappa$ *and* κ *is the smallest root of the determinantal equation*

$$\left| \tilde{\Xi}_0^* - \kappa \bar{\Sigma} \right| = 0$$

and

$$\tilde{\Xi}_0^* = (\widetilde{z_u} \lambda + z_V)' (\widetilde{z_u} \lambda + z_V).$$

(d)

$$\hat{\sigma}_{LIML} \xrightarrow{d} \sigma_u^2 S_1(b_{LIML})$$

where

$$S_1(b_{LIML}) = 1 - 2\rho' b_{LIML} + b'_{LIML} b_{LIML}$$

and

$$b_{LIML} = (v_1 - \kappa I_m)^{-1} (\widetilde{v_2} - \kappa\rho).$$

The limits of the TSLS estimator and the LIML estimator in Theorem 1 under weak instruments and near exogeneity are analogous to Theorem 1 and Theorem 2 in Staiger and Stock (1997) for weak instruments. We obtain their results by replacing $\widetilde{z_u}$ and $\widetilde{v_2}$ respectively by z_u and v_2 . Note that the difference in $\widetilde{v_2}$ and v_2 comes from the difference in $\widetilde{z_u}$ and z_u , which is $Q^{-1/2} C_2 \sigma_u^{-1}$, stemming from near exogeneity. We can obtain Staiger and Stock (1997)'s result by setting $C_2 = 0$. Theorem 3.1 shows that the additional terms stemming from near exogeneity can bring larger inconsistency and asymptotic bias in the estimation of β and σ_u

than those with weak instruments only.

Consider an interesting case with strong instruments ($\Pi_N = C_1$) but having the problem of near exogeneity. In that case, both the TSLS estimator and the LIML estimator are consistent. However, when the weak instruments are weakly correlated with the structural errors, Theorem 1 shows that the inconsistency and the asymptotic bias can increase very much.

The following corollary measures the bias of the TSLS estimator relative to the OLS estimator under weak instruments and near exogeneity. Let $\hat{\beta}_{OLS}$ denote the OLS estimator of β . Let $\Sigma_{YY} = p\lim(Y'Y/N)$. Let $h = E\left[v_1^{-1}(\lambda + z_v)'z_v\right]$ and $\Delta = E\left[v_1^{-1}(\lambda + z_v)'Q^{-1/2}C_2\sigma_u^{-1}\right]$.

Corollary 1 Suppose that Assumptions 1, 2 and 3 hold for a linear simultaneous equations model, then

$$B^2 = (E\hat{\beta}_{TSLS} - \beta_0)'\Sigma_{YY}(E\hat{\beta}_{TSLS} - \beta_0)/(E\hat{\beta}_{OLS} - \beta_0)'\Sigma_{YY}(E\hat{\beta}_{OLS} - \beta_0)$$
$$\rightarrow (h\rho + \Delta)'(h\rho + \Delta)/\rho'\rho.$$

The relative squared bias B^2 depends on ρ , h and Δ . Note that the squared bias of the OLS estimator is $\rho'\rho$, which stems from the correlation between u and V . Weak instruments lead to the bias based on h . According to part (e) of Theorem 1 in Staiger and Stock (1997), since v_1 is asymptotically proportional to the Wald statistic testing $\Pi = 0$, h and then the bias becomes very large when the strength of the instruments is very poor. The Δ results from near exogeneity. We obtain that $B^2 = \rho'h'h\rho/\rho'\rho$ under weak instruments in Staiger and Stock (1997) if we set $C_2 = 0$. The additional term Δ and the cross product terms between Δ and $h\rho$ can exaggerate the squared bias under weak instruments and near exogeneity.

3.3 INFERENCE WITH NEAR EXOGENEITY

In a linear simultaneous equations model, we test $H_0 : \beta = \beta_0$ versus $H_1 : \beta \neq \beta_0$. As Staiger and Stock (1997) and others show, most of the conventional test statistics, for example, the Wald statistic, do not work under weak instruments. Given the results in the previous section, these test statistics of course do not work under weak instruments and near exogeneity. Wang

and Zivot (1998) and Zivot, Startz, and Nelson (1998) find that the limiting distributions of the LM and LR statistics based on the TSLS and the LIML estimators are bounded by a chi-square distribution with degrees of freedom K, the number of instruments. Even these conservative statistics do not work under near exogeneity because the limiting distributions now depend on the unknown nuisance parameter C_2. It is also well known that overidentification tests do not work under weak instruments. As a result, it's not clear how to construct a pretest procedure for testing the exogeneity of instruments under weak instruments.

In this section, we examine some recently developed tests robust to weak instruments. We show that none of these tests is robust to near exogeneity and weak instruments simultaneously.

We first examine the Anderson-Rubin test (Anderson and Rubin, 1949) under near exogeneity. The test is given by

$$AR(\beta_0) = (y - Y\beta_0)'P_Z(y - Y\beta_0)/\frac{1}{N-K}(y - Y\beta_0)'M_Z(y - Y\beta_0)$$

The Anderson-Rubin test is robust to weak instruments since the test itself is asymptotically pivotal, which means that the limiting distribution of the test does not depend on the nuisance parameter $\hat{\Pi}$. The test converges, under the null hypothesis of $\beta = \beta_0$ and Assumptions 1 and 3, in distribution to a chi-square distribution with degrees of freedom K, the number of instruments. Moreira (2003) shows that the Anderson-Rubin test is uniformly most powerful among the class of unbiased tests when $K = m$. But this optimal property does not hold when $K > m$. The power of the Anderson-Rubin test becomes low when the number of instruments K is large.

The following theorem summarizes the asymptotic result of the test under near exogeneity.

Theorem 2 Suppose that Assumptions 1, 2 and 3 hold for a linear simultaneous equations model, then under the null hypothesis of $\beta = \beta_0$,

$$AR(\beta_0) \xrightarrow{d} \chi^2_K(\varsigma)$$

where $\chi^2_K(\varsigma)$ *is a noncentral chi-square distribution with degrees of freedom* K *and the noncentral parameter* $\varsigma = C_2'\Omega^{-1}C_2$.

Note that $\Omega = \sigma_u^2 \otimes Q$ is the variance covariance matrix of $\widetilde{\Psi}_{Zu}$ and Ψ_{Zu} .

Theorem 2 shows that the Anderson-Rubin test is not asymptotically pivotal any more under near exogeneity. The limit of the test depends on the nuisance parameter C_2 which comes from near exogeneity. We obtain the result of the test under weak instruments by letting $C_2 = 0$. So Theorem 2 is a more general result. Theorem 2 shows that, even under the null hypothesis, the Anderson-Rubin test with near exogeneity converges in distribution to a noncentral chi-square distribution depending on unknown nuisance parameters. Near exogeneity leads to a distortion in size when we use critical values from the chi-square distribution with degrees of freedom K .

The Kleibergen test (2002) is proposed to overcome the weakness of the Anderson-Rubin test that the power becomes low under a largely overidentified model. The Kleibergen test is given by

$$K(\beta_0) = (y - Y\beta_0)' P_{\widetilde{Y}(\beta_0)}(y - Y\beta_0) / \frac{1}{N-K}(y - Y\beta_0)' M_Z (y - Y\beta_0)$$

where $\widetilde{Y}(\beta_0) = Z\widetilde{\Pi}(\beta_0)$, and

$$\widetilde{\Pi}(\beta_0) = (Z'Z)^{-1} Z' [Y - (y - Y\beta_0) S_{\epsilon V}(\beta_0) / S_{\epsilon\epsilon}(\beta_0)],$$

where

$$S_{\epsilon\epsilon}(\beta_0) = \frac{1}{N-K}(y - Y\beta_0)' M_Z (y - Y\beta_0)$$

and

$$S_{\epsilon V}(\beta_0) = \frac{1}{N-K}(y - Y\beta_0)' M_Z Y.$$

The Kleibergen test is asymptotically pivotal, and converges under H_0 to the chi-square distribution with degrees of freedom m , the number of endogenous variables. Note that $\widetilde{\Pi}(\beta_0)$ in (171) is a consistent estimator of the reduced form coefficients Π and asymptotically independent of $Z'(y - Y\beta_0)$. Moreira (2003) shows that $Z'(y - Y\beta_0)$ and $\widetilde{\Pi}(\beta_0)$ are sufficient statistics for β and Π respectively. Note that the Kleibergen test is a LM type test

statistic, which is a quadratic form of $Z'(y - Y\beta_0)$ conditional on $\widetilde{\Pi}(\beta_0)$. Because of the asymptotic independence between $\widetilde{\Pi}(\beta_0)$ and $Z'(y - Y\beta_0)$, the Kleibergen test is asymptotically pivotal when instruments are valid, weak ($\Pi_N = C_1/\sqrt{N}$) or invalid ($\Pi_N = 0$). We extend the Kleibergen test to a more general situation combining weak instruments with near exogeneity, but we find that the nice property above does not hold any more.

Denote by Ψ_{ZU} the limit of $\frac{1}{\sqrt{N}} Z'[(Y - Z\Pi) - (y - Y\beta_0)S_{\epsilon V}(\beta_0)/S_{\epsilon\epsilon}(\beta_0)]$. Let $G = \frac{1}{\sqrt{N}} Z'(y - Y\beta_0)$ and $D = \gamma(N)Z'[Y - (y - Y\beta_0)S_{\epsilon V}(\beta_0)/S_{\epsilon\epsilon}(\beta_0)]$ where $\gamma(N)$ is some scale function of the sample size N to make " D " have a valid limit. Denote by $\overline{G}$ and $\overline{D}$ respectively the limiting distributions of G and D. Note that both G and D are valued at the true value β_0. The following theorem summarizes the asymptotic results of the Kleibergen test under near exogeneity and weak instruments.

Theorem 3 Suppose that Assumptions 1, 2 and 3 hold for a linear simultaneous equations model, then under the null hypothesis of $\beta = \beta_0$,

$$K(\beta_0) \xrightarrow{d} (\zeta + \Upsilon(C_2))'(\zeta + \Upsilon(C_2))$$

where

$$\zeta \sim N(0, I_m)$$

and

$$\Upsilon(C_2) = (\overline{D}'\Omega^{-1}\overline{D})^{-1/2'}\overline{D}\Omega^{-1}C_2$$

and furthermore, $\overline{D}$ *is different when the quality of instruments varies.*

(a) When the instruments are valid ($\Pi_N = C_1$ *),*

$$\overline{D} \xrightarrow{p} QC_1;$$

(b) When the instruments are weak ($\Pi_N = C_1/\sqrt{N}$ *),*

$\bar{D} \xrightarrow{d} \Psi_{ZU} + QC_1$;

(c) When the instruments are invalid ($\Pi_N = 0$ *),*

$\bar{D} \xrightarrow{d} \Psi_{ZU}$.

Note that C_1 is defined in Assumption 1, C_2 is defined in Assumption 2, and $Q = E[Z'Z]$. Unlike the result in Kleibergen (2003) that the Kleibergen test converges to a chi-square distribution robust to the quality of instruments, Theorem 3 shows that it tends to different nonstandard distributions when the quality of instruments varies. Although $\bar{D}$ varies with the quality of instruments, when $C_2 = 0$ (no near exogeneity), the Kleibergen test is asymptotically a quadratic form of a standard normal distribution ζ conditional on Ψ_{ZU} robust to the quality of instruments. Ψ_{ZU} is defined in the appendix. Since Ψ_{ZU} is asymptotically independent of this standard normal variable ζ, the Kleibergen test converges in distribution to the chi-square distribution with degrees of freedom m. When $C_2 \neq 0$, near exogeneity leads to an asymptotic bias $\Upsilon(C_2)$. So the distribution of the Kleibergen test conditional on Ψ_{ZU} is not pivotal, and varies with different $\bar{D}$ s.

Theorem 3 shows that the Kleibergen test converges to a nonstandard distribution depending on the nuisance parameter C_2. The nonstandard distribution is a quadratic form of the sum of a standard normal variable ζ and the deviation $\Upsilon(C_2)$ which is the asymptotic cost of near exogeneity. We obtain Kleibergen's (2003) result by setting $C_2 = 0$. So our theorem provides a more general result. Theorem 3 shows that even when the instruments are strong, the Kleibergen test with near exogeneity converges to a nonstandard distribution depending on unknown nuisance parameter C_2. Inference based on critical values from the chi-square distribution with degrees of freedom m can result in a size distortion.

Moreira (2003) develops a general method for *similar* tests based on the conditional distribution of nonpivotal statistics under weak instruments; for instance, the likelihood ratio test. He argues that there exist two asymptotically independent statistics which are invariant and sufficient for

the estimation and inference in a linear simultaneous equations model. One statistic depends on Π but the other does not. The asymptotic independence makes it possible to construct the conditional null distribution that does not depend on Π. As long as the conditional null distribution is continuous and does not depend on any unknown nuisance parameters, Moreira shows that its quantiles can be simulated and used to construct *similar* tests. Moreira (2003) proposes two likelihood ratio tests:

$$LR_1 = b_0'\bar{Y}'P_Z\bar{Y}b_0/b_0'\widehat{\Omega}b_0 - \bar{\lambda}^{\min}$$

where b_0 is the $(m+1)\times 1$ vector $(1\ ,\ -\beta_0')'$, $\widehat{\Omega} = \bar{Y}'M_Z\bar{Y}/(N-K)$ and $\bar{\lambda}^{\min}$ is the smallest eigenvalue of $\widehat{\Omega}^{-1/2\prime}\bar{Y}'P_Z\bar{Y}\ \widehat{\Omega}^{-1/2}$.

$$LR_2 = \frac{N}{2}\ln(1 + b_0'\bar{Y}'P_Z\bar{Y}b_0/b_0'\bar{Y}'M_Z\bar{Y}b_0) - \frac{N}{2}\ln(1 + \bar{\lambda}^{\min}/(N-K))$$

The LR_1 statistic is obtained by replacing the variance covariance matrix by a consistent estimator in a likelihood ratio test under assumptions of normality and known variance covariance matrix. The LR_2 statistic is the likelihood ratio test for the normal distribution with unknown variance covariance matrix. Both the LR_1 and LR_2 statistics are continuous functions that depend on two sufficient and asymptotically independent statistics S and $\widehat{T}$ where

$$S = Z'\bar{Y}b_0$$

and

$$\widehat{T} = Z'\bar{Y}\widehat{\Omega}^{-1}A_0$$

and A_0 is the $(m+1)\times m$ matrix $(\beta_0\ ,\ I_m)'$. Denote by $\psi(S,t,\widehat{\Omega},\beta_0)$ the conditional null distribution conditional on $\widehat{T} = t$, and by $c_\psi(t,\widehat{\Omega},\beta_0,\alpha)$ the $1-\alpha$ quantile of the null conditional distribution of $\psi(S,t,\widehat{\Omega},\beta_0)$. The test rejects the null if $\psi(S,t,\widehat{\Omega},\beta_0) > c_\psi(t,\widehat{\Omega},\beta_0,\alpha)$. Moreira shows that their conditional distributions conditional on $\widehat{T} = t$ do not depend on Π, and their quantiles are computable and can be used to

construct an exactly *similar* test. Recent studies (Andrews, Moreira and Stock, 2004) show that the conditional likelihood ratio test has good power under weak instruments.

Unfortunately, the conditional likelihood ratio test does not work under near exogeneity. Note that under near exogeneity

$$\begin{aligned}(1/\sqrt{N})S &= (1/\sqrt{N})Z'\bar{Y}b_0 \\ &= (1/\sqrt{N})Z'(y - Y\beta_0) \xrightarrow{d} \widetilde{\Psi}_{Zu}\end{aligned}$$

where $\widetilde{\Psi}_{Zu}$ is a normal distribution with mean C_2. So both the conditional null distribution $\psi(S, t, \widehat{\Omega}, \beta_0)$ and the critical value function $c_\psi(t, \widehat{\Omega}, \beta_0, \alpha)$ depend on the unknown parameter C_2. The simulation of the conditional null distribution needs the information of C_2, but C_2 cannot be consistently estimated because C_2 is a local to zero parameter. A conditional test based on the critical values obtained from simulating a conditional distribution ignoring the near exogeneity parameter C_2 cannot be *similar* in general.

To the best of our knowledge, no test exists in the literature that is robust simultaneously to the joint problem of weak instruments and near exogeneity. Although Bound, Jaeger, and Baker (1995) notice the possible serious inconsistency of the TSLS estimators, few econometricians pay attention to the performance of testing for $\beta = \beta_0$ under weak instruments and near exogeneity. On the one hand, overidentification tests for testing instrument exogeneity, for example, the Sargan test (Sargan, 1958) and the J test (Hansen, 1982; Newey, 1985), do not work under weak instruments. On the other hand, since the seminal paper by Staiger and Stock (1997), several tests have been developed in the literature robust to weak instruments, but we show that none of these tests is implementable under near exogeneity because the asymptotic distributions in each case are nonstandard and depend on the unknown nuisance parameters C_2. It's a big trouble in empirical studies when economists are confronted with the joint problem of weak instruments and near exogeneity.

In the next section, we consider resampling methods to approximate the Anderson-Robin test and the Kleibergen test under weak instruments and near exogeneity. The resampling method based Anderson-Robin and Kleibergen tests are constructed based on the data-depedent critical values obtained from the resampling.

3.4 RESAMPLING BASED TESTS

In this section, we employ the resampling based Anderson-Rubin and Kleibergen tests to cure the problem of size distortion under near exogeneity. In preceding sections, we show that the Anderson-Rubin test and the Kleibergen test are robust to the quality of instruments but have a size distortion under near exogeneity. The main reason is that near exogeneity brings a nuisance parameter C_2 into the asymptotic distributions of the tests. We obtain chi-square distributions for the Anderson-Rubin test and the Kleibergen test when $C_2 = 0$. The tests are no longer asymptotically chi-square distributions under near exogeneity, and as a result, size distortion occurs when we use critical values from the chi-square distribution. It is well known that the bootstrap does not work under weak instruments since generating bootstrap samples requires suitable estimates of β and Π (Dufour, 1997, 2003; MacKinnon, 2002). We consider the subsampling approach (Politis and Romano, 1994) and delete- d jackknife (Shao and Wu, 1989) as alternatives to bootstrapping. Instead of using critical values from chi-square distributions, we can use data-dependent critical values obtained from the resampling approaches. The resampling based tests are obtained by evaluating the same test statistics on each block of data, where the block size is much smaller than the sample size.

Consider resampling methods in a linear simultaneous equations model. Let $X_N = \{x_{N1}, x_{N2}, \ldots, x_{NN}\}$, a sample of N independent observations with a triangular array in the model. In order to employ the subsampling approach, let $X_{b,1}$, $X_{b,2}$, ..., X_{b,N_b} be blocks of $X_N = \{x_{N1}, x_{N2}, \ldots, x_{NN}\}$ with block size b. For independent data, we can construct blocks of X_N in any order. For the subsampling method, the blocks are generated randomly from sample observations without replacement and the number of blocks we can generate is $N_b = \binom{N}{b}$. For the delete- d jackknife method, we firstly delete d observations randomly from sample observations without replacement. Given d, the block size for each block is $N - d$ and $N_b = \binom{N}{d}$. For each $X_{b,j}$, $j = 1, 2, \ldots, N_b$, it includes $y_{j,b}$, $Y_{j,b}$ and $Z_{j,b}$ which are subvectors or submatrixes of y, Y and Z respectively. Note that $y_{b,j}$ is a $b \times 1$

vector, $Y_{b,j}$ is a $b \times m$ matrix, and $Z_{b,j}$ is a $b \times K$ matrix. These variables are denoted with subscript b because $y_{b,j}$, $Y_{b,j}$ and $Z_{b,j}$ represent randomly resampled data with block size b from sample observations without replacement.

Denote by $\wp$ the unknown probability distribution that generates the sample observations, and assume that $\wp$ belongs to a certain class of probability distributions P . Following Politis, Romano, and Wolf (1999), a general hypothesis testing procedure can be constructed as follows: the null hypothesis $H_0 : \wp \in P_0$ and the alternative hypothesis $H_1 : \wp \in P_1$ where $P_0 \cup P_1 = P$. In our case, the null hypothesis can be translated into a null hypothesis about a vector valued parameter $\beta(P)$ such that $H_0 : \beta = \beta_0 = \beta(P_0)$. Our goal is to construct a test with asymptotically correct size α ($\alpha \in (0, 1)$) based on a given statistic by using the resampling method. Define a test as

$$T_N = \tau_N t_N(x_{N1}, x_{N2}, \ldots, x_{NN})$$

where τ_N is a convergence rate such that $\tau_N \to \infty$ as $N \to \infty$.

The corresponding cumulative distribution function is defined as

$$G_N(z, \wp) = \Pr_{\wp}(T_N(x_{N1}, x_{N2}, \ldots, x_{NN}) \leq z)$$

We assume that there exists a continuous limiting law $G(., \wp)$ such that $G_N(z, \wp)$ converges weakly to $G(., \wp)$ under the null as $N \to \infty$. Note that (221) implies that $t_N(x_{N1}, x_{N2}, \ldots, x_{NN}) \xrightarrow{p} 0$ as $\tau_N \to \infty$. Correspondingly, the $1 - \alpha$ quantile of the continuous limiting law $G(., \wp)$ is defined as follows

$$c(1 - \alpha, \wp) = \inf\{z : G(z, \wp) \geq 1 - \alpha\}.$$

The idea of resampling is to approximate the sample distribution by the average of the corresponding empirical distributions obtained from the resampling methods. Denote by $t_{N,b,j}$ the corresponding version of t_N evaluated in a block $X_{b,j}$. The cumulative distribution function of $t_{N,b,j}$ is

$$G_{b,j}(z, \wp) = \Pr_{\wp}(\tau_b t_{N,b,j} \leq z)$$

The sample distribution $G_N(z\ ,\ \wp)$ is approximated by

$$\widehat{G}_{N,b}(z) = \frac{1}{N-b+1} \sum_{j=1}^{N-b+1} 1\{\tau_b t_{N,b,j} \leq z\}$$

where $1\{..\}$ is an indicator function which takes value 1 if the inside inequality holds true and 0 otherwise.

Correspondingly, the $1-\alpha$ quantile of the $\widehat{G}_{N,b}$ is defined as

$$c_{N,b}(1-\alpha) = \inf\{z : \widehat{G}_{N,b}(z) \geq 1-\alpha\}$$

Note that $c_{N,b}(1-\alpha)$ is a data-dependent critical value of the resampling based tests. The resampling based tests reject the null hypothesis when $T_N > c_{N,b}(1-\alpha)$.

In order to construct asymptotic results, the subsampling method requires that the block size $b \to \infty$ as the sample size $N \to \infty$ and $b/N \to 0$. For the delete- d jackknife method, the block size $b = N-d$ and it requires that $d = \gamma N$, where $0 < \gamma < 1$. In large samples, $\gamma = (N-b)/N \to 1$ for the subsampling method. In finite samples, the subsampling method is related to the choice of small blocks while the delete- d jackknife is related to the choice of relatively large blocks.

Consider the resampling based Anderson-Rubin test evaluated in a block $X_{b,j}$

$$AR(\beta_0)_{N,b,j} = (b-K)u'_{b,j}Z_{b,j}(Z'_{b,j}Z_{b,j})^{-1}Z'_{b,j}u_{b,j} / u'_{b,j}M_{Z_{b,j}}u_{b,j}$$

where $u_{b,j} = y_{b,j} - Y_{b,j}\beta_0$. We approximate the limiting distribution of $AR(\beta_0)$ by

$$\widehat{G}_{N,b}(z) = \frac{1}{N-b+1} \sum_{j=1}^{N-b+1} 1\{AR(\beta_0)_{N,b,j,n_b} \leq z\}$$

Define $c_{N,b}(1-\alpha)$ as the $1-\alpha$ corresponding quantile of the distribution $\widehat{G}_{N,b}(z)$. The subsampling based Anderson-Rubin test rejects the null hypothesis when $AR(\beta_0) > c_{N,b}(1-\alpha)$.

The following theorem provides the asymptotic results of the resampling based Anderson-Rubin

test.

Theorem 4 *Suppose that Assumptions 1, 2 and 3 hold for a linear simultaneous equations model. Let* $X_N = \{x_{N1}, x_{N2}, \ldots, x_{NN}\}$ *be independent observations in triangular array defined on a probability distribution* $\wp$ *. Define*

$$\sigma^2_{b,j} = E(u'_{b,j} u_{b,j}/b)$$

and

$$Q_{b,j} = E(Z'_{b,j} Z_{b,j}/b)$$

Assume the following conditions hold. For some $\delta > 0$,

(a) $E|z_{n,i} u_n|^{2+\delta} < \Delta_1 < \infty$ *for all* $1 \leq n \leq N$ *and all* $1 \leq i \leq K$

(b) $E|z_{n,i} z_{n,l}|^{1+\delta} < \Delta_2 < \infty$ *for all* $1 \leq n \leq N$ *and all* $1 \leq i, l \leq K$

(c) $E|u_n^2|^{1+\delta} < \Delta_3 < \infty$ *for all* $1 \leq n \leq N$

(d) $\sigma^2_{b,j} \to \sigma^2_u > 0$ *uniformly in* j *as* $b \to \infty$

(e) $Q_{b,j} \to Q$ *uniformly in* j *and uniformly positive definite as* $b \to \infty$

Then under the null hypothesis of $\beta = \beta_0$ and $b \to \infty$ as $N \to \infty$,

$$AR(\beta_0)_{N,b,j} \xrightarrow{d} \chi^2_K(\tilde{\varsigma})$$

where $\chi^2_K(\tilde{\varsigma})$ *is a noncentral chi-square distribution with degrees of freedom* K *and the noncentral parameter* $\tilde{\varsigma} = (1-\gamma) C'_2 \Omega^{-1} C_2$, $0 < \gamma \leq 1$.

The theorem gives asymptotic results of both the subsampling and the delete- d jackknife based Anderson-Rubin tests under the null hypothesis with weak instruments and near exogeneity. Conditions (a), (b) and (c) imposed by the theorem are required respectively to apply the triangular array central limit theorem and the weak law of large numbers for the independent heterogeneously distributed observations. Conditions (d) and (e) state that the resampling versions of the variance of the structural errors and the moments of the instruments are

asymptotically close to the whole sample version as $b \to \infty$. Note that conditions (d) and (e) are common requirements for heteroskedastic observations; see Politis, Romano and Wolf (1997, 1999). In our case, conditions (d) and (e) are satisfied trivially because of the i.i.d. assumption of u and Z.

When $\gamma = 1$, which implies that $b/N \to 0$ as $N \to \infty$, Theorem 4 shows that the subsampling based Anderson-Rubin test converges in distribution to a chi-square distribution with degree of freedom K. The subsampling method cannot replicate the near exogeneity effect described by Theorem 2. When $0 < \gamma < 1$, we obtain asymptotic results of the delete-d jackknife based Anderson-Rubin test. Theorem 4 shows that the delete-d jackknife based Anderson-Rubin test converges in distribution to a noncentral chi-square distribution. The noncentral parameter is a fraction of the noncentral parameter defined in Theorem 2, which means the delete-d jackknife based Anderson-Rubin test can partially replicate the near exogeneity effect in the limiting distribution. We observe from simulations that by the choice of the block size b, the delete-d jackknife based Anderson-Rubin test is slightly liberal. By increasing the block size we can expect to reduce the size distortion due to near exogeneity.

Now, consider the resampling based Kleibergen test evaluated in a block $X_{b,j}$,

$$K(\beta_0)_{N,b,j} = (b-K)u'_{b,j}P_{\widetilde{Y}_{b,j}(\beta_0)}u_{b,j}/u'_{b,j}M_{Z_{b,j}}u_{b,j}$$

where

$$\widetilde{Y}_{b,j}(\beta_0) = Z_{b,j}\widetilde{\Pi}_{b,j}(\beta_0)$$

and

$$\widetilde{\Pi}_{b,j}(\beta_0) = (Z'_{b,j}Z_{b,j})^{-1}Z'_{b,j}[Y_{b,j} - u_{b,j}S_{\epsilon V,b,j}(\beta_0)/S_{\epsilon\epsilon,b,j}(\beta_0)],$$

$$S_{\epsilon\epsilon,b,j}(\beta_0) = \frac{1}{b-K}u'_{b,j}M_{Z_{b,j}}u_{b,j},$$

$$S_{\epsilon V,b,j}(\beta_0) = \frac{1}{b-K}u'_{b,j}M_{Z_{b,j}}Y_{b,j}.$$

We approximate the limiting distribution of $K(\beta_0)$ by

$$\widehat{G}_{N,b}(z) = \frac{1}{N-b+1} \sum_{j=1}^{N-b+1} 1\{K(\beta_0)_{N,b,j} \leq z\}$$

Define $c_{N,b}(1-\alpha)$ as the $1-\alpha$ corresponding quantile of the distribution $\widehat{G}_{N,b}(z)$. The resampling based Kleibergen test rejects the null hypothesis when $K(\beta_0) > c_{N,b}(1-\alpha)$. The following theorem provides the asymptotic validity of the subsampling based Kleibergen test.

Theorem 5 Suppose that Assumptions 1, 2 and 3 hold for a linear simultaneous equations model. Let $X_N = \{x_{N1}, x_{N2}, \ldots, x_{NN}\}$ *be independent observations in triangular array defined on a probability distribution* $\wp$ *. Define*

$$\sigma^2_{b,j} = E(u'_{b,j} u_{b,j}/b),$$

$$\Sigma_{Vu,b,j} = E(u'_{b,j} V_{b,j}/b),$$

and

$$Q_{b,j} = E(Z'_{b,j} Z_{b,j}/b).$$

Assume the following conditions hold. For some $\delta > 0$,

(a) $E|z_{n,i} u_n|^{2+\delta} < \Delta_1 < \infty$ *for all* $1 \leq n \leq N$ *and all* $1 \leq i \leq K$

(b) $E|z_{n,i} z_{n,j}|^{1+\delta} < \Delta_2 < \infty$ *for all* $1 \leq n \leq N$ *and all* $1 \leq i,j \leq K$

(c) $E|u_n^2|^{1+\delta} < \Delta_3 < \infty$ *for all* $1 \leq n \leq N$

(d) $E|u'_n V_n|^{1+\delta} < \Delta_4 < \infty$ *for all* $1 \leq n \leq N$

(e) $\sigma^2_{b,j} \to \sigma^2_u > 0$ *uniformly in* j *as* $b \to \infty$

(f) $\Sigma_{Vu,b,j} \to \Sigma_{Vu}$ *uniformly in* j *as* $b \to \infty$

(g) $Q_{b,j} \to Q$ *uniformly in* j *and uniformly positive definite as* $b \to \infty$

Then under the null hypothesis of $\beta = \beta_0$ and $b \to \infty$ as $N \to \infty$,

$$K(\beta_0)_{N,b,j} \xrightarrow{d} [\zeta + \sqrt{1-\gamma}\,\Upsilon(C_2)]'[\zeta + \sqrt{1-\gamma}\,\Upsilon(C_2)]$$

where $\Upsilon(C_2)$ *is defined in Theorem 3 and* $0 < \gamma \leq 1$.

The theorem gives asymptotic results of both the subsampling and the delete- d jackknife based Kleibergen tests under the null hypothesis with weak instruments and near exogeneity. Note that Theorem 3 states that $\Upsilon(C_2) = (\overline{D}'\Omega^{-1}\overline{D})^{-1/2'}\overline{D}\Omega^{-1}C_2$ and $\overline{D}$ is different when the quality of instruments varies. Since the resampling approach is data-dependent, Theorem 5 can be applied to the case of near exogeneity robust to the quality of instruments.

When $\gamma = 1$, Theorem 5 shows that the subsampling based Kleibergen test converges in distribution to a chi-square distribution with degree of freedom m , the number of instruments, which implies that the subsampling method cannot replicate the near exogeneity effect described by Theorem 3. When $0 < \gamma < 1$, we obtain asymptotic results of the delete- d jackknife based Kleibergen test. Theorem 5 shows that the delete- d jackknife based Kleibergen test converges to a nonstandard distribution which is a square of sums of a standard normal variable and a random variable depending on C_2 . Theorem 5 implies that the delete- d jackknife based Kleibergen test can partially replicate the near exogeneity effect in the limiting distribution. We observe from simulations that by the choice of the block size b , the delete- d jackknife based Kleibergen test is slightly liberal. By increasing the block size we can expect to reduce the size distortion due to near exogeneity.

The next section conducts a Monte Carlo simulation to compare the size performance of the Anderson-Rubin test, the Kleibergen test and their corresponding resampling based versions in finite samples under various environments.

3.5 MONTE CARLO SIMULATION

We consider a linear simultaneous equations model defined above. Since there is only one

endogenous variable Y, we set $K = 1$ (the just-identified case) and $K = 2$ (the overidentified case). The β is the only structural parameter and we set the true value $\beta_0 = 0$. The N is the sample size and we set $N = 80$ to conduct comparisons of tests performance in finite samples. The data (Z_i, u_i, V_i) is iid which are generated from a joint normal distribution $N(0, \Lambda)$.

When $l = 1$,

$$\Lambda = \begin{pmatrix} 1 & covZ_i u_i & 0 \\ covZ_i u_i & 1 & covV_i u_i \\ 0 & covV_i u_i & 1 \end{pmatrix}$$

When $l = 2$,

$$\Lambda = \begin{pmatrix} 1 & 0 & covZ_{i,1} u_i & 0 \\ 0 & 1 & covZ_{i,2} u_i & 0 \\ covZ_{i,1} u_i & covZ_{i,2} u_i & 1 & covV_i u_i \\ 0 & 0 & covV_i u_i & 1 \end{pmatrix}$$

where $covV_i u_i$ measures the endogeneity of Y, which takes values of 0.25. We also do simulations when $covV_i u_i = 0.9$. Since the results are similar to those when $covV_i u_i = 0.25$, we just report results when $covV_i u_i = 0.25$. When $l = 1$, $covZ_i u_i$ measures the degree of near exogeneity which takes values of 0, 0.10, or 0.15. When $l = 2$, we set $covZ_{i,1} u_i = covZ_{i,2} u_i$ and it takes values of 0, 0.10, or 0.15. The data generated from (411) and (421) also differ over the value of Π. The vector Π controls the quality of instruments. We set $\Pi = 0$, 0.1, or 1 in all cells of the vector to respectively represent nonidentification, weak instruments and strong instruments.

We test $H_0 : \beta_0 = 0$ against $H_1 : \beta_0 \neq 0$. We study the size and power of the Anderson-Rubin test and the Kleibergen test defined in Section 4 and their corresponding resampling versions defined in Section 5 under various environments.

We conduct 1000 iterations to compare the sizes of the Anderson-Rubin test, the Kleibergen test

and their corresponding resampling versions under the null hypothesis $\beta_0 = 0$ at the 10% nominal level in a small sample ($N = 80$) under various environments. We consider the just-identified case ($l = 1$) and the overidentified case ($l = 2$) respectively. The simulation shows that results are similar to each other when $l = 1$ or $l = 2$ so we only report the just-identified case.

Tables 1 reports actual rejection probabilities of the Anderson-Rubin test under the null hypothesis ($\beta_0 = 0$) when the instruments are completely nonidentified ($\Pi = 0$), weak ($\Pi = 0.1$) or strong ($\Pi = 1$). Tables 2 reports actual rejection probabilities of the Kleibergen test under the null hypothesis ($\beta_0 = 0$) in all the three cases. When $covZ_i u_i = 0$ which means that there is no near exogeneity problem, for the Anderson-Rubin test, the actual sizes ranges from 9.1 to 10.1 . For the Kleibergen test, the actual sizes ranges from 9.1 to 10.6 . This means both tests work very well if there is no near exogeneity.

However, Theorems 4.1 and 4.2 show that the limits of both test statistics are not chi-square distributed under near exogeneity and using critical values from the chi-square distribution leads to large size distortion. The simulation results reflect the facts stated in theorems. For the Anderson-Rubin test with $covZ_i u_i = 0.10$, actual sizes are between 22 and 25 when $l = 1$. For the Kleibergen test with $covZ_i u_i = 0.10$, actual sizes are between 24.0 and 24.9 when $l = 1$. If $covZ_i u_i = 0.15$, the degree of size distortion increases very much. For the Anderson-Rubin test, actual sizes are around 24 when $l = 1$. For the Kleibergen test, the actual sizes range from 37.8 to 41.1 when $l = 1$. These results show that using chi-square critical values without taking account of near exogeneity is very misleading.

Table 3 compares size properties of resampling based Anderson-Rubin tests under near exogeneity in finite samples for various choices of the block size b . We choose the block size $b = \{5, 10, 15, 20, 25, 30, 40\}$. Roughly speaking, $b = \{5, 10\}$ can correspond to subsampling method and other b s correspond to delete- d jackknife method. Compared to Table 1, we observe the reduction in size distortion in Table 3 by using data-dependent critical values obtained from resampling. When instruments are strong and $b = 20$, the actual size is 9.9

when the correlation between instruments and structural errors is 0.10, and the actual size is 14.9 when the correlation between instruments and structural errors is 0.15. In Table 1, their corresponding actual sizes are 25.1 and 38.2 respectively. We can observe similar situations when instruments are weak or nonidentified.

Now, consider a realistic situation that the correlation between instruments and structural errors is between 0.10 and 0.15. When the block size is large, for example, $b = 40$, the resampling based Anderson-Rubin test is very conservative. For example, the actual sizes under strong instruments are 1.1 and 4.2 respectively when the correlation between instruments and structural errors are 0.1 and 0.15. We also observe similar situations when instruments are weak or nonidentified. When the block size shrinks, we can observe more rejections. When $b = 25$ and $b = 20$, we observe suitable actual sizes under near exogeneity. When $b = 10$ and $b = 5$, we observe overrejections. For example, when $b = 5$, the actual rejection probabilities under weak instruments are 16.3 and 27.8 respectively when the correlation between instruments and structural errors are 0.1 and 0.15. Note that $b = \{5, 10\}$ represents subsampling methods. Theorem 4 shows that the subsampling method cannot replicate the near exogeneity effect and converges to the same chi-square distribution defined in Theorem 2. The size distortion when $b = \{5, 10\}$ reflects this fact. Compared to the subsampling method, we observe that the delete- d jackknife is slightly liberal. They are conservative when the block size is large and obtain right actual rejection probabilities when b is in the range of 20 and 25. For example, when $b = 20$ and $covZ_i u_i = 0.10$, the actual rejection probabilities are 9.9, 9.7, and 9.5 respectively under strong instruments, weak instruments and nonidentification. When $b = 25$ and $covZ_i u_i = 0.15$, the actual rejection probabilities are 12.3, 13.9, and 12.6 respectively under strong instruments, weak instruments and nonidentification.

We also observe very conservative results when there is no exogeneity, that is, $covZ_i u_i = 0$. One possible reason is that when we resample the sample data, the moment condition

$(\sum_{i=1}^{b} Z_i u_i / b)$ in blocks is not demeaned in finite samples, which produces larger data-dependent critical values than those from chi-square distribution. This creates undersized results. Table 4 reports the resampling based Kleibergen test under near exogeneity in finite samples for various choices of the block size b. We choose the block size $b = \{5, 10, 15, 20, 25, 30, 40\}$. Roughly speaking, $b = \{5, 10\}$ can correspond to subsampling method and other b s correspond to delete- d jackknife method. Compared to Table 2, we observe the reduction in size distortion under resampling methods. We also observe similar relationship between actual sizes and choices of the block size to Table 3.

3.6 CONCLUSIONS

This paper studies the asymptotic properties of estimation and inference with weak instruments and near exogeneity in a linear simultaneous equations model. Weak instruments and near exogeneity are related to two important criteria of instrumental variables regressions. We show that the TSLS estimator and the LIML estimator with weak instruments and near exogeneity can have a relatively large asymptotic bias compared to the case where only weak instruments occur. We show that the limiting distributions of the Anderson-Rubin test and the Kleibergen test are no longer asymptotically pivotal under near exogeneity, and it leads to serious size distortion in hypothesis testing if we use the critical values from the chi-square distributions. We show that the conditional likelihood ratio test does not work in our case because the conditional distribution of the test under the null hypothesis depends on an unknown parameter which reflects near exogeneity. We propose delete- d jackknife based Anderson-Rubin and Kleibergen tests to correct size distortion in finite samples under weak instruments and near exogeneity.

3.7 APPENDIX

3.7.1 Appendix 1

Proof of Lemma 1 (a) First, substituting Y from the reduced form equation, we have

$$\begin{aligned} Y'u/N &= (Z\Pi + V)'u/N \\ &= V'u/N + \Pi'Z'u/N. \end{aligned}$$

Note that $V'u/N \xrightarrow{p} \Sigma_{Vu}$ by part (a) in Assumption 3. The weak law of large numbers and Assumption 2 give

$$Z'u/N \xrightarrow{p} E[Z_i'u_i] = C_2/\sqrt{N}$$

And note that $\Pi = C_1/\sqrt{N}$ from Assumption 1, so we have

$$Y'u/N \xrightarrow{p} \Sigma_{Vu}$$

since the second part tends to zero in probability.

To show

$$Y'Y/N \xrightarrow{p} \Sigma_{VV},$$

we follow the proof above, so

$$\begin{aligned} Y'Y/N &= (Z\Pi + V)'(Z\Pi + V)/N \\ &= V'V/N + \Pi'Z'Z\Pi/N + \Pi'Z'V/N + V'Z\Pi/N \end{aligned}$$

By part (a) in Assumption 3, we have

$$V'V/N \xrightarrow{p} \Sigma_{VV}.$$

Note that the last three parts tend to zero in probability because of part (b) and (c) in Assumption 3 and Assumption 1. The result then follows.

(b)First, we observe that

$$\begin{aligned} P_Z^{1/2}u &= (Z'Z)^{-1/2}Z'u \\ &= (Z'Z/N)^{-1/2}(Z'u/\sqrt{N}) \end{aligned}$$

Then, we have

$$Z'u/\sqrt{N} \Rightarrow \widetilde{\Psi}_{Zu} = \Psi_{Zu} + C_2$$

by part (c) in Assumption 3. Note that Ψ_{Zu} is a standard normal vector which is the same as the Ψ_{Zu} defined in Staiger and Stock (1997). Part (b) in Assumption 3 gives that

$$(Z'Z/N)^{-1/2} \Rightarrow Q^{-1/2}.$$

So

$$\begin{aligned}
& P_Z^{1/2} u \xrightarrow{d} Q^{-1/2}\widetilde{\Psi}_{Zu} \\
&= (Q^{-1/2}\widetilde{\Psi}_{Zu}\sigma_u^{-1})\sigma_u \\
&= (Q^{-1/2}(\Psi_{Zu} + C_2)\sigma_u^{-1})\sigma_u \\
&= z_u\sigma_u + Q^{-1/2}C_2 \\
&= (z_u + Q^{-1/2}C_2\sigma_u^{-1})\sigma_u \\
&= \widetilde{z_u}\sigma_u
\end{aligned}$$

Note that $\widetilde{z_u} = z_u + Q^{-1/2}C_2\sigma_u^{-1}$ stated in the lemma.

The proof of $P_Z^{1/2}V \xrightarrow{d} z_v\Sigma_{VV}^{1/2}$ is the same as that in Staiger and Stock (1997).

(c) Note that

$$\begin{aligned}
P_Z^{1/2}Y &= (Z'Z/N)^{-1/2\prime}(Z'Y/\sqrt{N}) \\
&= (Z'Z/N)^{-1/2\prime}(Z'Z\Pi/\sqrt{N} + Z'V/\sqrt{N}).
\end{aligned}$$

Then we know that $\Pi = C_1/\sqrt{N}$ in Assumption 1 and $Z'V/\sqrt{N} \xrightarrow{d} \Psi_{ZV}$ and $Z'Z/N \xrightarrow{p} Q$ by parts (c) and (b) in Assumption 3 respectively. The result directly follows.

(d) First, we observe that

$$Y'P_Zu = (Y'Z/\sqrt{N})(Z'Z/N)^{-1}(Z'u/\sqrt{N})$$

Note that $(Z'Z/N)^{-1} \xrightarrow{p} Q^{-1}$ from part (b) in Assumption 3 and $Z'u/\sqrt{N} \Rightarrow \widetilde{\Psi}_{Zu}$ from part (c) in Assumption 3.Then, substituting Y from Equation (111), we have

$$\begin{aligned}
Y'Z/\sqrt{N} &= (Z\Pi + V)'Z/\sqrt{N} \\
&= V'Z/\sqrt{N} + \Pi'Z'Z/\sqrt{N}.
\end{aligned}$$

From part (c) in Assumption 3, we have

$V'Z/\sqrt{N} \xrightarrow{d} \Psi'_{ZV}$.

Assumption 1 and part (b) in Assumption 3 give that

$\Pi'Z'Z/\sqrt{N} \xrightarrow{p} C_1'Q$.

So, we have

$$
\begin{aligned}
& Y'P_Z u \xrightarrow{d} (\Psi'_{ZV} + C_1'Q)Q^{-1}\widetilde{\Psi}_{Zu} \\
= \; & (\Psi_{ZV} + Q'C_1)'Q^{-1}(\Psi_{Zu} + C_2) \\
= \; & \Sigma_{VV}^{1/2\prime}(Q^{1/2}C_1\Sigma_{VV}^{-1/2} + Q^{-1/2\prime}\Psi_{ZV}\Sigma_{VV}^{-1/2})'(Q^{-1/2\prime}\Psi_{Zu}\sigma_u^{-1} + Q^{-1/2\prime}C_2\sigma_u^{-1})\sigma_u \\
= \; & \Sigma_{VV}^{1/2\prime}(\lambda + z_V)'(z_u + Q^{-1/2\prime}C_2\sigma_u^{-1})\sigma_u \\
= \; & \Sigma_{VV}^{1/2\prime}(\lambda + z_V)'\widetilde{z}_u\sigma_u \\
= \; & \Sigma_{VV}^{1/2\prime}\widetilde{v}_2\sigma_u
\end{aligned}
$$

where $\widetilde{v}_2 = (\lambda + z_V)'\widetilde{z}_u$ defined in the lemma.

To show $u'P_Z u \xrightarrow{d} \sigma_{uu}\widetilde{z_u}'\widetilde{z_u}$, note that

$$u'P_Z u = (u'P_Z^{1/2\prime})(P_Z^{1/2}u).$$

The result follows directly from part (b) in the lemma.

Note that the proof of $Y'P_Z^{1/2}Y \xrightarrow{d} \Sigma_{VV}^{1/2\prime}v_1\Sigma_{VV}^{1/2}$ is the same as that in Staiger and Stock (1997). *Q.E.D.*

Proof of Theorem 1 (a) First, we have

$$
\begin{aligned}
\hat{\beta}_{TSLS} - \beta_0 &= (Y'P_Z Y)^{-1}(Y'P_Z u) \\
&\xrightarrow{d} (\Sigma_{VV}^{1/2\prime}v_1\Sigma_{VV}^{1/2})^{-1}(\Sigma_{VV}^{1/2\prime}\widetilde{v}_2\sigma_u) \\
&= \sigma_u\Sigma_{VV}^{-1/2}v_1^{-1}\widetilde{v}_2
\end{aligned}
$$

Note that the second step is obtained from part (d) in Lemma 1.

(b) The result of part (b) follows Theorem 2 in Staiger and Stock (1997) by replacing b by

$b_{TSLS} = v_1^{-1}\tilde{v}_2.$

(c)First, replacing y by the structural equation in the LIML estimator, we have

$$\begin{aligned}
\hat{\beta}_{LIML} - \beta_0 &= [Y'(I - kM_Z)Y]^{-1}[Y'(I - kM_Z)u] \\
&\Rightarrow [Y'Y - (\kappa/N + 1)Y'M_ZY]^{-1}[Y'u - (\kappa/N + 1)Y'M_Zu] \\
&\xrightarrow{d} [Y'Y - Y'M_ZY - (\kappa/N)Y'M_ZY]^{-1}[Y'u - Y'M_Zu - (\kappa/N)Y'M_Zu] \\
&= [Y'P_ZY - (\kappa/N)Y'M_ZY]^{-1}[Y'P_Zu - (\kappa/N)Y'M_Zu]
\end{aligned}$$

The second step is obtained by the fact that $N(k-1) \Rightarrow \kappa$. Note that

$$Y'M_ZY/N = Y'Y/N + Y'P_ZY/N \xrightarrow{p} \Sigma_{VV}$$

from part (a) in Lemma 1 and

$$Y'P_ZY/N \xrightarrow{p} 0$$

from part (d) in Lemma 1. By the similar reason, we have

$$Y'M_Zu/N \xrightarrow{p} \Sigma_{Vu}.$$

So

$$\begin{aligned}
\hat{\beta}_{LIML} - \beta_0 &= [Y'P_ZY - \kappa(Y'M_ZY/N)]^{-1}[Y'P_Zu - \kappa(Y'M_Zu/N)] \\
&\xrightarrow{d} [\Sigma_{VV}^{1/2\prime} v_1 \Sigma_{VV}^{1/2} - \kappa\Sigma_{VV}]^{-1}[\Sigma_{VV}^{1/2\prime}\widetilde{v_2}\sigma_u - \kappa\Sigma_{Vu}] \\
&= \sigma_u\Sigma_{VV}^{-1/2}(v_1 - \kappa I_m)^{-1}(\tilde{v}_2 - \kappa\rho)
\end{aligned}$$

where $N(k-1) \Rightarrow \kappa$ and k is the smallest root of the determinantal equation $\left|\bar{Y}'\bar{Y} - k\bar{Y}'M_Z\bar{Y}\right| = 0$. To complete the proof, we follow the method used in Theorem 3 in Staiger and Stock (1997) to find the smallest root of the limit of the determinantal equation as $N \to \infty$. Let $J = \begin{pmatrix} 1 & 0 \\ -\beta & I_m \end{pmatrix}$ and note that $\bar{Y}J = (u \quad Y)$. Since J is a non-singular $(m+1) \times (m+1)$ matrix, the roots of the modified determinantal equation $\left|J'\bar{Y}'\bar{Y}J - kJ'\bar{Y}'M_Z\bar{Y}J\right| = 0$ are the same as the roots of the original determinantal equation.

Denote by

$$
\begin{aligned}
D(\kappa) &= J'\bar{Y}'\bar{Y}J - (\kappa/N+1)J'\bar{Y}'M_Z\bar{Y}J \\
&= J'\bar{Y}'P_Z\bar{Y}J - (\kappa/N)J'\bar{Y}'M_Z\bar{Y}J \\
&= \begin{pmatrix} u'P_Zu & u'P_ZY \\ Y'P_Zu & Y'P_ZY \end{pmatrix} - \kappa \begin{pmatrix} u'M_Zu/N & u'M_ZY/N \\ Y'M_Zu/N & Y'M_ZY/N \end{pmatrix} \\
&\xrightarrow{d} \begin{pmatrix} \sigma_u^2 \tilde{z}_u'\tilde{z}_u & \sigma_u \tilde{z}_u'(\lambda + z_v)\Sigma_{VV}^{1/2} \\ \Sigma_{VV}^{1/2}(\lambda + z_v)'\tilde{z}_u\sigma_u & \Sigma_{VV}^{1/2}(\lambda + z_v)'(\lambda + z_v)\Sigma_{VV}^{1/2} \end{pmatrix} - \kappa \begin{pmatrix} \sigma_u^2 & \Sigma_{Vu}' \\ \Sigma_{Vu} & \Sigma_{VV} \end{pmatrix} \\
&= \begin{pmatrix} \sigma_u & 0 \\ 0 & \Sigma_{VV}^{1/2} \end{pmatrix} \tilde{\Xi}_0^* \begin{pmatrix} \sigma_u & 0 \\ 0 & \Sigma_{VV}^{1/2} \end{pmatrix} - \begin{pmatrix} \sigma_u & 0 \\ 0 & \Sigma_{VV}^{1/2} \end{pmatrix} \kappa\bar{\Sigma} \begin{pmatrix} \sigma_u & 0 \\ 0 & \Sigma_{VV}^{1/2} \end{pmatrix}
\end{aligned}
$$

where $\tilde{\Xi}_0^*$ is defined in the statement of Theorem 1 and $\bar{\Sigma}$ is defined in the statement of Lemma 1. So κ is the smallest root of $\left|\tilde{\Xi}_0^* - \kappa\bar{\Sigma}\right| = 0$. Note that the above derivation is obtained from part (d) in Lemma 1.

(d) The result of part (d) follows Theorem 2 in Staiger and Stock (1997) by replacing b by

$$b_{LIML} = (v_1 - \kappa I_m)^{-1}(\widetilde{v_2} - \kappa\rho). \qquad \textit{Q.E.D.}$$

Proof of Corollary 1 Employing Theorem 1 and the Dominated Convergence Theorem, we get

$$
\begin{aligned}
& E\ \hat{\beta}_{TSLS} - \beta_0 \xrightarrow{d} \sigma_u \Sigma_{VV}^{-1/2} E v_1^{-1} \tilde{v}_2 \\
&= \sigma_u \Sigma_{VV}^{-1/2} E v_1^{-1} (\lambda + z_v)'(z_u + Q^{-1/2} C_2 \sigma_u^{-1}) \\
&= \sigma_u \Sigma_{VV}^{-1/2} E v_1^{-1} (\lambda + z_v)' z_u + \sigma_u \Sigma_{VV}^{-1/2} E v_1^{-1} (\lambda + z_v)' Q^{-1/2} C_2 \sigma_u^{-1}
\end{aligned}
$$

Note that z_u is asymptotically equivalent to $z_v\rho$, so we have

$$E\ \hat{\beta}_{TSLS} - \beta_0 \xrightarrow{d} \sigma_u \Sigma_{VV}^{-1/2}(h\rho + \Delta).$$

Note that

$$E\ \hat{\beta}_{OLS} - \beta_0 \xrightarrow{d} \sigma_u \Sigma_{VV}^{-1/2}\rho.$$

The result in Corollary 1 follows from using $\Sigma_{YY} = p\lim(Y'Y/N)$ and σ_u being a scalar. *Q.E.D.*

Proof of Theorem 2 The Anderson-Rubin test is given by

$$AR(\beta_0) = (y - Y\beta_0)'P_Z(y - Y\beta_0)/\frac{1}{N-K}(y - Y\beta_0)'M_Z(y - Y\beta_0)$$

We first observe that

$$\begin{aligned} &\frac{1}{N-K}(y - Y\beta_0)'M_Z(y - Y\beta_0) \\ &= \frac{1}{N-K}u'M_Zu \\ &= \frac{1}{N-K}u'u - \frac{1}{N-K}u'P_Zu \end{aligned}$$

By part (a) in Assumption 3, the first term converges in probability to σ_u^2 , and the last term tends to zero by part (d) in Lemma 1. So we have

$$\frac{1}{N-K}(y - Y\beta_0)'M_Z(y - \beta_0) \xrightarrow{p} \sigma_u^2.$$

Next, note that

$$(y - Y\beta_0)'P_Z(y - Y\beta_0) = u'P_Zu.$$

Define $\xi = P_Z^{1/2}u$. Part (b) in Lemma 1 gives that

$$\begin{aligned} &\xi \xrightarrow{d} \tilde{z}_u\sigma_u \\ &= z_u\sigma_u + Q^{-1/2}C_2 \sim N(Q^{-1/2}C_2, \sigma_u^2). \end{aligned}$$

So

$$\begin{aligned} &(y - Y\beta_0)'P_Z(y - Y\beta_0) \xrightarrow{d} \xi'\xi \\ &\xrightarrow{d} \chi_K^2(C_2'(\sigma_u^2 \otimes Q)^{-1}C_2) \\ &= \chi_K^2(C_2'\Omega^{-1}C_2). \qquad Q.E.D. \end{aligned}$$

Proof of Theorem 3 We first follow Kleibergen's (2003) idea to construct two asymptotically independent variables.

Note that

$$\{\frac{1}{\sqrt{N}}Z'(y-Y\beta_0), \frac{1}{\sqrt{N}}Z'(Y-Z\Pi)\} \xrightarrow{d} \{\widetilde{\Psi}_{Zu}, \Psi_{ZV}\}$$

and

$$\begin{pmatrix} \widetilde{\Psi}_{Zu} \\ vec(\Psi_{ZV}) \end{pmatrix} \sim N\left(\begin{pmatrix} C_2 \\ 0 \end{pmatrix}, \Sigma \otimes Q\right)$$

from Assumption 3.

Post-multiplying it by $\begin{pmatrix} 1 & -\Sigma'_{Vu}/\sigma_u^2 \\ 0 & I_m \end{pmatrix}$ gives

$$\{\frac{1}{\sqrt{N}}Z'(y-Y\beta_0), \frac{1}{\sqrt{N}}Z'[(Y-Z\Pi)-(y-Y\beta_0)\Sigma'_{Vu}]/\sigma_u^2\} \xrightarrow{d} \{\widetilde{\Psi}_{Zu}, \Psi_{ZU}\}$$

where

$$U = (Y-Z\Pi)-(y-Y\beta_0)\Sigma'_{Vu}/\sigma_u^2 = V - u\Sigma'_{Vu}/\sigma_u^2$$

and

$$(\widetilde{\Psi}_{Zu}, vec\Psi_{ZU}) \sim N\left(\begin{pmatrix} C_2 \\ -C_2\Sigma'_{Vu}/\sigma_u^2 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \Sigma_{VV}-\Sigma_{Vu}\Sigma'_{Vu}/\sigma_u^2 \end{pmatrix}\right)$$

So $\widetilde{\Psi}_{Zu}$ and Ψ_{ZU} are asymptotically independent. Note that in Kleibergen's proof, C_2 is zero.

Next, we have

$$S_{\epsilon V}(\beta_0) \xrightarrow{p} \Sigma'_{Vu}$$

and

$$S_{\epsilon\epsilon}(\beta_0) \xrightarrow{p} \sigma_u^2$$

by parts (a) and (d) in Lemma 1. So $S_{\epsilon V}(\beta_0)$ and $S_{\epsilon\epsilon}(\beta_0)$ are consistent estimators of Σ'_{Vu} and σ_{uu} respectively. This implies that $\frac{1}{\sqrt{N}}Z'[(Y-Z\Pi)-(y-Y\beta_0)S_{\epsilon V}(\beta_0)/S_{\epsilon\epsilon}(\beta_0)]$ has the

same limiting behavior as $\frac{1}{\sqrt{N}}Z'(Y - Z\Pi) - (y - Y\beta_0)\Sigma'_{Vu}/\sigma_u^2$. So

$$\frac{1}{\sqrt{N}}Z'[(Y - Z\Pi) - (y - Y\beta_0)S_{\epsilon V}(\beta_0)/S_{\epsilon\epsilon}(\beta_0)] \xrightarrow{d} \Psi_{ZU}$$

and it is asymptotically independent of Ψ_{Zu} .

Now consider

$$\begin{aligned}\widetilde{Y}(\beta_0) &= Z\widetilde{\Pi}(\beta_0)\\ &= P_Z[Y - (y - Y\beta_0)S_{\epsilon V}(\beta_0)/S_{\epsilon\epsilon}(\beta_0)]\end{aligned}$$

and we have

$$P_{\widetilde{Y}(\beta_0)} = Z(Z'Z/N)^{-1}D[D'(Z'Z/N)^{-1}D]^{-1}D'(Z'Z/N)^{-1}Z'/N$$

where

$$D = \gamma(N)Z'[Y - (y - Y\beta_0)S_{\epsilon V}(\beta_0)/S_{\epsilon\epsilon}(\beta_0)].$$

When the instruments are strong, we have $\gamma(N) = 1/N.$ We have $\gamma(N) = 1/\sqrt{N}$ when the instruments are weak or invalid. The Kleibergen test is given by

$$\begin{aligned}K(\beta_0) &= (y - Y\beta_0)'P_{\widetilde{Y}(\beta_0)}(y - Y\beta_0)/\frac{1}{N-K}(y - Y\beta_0)'M_Z(y - Y\beta_0)\\ &= \frac{1}{\sqrt{N}}(y - Y\beta_0)'Z(Z'Z/N)^{-1}D[D'(Z'Z/N)^{-1}D]^{-1}D'(Z'Z/N)^{-1}\\ &\quad\times \frac{1}{\sqrt{N}}Z'(y - Y\beta_0)\big/[\frac{1}{N-K}(y - Y\beta_0)'M_Z(y - Y\beta_0)]\\ &\quad\xrightarrow{d} \overline{G}'\Omega^{-1}\overline{D}[\overline{D}'\Omega^{-1}\overline{D}]^{-1}\overline{D}'\Omega^{-1}\overline{G}\end{aligned}$$

where $\overline{G}$ is the limit of $G = \frac{1}{\sqrt{N}}Z'(y - Y\beta_0)$ and $\overline{D}$ is the limit of D and D is defined in Section 4. Part (c) in Assumption 3 gives $\overline{G} \sim \widetilde{\Psi}_{Zu}$ where $\widetilde{\Psi}_{Zu} \sim N(C_2\ ,\ \Omega)$. Note that $\frac{1}{N-K}(y - \beta_0)'M_Z(y - \beta_0) \xrightarrow{p} \sigma_u^2$ from part (a) in Assumption 3 and part (d) in Lemma 1.

Next, consider

$$[\overline{D}'\Omega^{-1}\overline{D}]^{-1/2'}\overline{D}'\Omega^{-1}\overline{G} = [\overline{D}'\Omega^{-1}\overline{D}]^{-1/2'}\overline{D}'\Omega^{-1}(\Psi_{Zu} + C_2)$$

where Ψ_{Zu} is a normal distribution with zero mean and variance covariance matrix Ω and we have $\widetilde{\Psi}_{Zu} = \Psi_{Zu} + C_2$. The nonzero mean C_2 comes from near exogeneity. Note that

$$[\overline{D}'\Omega^{-1}\overline{D}]^{-1/2\prime}\overline{D}'\Omega^{-1}\Psi_{Zu} \xrightarrow{d} N(0, I_m) \sim \zeta.$$

So we have

$$K(\beta_0) \xrightarrow{d} (\zeta + \Upsilon(C_2))'(\zeta + \Upsilon(C_2)).$$

Next, we show that $\overline{D}$ is different when the quality of the instruments varies. The following statements provide the limits of $\overline{D}$ when the instruments are strong, weak or completely nonidentified.

(a) When the instruments are strong, $\Pi = C_1$,

$$\begin{aligned}
&\frac{1}{N}Z'[Y - (y - Y\beta_0)S_{\epsilon V}(\beta_0)/S_{\epsilon\epsilon}(\beta_0)] \\
&= \frac{1}{N}Z'[(Y - Z\Pi) - (y - Y\beta_0)S_{\epsilon V}(\beta_0)/S_{\epsilon\epsilon}(\beta_0)] + \frac{1}{N}Z'Z\Pi \\
&\to QC_1
\end{aligned}$$

where the first term in the first equation converges to zero since

$$\frac{1}{\sqrt{N}}Z'[(Y - Z\Pi) - (y - Y\beta_0)S_{\epsilon V}(\beta_0)/S_{\epsilon\epsilon}(\beta_0)] \xrightarrow{d} \Psi_{ZU}.$$

(b) When the instruments are weak, $\Pi = C_1/\sqrt{N}$,

$$\begin{aligned}
&\frac{1}{\sqrt{N}}Z'[Y - (y - Y\beta_0)S_{\epsilon V}(\beta_0)/S_{\epsilon\epsilon}(\beta_0)] \\
&= \frac{1}{\sqrt{N}}Z'[(Y - Z\Pi) - (y - Y\beta_0)S_{\epsilon V}(\beta_0)/S_{\epsilon\epsilon}(\beta_0)] + \frac{1}{\sqrt{N}}Z'ZC_1/\sqrt{N} \\
&\xrightarrow{d} \Psi_{ZU} + QC_1.
\end{aligned}$$

(c) When the instruments are completely nonidentified, $\Pi = 0$,

$$\begin{aligned}
&\frac{1}{\sqrt{N}}Z'[Y - (y - Y\beta_0)S_{\epsilon V}(\beta_0)/S_{\epsilon\epsilon}(\beta_0)] \\
&= \frac{1}{\sqrt{N}}Z'[(Y - Z\Pi) - (y - Y\beta_0)S_{\epsilon V}(\beta_0)/S_{\epsilon\epsilon}(\beta_0)] \\
&\xrightarrow{d} \Psi_{ZU} .
\end{aligned}$$

Q.E.D.

Proof of Theorem 4 The resampling based Anderson-Rubin test $AR(\beta_0)_{N,b,j}$ is defined as,

$$AR(\beta_0)_{N,b,j} = (b-K)u'_{b,j}Z_{b,j}(Z'_{b,j}Z_{b,j})^{-1}Z'_{b,j}u_{b,j} \big/ u'_{b,j}M_{Z_{b,j}}u_{b,j}$$

We accomplish the proof by three steps.

Step 1: We want to show

$$u'_{b,j}u_{b,j}/b \xrightarrow{p} \sigma_u^2$$

as $b \to \infty$. Note that

$$u'_{b,j}u_{b,j}/b = \frac{1}{b-K}\sum_{n\in\{1,2,..,N\}}^{b} u_n^2.$$

$\sum_{n\in\{1,2,..,N\}}^{b}$ represents the summation of b observations which are randomly picked from the sample observations $\{1,2,\ldots,N\}$. The law of large numbers and condition (c) give that

$$u'_{b,j}u_{b,j}/b \xrightarrow{p} \sigma_{b,j}^2$$

as $b \to \infty$ and the result follows by condition (d).

Step 2: We want to show

$$(Z'_{b,j}Z_{b,j}/b)^{-1} \xrightarrow{p} Q^{-1}.$$

Note that

$$Z'_{b,j}Z_{b,j}/b = \frac{1}{b}\sum_{n\in\{1,2,..,N\}}^{b} z'_n z_n$$

and denote its (p,q) -th entry by

$$(D_b^1)_{p,q} = \frac{1}{b}\sum_{n\in\{1,2,..,N\}}^{b} z_{n,p}z_{n,q}$$

where $1 \le p,q \le K$. Then the law of large numbers and condition (b) give that $(D_b^1)_{p.q} \xrightarrow{p} Q_{b,j}^{p,q}$ as $b \to \infty$ for all p and q where $Q_{b,j}^{p,q}$ is the (p,q) -th entry of $Q_{b,j}$. Since $Q_{b,j} \to Q$ uniformly in n as $b \to \infty$, we have $Z'_{b,j}Z_{b,j}/b \xrightarrow{p} Q$. The result in Step 2

follows by the continuous mapping theorem.

Step 3: We want to show

$$b^{-1/2} Z'_{b,j} u_{b,j} \xrightarrow{d} N(\sqrt{1-\gamma}\, C_2, \Omega)$$

for the delete- d jackknife method, where $\Omega = \sigma_u^2 \otimes Q$ and $0 < \gamma < 1$
Note that

$$b^{-1/2} Z'_{b,j} u_{b,j} = b^{-1/2} \sum_{n \in \{1,2,..,N\}}^{b} z'_n u_n$$

and denote its p th entry by

$$(D_b^2)_p = b^{-1/2} \sum_{n \in \{1,2,..,N\}}^{b} z_{n,p} u_n$$

where $1 \leq p \leq K$. By Step 1 and 2, we have

$$Var(D_b^2)_p = \sigma_{b,j} Q_{b,j}^{pp} \to \sigma_u^2 Q^{pp} > 0$$

where Q^{pp} is the (p,p) -th entry of Q . Then, we have

$$\begin{aligned} b^{-1/2} Z'_{b,j} u_{b,j} &= b^{-1/2} \sum_{n \in \{1,2,..,N\}}^{b} z'_n u_n \\ &= b^{-1/2} \sum_{n \in \{1,2,..,N\}}^{b} [z'_n u_n - E(z'_n u_n)] + b^{-1/2} \sum_{n \in \{1,2,..,N\}}^{b} E(z'_n u_n) \end{aligned}$$

For the first term, condition (a) provides a sufficient condition for the triangular array central limit theorem so we obtain

$$b^{-1/2} \sum_{n \in \{1,2,..,N\}}^{b} [z'_n u_n - E(z'_n u_n)] \xrightarrow{d} N(0, \Omega)$$

where $\Omega = \sigma_u^2 \otimes Q$. For the second term, from Assumption 2, we have

$$\begin{aligned} b^{-1/2} \sum_{n \in \{1,2,..,N\}}^{b} E(z'_n u_n) &\to \sqrt{\frac{b}{N}}\, C_2 \\ &= (\sqrt{1-\gamma}) C_2 \end{aligned}$$

where $\gamma = d/N = (N-b)/N = 1 - \frac{b}{N}$, $0 < \gamma < 1$.
Now, consider the case of subsampling method. We have

$$b^{-1/2} Z'_{b,j} u_{b,j} = b^{-1/2} \sum_{n \in \{1,2,..,N\}}^{b} z'_n u_n$$

$$= b^{-1/2} \sum_{n \in \{1,2,..,N\}}^{b} [z'_n u_n - E(z'_n u_n)] + b^{-1/2} \sum_{n \in \{1,2,..,N\}}^{b} E(z'_n u_n)$$

The first term converges to the same distribution defined in (501). For the second term,

$$b^{-1/2} \sum_{n \in \{1,2,..,N\}}^{b} E(z'_n u_n) \to \sqrt{\frac{b}{N}} C_2 \to 0$$

since the subsampling method requires that

$$\frac{b}{N} \to 0$$

as $b \to \infty$ and $N \to \infty$. So by the subsampling method we have

$$b^{-1/2} Z'_{b,j} u_{b,j} \xrightarrow{d} N(0, \Omega).$$

Since when $\frac{b}{N} \to 0$, $\gamma = 1 - \frac{b}{N} \to 1$. The two resampling methods can be written together when we allow $0 < \gamma \leq 1$. Note that $0 < \gamma < 1$ corresponds the delete- d jackknife and $\gamma = 1$ corresponds the subsampling.
Now, consider the resampling based Anderson-Rubin test,

$$AR(\beta_0)_{N,b,j} = (b-K) u'_{b,j} Z_{b,j} (Z'_{b,j} Z_{b,j})^{-1} Z'_{b,j} u_{b,j} / u'_{b,j} M_{Z_{b,j}} u_{b,j}$$

$$= (b^{-1/2} u'_{b,j} Z_{b,j})(Z'_{b,j} Z_{b,j}/b)^{-1} (b^{-1/2} Z'_{b,j} u_{b,j}) / (\frac{1}{b-K} u'_{b,j} M_{Z_{b,j}} u_{b,j})$$

Note that

$$\frac{1}{b-K} u'_{b,j} M_{Z_{b,j}} u_{b,j}$$

$$= \frac{1}{b-K} u'_{b,j} u_{b,j} - \frac{1}{b-K} u'_{b,j} P_{Z_{b,j}} u_{b,j}$$

By Step 2 and 3, $u'_{b,j} P_{Z_{b,j}} u_{b,j}$ converges to a distribution, so we have

$$\frac{1}{b-K} u'_{b,j} P_{Z_{b,j}} u_{b,j} \xrightarrow{d} 0$$

as $b \to \infty$ and K, the number of instruments, is fixed. By Step 1, we have

$$\frac{1}{b-K} u'_{b,j} u_{b,j} \xrightarrow{p} \sigma_u^2.$$

It follows that $AR(\beta_0)_{N,b,j} \xrightarrow{d} \chi_K^2(\tilde{\varsigma})$ where $\tilde{\varsigma} = (1-\gamma) C' \Omega^{-1} C$, $0 < \gamma \leq 1$. *Q.E.D.*

Proof of Theorem 5 The proof of Theorem 5 is very similar to the proof of Theorem 4. *Q.E.D.*

3.7.2 Appendix 2

Table 3-1: Sizes of the Anderson-Rubin test under near exogeneity

	$covZ_iu_i =$	$covZ_iu_i =$	$covZ_iu_i =$
$\Pi = 1$	9.9	25.1	38.2
$\Pi = 0.1$	10.1	23.5	38.2
$\Pi = 0$	9.1	22.5	38.3

Note: The data generating process of the simulation is based on Λ *and* $covV_iu_i = 0.25$ *. The sample size is* $N = 80$ *and the nominal size is* 10% *.* Π *is an indicator of the quality of instruments.* $\Pi = \{1, 0.1, 0\}$ *represents strong instruments, weak instruments and nonidentification respectively.*

Table 3-2: Sizes of the Kleibergen test under near exogeneity

	$covZ_iu_i =$	$covZ_iu_i =$	$covZ_iu_i =$
$\Pi = 1$	10.6	24.0	37.8
$\Pi = 0.1$	9.1	24.9	41.1
$\Pi = 0$	9.2	24.5	40.4

Note: The data generating process of the simulation is based on Λ *and* $covV_iu_i = 0.25$ *. The sample size is* $N = 80$ *and the nominal size is* 10% *.* Π *is an indicator of the quality of instruments.* $\Pi = \{1, 0.1, 0\}$ *represents strong instruments, weak instruments and nonidentification respectively.*

Table 3-3: Sizes of the resampling based AR test under near exogeneity

	$covZ_iu_i =$	$covZ_iu_i =$	$covZ_iu_i =$
Strong Instruments ($\Pi = 1$)			
$b = 5$	6.2	13.6	27.8
$b = 10$	4.5	13.4	26.8
$b = 15$	4.0	10.0	24.5
$b = 20$	1.9	9.9	14.9
$b = 25$	1.4	6.3	12.3
$b = 30$	0.5	4.3	8.5
$b = 40$	0.2	1.1	4.2
Weak Instruments ($\Pi = 0.1$)			
$b = 5$	6.8	16.3	27.8
$b = 10$	5.7	13.3	25.5
$b = 15$	3.1	11.4	19.4
$b = 20$	2.1	9.7	16.2
$b = 25$	1.8	5.4	13.9
$b = 30$	1.2	4.2	9.0
$b = 40$	0.3	1.6	2.3
Nonidentification ($\Pi = 0$)			
$b = 5$	5.2	15.7	28.5
$b = 10$	5.1	15.5	25.5
$b = 15$	3.2	10.5	20.9
$b = 20$	1.9	9.5	17.0
$b = 25$	1.4	6.9	12.6
$b = 30$	1.1	5.0	9.7
$b = 40$	0.3	1.2	4.1

Note: The data generating process of the simulation is based on Λ *and* $covV_iu_i = 0.25$ *. The sample size is* $N = 80$ *and the nominal size is* 10% *.* b *represents the block size used in simulations. We compute actual sizes when* $b = \{5, 10, 15, 20, 25, 30, 40\}$ *.*

Table 3-4: Sizes of the resampling based K test under near exogeneity

	$covZ_iu_i =$	$covZ_iu_i =$	$covZ_iu_i =$
Strong Instruments ($\Pi = 1$)			
$b = 5$	6.9	15.5	27.4
$b = 10$	4.7	14.5	27.9
$b = 15$	2.9	12.0	23.6
$b = 20$	2.4	8.4	17.4
$b = 25$	1.8	6.2	13.3
$b = 30$	1.1	3.5	8.4
$b = 40$	0.1	1.3	3.9
Weak Instruments ($\Pi = 0.1$)			
$b = 5$	4.4	13.3	27.5
$b = 10$	4.4	15.9	24.3
$b = 15$	3.8	11.4	24.2
$b = 20$	3.2	10.2	16.0
$b = 25$	1.2	6.1	14.4
$b = 30$	0.9	3.7	8.9
$b = 40$	0.0	1.6	4.5
Nonidentification ($\Pi = 0$)			
$b = 5$	5.3	14.9	26.3
$b = 10$	4.9	13.9	26.2
$b = 15$	3.6	9.2	20.4
$b = 20$	2.8	7.8	15.2
$b = 25$	1.9	6.8	14.7
$b = 30$	1.0	2.5	8.4
$b = 40$	0.1	0.7	3.1

Note: The data generating process of the simulation is based on Λ *and* $covV_iu_i = 0.25$ *. The sample size is* $N = 80$ *and the nominal size is* 10% *.* b *represents the block size used in simulations. We compute actual sizes when* $b = \{5, 10, 15, 20, 25, 30, 40\}$ *.*

4.0 GMM WITH WEAK IDENTIFICATION AND NEAR EXOGENEITY

This chapter studies the asymptotic properties of estimation and inference with weak identification and near exogeneity in a GMM framework with instrumental variables. GMM is a natural extension of a linear simultaneous equations model which allows a set of nonlinear and non-differentiable equations. The technique used in Chapter 1 which is mainly based on mean value theorem and the classic central limit theorem cannot be applied into a nonlinear and non-differentiable environment. We can benefit from empirical process theory and the functional central limit theorem to establish large sample properties. We obtained limiting results under weak identification and near exogeneity of general GMM estimators and some specific GMM estimators, such as one-step GMM estimator, two-step GMM estimator and continuous updating estimator. We also examine the asymptotic properties of the Anderson-Rubin type and the Kleibergen type tests under weak identification and near exogeneity.

This chapter is organized as follows. Section 4.1 decribes the model and assumptions. Section 4.2 examines the limiting results of GMM estimators under near exogeneity and weak identification. Section 4.3 studies inference under near exogeneity and weak identification, and Section 4.4 concludes. Appendix is included in Section 4.5.

4.1 THE MODEL AND ASSUMPTIONS

In this chapter, we consider a GMM framework with instrumental variables under weak identification and near exogeneity. Let $\theta = (\alpha', \beta')$ be an m -dimensional unknown parameter vector with true value $\theta_0 = (\alpha_0', \beta_0')'$ in the interior of the compact parameter space Θ . The true value θ_0 satisfies some conditional moment restrictions which can be explicitly written as

$$E\phi_i(\theta_0) = E[h(Y_i, \theta_0) \otimes Z_i] = C/\sqrt{N},$$

where $h(.)$ is a real valued $H \times 1$ vector of functions, Z_i is a $K \times 1$ vector of instrumental variables, and Y_i is the observation which possibly contains endogenous variables, for $i = 1, 2, \ldots, N$ and $HK \geq m$. The C is a $HK \times 1$ vector of constants. When C is a vector of zeros, this is the GMM model with instrumental variables defined by Stock and Wright (2000). When C is not all zeros, Equation (105) defines the GMM model with near exogeneity. The degree of near exogeneity is local to zero. When the sample size N grows to large, the correlation between $h(.)$ and the instruments Z_i tends to zero. The linear simultaneous equations model defined in Chapter 1 is a special case of Equation (105), where

$$E\phi_i(\theta_0) = E[Z_i'(y_i - Y_i\theta_0)] = C/\sqrt{N}.$$

So $h(.) = y_i - Y_i\theta_0$ is a linear function and $Y_i = (y_i, Y_i)$ contains only endogenous variables. But in this chapter, the $h(.)$ can be a set of general nonlinear functions with possible non-differentiability.

We follow Stock and Wright (2000)'s paper to consider a mixed case in which a subset of θ, say α, is weakly identified. Let $\Theta = A \times B$, where $\alpha \in A$ is an $m_1 \times 1$ vector, $\beta \in B$ is an $m_2 \times 1$ vector, and $m_1 + m_2 = m$. Also, let $\widetilde{m}_N(\alpha, \beta) = EN^{-1} \sum_{i=1}^{N} \phi_i(\alpha, \beta)$. Now, we can utilize the following identity,

$$\widetilde{m}_N(\alpha, \beta) = \widetilde{m}_N(\alpha_0, \beta_0) + \widetilde{m}_{1N}(\alpha, \beta) + \widetilde{m}_{2N}(\beta)$$

where

$$\widetilde{m}_{1N}(\alpha, \beta) = \widetilde{m}_N(\alpha, \beta) - \widetilde{m}_N(\alpha_0, \beta)$$

and

$$\widetilde{m}_{2N}(\beta) = \widetilde{m}_N(\alpha_0, \beta) - \widetilde{m}_N(\alpha_0, \beta_0)$$

The identification of θ requires whether the moment restrictions can be satisfied uniquely. If β is strictly identified, then $\widetilde{m}_{2N}(\beta)$ should be large when $\beta \neq \beta_0$. However, $\widetilde{m}_{1N}(\alpha, \beta)$

should be close to zero when $\alpha \neq \alpha_0$ and $\beta = \beta_0$ if α is weakly identified. We can use a local to zero model to define the weak identification of the α ,

$$\widetilde{m}_N(\alpha, \beta) - \widetilde{m}_N(\alpha_0, \beta) = m_{1N}(\alpha, \beta)/\sqrt{N}$$

where $m_{1N}(\alpha, \beta) : A \times B \to R^{HK}$ is a set of continuous functions such that $m_{1N}(\theta) \to m_1(\theta)$ uniformly on Θ as N grows to large. The $m_1(\theta) : A \times B \to R^{HK}$ is a set of continuous functions and is bounded on Θ . Also, let $\widetilde{m}_{2N}(\beta) : B \to R^{HK}$ be a set of continuous functions such that $\widetilde{m}_{2N}(\beta) \to m_2(\beta)$ uniformly on B as N grows to large, where $m_2(\beta) : B \to R^{HK}$ is a set of continuous functions such that $m_2(\beta_0) = 0$ and $m_2(\beta) \neq 0$ for $\beta \neq \beta_0$. By taking into account a joint case of near exogeneity and weak identification, Equation (135) can be rewritten as

$$\begin{aligned}\widetilde{m}_N(\alpha, \beta) &= \widetilde{m}_N(\alpha_0, \beta_0) + \widetilde{m}_{1N}(\alpha, \beta) + \widetilde{m}_{2N}(\beta) \\ &= C/\sqrt{N} + m_{1N}(\alpha, \beta)/\sqrt{N} + \widetilde{m}_{2N}(\beta)\end{aligned}$$

because of Equation (105). When $C = 0$, we can obtain the result of Stock and Wright (2000), in which case they don't consider the problem of near exogeneity. Now, we can give assumptions that formally define near exogeneity and weak identification.

Assumption 1 The true parameter $\theta_0 = (\alpha_0', \beta_0')'$ is in the interior of the compact space $\Theta = A \times B$, $A \subset R^{m_1}$, $B \subset R^{m_2}$, and $m = m_1 + m_2$. The true parameter θ_0 satisfies the moment conditions defined by Equation (105).

Assumption 2

$$EN^{-1} \sum_{i=1}^{N} \phi_i(\alpha, \beta) = C/\sqrt{N} + m_{1N}(\alpha, \beta)/\sqrt{N} + \widetilde{m}_{2N}(\beta), \text{ where}$$

(2.1) $m_{1N}(\theta) \to m_1(\theta)$ uniformly on Θ, $m_1(\theta_0) = 0$, and $m_1(\theta)$ is continuous in θ and is bounded on Θ;

(2.2) $\widetilde{m}_{2N}(\beta) \to m_2(\beta)$ uniformly on Θ, $m_2(\beta) = 0$ if and only if $\beta = \beta_0$. Define $R(\beta) = \partial m_2(\beta)/\partial\beta'$ which is a $HK \times m_2$ matrix. $R(\beta)$ is continuous in β and $R(\beta_0)$ has a full column rank.

We can apply the above assumptions into the linear simultaneous equations model defined in Chapter 2. In Chapter 2, all parameters in θ are weakly identified. The identity defined above can be rewritten as

$$\begin{aligned}\widetilde{m}_N(\theta) &= \widetilde{m}_N(\theta_0) + [\widetilde{m}_N(\theta) - \widetilde{m}_N(\theta_0)] \\ &= \widetilde{m}_N(\theta_0) + m_{1N}(\theta)/\sqrt{N}\end{aligned}$$

where $\widetilde{m}_N(\theta_0) = EN^{-1}\sum_{i=1}^{N}\phi_i(\theta_0) = C/\sqrt{N}$ by the near exogeneity in Assumption 2. In the linear simultaneous equations model,

$$\begin{aligned}EN^{-1}\sum_{i=1}^{N}\phi_i(\theta) &= EN^{-1}\sum_{i=1}^{N}[Z_i'(y_i - Y_i\theta)] \\ &= EN^{-1}\sum_{i=1}^{N}[Z_i'(y_i - Y_i\theta_0) - Z_i'Y_i(\theta - \theta_0)] \\ &= EN^{-1}\sum_{i=1}^{N}\{[Z_i'(y_i - Y_i\theta_0)] - [Z_i'Z_i\Pi(\theta - \theta_0)]\}\end{aligned}$$

By the above equation, we obtain

$$EN^{-1}\sum_{i=1}^{N}[Z_i'(y_i - Y_i\theta_0)] = C/\sqrt{N}$$

Since $\Pi = \Pi_N = C_1/\sqrt{N}$ defined by Assumption ID in Chapter 2, we have

$$\widetilde{m}_N(\theta) = C/\sqrt{N} + m_{1N}(\theta)/\sqrt{N}$$

where $m_{1N}(\theta) = EN^{-1}\sum_{i=1}^{N}[Z_i'Z_iC_1(\theta - \theta_0)]$. The first term in (245) is due to near

exogeneity and the second term is used to define the weak identification of θ .

Next, we consider the GMM estimator that minimizes the objective function $S_N(\theta\ ,\ \bar{\theta}_N(\theta))$ for $\theta \in \Theta$, where

$$S_N(\theta, \bar{\theta}_N(\theta)) \ = [N^{-1/2} \sum_{i=1}^{N} \phi_i(\theta)]' W_N(\bar{\theta}_N(\theta))[N^{-1/2} \sum_{j=1}^{N} \phi_j(\theta)]$$

where $W_N(\bar{\theta}_N(\theta))$ is a positive definite $HK \times HK$ weighting matrix and bounded in probability. Different GMM estimators depend upon the adoption of different weighting matrix. For a one-step GMM estimator, the weighting matrix is usually an identity matrix so $W_N(\bar{\theta}_N(\theta))$ doesn't depend upon the data and the unknown parameter θ . For a two-step efficient GMM estimator (Hansen, 1982), the weighting matrix is computed by using a one-step GMM estimator. For a continuously updating GMM estimator (Hansen, Heaton and Yaron, 1996), the weighting matrix is changed with each choice of the unknown parameter θ , so $W_N(\bar{\theta}_N(\theta))$ can be written as $W_N(\theta)$. In order to establish the large sample properties of the GMM estimators, we need the uniform convergence of the weighting matrix $W_N(\theta)$. This is also the assumption used by Stock and Wright (2000).

Assumption 3 $W_N(\theta) \xrightarrow{p} W(\theta)$ uniformly on Θ , where $W(\theta)$ is a $HK \times HK$ symmetric positive definite matrix and is continuous in θ .

Next, following Andrews (1994) and Stock and Wright (2000), we define an empirical process $\Psi_N(\theta)$ by

$$\Psi_N(\theta) \ = \ N^{-1/2} \sum_{i=1}^{N} [\phi_i(\theta) - E\phi_i(\theta)] \ \text{ for } \theta \in \Theta$$

Note that $\phi_i(\theta) = \phi_i(Y_i, Z_i, \theta)$ where Y_i and Z_i are independent observations. $\phi_i(\theta)$ can be regarded as a class of R^{HK} valued functions defined on Y_i and Z_i indexed by $\theta \in \Theta$. Let " $\Rightarrow$ " denote weak convergence of a sequence of empirical processes. By Andrews (1994) and Vaart and Wellner (1996), weak convergence of the empirical process in Equation (265) can be defined as

$$\Psi_N(\theta) \Rightarrow \Psi(\theta) \text{ if } E^* f(\Psi_N(.)) \rightarrow Ef(\Psi(.))$$

for all bounded, uniformly continuous real functions f on $\mathbf{B}(\mathbf{\Theta})$, where $\mathbf{B}(\mathbf{\Theta})$ is the set of all continuous, bounded functions $f: \Theta \rightarrow R$. Note that " E^* " is the expectation over the empirical process. Let $\Omega(\theta_1, \theta_2) = \lim_{N\rightarrow\infty} E\Psi_N(\theta_1)\Psi_N(\theta_2)'$. The following assumption of weak convergence is mainly based on Pollard (1984, 1990), Andrews (1994) and Vaart and Wellner (1996). It's similar to Assumption A and B used in Stock and Wright (2000).

Assumption 4 $\Psi_N(\theta) \Rightarrow \Psi(\theta)$, where $\Psi(\theta)$ is a Gaussian limit stochastic process on Θ with zero mean and covariance $\Omega(\theta_1, \theta_2)$.

Assumption 4 is a kind of high level assumption which follows from three sufficient conditions (Andrews, 1994): (1) Θ is a totally bounded space; (2) finite dimensional convergence holds: $\forall(\theta_1, \ldots, \theta_J) \in \Theta$, $(\Psi_N(\theta_1)', \ldots, \Psi_N(\theta_J)')'$ converges in distribution; (3) $\Psi_N(\theta)$ is stochastic equicontinuity. Condition (1) is satisfied by Assumption 1 that Θ is a compact space. Condition (2) is easily to verified by multivariate central limit theorem. For example, we can use univariate triangular array central limit theorem (Liapunov Theorem, see Davidson, 1994) to obtain the normal limit of the stochastic process $\Psi_N(\theta)$ at $\theta = \theta_0$, say

$$\Psi_N(\theta_0) \xrightarrow{d} N(0, \Omega(\theta_0, \theta_0)$$

by imposing the moment condition such that $E \mid \phi_i(\theta_0) \mid^{2+\delta} < \Delta < \infty$ for some $\delta > 0$.

For the finite dimensional convergence, we can assume a similar moment condition which holds uniformly on Θ . Condition (3) stochastic equicontinuity relies on a condition which is referred as entropy condition (Pollard, 1990). By Theorem 1 and 2 in Andrews (1994), $\phi_i(\theta)$ falls into a type II class of functions so that the Pollard's entropy condition follows from the Lipschitz continuity. To be summarized, Assumption 4 follows from the following primitive assumptions.

(i) Θ is a compact parameter space;
(ii) $\phi_i(\theta)$ is independent;
(iii) $E \mid \phi_i(\theta) \mid^{2+\delta} < \Delta < \infty$ uniformly over Θ for some $\delta > 0$;
(vi) Lipschitz in θ : $\mid \phi_i(\theta_1) - \phi_i(\theta_2) \mid \leq B_i(.)\|\theta_1 - \theta_2\| \quad \forall \theta_1, \theta_2 \in \Theta$, and $B_i(.)$ satisfies $\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} E[B_i(.)^{2+\delta}] < \infty$ for some $\delta > 0$.

Assumption (i) implies totally boundedness. Assumptions (ii) and (iii) imply finite dimensional convergence. Assumptions (i) and (vi) imply stochastic equicontinuity. It's very easy to verify that the $\phi_i(\theta)$ defined in the linear simultaneous equations model in Chapter 1 satisfies these assumptions.

4.2 ESTIMATION: LIMITING RESULTS OF GMM ESTIMATORS

In this section, we derive the asymptotic results of GMM estimators under near exogeneity and weak identification. We firstly derive general limiting results of GMM estimators and then derive limiting results of some specific GMM estimators, such as one-step estimator, two-step efficient estimator and continuously updating estimator. In each case, we examine the limiting results of the weakly identified parameter α and the well identified parameter β .

4.2.1 General Limiting Results of GMM Estimators

We derive the general asymptotic results of GMM estimators in this subsection. First, we examine the limiting results of the well identified parameter β. The following lemma shows that the GMM estimator $\widehat{\beta}$ is consistent under near exogeneity and the convergence rate is square root of the sample size N.

Lemma 1 $\sqrt{N}(\widehat{\beta} - \beta_0) = O_p(1)$.

All proofs are given in the appendix.

Lemma 1 shows that near exogeneity doesn't affect the convergence of a well identified parameter. Intuitively, the drift term in Equation (105) shrinks toward zero as the sample size N grows to large. We have a similar story in the linear case. In the linear simultaneous equations model defined in Chapter 1, when there only exists the problem of near exogeneity, both the TSLS estimator and the LIML estimator are consistent. However, situations are a little complicated in this chapter. There are two parameters, of which one is weakly identified and the other is well identified. One natural question is whether the weakly identified parameter α affect the limiting results of the well identified parameter β. A joint limiting result of α and β is necessary to answer such a question. The following theorem gives the joint limits of both parameters under near exogeneity and weak identification for a general GMM estimator.

Theorem 1 Suppose that Assumptions 1-4 hold,
then

$$(\widehat{\alpha}, \sqrt{N}(\widehat{\beta} - \beta_0)) \xrightarrow{d} (a^*, b^*)$$

where

$$a^* = \arg\min_{\alpha \in A} S^*(\alpha; \bar{\theta}(\alpha, \beta_0))$$

$$\begin{aligned} b^* &= -[R(\beta_0)' W(\bar{\theta}(a^*, \beta_0) R(\beta_0)]' R(\beta_0)' W(\bar{\theta}(a^*, \beta_0) \\ &\quad \times [\Psi(a^*, \beta_0) + C + m_1(a^*, \beta_0)] \end{aligned}$$

$$\begin{aligned} S^*(\alpha; \bar{\theta}(\alpha, \beta_0)) &= [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)]' M(\alpha, \beta_0, \bar{\theta}(\alpha, \beta_0)) \\ &\quad \times [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)] \end{aligned}$$

where

$$\begin{aligned} M(\alpha, \beta_0, \bar{\theta}(\alpha, \beta_0)) &= W(\bar{\theta}(\alpha, \beta_0) - W(\bar{\theta}(\alpha, \beta_0) R(\beta_0) \\ &\quad \times [R(\beta_0)' W(\bar{\theta}(\alpha, \beta_0) R(\beta_0)]^{-1} \\ &\quad \times R(\beta_0)' W(\bar{\theta}(\alpha, \beta_0) \end{aligned}$$

The above theorem is similar to Theorem 1 in Caner (2005) and is analogous to Theorem 1 in Stock and Wright (2000) and Theorem 2 in Guggenberger and Smith (2005). We can obtain Stock and Wright's result by setting $C = 0$. It's not surprising that $\hat{\alpha}$ is not consistent since α is a weakly identified parameter. Like the case of the linear simultaneous equations model, the estimator of the weakly identified parameter convergences to a nonstandard distribution a^*. The joint limits given in the above theorem can explain why the estimator $\hat{\beta}$ of the well identified parameter also convergence to a nonstandard distribution b^*. The distribution of $\hat{\beta}$ depends on a^* but we cannot estimate α consistently. When we set $C = 0$ and $\alpha = \alpha_0$, Equation (314) can be simplified as

$$\begin{aligned} b^* &= -[R(\beta_0)' W(\bar{\theta}(\alpha_0, \beta_0) R(\beta_0)]' R(\beta_0)' W(\bar{\theta}(\alpha_0, \beta_0) \Psi(\alpha_0, \beta_0) \\ &\xrightarrow{d} N(0, (R(\beta_0)' \Omega^{-1}(\alpha_0, \beta_0) R(\beta_0)) \end{aligned}$$

since $m_1(\alpha_0, \beta_0) = 0$ by Assumption 2 and $\Psi(\alpha_0, \beta_0) \xrightarrow{d} N(0, \Omega(\alpha_0, \beta_0))$ by triangular array central limit theorem. Near exogeneity doesn't affect the convergence rate of β but it

shifts the distribution of the estimator. When the drift term $C \neq 0$, we have

$$b^* \xrightarrow{d} N(C, (R(\beta_0)'\Omega^{-1}(\alpha_0, \beta_0)R(\beta_0))$$

To the weakly identified parameter α, near exogeneity can enlarge the bias term which is obtained by Stock and Wright (2000).

4.2.2 Limiting Results for Specific GMM Estimators

We first consider a one-step GMM estimator with an identity weighting matrix. Denote by $(\widehat{\alpha}_1, \widehat{\beta}_1)$ the one-step GMM estimator which minimizes the following objective function

$$S_{1N}(\theta) = [N^{-1/2} \sum_{i=1}^{N} \phi_i(\theta)]'[N^{-1/2} \sum_{j=1}^{N} \phi_j(\theta)].$$

The following corollary gives the joint limits of $(\widehat{\alpha}_1, \sqrt{N}(\widehat{\beta}_1 - \beta_0))$ under near exogeneity and weak identification.

Corollary 1 Suppose that Assumptions 1, 2, 4 holds, then

$$(\widehat{\alpha}_1, \sqrt{N}(\widehat{\beta}_1 - \beta_0)) \xrightarrow{d} (a_1^*, b_1^*)$$

where

$$a_1^* = \arg\min_{\alpha \in A} S_1^*(\alpha, C)$$

$$b_1^* = -[R(\beta_0)'R(\beta_0)]^{-1}R(\beta_0)'[\Psi(a_1^*, \beta_0) + C + m_1(a_1^*, \beta_0)]$$

$$S_1^*(\alpha, C) = [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)]'M_1(\alpha)[\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)]$$

where

$$M_1(\alpha) = I - R(\beta_0)[R(\beta_0)'R(\beta_0)]^{-1}R(\beta_0)'.$$

The two-step efficient GMM estimator is obtained by using the one-step GMM estimator $(\widehat{\alpha}_1, \widehat{\beta}_1)$ to establish an estimate of the weighting matrix. Denote by $(\widehat{\alpha}_2, \widehat{\beta}_2)$ the two-step efficient GMM estimator which minimizes the following objective function

$$S_{2N}(\theta) = [N^{-1/2} \sum_{i=1}^{N} \phi_i(\theta)]' W_N(\widehat{\alpha}_1, \widehat{\beta}_1)[N^{-1/2} \sum_{j=1}^{N} \phi_j(\theta)]$$

The following corollary establishes the joint limits of $(\widehat{\alpha}_2, \sqrt{N}(\widehat{\beta}_2 - \beta_0))$ under near exogeneity and weak identification.

Corollary 2 Suppose that Assumptions 1-4 hold, then

$$(\widehat{\alpha}_2, \sqrt{N}(\widehat{\beta}_2 - \beta_0)) \xrightarrow{d} (a_2^*, b_2^*)$$

where

$$a_2^* = \arg\min_{\alpha \in A} S_2^*(\alpha, a_1^*, C)$$

$$\begin{aligned} b_2^* = & -[R(\beta_0)'\Omega^{-1}(a_1^*, \beta_0)R(\beta_0)]^{-1} R(\beta_0)'\Omega^{-1}(a_1^*, \beta_0) \\ & \times [\Psi(a_2^*, \beta_0) + C + m_1(a_2^*, \beta_0)] \end{aligned}$$

$$\begin{aligned} S_2^*(\alpha, a_1^*, C) = & [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)]' M_1(\alpha, a_1^*) \\ & \times [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)] \end{aligned}$$

where

$$\begin{aligned} M_1(\alpha, a_1^*) = & \Omega^{-1}(a_1^*, \beta_0) \\ & - \Omega^{-1}(a_1^*, \beta_0)R(\beta_0)[R(\beta_0)'\Omega^{-1}(a_1^*, \beta_0)R(\beta_0)]^{-1} \\ & \times R(\beta_0)'\Omega^{-1}(a_1^*, \beta_0) \end{aligned}$$

In the two-step efficient GMM estimator, the weighting matrix $W_N(\widehat{\alpha}_1, \widehat{\beta}_1)$ is based on the one-

step GMM estimator $\widehat{\alpha}_1$ and $\widehat{\beta}_1$, and so the weighting matrix converge to $\Omega^{-1}(a_1^*, \beta_0)$ in the limiting concentrated objective function $S_2^*(\alpha, a_1^*, C)$.

In the case of the linear simultaneous equations model defined in Chapter 2, when the conditional homoskedasticity of the errors is assumed, the objective function of the two-step efficient GMM estimator can be rewritten as

$$S_{2N}(\theta) = (y - Y\theta)' P_Z (y - Y\theta) / \widehat{\Sigma}_{hh}(\widehat{\theta}_1)$$

where

$$\widehat{\Sigma}_{hh}(\widehat{\theta}_1) = N^{-1} \sum_{i=1}^{N} E\{[h_i(\widehat{\theta}_1) - Eh_i(\widehat{\theta}_1)] \times [h_i(\widehat{\theta}_1) - Eh_i(\widehat{\theta}_1)]'\}$$

and

$$P_Z = Z(Z'Z)^{-1} Z'.$$

In the linear simultaneous equations model, $h_i(\theta) = y_i - Y_i\theta$ and all parameters in θ are weakly identified. Since θ is quadratic in $S_{2N}(\theta)$, we can derive an analytical solution which yields

$$\widehat{\theta} = (Y' P_Z Y)^{-1} (Y' P_Z y)$$

We know this is just the TSLS estimator.

The continuously updating estimator is obtained when the weighting matrix is continuously updated at the parameter value θ . Denote by $(\widehat{\alpha}_c, \widehat{\beta}_c)$ the continuously updating estimator that minimizes the following objective function

$$S_{cN}(\theta) = [N^{-1/2} \sum_{i=1}^{N} \phi_i(\theta)]' W_N(\theta) [N^{-1/2} \sum_{j=1}^{N} \phi_j(\theta)]$$

The following corollary establishes the joint limits of the continuously updating estimator $(\widehat{\alpha}_c, \widehat{\beta}_c)$ under near exogeneity and weak identification.

Corollary 3 Suppose that Assumptions 1-4 hold, then

$$(\widehat{\alpha}_c, \sqrt{N}(\widehat{\beta}_c - \beta_0)) \xrightarrow{d} (a_c^*, b_c^*)$$

where

$$a_c^* = \arg\min_{\alpha \in A} S_c^*(\alpha, C)$$

$$\begin{aligned} b_c^* &= -[R(\beta_0)'\Omega^{-1}(a_c^*, \beta_0)R(\beta_0)]^{-1}R(\beta_0)'\Omega^{-1}(a_c^*, \beta_0) \\ &\quad \times [\Psi(a_c^*, \beta_0) + C + m_1(a_c^*, \beta_0)] \end{aligned}$$

$$\begin{aligned} S_c^*(\alpha, C) &= [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)]'\Omega^{-1}(a_c^*, \beta_0) \\ &\quad \times \{I - R(\beta_0)[R(\beta_0)'\Omega^{-1}(a_c^*, \beta_0)R(\beta_0)]^{-1}R(\beta_0)'\Omega^{-1}(a_c^*, \beta_0)\} \\ &\quad \times [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)]. \end{aligned}$$

Consider a special case of Corollary 3: the linear simultaneous equations model with all weakly identified parameters and conditional homoskedasticity defined in Chapter 2. Since

$$\phi_i(\theta) = Z_i'(y_i - Y_i\theta)$$

and

$$W_N(\theta) = [N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_i(\theta)\phi_j(\theta)']^{-1},$$

the objective function $S_{cN}(\theta)$ defined in (545) can be simplified as

$$\begin{aligned} S_{cN}(\theta) &= [N^{-1/2}\sum_{i=1}^{N}\phi_i(\theta)]'[N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_i(\theta)\phi_j(\theta)']^{-1} \\ &\quad \times [N^{-1/2}\sum_{j=1}^{N}\phi_j(\theta)] \\ &= (y - Y\theta)'Z(Z'Z)^{-1}Z'(y - Y\theta)/u(\theta)'u(\theta) \\ &= N[1 + \kappa^{-1}(\theta)]^{-1} \end{aligned}$$

where

$$u(\theta) = y - Y\theta$$

$$\kappa(\theta) = (y - Y\theta)' P_z (y - Y\theta)/(y - Y\theta)' M_z (y - Y\theta)$$

and

$$M_Z = I - P_Z.$$

Note that the above equation is obtained since we have

$$T^{-1}(y - Y\theta)' M_z (y - Y\theta) \xrightarrow{p} u(\theta)' u(\theta)$$

The continuously updating estimator in the linear case is identical to minimize $\kappa(\theta)$, which is just the LIML estimator; see Davidson and MacKinnon (1993).

4.3 INFERENCE WITH NEAR EXOGENEITY AND WEAK IDENTIFICATION

In a GMM framework with instrumental variables, we want to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ under near exogeneity and weak identification. Staiger and Wright (2000) examined several conventional test statistics under weak identification, such as Wald statistic and likelihood ratio statistic. These conventional test statistics do not work in general under weak identification. The exogeneity tests of instruments, like J-test (Hansen, 1982; Newey, 1985), cannot be valid in general under weak identification either.

In this section, we firstly consider some robust test statistics which have been recently developed against weak identification in the literature, and then examine their performance under near exogeneity.

We first consider an Anderson-Rubin type test proposed by Stock and Wright (2000). The test is given by

$$S_N(\theta_0;\theta_0) = [N^{-1/2} \sum_{i=1}^{N} \phi_i(\theta_0)]' W_N(\theta_0) [N^{-1/2} \sum_{j=1}^{N} \phi_j(\theta_0)].$$

Since the moment function is generally nonlinear, it's easier to work on the objective function

rather than on the estimator as we did in the case of the linear simultaneous equations model. The Anderson-Rubin type test is just the objective function $S_{cN}(\theta)$ of the continuously updating estimator when $\theta = \theta_0$. Since it utilizes the objective function $S_{cN}(.)$, it was called " S statistic" by Stock and Wright (2000). The S statistic is robust to weak identification because the test itself is asymptotically pivotal and convergence in distribution to a chi-square distribution under the null hypothesis. Note that we cannot establish an Anderson-Rubin type test based on the objective function of the two-step GMM estimator. The objective function of the two-step GMM estimator is not asymptotically pivotal because the weighting matrix in the objective function is derived through the one-step estimator, which is not consistent under weak identification.

To examine the asymptotic property of the S statistic under near exogeneity, we can work under a much weaker assumption than Assumption 4. The following theorem summarizes the asymptotic result of the S statistic under near exogeneity.

Theorem 2 Suppose Assumptions 1-3 hold under the null hypothesis of $\theta = \theta_0$ *, then*

$$S_N(\theta_0;\theta_0) \xrightarrow{d} \chi^2_{HK}(C'\Omega^{-1}(\theta_0;\theta_0)C)$$

where $\chi^2_{HK}(C'\Omega^{-1}(\theta_0;\theta_0)C)$ *is a noncentral chi-square distribution with noncentral parameter* $C'\Omega^{-1}(\theta_0;\theta_0)C$ *and the degree of freedom* HK *.*

Theorem 2 shows that the S statistic is not asymptotically pivotal under near exogeneity. The limit of the test statistic depends on the nuisance unknown parameter C which comes from near exogeneity. We obtain a chi-square distribution with degree of freedom HK when we set $C = 0$. It leads to a size distortion under near exogeneity when we use critical values from the chi-square distribution. In empirical practice, it'll overreject a true hypothesis.

Kleibergen (2005) proposes a GMM version K statistic. The K statistic is also based on the objective function of the continuously updating GMM estimator. To establish the limits of the K statistic, we need two more assumptions. Denote by $q_i(\theta_0)$ the first order derivative of $\phi_i(\theta)$ with respect to θ which is evaluated at $\theta = \theta_0$, and let

$$J_\theta(\theta_0) = \lim_{N\to\infty} E[N^{-1} \sum_{i=1}^{N} q_i(\theta_0)]$$

Assumption 5 Let

$$q_{i,j}(\theta_0) = \partial\phi_i(\theta)/\partial\theta_j' \mid_{\theta=\theta_0} \quad j = 1, 2, \ldots, m.$$

and $q_i(\theta_0) = (q_{i,1}'(\theta_0), q_{i,2}'(\theta_0), \ldots, q_{i,m}'(\theta_0))'$. We assume the following limits hold jointly

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \begin{pmatrix} \phi_i(\theta_0) - E[\phi_i(\theta_0)] \\ q_i(\theta_0) - E[q_i(\theta_0)] \end{pmatrix} \xrightarrow{d} (\Psi_\phi', \Psi_q')'$$

where

$$\begin{pmatrix} \Psi_\phi \\ \Psi_q \end{pmatrix} \sim N(0, V(\theta))$$

and $V(\theta)$ is a positive semi-definite symmetric $(HK + mHK) \times (HK + mHK)$ matrix

$$V(\theta) = \begin{pmatrix} V_{\phi\phi} & V_{\phi q} \\ V_{q\phi} & V_{qq} \end{pmatrix}$$

and

$$V(\theta) = \lim_{N\to\infty} EN^{-1} \sum_{i=1}^{N} \sum_{l=1}^{N} \begin{pmatrix} \phi_i(\theta_0) - E[\phi_i(\theta_0)] \\ q_i(\theta_0) - E[q_i(\theta_0)] \end{pmatrix} \begin{pmatrix} \phi_l(\theta_0) - E[\phi_i(\theta_0)] \\ q_l(\theta_0) - E[q_i(\theta_0)] \end{pmatrix}'.$$

Assumption 6 Assume that the estimator of the covariance matrix $V(\theta_0)$ and the estimator of the derivative of $W(\theta_0) = V_{\phi\phi}^{-1}(\theta_0)$ with respect to θ have the limits that hold jointly

$$\widehat{V}(\theta_0) \xrightarrow{p} V(\theta_0)$$

and

$$\partial vec(\widehat{V}_{\phi\phi}(\theta_0))/\partial\theta' \xrightarrow{p} \partial vec(V_{\phi\phi}(\theta_0))/\partial\theta'$$

where

$$V_{\phi\phi}(\theta_0) = \lim_{N\to\infty} E\{N^{-1}\sum_{i=1}^{N}\sum_{l=1}^{N}(\phi_i(\theta_0) - E[\phi_i(\theta_0)])(\phi_i(\theta_0) - E[\phi_i(\theta_0)])'\}.$$

The K statistic is based on the first order derivative of Equation (670) with respect to θ. The K statistic is given by

$$K(\theta_0) = \frac{1}{4N}(\partial S_N(\theta_0;\theta_0)/\partial\theta)[\widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)\widehat{D}_N(\theta_0)]^{-1}$$
$$\times (\partial S_N(\theta_0;\theta_0)/\partial\theta)'$$

where

$$\frac{1}{2}\partial S_N(\theta_0;\theta_0)/\partial\theta = \phi_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)\widehat{D}_N(\theta_0)$$

$$\widehat{D}_N(\theta_0) = [q_{N,1}(\theta_0) - \widehat{V}_{q\phi,1}(\theta_0)\widehat{V}_{\phi\phi}^{-1}(\theta_0)\phi_N(\theta_0)\ \ldots$$
$$\ldots q_{N,m}(\theta_0) - \widehat{V}_{q\phi,m}(\theta_0)\widehat{V}_{\phi\phi}^{-1}(\theta_0)\phi_N(\theta_0)]$$

and $\widehat{V}_{q\phi}(\theta_0) = (\widehat{V}_{q\phi,1}(\theta_0)', \widehat{V}_{q\phi,2}(\theta_0)', \ldots, \widehat{V}_{q\phi,m}(\theta_0))'$.

Note that $\widehat{D}_N(\theta_0)$ is a consistent estimator of $J_\theta(\theta_0)$ even in the case of weak identification. Either under strong identification or weak identification, the K statistic is an asymptotically pivotal distribution conditional on $\widehat{D}_N(\theta_0)$. Because of the asymptotic independence between $\widehat{D}_N(\theta_0)$ and Ψ_ϕ , the K statistic converges unconditionally to a chi-square distribution with degree of freedom m under weak identification. The following theorem summarizes the asymptotic results of the K statistic under near exogeneity and weak identification.

Theorem 3 Suppose that Assumptions 1, 2, 5 and 6 hold under the null hypothesis of $\theta = \theta_0$ *, then*

$$K(\theta_0) \xrightarrow{d} (\xi + \Xi(C))'(\xi + \Xi(C))$$

where

$$\xi \sim N(0, I_{HK})$$

$$\Xi(C) = [D'V_{\phi\phi}^{-1}(\theta_0)D]^{-1/2}D'V_{\phi\phi}^{-1}(\theta_0)C$$

and D *is the limit of* $\gamma(N)\widehat{D}_N(\theta_0)$ *, and further* D *varies when*

(i) θ *is well identified,* $D \xrightarrow{d} C_q$

(ii) θ *is weakly identified,* $D \xrightarrow{d} C_q + \Psi_{q.\phi}$

(iii) θ *is nonidentified,* $D \xrightarrow{d} \Psi_{q.\phi}$

where $C_q = J_\theta(\theta_0)$ *which has a fixed full rank value, and* $\Psi_{q.\phi}$ *is a limiting distribution such that*

$$N^{-1/2}vec[\widehat{D}_N(\theta_0) - J_\theta(\theta_0)] \xrightarrow{d} \Psi_{q.\phi}.$$

Theorem 3 shows that the K statistic converges to a nonstandard distribution under near exogeneity. The nonstandard distribution is a quadratic form of the sum of a standard normal variable ξ and the drift term $\Xi(C)$ which comes from near exogeneity. When the identification condition varies, we obtain different limits of $\Xi(C)$. We can obtain a chi-square distribution with degree of freedom m when $C = 0$. So Theorem 3 provides a general result. Theorem 3 also implies that inference based on the critical value from chi-square distribution can result in a large size distortion.

4.4 CONCLUSIONS

This chapter studies the asymptotic properties of estimation and inference under near exogeneity and weak identification in a GMM framework with instrumental variables. We derive the limits of the one-step GMM estimator, the efficient two-step GMM estimator and the continuously updating estimator under near exogeneity and weak identification. We consider a mixed case where some parameters are weakly identified and others are well identified. The GMM estimators of the well identified parameters are consistent but converge to a nonstandard distribution. In all cases, near exogeneity can bring a relatively large asymptotic bias for GMM estimators compared to the case where only weak identification occurs. We show that the Anderson-Rubin type S statistic and the Kleibergen type K statistic are no longer asymptotically pivotal under near exogeneity. It leads to a serious size distortion when using critical values from chi-square distribution.

4.5 APPENDIX

Proof of Lemma 1 First, we show that β is consistent. Consider the objective function $S_N(\theta, \bar{\theta}_N(\theta))$ the first term can be rewritten as

$$N^{-1/2} \sum_{i=1}^{N} \phi_i(\theta) = N^{-1/2} \sum_{i=1}^{N} [\phi_i(\theta) - E\phi_i(\theta)] + N^{-1/2} \sum_{i=1}^{N} E\phi_i(\theta).$$

The first term converges to $\Psi(\theta)$ by Assumption 4 and the second term can be rewritten as

$$\begin{aligned} N^{-1/2} \sum_{i=1}^{N} E\phi_i(\theta) &= \sqrt{N} E N^{-1} \sum_{i=1}^{N} \phi_i(\theta) \\ &\to C + m_1(\alpha, \beta) + \sqrt{N} m_2(\beta) \end{aligned}$$

by Assumption 2. By Assumption 3, we have

$$\begin{aligned} S_N(\theta, \bar{\theta}_N(\theta)) &\xrightarrow{p} [\Psi(\theta) + C + m_1(\alpha, \beta) + \sqrt{N} m_2(\beta)]' W(\bar{\theta}(\theta)) \\ &\times [\Psi(\theta) + C + m_1(\alpha, \beta) + \sqrt{N} m_2(\beta)]. \end{aligned}$$

Scale the above equation by N^{-1}, we obtain

$$N^{-1}S_N(\theta,\bar{\theta}_N(\theta)) \xrightarrow{p} m_2(\beta)'W(\bar{\theta}(\theta))m_2(\beta)$$

uniformly in β. Since $W(\bar{\theta}(\theta))$ is positive definite by Assumption 3 and $m_2(\beta) = 0$ if and only if $\beta = \beta_0$, the consistency of β follows by the continuity of the $\arg\min$ operator. The rate of convergence follows from the proof of Lemma A1 in Stock and Wright(2000). *Q.E.D.*

Proof of Theorem 1 To derive the limiting results in the theorem, we work on the objective function $S_N(\alpha,\beta,\bar{\theta}_N(\theta))$ directly. First, we define

$$b = \sqrt{N}(\beta - \beta_0).$$

By Lemma 1, we know that $b = O_p(1)$. The objective function then can be written as

$$\begin{aligned} S_N(\alpha,\beta,\bar{\theta}_N(\theta)) &= S_N(\alpha,\beta_0 + b/\sqrt{N},\bar{\theta}_N(\theta)) \\ &= [N^{-1/2}\sum_{i=1}^{N}\phi_i(\theta)]'W_N(\bar{\theta}_N(\theta))[N^{-1/2}\sum_{j=1}^{N}\phi_j(\theta)]. \end{aligned}$$

The first and last terms in above equation can be written as

$$\begin{aligned} & N^{-1/2}\sum_{i=1}^{N}\phi_i(\alpha,\beta_0 + b/\sqrt{N}) \\ &= N^{-1/2}\sum_{i=1}^{N}[\phi_i(\alpha,\beta_0 + b/\sqrt{N}) - E\phi_i(\alpha,\beta_0 + b/\sqrt{N})] + N^{-1/2}\sum_{i=1}^{N}E\phi_i(\alpha,\beta_0 + b/\sqrt{N}). \end{aligned}$$

By Assumption 4 and Lemma 1, we have

$$N^{-1/2}\sum_{i=1}^{N}[\phi_i(\alpha,\beta_0 + b/\sqrt{N}) - E\phi_i(\alpha,\beta_0 + b/\sqrt{N})] \Rightarrow \Psi(\alpha,\beta_0).$$

The second term in Equation (920) can be written as

$$\begin{aligned} N^{-1/2}\sum_{i=1}^{N}E\phi_i(\alpha,\beta_0 + b/\sqrt{N}) &= \sqrt{N}EN^{-1}\sum_{i=1}^{N}\phi_i(\alpha,\beta_0 + b/\sqrt{N}) \\ &= C + m_{1N}(\alpha,\beta_0 + b/\sqrt{N}) + \sqrt{N}m_{2N}(\beta_0 + b/\sqrt{N}) \end{aligned}$$

which follows from Assumption 2. Note that $m_{1N}(\theta) \to m_1(\theta)$ uniformly in θ and by Lemma 1, we have

$$m_{1N}(\alpha,\beta_0 + b/\sqrt{N}) \xrightarrow{p} m_1(\alpha,\beta_0).$$

We apply the mean value theorem to the last term in above equation. We can obtain

$$\sqrt{N}m_{2N}(\beta_0 + b/\sqrt{N}) = \sqrt{N}m_{2N}(\beta_0) + R(\tilde{\beta})b$$

where $\widetilde{\beta} \in [\beta_0, \beta_0 + b/\sqrt{N}]$ and $R(\beta) = \partial m_2(\beta)/\partial\beta'$ which is defined in Assumption 2. By Assumption 2, $m_{2N}(\beta_0) \to m_2(\beta_0) = 0$ and $\widetilde{\beta} \xrightarrow{p} \beta$ by Lemma 1. So we have

$$\sqrt{N}\, m_{2N}(\beta_0 + b/\sqrt{N}) \to R(\beta_0)b.$$

By Assumption 3, we have

$$W_N(\bar{\theta}_N(\theta)) \xrightarrow{p} W(\bar{\theta}(\alpha, \beta_0)).$$

So the objective function has the following limits

$$\begin{aligned} S_N(\alpha, \beta, \bar{\theta}_N(\theta)) \Rightarrow\ & [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0) + R(\beta_0)b]' \\ & \times W(\bar{\theta}(\alpha, \beta_0))[\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0) + R(\beta_0)b]. \end{aligned}$$

Next, we fix α and differentiate it with respect to b. By solving the first order condition, we denote the solution by b^*,

$$\begin{aligned} b^*(\alpha) = & -[R(\beta_0)'W(\bar{\theta}(\alpha, \beta_0))R(\beta_0)]^{-1}R(\beta_0)'W(\bar{\theta}(\alpha, \beta_0)) \\ & \times [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)] \end{aligned}$$

Plug $b^*(\alpha)$ into the objective function to yield the concentrated limiting objective function $S^*(\alpha; \bar{\theta}(\alpha, \beta_0))$. To see this, note that

$$\begin{aligned} R(\beta_0)b^* = & -R(\beta_0)[R(\beta_0)'W(\bar{\theta}(\alpha, \beta_0))R(\beta_0)]^{-1}R(\beta_0)'W(\bar{\theta}(\alpha, \beta_0)) \\ & \times [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)]. \end{aligned}$$

So we have

$$\begin{aligned} & \Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0) + R(\beta_0)b^* \\ = & [I - R(\beta_0)(R(\beta_0)'W(\bar{\theta}(\alpha, \beta_0))R(\beta_0))^{-1}R(\beta_0)'W(\bar{\theta}(\alpha, \beta_0))] \\ & \times [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)]. \end{aligned}$$

Plug it into the objective function,

$$\begin{aligned} & S^*(\alpha; \bar{\theta}(\alpha, \beta_0)) \\ = & [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)]' \\ & \times [I - R(\beta_0)(R(\beta_0)'W(\bar{\theta}(\alpha, \beta_0))R(\beta_0))^{-1}R(\beta_0)'W(\bar{\theta}(\alpha, \beta_0))]' \\ & \times W(\bar{\theta}(\alpha, \beta_0)) \\ & \times [I - R(\beta_0)(R(\beta_0)'W(\bar{\theta}(\alpha, \beta_0))R(\beta_0))^{-1}R(\beta_0)'W(\bar{\theta}(\alpha, \beta_0))] \\ & \times [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)]. \end{aligned}$$

Note that

$$
\begin{aligned}
& [I - R(R'WR)^{-1}R'W]'W[I - R(R'WR)^{-1}R'W] \\
= \; & [I - R(R'WR)^{-1}R'W]'[W - WR(R'WR)^{-1}R'W] \\
= \; & [I - R(R'WR)^{-1}R'W]'[I - WR(R'WR)^{-1}R']W \\
= \; & [I - WR(R'WR)^{-1}R']W \\
= \; & M(\alpha, \beta_0, \bar{\theta}(\alpha, \beta_0)).
\end{aligned}
$$

So we obtain that

$$
\begin{aligned}
S^*(\alpha; \bar{\theta}(\alpha, \beta_0)) \; &= \; [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)]' \\
& \quad \times M(\alpha, \beta_0, \bar{\theta}(\alpha, \beta_0))[\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)].
\end{aligned}
$$

and $\alpha^* = \arg\min_{\alpha \in A} S^*(\alpha; \bar{\theta}(\alpha, \beta_0))$. Substituting α^* into $b^*(\alpha)$, we can obtain $b^*(\alpha^*)$ defined in the theorem.

Since $\arg\min$ is a continuous mapping and α^* is a unique minimum over A , by Theorem 3.2.2 of Vaart and Wellner (1996), it follows that $(\hat{\alpha}, \sqrt{N}(\hat{\beta} - \beta_0)) \xrightarrow{d} (a^*, b^*)$. *Q.E.D.*

Proof of Corollary 1 The result in the corollary follows by Theorem 1 when we replace the general objective function $S_N(\alpha, \beta, \bar{\theta}_N(\theta))$ by the one-step objective function $S_{1N}(\theta)$ defined in (365). *Q.E.D.*

Proof of Corollary 2 The two-step efficient GMM estimator depends on an estimate of the weighting matrix which utilizes the first-step GMM estimator. By Assumption 3, Lemma 1, and the definition of the two-step efficient GMM estimator, we have

$$
W_N(\hat{\alpha}_1, \hat{\beta}_1) \xrightarrow{p} \Omega^{-1}(a_1^*, \beta_0).
$$

Following Theorem 1 by replacing the general objective function $S_N(\alpha, \beta, \bar{\theta}_N(\theta))$ by the two-step objective function $S_{2N}(\theta)$ defined in (420), we can obtain the results in the corollary. Note that in this case the b_2^* depends on both the one-step estimator a_1^* and the two-step estimator a_2^* . *Q.E.D.*

Proof of Corollary 3 The continuously updating estimator depends on a weighting matrix which is continuously updated by the value of the estimator. But, we can simplify the limiting weighting matrix by Lemma 1 and Assumption 3,

$$W_N(\alpha,\beta) = W_N(\alpha,\beta_0 + b/\sqrt{N})$$
$$\xrightarrow{p} \Omega^{-1}(a,\beta_0).$$

The limiting weighting matrix doesn't depend on b. Then we can follow Theorem 1 by replacing the general objective function $S_N(\alpha,\beta,\bar{\theta}_N(\theta))$ by the continuously updating objective function $S_{cN}(\theta)$ defined in (545). *Q.E.D.*

Proof of Theorem 2 We have

$$S_N(\theta_0;\theta_0) = [N^{-1/2}\sum_{i=1}^{N}\phi_i(\theta_0)]' W_N(\theta_0)[N^{-1/2}\sum_{j=1}^{N}\phi_j(\theta_0)]$$

The first and the last terms can be rewritten as

$$N^{-1/2}\sum_{i=1}^{N}\phi_i(\theta_0) = N^{-1/2}\sum_{i=1}^{N}[\phi_i(\theta) - E\phi_i(\theta)] + \sqrt{N}EN^{-1}\sum_{i=1}^{N}\phi_i(\theta)$$
$$\Rightarrow \Psi(\theta_0) + C + m_1(\theta_0) + \sqrt{N}m_2(\beta_0)$$

by Assumptions 2 and 4. Since $m_1(\theta_0) = 0$ and $m_2(\beta_0) = 0$ from Assumption 2, we have

$$N^{-1/2}\sum_{i=1}^{N}\phi_i(\theta_0) \xrightarrow{d} \varrho = N(C,\Omega(\theta_0,\theta_0)).$$

By Assumption 3, we have

$$W_N(\theta_0) \xrightarrow{p} \Omega^{-1}(\theta_0,\theta_0)$$

So we obtain that

$$S_N(\theta_0;\theta_0)$$
$$\xrightarrow{d} \varrho'\Omega^{-1}(\theta_0,\theta_0)\varrho$$
$$\xrightarrow{d} \chi^2_{HK}(C'\Omega^{-1}(\theta_0,\theta_0)C). \qquad Q.E.D.$$

Proof of Theorem 3 We follow Kleibergen's (2005) idea to construct two asymptotically independent variables. By Assumption 5, we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix}\phi_i(\theta_0) - E[\phi_i(\theta_0)] \\ q_i(\theta_0) - E[q_i(\theta_0)]\end{pmatrix} \xrightarrow{d} (\Psi'_\phi, \Psi'_q)'$$

where

$$\begin{pmatrix} \Psi_\phi \\ \Psi_q \end{pmatrix} \sim N(0, V(\theta)).$$

Pre-multiplying it by

$$\begin{pmatrix} I_{HK} & 0 \\ -\widehat{V}_{q\phi}(\theta_0)\widehat{V}_{\phi\phi}(\theta_0)^{-1} & I_{mHK} \end{pmatrix},$$

and by Assumption 6, we have

$$\begin{pmatrix} I_{HK} & 0 \\ -\widehat{V}_{q\phi}(\theta_0)\widehat{V}_{\phi\phi}(\theta_0)^{-1} & I_{mHK} \end{pmatrix}$$
$$\xrightarrow{p} \begin{pmatrix} I_{HK} & 0 \\ -V_{q\phi}(\theta_0)V_{\phi\phi}(\theta_0)^{-1} & I_{mHK} \end{pmatrix}.$$

Let

$$\overline{\phi}_N(\theta_0) = \sum_{i=1}^{N} \{\phi_i(\theta_0) - E[\phi_i(\theta_0)]\}$$

and

$$\overline{q}_N(\theta_0) = \sum_{i=1}^{N} \{q_i(\theta_0) - E[q_i(\theta_0)]\}.$$

Then, we can obtain that

$$\sqrt{N}\begin{pmatrix} I_{HK} & 0 \\ -\widehat{V}_{q\phi}(\theta_0)\widehat{V}_{\phi\phi}(\theta_0)^{-1} & I_{mHK} \end{pmatrix}\begin{pmatrix} N^{-1}\overline{\phi}_N(\theta_0) \\ N^{-1}\overline{q}_N(\theta_0) \end{pmatrix}$$
$$= \sqrt{N}\begin{pmatrix} N^{-1}\overline{\phi}_N(\theta_0) \\ N^{-1}\overline{q}_N(\theta_0) - N^{-1}\widehat{V}_{q\phi}(\theta_0)\widehat{V}_{\phi\phi}(\theta_0)^{-1}\overline{\phi}_N(\theta_0) \end{pmatrix}$$
$$\xrightarrow{d} \begin{pmatrix} \Psi_\phi \\ \Psi_{q.\phi} \end{pmatrix}$$

where

$$\Psi_{q.\phi} = \Psi_q - V_{q\phi}(\theta_0)V_{\phi\phi}(\theta_0)^{-1}\Psi_\phi$$

and

$$\begin{pmatrix} \Psi_\phi \\ \Psi_{q.\phi} \end{pmatrix} \sim N(0, \begin{pmatrix} V_{\phi\phi}(\theta_0) & 0 \\ 0 & V_{qq.\phi}(\theta_0) \end{pmatrix})$$

Note that

$$V_{qq.\phi}(\theta_0) = V_{qq}(\theta_0) - V_{q\phi}(\theta_0)V_{\phi\phi}(\theta_0)^{-1}V_{\phi q}(\theta_0)$$

So $(\Psi'_\phi, \Psi'_{q.\phi})'$ has a joint normal distribution with zero correlation which means the asymptotic independence between Ψ_ϕ and $\Psi_{q.\phi}$.
Next, note that

$$\begin{aligned} & N^{-1}\overline{q}_N(\theta_0) - N^{-1}\widehat{V}_{q\phi}(\theta_0)\widehat{V}_{\phi\phi}(\theta_0)^{-1}\overline{\phi}_N(\theta_0) \\ = & [N^{-1}q_N(\theta_0) - N^{-1}\widehat{V}_{q\phi}(\theta_0)\widehat{V}_{\phi\phi}(\theta_0)^{-1}\overline{\phi}_N(\theta_0)] - EN^{-1}q_N(\theta_0) \\ = & N^{-1}\widehat{D}_N(\theta_0) - J_\theta(\theta_0). \end{aligned}$$

So we have

$$\begin{aligned} & \sqrt{N}\begin{pmatrix} N^{-1}\overline{\phi}_N(\theta_0) \\ vec(N^{-1}\widehat{D}_N(\theta_0) - J_\theta(\theta_0)) \end{pmatrix} \\ & \xrightarrow{d} \begin{pmatrix} \Psi_\phi \\ \Psi_{q.\phi} \end{pmatrix}. \end{aligned}$$

Now, consider the K statistic,

$$\begin{aligned} K(\theta_0) = & \frac{1}{4N}(\partial S_N(\theta_0;\theta_0)/\partial\theta)[\widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)\widehat{D}_N(\theta_0)]^{-1} \\ & \times \times(\partial S_N(\theta_0;\theta_0)/\partial\theta)' \\ = & N^{-1/2}\phi_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)\widehat{D}_N(\theta_0)[\widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)\widehat{D}_N(\theta_0)]^{-1} \\ & \times \widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)N^{-1/2}\phi_N(\theta_0). \end{aligned}$$

Let

$$\widehat{\xi} = [\widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)\widehat{D}_N(\theta_0)]^{-1/2}\widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)N^{-1/2}\overline{\phi}_N(\theta_0).$$

and

$$\begin{aligned} \widehat{\Xi} = & [\widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)\widehat{D}_N(\theta_0)]^{-1/2}\widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0) \\ & \times \sqrt{N}EN^{-1}\sum_{i=1}^{N}\phi_i(\theta_0). \end{aligned}$$

By Assumption 2 and Assumption 4, we have

$$\widehat{\xi} \Rightarrow \xi \sim N(0, I_{HK})$$

and

$$\widehat{\Xi} \xrightarrow{p} \Xi[C]$$

where

$$\Xi[C] = [D' V_{\phi\phi}^{-1}(\theta_0) D]^{-1/2} D' V_{\phi\phi}^{-1}(\theta_0) C$$

and $\gamma(N)\widehat{D}_N(\theta_0) \xrightarrow{d} D$.

When θ is well identified, $J_\theta(\theta_0)$ has full rank. We set $\gamma(N) = 1/N$, then

$$\begin{aligned} N^{-1}\widehat{D}_N(\theta_0) &= \frac{1}{\sqrt{N}}\{\sqrt{N}[N^{-1}\widehat{D}_N(\theta_0) - J_\theta(\theta_0)]\} + J_\theta(\theta_0) \\ &\xrightarrow{p} C_q \end{aligned}$$

because $\sqrt{N}[vec(N^{-1}\widehat{D}_N(\theta_0) - J_\theta(\theta_0))] \xrightarrow{d} \Psi_{q.\phi}$.

When θ is weakly identified, $J_\theta(\theta_0) = J_{\theta,N}(\theta_0) = C_q/\sqrt{N}$. We set $\gamma(N) = 1/\sqrt{N}$, then

$$\begin{aligned} N^{-1/2}\widehat{D}_N(\theta_0) &= \sqrt{N}[N^{-1}\widehat{D}_N(\theta_0) - J_\theta(\theta_0)] + \sqrt{N} J_\theta(\theta_0) \\ &\xrightarrow{d} C_q + \Psi_{q.\phi} \end{aligned}$$

When is totally nonidentified, $J_\theta(\theta_0) = 0$. We set $\gamma(N) = 1/\sqrt{N}$, then

$$N^{-1/2}\widehat{D}_N(\theta_0) \xrightarrow{d} \Psi_{q.\phi}. \qquad Q.E.D.$$

5.0 CONCLUSIONS

Empirical economic studies are often confronted by the joint problem of weak instruments and near exogeneity, such as labor economics and empirical economic growth theory. This dissertation presents new evidence and solutions on estimation and inference with weak instruments and near exogeneity. Chapter 1 reexamines the effect of institutions on economic performance in Acemoglu, Johnson and Robinson (2001) where the measurement of current institutions is instrumented by European settler mortality rates. Since many economists argue that the settler mortality rates can possibly affect economic performance through other channels, I reexamine the effect of institutions by considering near exogeneity. I provide some evidence to show that the effect of institutions is not significant in many regression specifications when the settler mortality rates are used as the main instrument. Chapter 2 studies estimation and inference with weak instruments and near exogeneity in a linear simultaneous equations model. I show that near exogeneity can exaggerate asymptotic bias of the TSLS and the LIML estimators. When using critical values from chi-square distributions, Anderson-Rubin and Kleibergen tests under exogeneity have a large size distortion. I propose the delete-d jackknife based Anderson-Rubin and Kleibergen tests to automatically reduce the size distortion in finite samples without a need for any pretest of exogeneity. Chapter 3 extends estimation and inference with weak identification and near exogeneity into a GMM framework with instrumental variables. A GMM framework allows nonlinear and nondifferentiable moment conditions. I examine asymptotic results of one-step GMM estimator, two-step efficient GMM estimator and continuously updating estimator with weak identification and near exogeneity. Near exogeneity can produce relatively large bias for all these estimators. The Anderson-Rubin type and the Kleibergen type tests under near exogeneity converge in distribution to nonstandard distributions, which creates large size distortion when using critical values from chi-square distributions. The delete-d jackknife based approach can reduce the size distortion

BIBLIOGRAPHY

Acemoglu, D., Johnson, S., Robinson, J.A., (2001) The Colonial Origins of Comparative Development: An Empirical Investigation, *American Economic Review*, 91, 1369-1401

Acemoglu, D., Johnson, S., Robinson, J.A., (2002) Reversal of Fortune: Geography and Development in the Making of the Modern World Income Distribution, *Quarterly Journal of Economics*, 117, 1231-1294.

Andersen, T.W., Rubin, H., (1949) Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations, *The Annals of Mathematical Statistics*, 21, 570-582.

Andrews, D.W.K., (1994) Empirical Process Methods in Econometrics, Ch.37 in : Griliches and Intriligator, eds., *Handbook of Econometrics*, Amsterdam: North-Holland. .

Andrews, D.W.K., (2000) Inconsistency of the Bootstrap with a Parameter Is on the boundary of the Parameter Space, *Econometrica*, 68, 399-405.

Andrews, D.W.K., Moreira, M., Stock, J., (2004) "Optimal Invariant Similar Tests for Instrumental Variables Regression," working paper, Harvard University and Yale University.

Angrist, J.D., (1990) Lifetime Earnings and Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records, *American Economic Review*, 80, 313-336.

Angrist, J.D., Krueger, A.B., (1991) Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics*, 106, 979-1014.

Bound, J., Jaeger, D.A., Baker, R.M., (1995) Problems with Instrumental Variable Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak, *Journal of the American Statistical Association*, 90, 443--450.

Caner, M., (2005) Near Exogeneity and Weak Identification in Generalized Empirical Likelihood Estimators: Fixed and Many Moment Asymptotics, working paper, University of Pittsburgh.

Davidson, J., (1994) *Stochastic Limit Theory: An Introduction for Econometricians*, Cambridge:

Oxford University Press.

Dufour, J.M., (1997) Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models, *Econometrica*, 65, 1365-1387.

Dufour, J.M., (2003) "Identification, Weak Instruments, and Statistical Inference in Econometrics," *Canadian Journal of Economics*, 36, 767-808.

Engle, R.F., Hendry, D.F., Richard, J.F., (1983) Exogeneity, *Econometrica*, 51, 277--304.

Glaeser, E.L., La Porta, R., Lopez-De, Silanes, F., Shleifer, A., (2004) Do Institutions Cause Growth? *Journal of Economic Growth*, 9, 271-303.

Guggenberger, P., Smith, R.J., (2005) Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification, *Econometric Theory*, 21, 667-709.

Guggenberger, P., Wolf, M., (2004) Subsampling Tests of Parameter Hypotheses and Overidentifying Restrictions with Possible Failure of Identification, working paper, UCLA.

Hall, A.R., Inoue, A., (2003) The Large Sample Behavior of the Generalized Method of Moments Estimator in Misspecified Models, *Journal of Econometrics*, 114, 361-394.

Hall, R.E., Jones, C.I., (1999) "Why Do Some Countries Produce So Much More Output Per Worker than Others?" *Quarterly Journal of Economics*, 114, 83-116.

Hansen, Lars Peter, (1982) "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

Hausman, J.A., (1984) "Specification and Estimation of Simultaneous Equations Models," Ch.7 in : Griliches and Intriligator, eds., *Handbook of Econometrics*, Amsterdam: North-Holland.

Kleibergen, F., (2002) Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression, *Econometrica*, 70, 1781-1803.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R., (1999) "The Quality of Government," *Journal of Law, Economics and Organization*, 15, 222-279.

MacKinnon, J.G., (2002) "Bootstrap Inference in Econometrics," *Canadian Journal of Economics*, 35, 615-645.

McArthur, J.W., Sachs, J.D., (2000) "Institutions and Geography: Comments on Acemoglu, Johnson and Robinson," *NBER working paper*, no. 8114, Cambridge, MA.

Moreira, M.J., (2003) A Conditional Likelihood Ratio Tese for Structural Models, *Econometrica*, 71, 1027-1048.

Newey, W.K., (1985) Generalized Method of Moments Specification Testing, *Journal of*

Econometrics, 29, 229-256.

Phillips, Peter C.B., (1984) "Exact Small Sample Theory in the Simultaneous Equations Model," Ch8 in: Griliches and Intriligator, eds., *Handbook of Econometrics*, Amsterdam: North-Holland.

Politis, D.N. and Romano, J.P., (1994) "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions," *Annals of Statistics*, 22, 2031-2050.

Politis, D.N., Romano, J.P., Wolf, M., (1997) Subsampling for Heteroskedastic Time Series, *Journal of Econometrics*, 81, 281-319.

Politis, D.N., Romano, J.P., Wolf, M., (1999) *Subsampling*, New York: Springer.

Pollard, D. (1984) *Convergence of Stochastic Process*, New York: Springer Verlag.

Pollard, D. (1990) *Empirical Process: Theory and Applications*, Hayward CA: Institute of Mathematical Statistics.

Sargan, J.D., (1958) "Estimation of Economic Relationship Using Instrumental Variables" *Econometrica*,26, 393-514.

Staiger, D., Stock, J.H., (1997) Instrumental Variables Regression with Weak Instruments, *Econometrica*, 65, 557-586.

Stock, J.H., Wright, J.H., (2000) "GMM with Weak Identification," *Econometrica*, 68, 1055-1096.

Stock, J.H., Wright, J.H., Yogo, M., (2000) A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments, *Journal of Business and Economic Statistics*, 20, 518-529.

Van der Vaart, A.W., Wellner, J., (1996) *Weak Convergence and Empirical Process*, New York: Springer-Verlag.

Wang, J. and Zivot, E., (1998) "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments," *Econometrica*, 66, 1389-1404.

Wooldridge, J.W., (2002) *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press.

Wu, C.F.J., (1989) "On the Asymptotics Properties of the Jackknife Histogram," *The Annals of Statistics*, 18, 1438-1452.

Zivot, E., Startz, R., and Nelson, C.R., (1998) "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39, 1119-1144.